IDRBT JOURNAL OF BANKING TECHNOLOGY

Volume 1 | Number 1 | Jul - Dec 2017

Explore, Enable, Excel

Institute for Development and Research in Banking Technology
(Established by Reserve Bank of India)

IDRBT Journal of Banking Technology

Aims and Scope

The aim of the IDRBT Journal of Banking Technology (IJBT) is to promote new thinking, conceptual frameworks and research/practice-oriented innovative ideas on topics of current interest and broad relevance to application of Technology in Banking and Financial Services.

The scope of the journal covers all aspects of technology, which directly or indirectly contributes to the technological growth of the banking and financial sector, both from researchers' as well as practitioners' perspectives. It publishes original research/practice articles on all aspects of computing and communication technologies, which are/can be used in banking and finance, including case studies, experimental and survey articles.

IDRBT Journal of Banking Technology (IJBT)

Journal Website

http://www.idrbt.ac.in/ijbt.html

Subscription Information

The IDRBT Journal of Banking Technology is published twice a year in January and July. For information on subscription rates, please contact: publisher@idrbt.ac.in

Advertisements

Please contact publisher@idrbt.ac.in for advertisement rates.

Office of Publication

Publications Office,
Institute for Development and Research in Banking Technology,
Castle Hills, Road No.1, Masab Tank,
Hyderabad-500 057, India.
e-mail: publisher@idrbt.ac.in

CONTENTS

Dr. A. S. Ramasastri

Banks have been adopting various technology solutions for carrying out their functions. All such technologies leveraged by banks can be broadly considered to constitute banking technology.

In view of the large scale operations, which were to some extent routine, banks first deployed the early versions of manual calculators. Gradually, they moved to electronic calculators and then to various generations of computers. Presently, banking sector is one of the biggest consumers of computing infrastructure. Similarly, banks have moved from the days of telegram to today's digital networks for communications. Banking is moving closer to its customers through desktops and mobiles. Instead of banking, it is digital banking today.

Adoption of technology is bringing several benefits to banking; some of them being – better customer service, operational efficiency and cost optimisation. Such technology adoption is being made possible by the work carried out by several individuals and institutions across the globe. While on one hand there have been advancements in computer, communication and other relevant technologies due to the research carried out in academic institutions; on the other hand, bankers and technology companies are working together to ensure that newer technologies are adopted for better banking. It is the combination of fundamental and applied research that is pushing banking technology forward.

The Institute for Development and Research in Banking Technology (IDRBT) was established by Reserve Bank of India over two decades ago. It is a visionary and forward looking initiative of the central bank of the country to establish a unique Institute exclusively for research and development in the area of banking technology.

The Institute has been carrying out research and development in the area of technology useful to banks and other financial institutions in India. Towards this objective, the Institute has been bringing out various publications that present the latest technological innovations and developments. The Institute hosts international conferences every year with participation of renowned researchers from across the world.

Dr. A. S. Ramasastri (✉)
Director, Institute for Development and Research in Banking Technology, Hyderabad, India
e-mail: asramasastri@idrbt.ac.in

The IDRBT Journal of Banking Technology is the culmination of the Institute's efforts to bring together high quality research output across all areas of technology that are directly or indirectly relevant to banking. The aim of the journal is to publish current research from academia and also present the emerging perspectives of practitioners.

I am happy that researchers and professionals of very high standing contributed to this inaugural issue of the Journal. It is an honour to receive an article by Dr. Duvvuri Subbarao, Former Governor, Reserve Bank of India on "Disruptive Innovation in the Financial Sector" for the very first issue. The article is an extension of his talk on the same topic at the 18th International Conference on Distributed Computing and Networking, held at IDRBT from January 05-07, 2017

The three research articles – "*Artificial Neural Networks for Artificial Intelligence*" by Prof. Nikola Kasabov, "*Anomaly Detection in Banking Operations*" by Prof. Chilukuri K. Mohan and Prof. Kishan G. Mehrotra and "*Indoor Self-Localization via Bluetooth Low Energy Beacons*" by Prof. Marco Aiello et al - have dwelt on artificial intelligence, anomaly detection and location information, which are important areas of focus of banking technology. The article titled "*Upcoming Research Challenges in the Financial Services Industry: A Technology Perspective*" is an apt contribution by Raghu Krishnapuram, Anirban Mondal for the first issue of the Journal as it details the various current research areas in the financial services industry and draws the broad contours to the scope of the journal.

We have set ourselves the goal to make this journal a repository of high quality research and application articles on technology that is useful to banking. We feel, we have half-achieved the goal with the receipt of articles from reputed researchers and professionals from various parts of the world – making it a truly international journal.

We endeavour to grow with every issue under the able guidance of the top-rated Editorial Board members and continued support from researchers and practitioners globally.

Dr. A. S. Ramasastri
Editor-in-Chief

IJBT (2017) 01: 1-15
RESEARCH ARTICLE

Indoor self-localization via bluetooth low energy beacons

Azkario Rizky Pratama[1,2] · Widyawan[2] · Alexander Lazovik[1] · Marco Aiello[1]

Abstract Indoor localization is concerned with mapping sensory data to physical locations inside buildings. Location of a user or a mobile device is an essential part of the context, and is therefore very useful for pervasive computing applications. Many proposals exist for solving the localization problem, typically based on image or radio signal processing, though the problem is still generally considered to be open, especially when costs and privacy constraints play an important role. In this paper, we propose a solution based on the emerging Bluetooth Low Energy (BLE) standard and off-the-shelf hardware. Such approach proves to satisfy economic constraints, while challenging in terms of accurate location. To translate beacon signals into locations, we consider several approaches, i.e., cosine similarity, nearest neighbourhood classification, and the nearest beacon. Our experiments indicate a vector based approach as the most suited one. In fact, we show its effectiveness in an actual office deployment consisting of five indoor areas: three multiuser offices, a social corner, and a hallway. We achieve 90% and 80% for accuracy and F-measure, respectively.

Keywords: Ubiquitous computing · Smart buildings · Indoor localization · Bluetooth · Low energy beaconing

1 Introduction

Location information is a central component of the context and in turn of most modern pervasive system applications. Typical scenario for its uses are emergency response,

Azkario Rizky Pratama:
a.r.pratama@rug.nl

Widyawan:
widyawan@ugm.ac.id

Alexander Lazovik:
a.lazovik@rug.nl

✉ Marco Aiello:
a.lazovik@rug.nl

[1] Distributed Systems Group, Johann Bernoulli Institute for Mathematics and Computer Science, University of Groningen, The Netherlands

[2] Department of Electrical Engineering and Information Technology, Universitas Gadjah Mada, Indonesia

user navigation, environment optimization, and smart buildings. Depending on the application, the requirements for localization can be quite diverse. Precision in the localization can be traded off with costs of the sensors required for determining location (e.g., for energy saving in smart buildings), while this is not the case for emergency response systems that need reliable context information.

Our main research interest is in the area of smart building with particular emphasis on energy saving [1,2]. In particular, we are interested on approaches that are affordable, least invasive as possible, and sufficiently reliable [3].

In typical current office buildings, Wi-Fi access points (APs) are standard to provide wireless connection to occupants, Received Signal Strength (RSS) from each AP can be used further as an indication of where people are located. While Wi-Fi has been successfully used in buildings over the past two decades, new standards are emerging that can help address the localization problem. Bluetooth Low Energy (BLE) standard, first announced as iBeacon by Apple in 2013, is a low range, low energy communication protocol that extends the widely diffuse bluetooth standard. In BLE, a tiny packet is broadcasted periodically from each node to the surrounding environment, and mobile phones receive packets containing unique IDs and signal transmission strengths. Receiving phones need not authenticate nor be identified, thus helping maintaining privacy. Compared to Wi-Fi, the BLE beacon hardware is of smaller size and much lower energy requirements, thus facilitating its deployment and adoption. Fixed emitting stations, such as Wi-Fi and BLE, act as beacons useful for localization.

1.1 Landmark-based localization

Landmark-based localization utilizes unique features of the physical environment for the purpose of identifying the navigated space. The features can be natural or human artefacts; the features can be passive or actively signalling their unique identifier. Radio based beacons send radio signals on a regular basis with their unique ID. The received signal strength (RSS) is then used by the receiver to infer the location.

The most adopted method to utilize radio based beacons is that of constructing dense fingerprint maps of RSS to have comprehensive picture of distribution of signals in relation to the map of the location. Such an approach is unfortunately not scalable and portable, as it requires major time investments in creating the fingerprint maps.

1.2 Approaches

Given the requirement of having a low-cost, scalable and portable solution to the indoor localization problem, we consider several approaches based on BLE. In particular, we compare three – one, based on the strongest signal strength; second, based on magnitude distance between points; and third, based on angular distance among signal strengths vectors.

1.2.1 Nearest beacon

The simplest to beacon based localization is to consider the strongest signal as the indication of location, thus as the Nearest Beacon (NB). Assuming there are n-beacons deployed in the environment. Given an observation o formed by the readings of all known beacons $o = \{b_1, b_2, \ldots, b_n\}$ where $b_i \in \mathbb{R}^n$ is a RSS, then the location is the location assigned to $max(o)$. This approach is based on the fact that, in open space environments, the RSS has a direct correlation with the distance.

1.2.2 K-Nearest neighbor

The Nearest Neighbor (NN) algorithm considers the Euclidean distance in a feature space and classifies according to the minimal distance of the observation with known feature points. In our case, assuming that there are c-room classes, for each class, we arbitrarily collect k-instances as reference data, where each instance is a set of beacon readings $r_i^c = [b^1{}_i, b^2{}_i \ldots b^n{}_i]$ each one of which is associated with a location.

Then, given observation o, we compute the Euclidean distance for all elements of k and take (1).

$$\min\big(distance\ (o,r)\big) = \sqrt{\sum_{i=1}^{n}(o_i - r_i)^2} \tag{1}$$

1.2.3 Cosine similarity

Considering the feature space formed by all possible beacon values, one can consider reading as a vector. By considering the angle between the read vector and vectors representing locations, one can classify the location of the user.

Let $v = [v_1, v_2, \ldots, v_m]$ be a set of vectors representing certain locations and each location can be represented by one or more vectors. Given observation o, we can compute the cosine distance using (2).

$$\theta_{o,v_i} = \cos^{-1}\left(\frac{o.v_i}{|o||v_i|}\right) \tag{2}$$

From experimentation, one then defines a threshold value γ used to decide location. This is used to define a step function of location membership based on the cosine distance between the read vector and the references in the following way:

$$\cos\theta\ (vi\ , vj\ ) = \begin{cases} \text{united, if } (\gamma > \cos\theta \geq 1) \\ \text{spreaded, if } (0 \geq \cos\theta \geq \gamma) \end{cases} \tag{3}$$

1.3 Contribution and paper organization

We demonstrate a scalable and portable indoor localization solution that exploits readily available resources and features of the emerging BLE standard. To increase its portability, we do not specify predetermined rules for reference data collection. Thus, we accept a lower accuracy in the classification, to avoid long and tedious predefined mapping of the deployment spaces. We address the possible shortcomings of such an approach by investigating an approach to classification based on the cosine distance.

The paper is structured as follows. In Section 2, we present an overview and the design of the proposed system as well as its concrete implementation. In Section 3 and Section 4, we elaborate on the case study set-ups; and provide results and discussions, respectively. Related work is presented in Section 5 and conclusions are presented in Section 6.

2 Design and implementation

Using an office space actually in use, we deploy a set of BLE beacons and design a system for the localization with the intent of testing several approaches to self-localization.

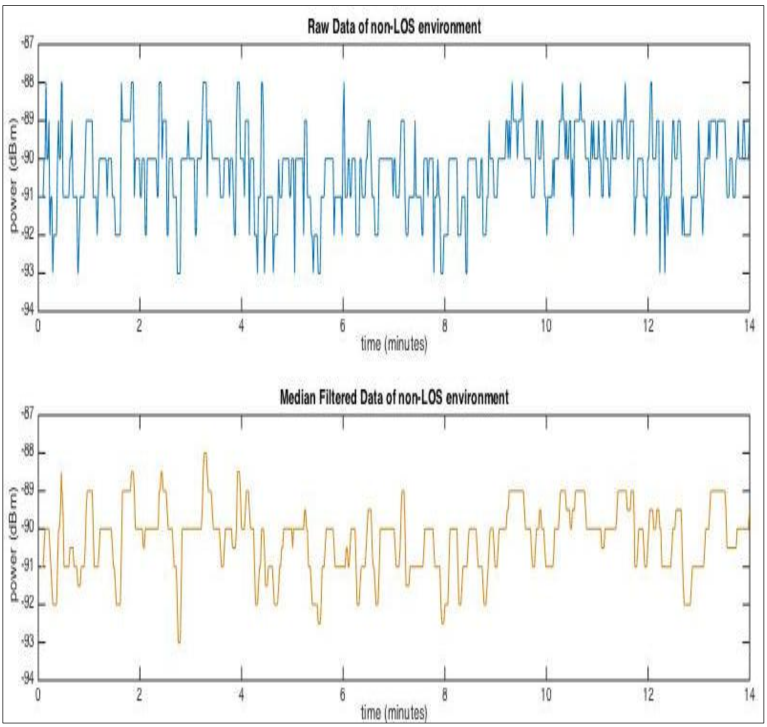


Fig. 1 (a). Raw data; (b). Median filtered data with window size of 6 samples

2.1 Design and assumptions

We set working desks as *observation points* for room offices, that is, locations where the user is likely to reside for a longer period of time. Each class can have one, more than one, or even no observation point. With respect to classes that have no observation points (e.g., a hallway or a social corner), we define an *observation area,* where reference data is collected while the occupant is moving around.

The underlying assumption is that a person can be represented by one device that is always brought by the individual wherever s/he goes. The device runs a dedicated application and the Bluetooth interface must be on.

2.2 Bernoulliborg implementation

We test our system and several approaches to localization in our own office building, the Bernoulliborg, in the Zernike campus of the University of Groningen. We select our offices on the *5th* floor, as shown in Figure 2. There are three office spaces (i.e.: R1, R2, R3), a Social Corner (SC), and Hallway (HW) interconnected among the spaces. To collect raw data, we write an android-based application. The mobile phone measures beacon's signal strength according to the iBeacon protocol [4]. These data are transmitted over Wi-Fi and stored in a centralized database.

We begin by copiously collecting BLE signals in the frequency 1 Hz from a single beacon node in the environment for 14 minutes length. It consists of about 800 samples. Such data set is sufficient to illustrate the signal fluctuation. We find that even though a device is positioned at a fixed position, the RSS fluctuates often (see Figure 1(a). To have a more stable value, we apply low pass filtering with a specific window size to eliminate outliers from the raw signal, as shown in Figure 1(b). The window size is determined on the basis of the expected time a person will need to transition to a new room. As the rooms are about 7x4 meters, it is reasonable to assume that people need 20-30 seconds to move to another space. Given the 0.25Hz sampling rate, a window size of 6 instances, corresponding to approximately 24 seconds, is a reasonable size.

In addition to using the RSS, we consider the steps count as a feature that detects occupant's movements.

Table 1 Defined parameter

Parameter	Value
Transmitted power	20*dBm*
Beacon density	2 beacons per-30*m*
Broadcast	950*ms*
Listening	2500*ms*
Waiting	1500*ms*

We access the value that represents step occurrence in a specific time, value that is typically available in modern mobile OS.

3 Experiment

We perform an experimental evaluation in the Bernoulliborg offices to determine the feasibility and accuracy of localization at the room level based on BLE.

3.1 Setup

As proximity based localization system, Bluetooth Low Energy offers some configurable parameters that need adjusting according to the specific utilization. Our goal in determining the parameter values is to preserve energy of both beacon nodes and user's mobile phone, while gaining a sufficient accuracy. Our choice of parameters for the experiment is shown in Table 1.

Broadcasting power is set to *-20dBm* (instead of the default value of *-12dBm*). This is to extend beacons battery life expectancy to more than 3 years. With such power configuration, each beacon theoretically will be able to reach devices within a 3.5 meter range. This shorter coverage entails the deployment of more beacons per unit of space. In our experiment we used 12 beacons to cover 3 working offices, a SC, and HW, totally approximately 170 m^2. Each beacon has the same *950ms* broadcasting interval, while a mobile phone will listen for beacon messages for *2500ms* and wait for next batch for *1500ms*. We then collect reference data in 10 observation points and an observation area for each HW and SC, shown as red dots in Figure 2.

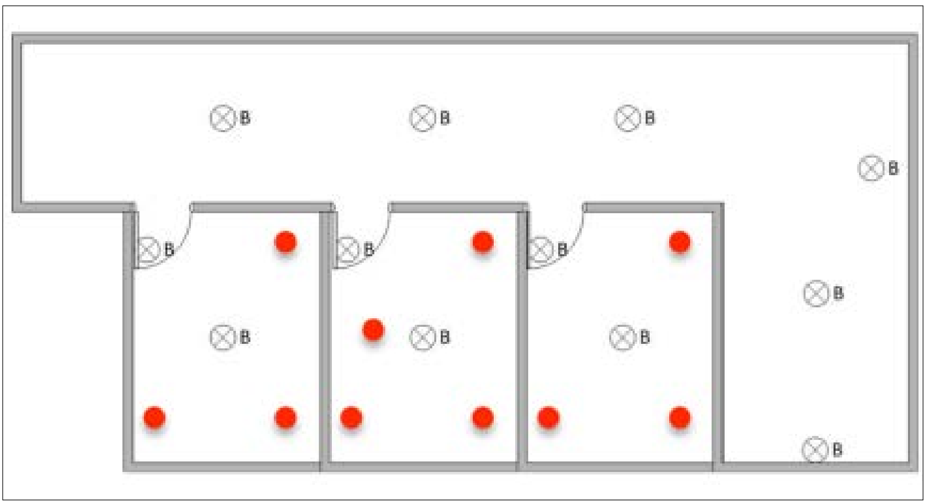


Fig. 2 Beacon deployment and observation points

To evaluate the techniques, we conduct 3 different movement scenes consisting of a total of 30 room transitions. Namely,

- HW - R1 - HW - SC - HW - R2 - HW - SC - HW - R3 - HW
- HW - R1 - HW - R3 - HW - SC - HW - R2 - HW - SC - HW
- HW - SC - HW - R1 - HW - R2 - HW - R3 - HW - SC - HW

Since each room has several observation points, as illustrated in Figure 2, the occupancy points of a room in one scene are different from those of the same room in other scenes.

For the purpose of classifier performance checking, we provide information on the actual location when measuring, that is, some form of ground truth against which we compare the measured values and their classification. The ground truth values are collected semi-manually by pushing a physical button of the phone's headset to trigger a confirmation dialog box of current position that the user has to confirm. Such value is stored together with the device's clock time. When a user switches from one room to another one, he is expected to press the button and select the new entered room, among the possible ones. Such real location value is used to evaluate the performance by comparing to classification outputs.

3.2 Metrics

The True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN) are the basic ingredients of the evaluation metrics to quantify the performance of a classification method. These represent the number of actual (Positive) or non-occurred (Negative) events that are correctly or incorrectly recognized. Based on these, we have two metrics:

1. *Accuracy*

The accuracy metric is the number of correct predictions made, divided by the total number of predictions; that is,

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

2. *F-measure*

The F-measure metric gives the harmonic mean of the number of correct estimated events divided by the total number of estimated events (a.k.a. precision) and the number of correct estimated events divided by the total number of real or relevant events (a.k.a. recall). Precision, Recall, and F-measure are defined in Equations (5), (6) and (7), respectively.

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

$$F - measure = \frac{(1 + \beta^2) * recall * precision}{\beta^2 * recall + precision} \tag{7}$$

4 Results and discussions

4.1 Results: Classification approach comparison

Several approaches can be used to map the beacon's RSS into a location. The first step we take is to compare three well-known classification techniques to identify the most promising one. These are Nearest Neighbor (NN), Nearest Beacon (NB), and one based on the cosine distance.

We consider the NB as a baseline as this is the simplest approach. It works by associating the location of a device with the strongest RSS signal among all the available signals. We also consider the Euclidean NN for the classification technique. NN makes use of the straight-line distance between two points (i.e. reference and query signals). Finally, the cosine distance works by translating RSS into elements of a vector space and measuring the angle among them considered as a degree of similarity between pair of vectors.

The comparison of these approaches using the test data of scene-2 is shown in Figure 3. The baseline is about 82% F-measure (i.e.: indicated by the strongest signal). This is higher than the Euclidean NN classification. The reason for the NB remaining stable in the graph is because the approach immediately sets the location based on the strongest signal. Hence, different window sizes do not affect the performance as no low pass filter is applied. One can see that the Euclidean NN and the cosine distance provide around 69% and 83% F-measure, respectively, without passing through a low pass filter. Such a filter could relatively enhance the performance of both approaches. The cosine distance undoubtedly outperforms the NN for all window sizes. It also exceeds the baseline for a meaningful interval, namely for time windows of size up to 25 units. There is unnoticeable benefit of considering window sizes larger than 25.

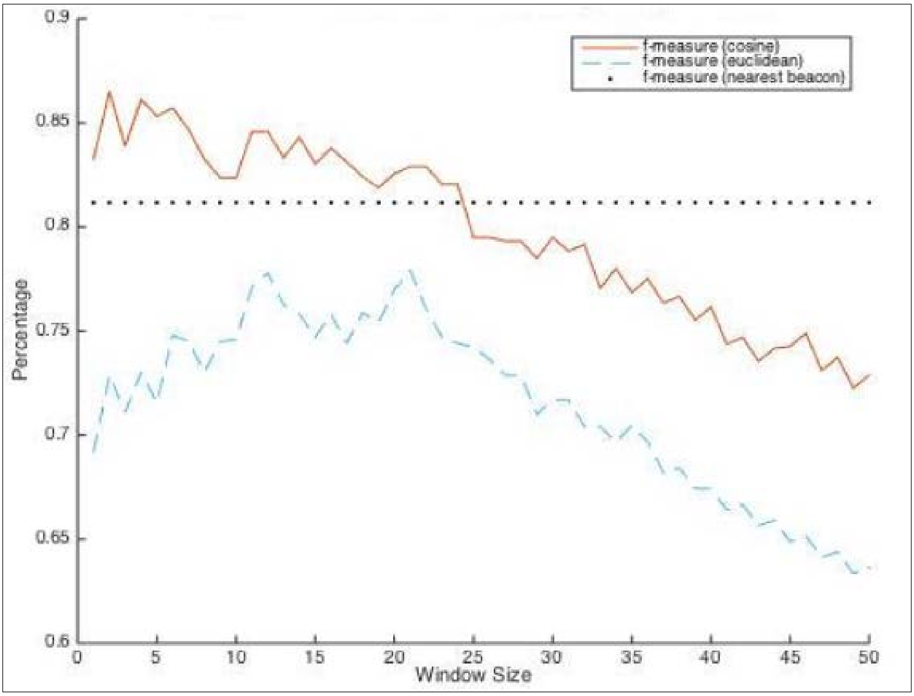


Fig. 3 Comparison of F-measure for increasing time windows

The average performance of the three scenes consisting of 30 room transitions is summarized in Table 2. We evaluate the approaches by considering either only the RSS or a combination of RSS and steps count. The NB approach, however, only takes the RSS into account. Interestingly, the steps count feature does not appear to bring an increase in the accuracy of classification. It even somewhat reduces the performance of both the Euclidean NN and cos-distance approaches. The Euclidean NN method is the least performing in terms of F-measure, with only a value of 68.85%. We take the F-measure into account since our dataset consists of unbalanced classes (the hallway, for example, is less occupied than other classes). Then the classification performance is only partially described by the accuracy metric.

Table 2 Performance of classification techniques

Method	Accuracy (%)	F-Measure (%)
Euclidean NN	81.83	68.85
Nearest Beacon	87.50	75.75
Cos-distance	89.90	80.41
Euclidean NN (+ steps count)	81.62	68.51
Cos-distance (+ steps count)	89.78	80.34

4.2 Results: Cosine distance

Given the comparison results, we focus on the cosine similarity on vectors as the classification approach.

We consider each observation point as a location reference. We have *m*-observation points with readings from *n*-beacons. So, we can define a $m \times n$ prototype matrix as:

	b_1	b_2	b_3	...	b_n
v_1	RSS_1^1	RSS_1^2	RSS_1^3	...	RSS_1^n
v_2	RSS_2^1	RSS_2^2	RSS_2^3	...	RSS_2^n
...	...	...	...	...	...
v_m	RSS_m^1	RSS_m^2	RSS_m^3	...	RSS_m^n

Each prototype vector v_i is computed as the median of all instances measured at an observation point. The median is taken to represent the observation point and the area it is in. From the practical point of view, we find that some beacons in the hallway (HW) do not contribute to these reference vectors. Therefore, there is an opportunity to reduce the dimensionality of the vector space by discarding unused beacons. To verify the validity of such reduction, we perform a classification experiment with and without these beacons taken into account in the vector space. The result is shown in Figure 4. Both accuracy and F-measures slightly decline for some window sizes, up to 13 units. For the rest, no different performance is noticeable, thus indicating that such space reduction does not hinder the quality of the classification in any significant way.

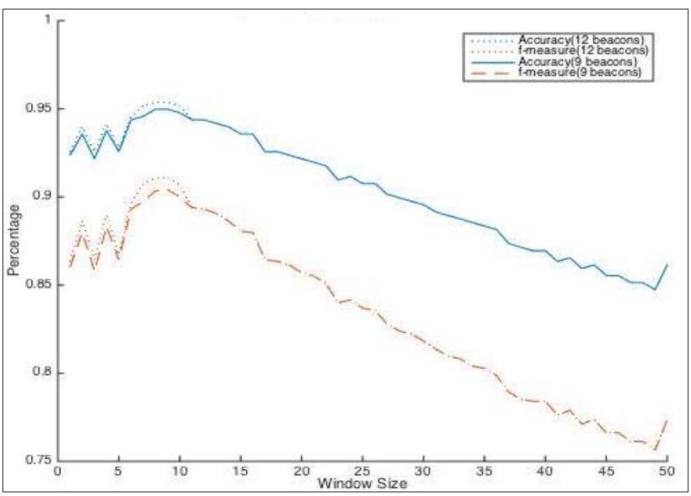


Fig. 4 Impact of discarding beacons in Hallway

The next step is to compute the angle among pairs of vectors. Our aim is to evaluate whether such vectors are a valid representation of a class, that is, of a location. The idea is that the vector representing a class should indicate different direction with vectors that represent other classes. We use rule expressed by Equation 3 to determine whether two vectors have a close enough direction, and therefore represent the same location (at the level of a room).

The cosine distance between two prototype vectors is shown in Table 3. Vector v_1, which represents the Hallway, has a similar direction to three vectors representing class R2, namely v_5, v_6, and v_7. Vector v_8, which represents class R2, has very similar direction to v_9 and v_{10} that in turn represent class R3. We rerun the experiment discarding v_8. Such an elimination improves both accuracy and F-measure by up to 10%, as shown in Figure 5. Based on the cosine similarity approach preceded by these analysis, we can achieve the results shown in Table 2. The average accuracy is 90% and the F-measure of 80%. The breakdown result of this approach for each scene is shown in Table 4. One can see that this approach delivers more than 93% and 85% accuracy and F-measure, respectively, for both scene-2 and scene-3. The most difficult scene to be localized is scene-1 with F-measure value of 65%. The taken observation points on this scene have dropped classification performance.

Table 3 Cosine-distance between pair of vectors

		v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9	v_{10}	v_{11}	v_{12}
HW	v_1	1	0.4 38	0	0	0.831	0.757	0.894	0	0	0	0	0.056
	v_2		1	0	0	0.455	0.414	0.490	0	0	0	0	0.547
R1	v_3			1	1	0	0.533	0	0	0	0	0	0
	v_4				1	0	0.533	0	0	0	0	0	0
	v_5					1	0.787	0.939	0	0	0	0	0
R2	v_6						1	0.846	0	0	0	0	0
	v_7							1	0	0	0	0	0
	v_8								1	0.970	0.894	0	0
	v_9									1	0.976	0.232	0
R3	v_{10}										1	0.427	0
	v_{11}											1	0
SC	v_{12}												1

Table 4 Classification with cosine distance approach

Scene	Accuracy (%)	F-Measure (%)
1	81.20	65.66
2	93.35	85.32
3	94.38	89.31
Average	89.64	80.10

4.3 Discussion

Our experiments showed that counting steps does not improve the room-level localization for an office space. There is a potential benefit to use steps-count feature for hierarchical classification, such as classifying binary states (i.e. in moving- or static-state) followed by more detail room classification (e.g. static inside room R2).

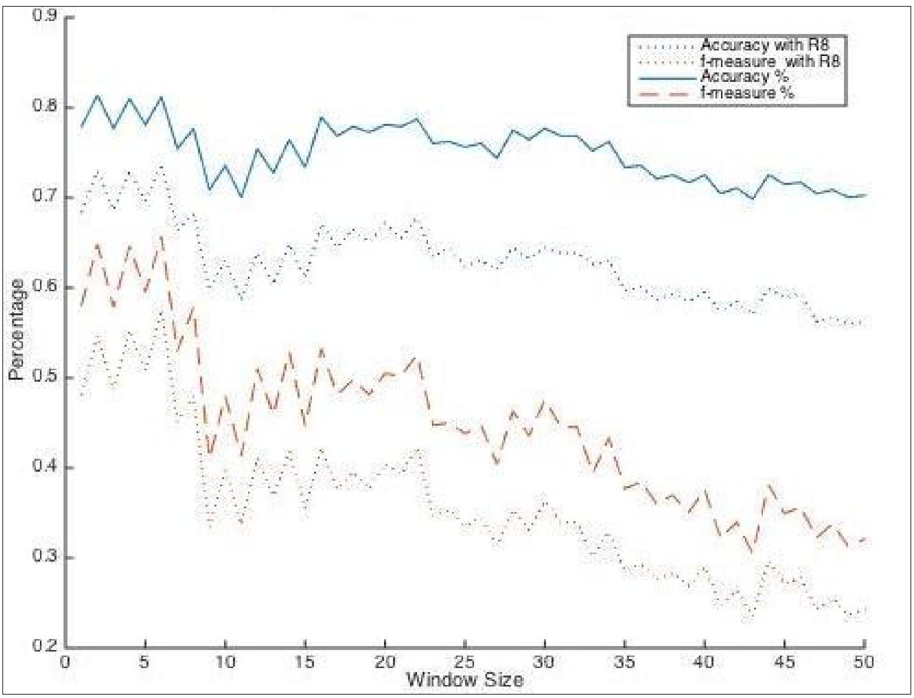


Fig. 5 Impact of discarding prototype vector v_8

The experiment shows that the cosine distance with appropriate thresholding and time windowing offers very good classification results, going up to 89.78%. Furthermore, the contribution of the signals from areas which are rarely used is low and can even be disregarded in the classification process. The reason is that there is a delay between the signal being transmitted from beacons and it being received by a mobile phone. Both period of signal propagation time and listener's sampling interval take part in such delay. Hence, in the case of the place where people rarely drop-by for longer periods (in Hallway, for example), the system is simply not gathering sufficient information from landmarks installed in that place. In other terms, the analysis also suggests where there should be more and where less density of beacons.

The experiment also shows that based on vector evaluation as expressed in Equation 3, we can differentiate between vectors that have similar direction and the others. Vectors v_1 and v_8 are two vectors which have a very similar direction to the vectors that represent other classes, but we cannot ignore both of them. We argue that v_1 is the vector that can’t be removed as it represents the Hallway area which is interconnected with all other areas. Thus, it has a similar direction with the other vectors representing other classes/spaces. With respect to v_8, this is intended to represent class R2 from a fixed observation point in the R2. It is expected to only have similar direction to other vectors representing class R2, but this is not the case. Hence, we decide this to be a removable vector.

5 Related work

Several efforts to obtain reliable location information through Radio Frequency (RF) technologies have been proposed. In [5], Liu *et. al.* estimate location of a tracking tag by comparing the signal strength received by the tag to the RSS of reference tags, both relative to active RFID readers. However, the readers price is expensive and the requirement of carrying additional tag make this inconvenient for our office scenario. Other approaches make use of existing Wi-Fi infrastructures. Some of them take advantage of the signal propagation model [6], of sniffing packets of wireless traffic from occupant devices [7], or by constructing radio map fingerprints [8]. Such approaches provide localization information, though suffer from being dependant on specific physical building construction, require major changes in the access points (e.g. for sniffing communication packets), and have the prerequisite of extensive radio map fingerprints collection, respectively. In other words, they are all hardly portable and scalable. Even though several studies propose methods for building fingerprint maps automatically, these usually utilize some additional resources (e.g.: accelerometer [9, 10]), making system performance really dependent on the step detection accuracy and consuming additional battery resources.

Bluetooth Low Energy standard is designed for marketing purposes. Though it has the potential to be used in personal localization, studies in this field have not been as intensive as Wi-Fi-based localization. In their study, Faragher *et. al.* [11] have shown potential improvements in BLE-based fingerprinting over Wi-Fi technology. Nevertheless, BLE configurations is set with high advertising rates and high transmission powers. This is certainly impractical in daily use due to power constraints, even though we have shown that BLE-based communication is more energy-efficient than Wi-Fi [12]. Room-level localization with BLE beacons has been experimented also by [13], utilizing RSS as feature for room classification. The claim is that 83% accuracy is achieved by implementing KNN and Decision tree techniques. The authors of [14] asserted that 97.22% accuracy could be achieved even only using the strongest beacon signals. However, in their experiment, they only considered two rooms and did not take more adjacent rooms. Taking more adjacent rooms will become a more challenging case since beacon signals from neighbors can be discovered and would interfere in the localization.

They were also not actively moving in the Hallway while taking experiments, instead, they take data in static points in the Hallway. Moreover, these authors are not interested in the battery life of the beacons and accept frequent battery replacement [14] or even add wired power sources to their Arduino-based beacons [13]. In contrast, authors in [15] have experimented with a special purpose BLE on chip that exploits power harvester in supplying energy to the chip. Nevertheless, this system is intended to demonstrate the possibility of auto-powered indoor localization system that is almost impossible to be used daily due to its dependencies on specific hardwares.

In this study, we propose a scalable and portable room-level localization system, with bearing in mind the energy saving configuration such as increasing the density of battery-powered beacons deployment as a compensation for broadcasting power reduction. We arbitrarily (i.e. without any predefined rules) collect RSS as reference in rooms, similar to [10], but instead of matching this to step detection, we analyze the collected references based on cosine similarity approach. To the best of our knowledge, none of the earlier indoor localization systems have considered cosine similarity in choosing proper vectors and in classifying a room location. In our case, where a room is represented by multiple reference beacons (thus consisting many dimensional data), the properties of received signals are more appropriate to be evaluated in the cosine similarity rather than the Euclidean distance approach.

6 Conclusion

We introduced a room-level indoor localization system and its implementation. The system is based on the emerging BLE standard offering an overall affordable and scalable solution. We address the problem of location accuracy based on three lightweight approaches to classification. We provide an analysis of signal strength vectors in terms of vector component and vector direction in order to support the cosine distance approach, as experimental evidence indicates it as the most suited one.

According to our experimental results in an office building, the cosine distance approach provides greater result by 12% and 5% F-measure over the Euclidean nearest neighbor and the nearest beacon approaches, respectively. Overall, we achieve room-level localization with accuracy of up to 90%.

The approach presented here promises to be highly portable and potentially be brought to other buildings with varying indoor configurations and higher number of rooms. Next step we plan to undertake involves the automatic elimination of vectors that do not help effectiveness in terms of classifying a label. Collecting more instances will also give a benefit to investigate more general patterns and exploit probability-based approaches.

References

1. Georgievski, I., Degeler, V., Pagani, G. A., Nguyen, T. A., Lazovik, A., Aiello, M.: Optimizing energy costs for offices connected to the smart grid. IEEE Trans. Smart Grid. **3**(4), 2273-2285 (2012)
2. Nguyen, T. A., Aiello, M.: Energy intelligent buildings based on user activity: A survey. Energy and buildings. **56**, 244-257 (2013)

3. Setz, B., Nizamic, F., Lazovik, A., Aiello, M.: Power management of personal computers based on user behaviour. In: Proceedings of 5th International Conference on Smart Cities and Green ICT Systems (SMARTGREENS). 1-8. IEEE (2016)
4. Apple Inc: Getting Started with iBeacon. https://developer.apple.com/ ibeacon/ Getting-Started-with-iBeacon.pdf. 1-11 (2014)
5. Ni, L. M., Liu, Y., Lau, Y. C., Patil, A. P.: LANDMARC: Indoor location sensing using active RFID. Wireless networks. **10**(6), 701-710 (2004)
6. Bose, A., Foh, C. H.: A practical path loss model for indoor WiFi positioning enhancement. In: Proceedings of 6th International Conference on Information, Communications & Signal Processing. 1-5. IEEE (2007)
7. Lu, X., Wen, H., Zou, H., Jiang, H., Xie, L., Trigoni, N.: Robust occupancy inference with commodity WiFi. In: Proceedings of Wireless and Mobile Computing, Networking and Communications (WiMob). 1-8. IEEE (2016)
8. Bahl, P., Padmanabhan, V. N.: RADAR: An in-building RF-based user location and tracking system. In: Proceedings of 19th Annual Joint Conference of the IEEE Computer and Communications Societies. **2**, 775-784. IEEE (2000)
9. Jiang, Y., Pan, X., Li, K., Lv, Q., Dick, R. P., Hannigan, M., Shang, L.: Ariel: Automatic wi-fi based room fingerprinting for indoor localization. In: Proceedings of the ACM Conference on Ubiquitous Computing. 441-450. ACM ((2012)
10. Yang, Z., Wu, C., Liu, Y.: Locating in fingerprint space: wireless indoor localization with little human intervention. In: Proceedings of the 18th annual international conference on Mobile computing and networking. 269-280. ACM (2012)
11. Faragher, R., Harle, R.: An analysis of the accuracy of bluetooth low energy for indoor positioning applications. In: Proceedings of the 27th International Technical Meeting of the Satellite Division of the Institute of Navigation. 201-210 (2014)
12. Putra, G. D., Pratama, A. R., Lazovik, A., Aiello, M.: Comparison of energy consumption in Wi-Fi and bluetooth communication in a Smart Building. In: Proceedings of Computing and Communication Workshop and Conference (CCWC). 1-6. IEEE (2017)
13. Conte, G., De Marchi, M., Nacci, A. A., Rana, V., Sciuto, D.: BlueSentinel: A first approach using iBeacon for an energy efficient occupancy detection system. In: Proceedings of BuildSys@ SenSys. 11-19 (2014)
14. Lin, X. Y., Ho, T. W., Fang, C. C., Yen, Z. S., Yang, B. J., Lai, F.: A mobile indoor positioning system based on iBeacon technology. In: Proceedings of 37th Annual International Conference of the Engineering in Medicine and Biology Society (EMBC). 4970-4973. IEEE (2015)
15. Liu, Q., Pawelczak, P., Zuniga, M., Ijntema, W., Drif, A.: Wiploc: Perpetual indoor localization with rf wireless power transfer. Under submission

IJBT (2017) 01: 16-48
RESEARCH ARTICLE

Anomaly detection in banking operations

Chilukuri K. Mohan[1] · Kishan G. Mehrotra[1]

Abstract This paper presents an overview of anomaly detection algorithms and methodology, focusing on the context of banking operations applications. The main principles of anomaly detection are first presented, followed by listing some of the areas in banking that can benefit from anomaly detection. We then discuss traditional nearest-neighbor and clustering-based approaches. Time series and other sequential data analysis approaches are described. The problems posed by categorical data are also discussed, along with the methods proposed in the literature to address the same. The ensemble methods are presented, followed by mathematical perspectives on anomaly detection.

Keywords: Anomaly detection · Banking · Time-stamped data · Categorical data · Ensemble methods

1 Introduction

Banking operations include many daily, periodic, and aperiodic activities and transactions performed by or affecting numerous stakeholders such as employees, customers, debtors, and external entities. The complex nature of these activities and transactions necessitate constant monitoring to ensure that neither the bank nor its stakeholders are adversely affected by various events that may be malicious, random, or occurring due to inevitable business cycles. Events may unfold over time, and early detection can significantly ameliorate potential ill-effects, and in some cases actively prevent the same. This paper develops a framework for understanding and addressing such events using various anomaly detection algorithms.

Anomalies (or outliers) are variations from expected norms, or from prior data, or from predictions based on process models. Anomalies are contrasted with *inliers* or non-anomalous data. The variations from norms may be exhibited in terms of multiple data points or in patterns of data, rather than individual data points themselves. For instance, a customer's abnormal spending may be indicated by a collection of closely spaced purchases, rather than by a single purchase.

✉ Chilukuri K. Mohan
mohan@syr.edu

Kishan G. Mehrotra
mehrotra@syr.edu

[1] Department of Electrical Engineering and Computer Science, 4-206, Center for Science and Technology Syracuse University Syracuse, NY 13244-4100, USA

In some cases, norms may be derived from the knowledge of experts who have examined past data; in other cases, machine learning or data mining algorithms may have been applied to prior data. In the supervised learning context (also known as the signature-based approach), the data used to build the model is pre-characterized by experts (or prior observations), so that the anomaly detection task merely involves measuring the variation of new data from such models. However, many practical applications involve the unsupervised learning context, in which no prior labels conclusively distinguish anomalous data from normal data.

Unsupervised fraud detection approaches are particularly important since new fraudulent attacks are being invented every day and signature based techniques are unable to detect them. Researchers have also explored semi-supervised problems in which some data identified as belonging to the majority (normal) classes is available, although not all data is labelled.

Classical unsupervised anomaly detection algorithms focus on identifying anomalies from a finite set $\boldsymbol{D}$ of data points $\boldsymbol{x}_i$ that are described using d numerical dimensions or attributes. First, we define a distance measure between two data points, such as the Euclidean distance measure,

$$d(x_i, x_j) = \frac{1}{d}\sum_{k=1}^{d}(x_{i,k} - x_{j,k})^2$$

Such a distance measure is often extended to define a related distance measure $d(x_i, S)$ between a data point $\boldsymbol{x}_i$ and a set $\boldsymbol{S}$ of data points, such as the average distance between $\boldsymbol{x}_i$ and all points in S. Often, though not always, the data is first clustered (using an algorithm such as the k-means algorithm), and S is chosen to be the cluster of points nearest to $\boldsymbol{x}_i$. This distance measure is then used to estimate an "outlierness" or *anomaly score* indicator α, such as the distance between a point and its k nearest neighbors, often linearly normalized to ensure that $0 \leq \alpha(\boldsymbol{x}_i) \leq 1$. A point $\boldsymbol{x}_i$ is then defined to be anomalous if $\alpha(\boldsymbol{x}_i)$ exceeds a predetermined threshold, or is among the m highest anomaly scores among points in S, if the goal is to find m such points.

There is a vast literature on such anomaly detection algorithms, as well as on clustering algorithms on which they are based. Markou and Singh [53, 54], Hodge and Jim [37], Agyemang *et al.* [5], Chandola *et al.*[17, 18] and Pimentel et al. [60] provide extensive surveys for particular domains from different point of views, but the focus is mostly on signature-based fraud detection techniques. Sabau [64, 8], Patcha *et al.* [58] also present various anomaly detection techniques based on supervised, unsupervised and clustering methods.

This paper focuses on anomaly detection problems that arise in the banking context, and develops a framework to understand the same, addressing how such problems differ from the classic anomaly detection problems described in the preceding paragraph. In particular, we focus on problems where:

- Some data attributes are categorical (making the formulation of numerical distance measures infeasible or awkward);
- Some domain knowledge is available (so that the problem is neither fully supervised nor unsupervised, not even semi-supervised in the formal sense);

- Data arrives over time (so that anomalousness must be judged in the context of immediately preceding data); and
- May be inherently heterogeneous (so that anomalies of different kinds may appear in the data).

We categorize anomaly detection problems depending on whether the data being examined is a set, a sequence, or a structure of inherently higher dimensionality, leading to three major categories:

1. Point anomaly: One or more anomalous points in a collection of data points;

2. Contextual anomaly: A point that may be anomalous with respect to its neighbors, or to other data points that share a context or some features; and

3. Collective anomaly: A collection of (similar) data points that behave differently from other collections in the entire dataset. An individual data point may not then be considered anomalous by itself, but is considered anomalous in conjunction with other data points in a collection.

The first category is addressed by traditional algorithms based on separating a single outlier data point from others, e.g., whether an item purchased by an accountholder is substantially different from anything purchased by him in the past. The second emerges in analyzing data that arrives over time, e.g., higher credit card expenses may not be anomalous if incurred during a festive month, but may be anomalous during a non-festive month. A collection of numerous credit card purchases during a single day may be considered anomalous, although each purchase by itself cannot be flagged as anomalous; this illustrates the third category, often ignored by the literature.

A major problem in anomaly detection is a lack of effective general purpose anomaly detection techniques because an anomaly detection technique in one domain may not be suitable for other domains; both the normal and abnormal behaviours vary from domain to domain. For example, the techniques used to finding out anomalies in stock exchange transactions may not work well for network traffic analysis, although they both make use of transaction data [45].

Section 2 discusses banking applications where anomaly detection is useful. Section 3 presents traditional anomaly detection approaches, particularly those based on clustering. Time series approaches are sketched in Section 4. Section 5 discusses anomaly detection in the context of categorical data. Ensemble approaches are discussed in Section 6. This is followed by simple mathematical formulations of the problems addressed, in Section 7. Concluding remarks in Section 8 summarizes the discussions.

2 Banking applications of anomaly detection

Many banking applications require the use of anomaly detection methodologies. For instance, credit card and mobile phone fraud occurrences include a variety of scams that steal data and wealth from individuals and organizations. To prevent the misuse of any account, it is necessary to detect any unusual usage pattern by monitoring the customer's usage pattern and looking for any deviation from prior usage patterns.

Insurance fraud includes cases where criminals manipulate the claim processing system by submitting documents whose contents contain untruths. In insider trading, criminals make use of crucial information before it is made public, to make illicit profits and manipulate stock prices; early detection is important.

Three broad classes of applications can be distinguished: behavior tracking, situation awareness, and fraud. Most of this section, and the focus of most work in this area, is on fraud; we first describe the other two areas.

2.1 Behavior tracking

Customer behavior tracking can provide clues to assist marketing efforts, and to enable advising processes aimed at preventing deterioration of the financial status of any customer. Examples include cases where spending patterns of customers begin to diverge significantly from prior history, as well as information regarding purchase behaviors that may suggest increased probability of specific other purchases in the future, perhaps signaling changes in lifestyle.

Employee behavior tracking can be helpful to identify productive and unproductive patterns of activity, which can directly lead to reinforcing or corrective actions by managers or decision-makers.

In these contexts, we may seek to identify outliers of various kinds, comparing individual behaviors to their own past (long-term) behavior, to the behavior of the peer group to which they belong, and when compared to larger populations of individuals (e.g., all customers, or all employees in a geographical region).

2.2 Situation awareness

Tracking external financial parameters and their changes over time can alert banks about how they can expect their customers and employees to behave in the future. For example, sudden changes in market indicators may lead to predictions about expected withdrawals or spending by customers. Similarly, anomalous events in overseas markets or unexpected fluctuations in currency markets may be expected to influence some day-to-day activities by large institutions that conduct business with a bank.

2.3 Fraud

“Fraud” is a very broad term that is used to describe many kinds of activities, distinguished by the nature of the data and the approaches needed to detect and prevent such activities. Broadly, such activities may be classified into three groups of activities, based on the perpetrators of the same: clients/customers and their representatives, employees/officers of the bank or financial institution, and outsiders.

1. Fraudulent activities by customers or clients include money-laundering, trading in illegal goods, and deliberate attempts to mislead the bank or others *via* multiple transactions. In some cases, an account may be administered by an authorized individual who undertakes a series of transactions resulting in a loss for the primary account-holder or beneficiary. Analysis of transaction sequences over time (and compared to patterns for other accounts) can reveal anomalies that require careful analysis and further investigation.

2. Bank employees have access to enormous amounts of data and are trusted to take care of their customers' accounts with the highest standards of ethics. Unfortunately, some cases have emerged in the past where employees have exploited internal knowledge of the system to obtain personal benefits, or to indulge in risky trading behavior contrary to the mission and principles of their employer. Many such cases can be detected by looking for anomalies in the behaviors (over time) of employees authorized to access customer accounts.
3. Most serious are cases of purchases and cash withdrawals by unauthorized individuals, and by illegal trading of customer information, violating privacy and exposing customers to increased risk. Unsurprisingly, these have attracted the greatest attention in the industry and the research community. In particular, examples of credit card fraud (and data compromises) are considered to be prototypical of situations in which anomaly detection algorithms can help, and we discuss this example in greater detail below.

   When a debit or credit card is illegally used by an outsider, the nature of the purchase (amount, location, timing, etc.) can be compared to other purchases by the legitimate credit card user and flagged as an anomaly. Sometimes, the anomaly may consist in identifying a sudden increase in the (short-term) frequency of purchases (e.g., a flurry of successive transactions, none of which is a large amount). The nature of the item purchased may also signify increased probability of fraud; e.g., online purchases of electronics and expensive clothes may indicate fraud, if the nature of the items purchased is substantially different from prior purchases.

Irrespective of which anomaly detection algorithm is applied, there is always uncertainty in the decision-making, with both false positives (legitimate transactions being flagged as fraudulent) and false negatives (fraudulent transactions that are not caught by the anomaly detection algorithm) occurring with some probability. The direct cost involved in false negatives is easily measurable (the dollar amount), along with the potential for future fraudulent transactions with the same account: catching the first fraudulent transaction can help prevent others. With false positives, on the other hand, we must consider other intangible costs: most importantly, unpleasant interactions with the customer may ensue, with reduced customer satisfaction and potential loss of future business with that customer, as well as members of their organizations and social circles.

3 Principles of traditional anomaly detection

Traditional anomaly detection algorithms are formulated as solutions to the problem of identifying a small subset $\boldsymbol{s}$ of data points from a given set $\boldsymbol{D}$, such that points in $\boldsymbol{s}$ are significantly different from those in $\boldsymbol{D} \setminus \boldsymbol{s}$ (i.e., points in $\boldsymbol{D}$ but not in $\boldsymbol{s}$). Abstractly, the problem of finding m most anomalous data points in a numerical multi-dimensional data set $\boldsymbol{D}$, with a specific distance measure, $\boldsymbol{d}$, may be viewed as an optimization problem with the goal of finding $\boldsymbol{s}$, a subset of $\boldsymbol{D}$ such that $|\boldsymbol{s}|=m$, maximizing;

$$\sum_{x \in s} \min(\boldsymbol{d}(\boldsymbol{x}, \boldsymbol{y}) | \boldsymbol{y} \in (\boldsymbol{D} \setminus s)) .$$

Specific algorithms often focus on formulating an anomaly score (or outlierness) function which can be used to compare pairs of points, such that $\alpha(\boldsymbol{x}) > \alpha(\boldsymbol{y})$ whenever a data point $\boldsymbol{x}$ is to be considered more anomalous than another data point $\boldsymbol{y}$. We can then present the results of the application of the algorithm in one of two forms, as required:

a. find the *m* most anomalous points; or
b. find all points whose anomaly score exceeds a specified threshold.

Cost-benefit analysis may dictate further refinements of such a procedure, as in the following example.

Example: Thousands of credit card transactions may occur in a minute, and computational constraints may limit the system's capacity to process more than 100 of them. Then the system should select the top 100 transactions, and a naive approach is to select these based on the anomaly score $\alpha(\boldsymbol{x})$ alone. But other considerations may also be important, such as the false-negative cost $\beta(\boldsymbol{x})$. In addition, the dollar amount of the transaction together with a quantification of other intangible costs (e.g., loss of trust in the system, and follow-up costs after fraud discovery); the cost $\gamma(\boldsymbol{x})$ of following up the transaction as a possible fraud; and the misclassification cost $\delta(\boldsymbol{x}, y)$ for a false positive, where *y* indicates the customer, since the cost may be higher for some customers (e.g., a large account may be lost by the bank if the CEO of a company is upset that s/he is unable to conduct a transaction while traveling). Rational decision theory now dictates that the anomaly score $\alpha(\boldsymbol{x})$ be translated into a probability $p(\boldsymbol{x})$ that describes the likelihood of the transaction $\boldsymbol{x}$ being fraudulent, which can only be done if a substantial amount of prior data is available, mapping anomaly scores of previous transactions to ground truth (indicating which transactions are fraudulent). Changes to the algorithm necessitate recalibrating the relationship between $\alpha(\boldsymbol{x})$ and $p(\boldsymbol{x})$. Then the expected cost of considering a transaction as fraudulent is $\gamma(\boldsymbol{x}) + \big(1 - p(\boldsymbol{x})\big)\delta(\boldsymbol{x}, \boldsymbol{y})$ and the expected benefit is $p(\boldsymbol{x})\,\beta(\boldsymbol{x})$. Hence the main criterion for selecting transactions for further processing should be:

$$C(\boldsymbol{x}, \boldsymbol{y}) = p(\boldsymbol{x})\beta(\boldsymbol{x}) - \left(\gamma(\boldsymbol{x}) + \big(1 - p(\boldsymbol{x})\big)\,\delta(\boldsymbol{x}, \boldsymbol{y})\right)$$

and the resource-constrained system should select the top 100 transactions, as compared using the overall cost measure $C(\boldsymbol{x}, \boldsymbol{y})$.

We now consider traditional approaches to measuring the anomaly score of a transaction $\boldsymbol{x}$, which essentially transforms data points into a single numerical measure.

Distance measures Traditional anomaly detection algorithms, applied to points in a *d*-dimensional Euclidean space with numerical coordinates, usually begin with the Euclidean distance between two points,

$$d(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{\sum_i (x_i - y_i)^2}$$

In many applications involving incommensurate dimensions, it is preferable to use the Mahalanobis distance or the normalized Euclidean distance,

$$d(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{\sum_i \left(\frac{x_i - y_i}{s_i}\right)^2}$$

where each dimension is separately normalized using the standard deviation s_i for that specific (i^{th}) dimension. For example, if one dimension of a bank account is a customer's income (measured in units of currency), and another dimension is age (measured in unit of time), then the Mahalanobis measure circumvents the strange question of whether a difference in income of one rupee is equivalent to a difference in age of one year, one month, or one second.

The next question to be addressed is how we may evaluate the distance ***d(x, S)*** between a point ***x*** and a set of points ***S,*** e.g., in defining what we mean by the distance between a potentially anomalous point and a cluster to which it does not belong. A few alternatives are possible, with definite implications for the notion of anomaly that is based on the point-to-set distance measure:

a) Euclidean distance between the given point ***x*** and the centroid μ of the set ***S***;

b) Mahalanobis distance $\sqrt{(x-\mu)^T\, V^{-1}(x - \mu)}$ where V is the covariance matrix between dimensions for points in ***S***;

c) Minimum Euclidean distance between ***x*** and the points in ***S***, i.e., distance from ***x*** to the nearest point in ***S***; and

d) Minimum Mahalanobis distance between ***x*** and the points in ***S***.

Example: In attempting to determine whether the online purchase of an expensive pair of shoes via a credit card transaction is fraudulent, we may compare it to other similar transactions that are believed to be non-fraudulent. For instance, if the customer had made prior online purchases from the same retailer, or had purchased another pair of shoes, of the same shoe size and similar cost, from another retailer, we may believe that the purchase is non-fraudulent. But if even the most similar prior purchase by that customer was significantly less expensive, or if the "shoe does not fit", then a minimum distance criterion would suggest that this purchase should be flagged with a high anomaly score.

k-nearest neighbor approaches: Nearest neighbor methods are among the simplest anomaly detection algorithms, and these may be used in a few different ways for anomaly detection. First, the distances $d_1, d_2, \ldots, d_k$ of a point ***x*** from its k nearest neighbors are calculated, where $k > 0$ is a predetermined integer. We may then adopt one of the following strategies to consider anomalousness of ***x***:

- A majority $\left(\left\lceil\frac{k}{2}\right\rceil\right)$ of these distances exceed a predefined threshold;
- The mean $\left(\frac{1}{k}\sum_i d_i\right)$ of these distances exceeds a predefined threshold;
- The smallest ($min_i\ d_i$) of these distances exceeds a predefined threshold.

When the value of k is very small (e.g., $k = 1$), these approaches may fail to detect anomalies if there is a small cluster of anomalous points that are near each other but very far from a vast majority of other points in the data set. At the other extreme, when *k* is too high, non-anomalous points lying in a small cluster may be inaccurately tagged as anomalous.

Instead of a yes/no anomaly decision at this stage, if the goal is to compute an anomaly score, then the latter should be a quantity that increases with one of the distance measures (mentioned above) e.g.,

$$\alpha(\boldsymbol{x}) = \frac{\boldsymbol{d}(x)}{\boldsymbol{d}_{max}}.$$

which lies between 0 and 1 whenever an upper bound $\boldsymbol{d}_{max}$ is known for $\boldsymbol{d}$. When a data set grows with time, *k*-Nearest Neighbor approaches require that the effect of each new data point on (the neighborhoods of) all other data points must be determined. But no global or local properties of subsets of data are retained or updated, unlike clustering approaches that must retain information such as cluster centroid locations.

3.1 Clustering

A large number of unsupervised anomaly detection algorithms rely on clustering the given data set, identifying points that are outside (or at the boundaries of) clusters, and computing their anomaly scores.

The most well-known clustering algorithm is the k-means algorithm. The essence of the algorithm is that *k* seeds (data points) are randomly chosen to be the candidate cluster centroids, other points are allocated to the clusters associated with these centroids based on a minimum distance-to-centroid criterion, then the centroids are updated by arithmetic averaging within each cluster. The last two steps (allocation of points and updating centroids) are then repeated until results converge, i.e., remain almost unchanged.

The k-means algorithm dominates applications due to its simplicity, despite known problems, e.g., the results of the algorithm are not guaranteed to be optimal, and presenting the same data set in a different sequence may result in a different set of clusters. This algorithm also implicitly assumes that clusters are radially symmetric, which may not be appropriate.

X-means [59] is a variant of k-means; an algorithm which is advantageous over basic k-means algorithm to automatically determine the number of clusters. Chang et al. [19] used X-means algorithm to analyze the behavior changes of online auction fraudsters in Yahoo! Taiwan. Several other variations have also been suggested in the literature.

Issa and Vasarhelyi [42] proposed an anomaly detection method, based on k-means, to identify fraudulent refunds. Thiprungsri et al. [69] applied it to detect fraudulent life insurance claims by identifying small clusters with large beneficiary payment, huge interest amount and long processing. Nhien et al. [48] applied k-means clustering to study anti-money laundering detection.

An important concern is that the anomaly detection context justifies the non-allocation of some points to any cluster. This necessitates a modification to clustering algorithms, where a predetermined distance threshold may be established in order to consider a point to belong to a cluster, so that no cluster is allowed to contain "faraway" points. Variations of the Adaptive Resonance Theory approach [14] allocate such points to new clusters, not requiring a fixed number k of clusters; in the last step, points in very small clusters (with very few points) may be considered to be anomalies.

Asymmetrical cluster formation may be permitted by using a nearest-neighbor clustering approach instead of k-means: a point is then added to a cluster which contains its nearest neighbor, instead of the cluster with the nearest cluster centroid. Extending this approach, k-nearest-neighbor clustering allocates a data point to the cluster containing the largest number of its *k*-nearest neighbors that have been allocated already to clusters. As mentioned earlier, a distance threshold may be used to avoid forcing a data point into a distant cluster.

Another important class of clustering algorithms is *hierarchical*; they construct\dendrograms (trees) that preserve much more of the relative distance relationships than algorithms such as k-means. In the top-down approach, we begin with the entire set of data, and successively split it into two subsets (at each level of the tree) that are maximally distant from each other. This process is continued until subset sizes reach a pre-specified threshold (e.g., a threshold of 1 implies that each leaf node of the tree contains one data point). Alternatively, hierarchies may be constructed bottom-up, starting with singleton sets and successively merging sets that are nearest to each other. When new data arrives over time, the dendrogram may change substantially. For anomaly detection purposes, points whose distances are largest from others are of most interest, i.e., points which are separated early in the top-down approach, or added very late in the bottom-up approach. An anomaly score may hence be computed based on the stage at which a data point is removed (for top-down) or added (for bottom-up) with respect to the current tree.

Commonly used hierarchical clustering algorithms include BIRCH [57], CURE [58], and ROCK [59]. Although the computational complexity of BIRCH is low, it is sensitive to the data ordering and lacks the ability to handle combinations of numerical and categorical data. CURE [58] can recognize arbitrarily shaped clusters better than BIRCH but has higher computational complexity. ROCK can handle categorical data but its computational complexity is higher than CURE.

Rui et al. [69] used a combination of BIRCH [57] and k-means [34] to identify money laundering activities. Panigrahi et al. [70] used DBSCAN [30], a density based clustering algorithm, for credit card fraud detection.

Neural network models called self-organizing [feature] maps (SOM) [49, 50, 7] have also been used for partitioning data space into Voronoi regions, where each node in the SOM is analogous to a cluster centroid in the k-means algorithm, and each region includes the data points that are nearest to that node. Another important aspect of SOMs is that the network nodes are pre-configured into a topological structure (e.g., a two-dimensional mesh) that determines a neighborhood relation different from the distance relation between data points. Node positions are iteratively updated, focusing on the nodes nearest to each data point being presented, as well as their topological neighbors that are also adapted, though to a smaller degree. Variations of this approach, such as Growing Cell Structures [31], permit the number of nodes in the network to increase or decrease in successive iterations, thereby capturing data sets in which clusters appear to be clearly separated, unlike SOMs in which the predefined topological map (nodes with their neighborhood relations) do not change.

Deng et al. [24] proposed a clustering model VKSOM combining Self Organizing Map and k-means clustering for fraud detection in financial statements. The model enjoys the benefit of unsupervised self-learning SOM [55]; the k-means clustering algorithm is applied to the results of SOM.

When the density of data distribution varies in different parts of the data space, then the anomaly detection algorithms that rely on clustering may not perform well; a data point may be far away from its neighbors from the perspective of the high-density region, but not from the perspective of the low density region. This issue has been addressed by algorithms such as LOF [12] and its variants COF [68] and INFLO [43], which compute an anomaly score for a point ***x*** that is sensitive to the density of the neighborhood of ***x***, e.g., by appropriately normalizing the distance measure. The relevant question is the following: how far is a point from its neighbors, when compared to the distances of those neighbors from their own neighbors? Of course, the answer is not obvious if a point is “in between” a high density and a low density region, so that the nearest neighbors of the point differ in their local density.

Instead of directly addressing density using inter-point distance, rank-based anomaly detection algorithms such as RBDA [40] formulate an approach drawn from sociological considerations in identifying loners and socially marginalized people: *Do the friends of* ***x*** *consider* ***x*** *to be their friend?* For anomaly detection, this question is posed in terms of a relative ranking of nearest neighbors, e.g., the third nearest neighbor of ***x*** has rank 3 with respect to ***x.*** Given $k>0$, if the k nearest neighbors of point ***x*** are in set S, then we can compute the sum of the ranks of ***x*** with respect to each element of S. A true outlier would have a relatively poor rank (i.e., high value) with respect to most of its nearest neighbors, even if it is in the immediate neighborhood of one or two other outliers.

3.2 Learning with available class information

When some information about normal or anomalous data is available, anomaly detection has been modeled as follows:

1. Two-class classification (supervised learning) – Two classes of data, normal and abnormal, are assumed to be available. Almost by definition, there would be very few representatives of the anomaly class, which makes learning difficult. To mitigate this class imbalance, a cost-sensitive learning algorithm may be applied that assigns a much higher cost for misclassifying an anomalous observation than for misclassifying a normal observation [28]. However, determining these costs is not a trivial task and often demands an expert’s knowledge.
2. When representatives of only the normal class are given, an algorithm must be trained to recognize them; any points that cannot be assigned to the normal class are considered to be anomalous [8]. This is also known as one-class classification problem. This approach does not need a prior cost specification, and is hence of great interest.
3. When unlabeled data is utilized to complement labeled data and both types of data are used for training a predictive algorithm; the learning is known as semi-supervised learning, as discussed by Zhu [74] and Elkan and Noto [27].

In supervised anomaly detection, labeled data is available based on past history, and this can be advantageously used for anomaly detection. The simplest approach would be to evaluate a new data point with respect to its distances from the labeled elements of “normal” and “abnormal” classes. Points that are near the labeled normal data are considered to be normal, and those near the labeled abnormal data are considered anomalous. As before, the newly assigned label to the data point in question may depend on a majority of labels from among *k* nearest neighbors. Several additional tools are also available to address this problem when viewed as a two-class problem, e.g., Support Vector Machines (SVMs) [11, 38] and Feedforward neural networks [55] or Radial-basis-function (RBF) Neural Networks [33, 57, 26] can be applied to such data.

As per empirical findings, different anomaly detection algorithms work well for different data sets. This suggests the use of ensemble as well as multi-objective approaches. Ensemble methods apply multiple individual anomaly detection algorithms, and combine the anomaly scores obtained by each; more details are given in a later section. Multi-objective methods instead obtain a collection of data points that constitute the non-dominated elements, from the anomaly scores obtained using different component algorithms. In this context, a data point $\boldsymbol{x_1}$ weakly dominates $\ $\boldsymbol{x_2}$ if $\boldsymbol{x_2}$ is not considered more anomalous than $\boldsymbol{x_1}$ by any component algorithm; also, $\boldsymbol{x_1}$ strongly dominates $\boldsymbol{x_2}$ if $\boldsymbol{x_1}$ weakly dominates $\boldsymbol{x_2}$ but $\boldsymbol{x_2}$ does not weakly dominate $\boldsymbol{x_1}$. Among the most successful multi-objective algorithms are evolutionary algorithms such as NSGA-II [23]; the end result could, however, be a large set of non-dominated data points, and the selection of a smaller subset may involve the application of other heuristics, such as a preference for data points (from the non-dominated set) that are considered anomalous by a majority of the component algorithms.

4 Anomaly detection with time-stamped data sequences

So far, we have discussed the problem of anomaly detection for sets of data points, with no implicit ordering among them. However, many practical problems involve data that arrive over time, and are hence in a strict temporal sequence; treating the data as a set (ignoring the time-stamp) loses information essential to the problem. Treating the time-stamp as just another dimension (on par with other relevant dimensions such as dollar amounts) can only confuse the matter. The occurrence of a set of (other) attribute values at a specific time instant can mean something quite different from the same attribute values occurring at another time, depending on the immediately preceding values; these dependencies require explicitly modeling time as a special aspect of the data, and treating the data as a sequence rather than a set. Hence, anomaly detection for time-sequenced data requires algorithms that are substantially different from those discussed in the preceding section. Outlier detection for temporal data has been extensively surveyed by [20, 36]. Several anomaly detection techniques have been proposed for symbolic data sequences, and a recent survey is provided by Chandola, et al. [18].

Some researchers use model-based methods to find abnormal sub-sequences. First a model is generated to predict the behavior of the time series. Using this model, the predicted values are calculated and compared with the observed values. The cumulative score of the differences is defined as the anomaly score of each observed data object. These models include Regression, Auto-Regression [32], ARMA [56], ARIMA [56], and Support Vector Regression [67, 52]. These methods were mainly designed for individual outlier detection, not for abnormal subsequence detection, and the results are impacted significantly by model parameters or distributions of data sets. These deficiencies limit their practical application.

Applications of time series anomaly detection have emerged in many fields. For financial data, example time series are shown in Figure 1, representing the stock prices of the oil and gas companies for 2010 to 2012 in which the behavior of one series (dashed red line) is different from the rest.

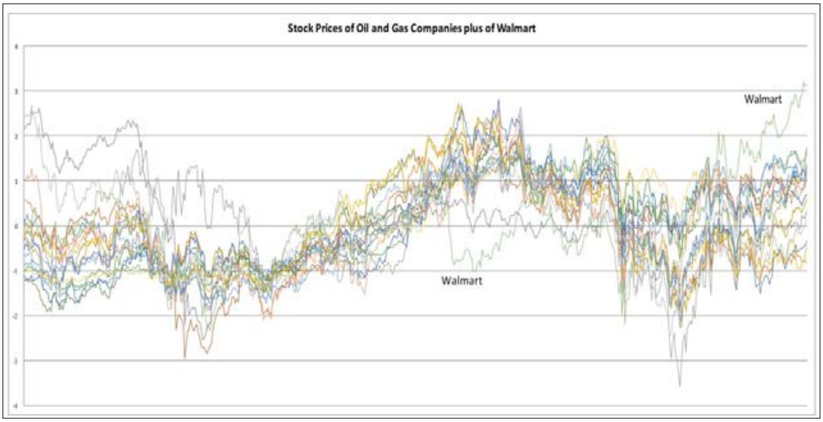


Fig. 1 Stock prices for some oil and gas companies and an outlier series. The anomalous series is stock price of Walmart (see right top corner, series in light green color).

Anomaly detection algorithms for data sequences fall into two categories, *viz.*, Procedural and Transformation approaches, as discussed below:

- In procedural algorithms, such as Regression and Hidden Markov Models (HMMs), a parametric model is built using the training data and an anomaly score is assigned to a test time series with associate probability.
- In the transformation approach, the data is transformed prior to anomaly detection. Transformations can be categorized into three major types:
    1. Aggregation approach focuses on dimensionality reduction by aggregating consecutive values.
    2. Discretization approach converts numerical attributes into a small number of discrete values, in order to utilize symbolic sequence anomaly detection algorithms and to reduce computational effort.
    3. Signal processing based transformations (e.g., Fourier and Wavelets) transform the data to a different space, and reduce the dimensionality of the data.

4.1 Finding anomalies within a sequence

Several kinds of anomalies have been categorized in the context of data sequences (over time). We first consider anomalies observed within a single (possibly multivariate) time series:

- The first is the occurrence of an *event* or *point anomaly*, characterized by a substantial variation in the value of a data point from preceding data points. For example, in the credit card fraud detection context, if a purchase is associated with a much higher cost than prior purchases by the same customer, an anomaly would be signaled. Sometimes, additional problem dimensions and refined analysis are important to avoid false negatives, e.g., the purchase over the Internet of a shirt for a hundred dollars may be flagged as anomalous even though the customer has made other purchases of higher cost; to characterize this transaction as an anomaly, prior data needs to be analyzed regarding the cost of prior purchases by this customer of clothing items over the Internet.

- Sometimes, it is important to detect anomalous sub-sequences within a given time series called *discords*, defined by Keogh et. al [47, 48] as "the subsequences of a longer time series that are maximally different from the rest of the sequence". The problem is difficult because the exact length of the anomalous subsequence may be unknown. For example, in detecting cardiac arrhythmias from an electrocardiograph (ECG), no individual reading may be out of range, but the sequence of a collection of successive values may not be consistent with the regular and periodic form of the data. We expect a certain (previously observed) regular waveform to be repeated, and variation from this expectation constitutes an anomaly. Another example consists of overseas money transfers; past history might indicate these occur about once a month for a customer (e.g., to financially support the customer's family living abroad), representing a simple periodic sequence over time, but the occurrence of multiple such transactions within a month then constitutes an anomaly. Keogh *et al.* [48] proposed a method to convert a series to multiple shorter subsequences, and used suffix trees to create an index structure for discretized sub-sequences; anomaly scores were computed by comparing trees generated by each subsequence. They also proposed an optimized dynamic time warping and SAX representation for time series [46, 47]. Wei *et al.* [71] proposed a time series BITMAP method based on comparison between the frequencies of SAX words of current data and past data.
- Sometimes the individual values may be within an acceptable range, but the rate of change over time may be anomalous, and we refer to this as a *rate anomaly*. For instance, in the credit card fraud detection problem, the balance (total amount owed by the customer) may suddenly increase within a day, due to a large number of small transactions made during a short amount of time; this should signal an anomaly even if each individual transaction is unremarkable, and the total balance remains within the range of prior monthly expenditures.
- Rate anomalies may be generalized to *contextual anomalies*, wherein a data point is anomalous with respect to the immediately preceding values, though not with respect to the entire range of possible values from past history. For instance, a customer's purchase transaction may be from a different country or city than the same customer's immediately preceding transaction that occurred a short while ago; this would be inconsistent with the expectation that the customer could not have suddenly travelled a large distance within a short period of time. Some false positives could be prevented by paying attention to other features of the context, e.g., if the prior purchase was from an airport, then there is a higher probability that the next purchase could be from a different geographical location.

- Other aspects of a data sequence may be considered anomalous because of multiple attribute values, requiring prior analysis of normative patterns that adequately describe the data sequence. For example, a customer's history may indicate that certain kinds of purchases are made singly and not in groups of successive transactions, so that multiple successive clothing purchases over the Internet may be considered anomalous, even though each such purchase is normal. Another customer's prior history, on the other hand, may indicate that it is normal for that customer to make 1-5 successive clothing purchases in the same day, so the occurrence of anomalies for that customer need to be determined differently.

Detection of an abnormal sub-sequence can be recast as the problem of comparing each sub-sequence of a given length *w* to other sub-sequences (also of length *w*) in the time series of length *n*. More precisely: given a time series *X*, the set of extracted sub-sequences, $X_w = \{X_{p,w}\ ;\ 1 \leq p \leq (n-w+1)\}$, consists of $X_{1,w} = \{\,x(1), x(2), \dots \dots .. x(w)\};\quad X_{2,w} = \{\,x(2), x(3), \dots \dots .. x(w+1)\}; \dots \dots .. \dots;\ X_{n-w+1,w} = \{\,x(n-w+1), x(n-w+2), \dots \dots .. x(n)\}$. One example is shown in Figure 2. We must determine if any $X_{i,w}$ is substantially different from the others, e.g., by evaluating whether its average distance to its *k* nearest neighbors is much larger than is the average distance for other sub-sequences.

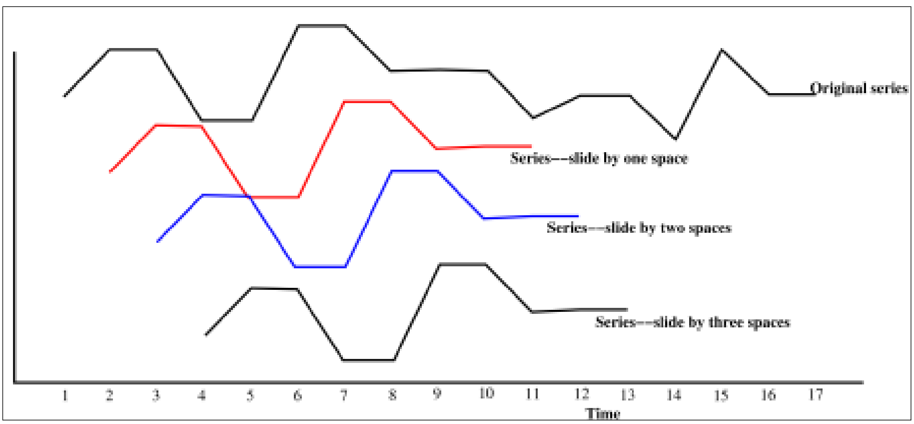


Fig. 2 Illustration for sliding window concept, with *w=10*

We observe that any sequence $X_{i,w}$ is substantially similar to $X_{i+1,w}$, since consecutive series share most elements, and are likely to be the nearest neighbors with only a small distance between them. This *self-match* problem makes it difficult to find anomalous sub-sequences, addressed by comparing a sub-sequence only to others with no overlap, as in Keogh, *et al.* [46]. Hence, we consider the set of series in X_w, finding the *k* nearest neighbor subsequences that do no overlap with each sub-sequence $X_{p,w} \in X_w$. Several methods have been proposed to improve computational efficiency; e.g., to find the nearest neighbor of a subsequence [63] in X_w, we may use Euclidean distance and instead of *k>1* nearest neighbors, we may use only the nearest neighbor, i.e., *k=1*.

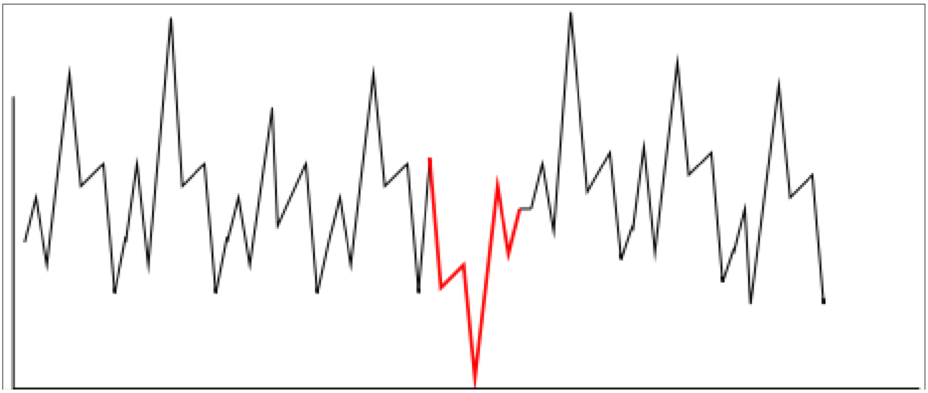


Fig. 3 Sub-sequence frequency – Red bold line represents a sub-sequence considered abnormal since it appears only once in this series, whereas other sub-sequence patterns occur much more frequently

Another approach, called *SAX words based frequency ratio (SAXFR)*, is based on assuming that the occurrence frequencies of abnormal sub-sequences are far lower than the frequencies of normal sub-sequences if length w is properly chosen. An example is shown in Figure 3. If we calculate the ratio between frequencies of SAX words generated from an abnormal sub-sequence with frequencies of these words in the entire series, then the ratio computed for an abnormal sub-sequence is much higher than the ratio computed for a normal series. The anomalousness measure is then defined to be α_{SAXFR} (X) = average of the reciprocal of the frequencies of SAX words contained in X.

For example, if three SAX words *abc, aac, abd* are generated from an abnormal subsequence, and these words appear *2, 3, 2* times respectively within the entire series, then the corresponding reciprocals are *1/2, 1/3, 1/2*. The average value can be calculated for such reciprocal frequencies, in this case, it is α_{SAXFR} (X)= *(0.50 + 0.33 + 0.50) / 3 = 0.44*. For another sub-sequence each of whose SAX words appear 10 times, this average would instead be 0.1, a much lower value, indicating that the sub-sequence is not anomalous.

To ensure that a SAX word is properly set to capture the outlierness of sub-sequence, the window size of SAX is set to *(w/2)*. If the window size of each word in SAX is too short, then the shorter sub-sequence represented by each SAX word may not be anomalous. If window size is too long, then the number of SAX words obtained in each sub-sequence (of length w) is less, and this may impact the results of the SAXFR approach.

The advantages of SAXFR are:

- Since the ratio is compared between a sub-sequence and the entire series, there is no need for nearest neighbor computation, nor for additional parameters.
- Frequencies can be pre-computed over the entire sequence, and sub-sequence SAX frequencies can be updated incrementally, so the computation is very fast.

In many algorithms, it may be difficult for users to find the right choices for parameter values. Another problem is that most algorithms focus on a single outlierness measure. Huang *et al*. [41] combine different measures, proposing an algorithm that requires little domain knowledge and can detect multiple abnormal sub-sequences in one run;

it also requires fewer parameters from users. Huang et al. proposed to use three measures – (1) SAXBAG, (2) STREND, and (3) DIFFSTD, described in Section 4.2, in the Multiple Measure Based Abnormal Subsequence Detection (MUASD) algorithm.

4.2 Finding anomalies between sequences

Another class of anomalies involves inter-time series anomalies, where multiple time sequences are available, of which a small number exhibit substantially different behavior. In other words, we are interested in determining whether the behavior of an individual time series differs substantially from most of those in a set of time series. For example, we may like to determine whether the sequence of overseas transactions being made by a customer is unusually high compared to other customers.

In many cases, the entire set of time series may be very large, with very little cohesion, so that it is difficult to identify anomalies with respect to the entire set. However, it may be possible to group the time series into clusters within each of which there is substantial similarity, so that a time series would be compared to those in the most similar subset. The grouping may be either on the basis of data attributes, or similarity between time series. For example, the number of overseas transactions made by a customer may be much higher than those made by most others with the same balance amount or income; this would be anomalous, even though it is normal for many other (wealthier) customers to make similar numbers of overseas transactions in the same period of time.

In the aforementioned cases, we encounter several problems. One of the most critical among them is the determination of an appropriate distance measure; it is not easy to determine the best similarity or distance measures that can be used for different types of time series. Euclidean distance is highly sensitive to outliers and also cannot be used when the time series are of different lengths. Another concern is the presence of random noise, naturally inherent in many time series; distinguishing noise from anomalies is a challenging task. Additional complications arise due to practical considerations, e.g., the length of the series may be very large and, consequently, the computational complexity of the proposed algorithm is prohibitively large.

Suppose we are given a time series data set $D = \{ x_i(t) \mid 1 \leq t \leq n; i = 1, 2, \ldots\ldots, m\}$, where $x_i(t)$ represents the data object of the i^{th} time series at time t, n is the length of the time series, and m is the number of time series in the dataset. The goal is to calculate $O(x_i)$, the outlierness of a series x_i, and $O_{threshold}$ a threshold such that if x_i is an anomalous series, then $O(x_i) \geq O_{threshold}$; *otherwise* $O(x_i) < O_{threshold}$.

The first step is to obtain a compact representation to capture important information contained in the series. Three main categories of approaches have been identified, by Ding *et al.* [25] and Chandola *et al.* [17], to reduce the dimensionality of time series: *model based*, *data-adaptive*, and *non-data adaptive* approaches. The compact representations obtained using these approaches must then be compared using suitable distance measures, which have been categorized into three groups [25]:

- The lock-step measures are based on one-to-one mapping between two-time series. Examples include the Cross-Euclidean distance (EUC) [29], the Cross-Correlation Coefficient-based measure, defined as

$\sqrt[2]{2(\,1 - corrcoef(d_x - \,d_y))}$ [10], SameTrend (STREND), and standard deviation of differences (DIFFSTD).

The elastic measures are based on one to many mapping between two time series, e.g., Dynamic Time Warping (DTW) [9, 44], and Edit Distance on Real sequence (EDR) [21]. For example, the similarity measure used by DTW addresses differences in two series, possibly due to acceleration or deceleration during the course of measurements, compounded by time lags caused by real-life processes such as the sensor measurements upstream and downstream in a manufacturing process. A time series *X* may be very similar to the series *Y*, except that they are out of sync in some places in the time domain. DTW attempts to accommodate such discrepancies in *X* and *Y* series. Consider $C(t, l) = |\,x(t) - y(l)|$ which represents the cost of aligning *x(t)* with *y(l)*. Then, the goal of DTW is to

$$minimize \sum\sum C(t,l).$$

This can be accomplished by using the classical dynamic programming algorithm developed to align two sequences.

- A time series can be transformed into another space; measures in the transformed space include TQuEST [1] distance, and Spatial Assembling Distance SpADe [22]. Other examples include Symbolic Aggregate approXimation (SAX) proposed by Keogh and Lin [46] with and without sliding window (SAXSL and SAXNW respectively); SAX with bag-of-pattern (SAXBAG) [51], Discrete Wavelet Transform [16], and Discrete Fourier Transform [29].

Many variations on these approaches have been proposed in recent years. For example, Wei *et al*. [71] proposed an algorithm to find unusual shapes by using SAX representation to speed up the process for selecting the best candidates; Lin *et al.* proposed using the distance between *bag-of-pattern* [51] based on the frequency of each unique SAX word, which can be applied for anomalous time series detection, Keogh *et al.* [44] suggested indexing techniques for dynamic time warping by using a lower bounding measure, which can be used for detecting distance measures of time series, Protopapas [62] proposed an outlierness measure based on the average of correlations between time series. Chan [16] showed that Euclidean distance in the Haar-wavelet transformed domain can be effective for finding time series matches, and reported that this approach outperformed the Discrete Fourier Transformation proposed by Faloutsos [29], and Bu *et al.* [13] applied Haar wavelet transform on time series and then built an augmented trie to find the top *k* discords in the time series database.

In Table 1, we summarize the advantages and disadvantages associated with using some of the distance measures. We observe that no single measure is capable of capturing different types of perturbations that may make a series anomalous, hence multiple measures should be considered together. For the best utilization of limited computational resources, we may select measures that are orthogonal to each other, to minimize redundancy. One approach minimizes redundancy by selecting three measures that are least correlated with each other: (a) SAXBAG, (b) Same Trend (STREND), and (c) Standard deviation of differences between two time series

(DIFFSTD); described below. These measures capture different aspects of the time series, and a combination of these gives a comprehensive measure of how isolated is a time series from others in the comparison set. SAXBAG captures behavior of a time series using a histogram of possible patterns, STREND identifies the degree of synchronization of a series compared with another series, and DIFFSTD measures the amount of deviation, as illustrated in Figure 4. Combining these three metrics produces a comprehensive and balanced distance measure that is more sensitive than individual measures.

- SAXBAG (proposed by Lin *et al.* [51]: Given a time series, a sub-sequence of size l is obtained using a sliding window. Each sub-sequence is reduced to w-dimensional discrete representation in two steps. First, the sub-sequence is further divided into $u = \frac{l}{w}$ equal size sub-subsequences and the average value of the data falling within a sub-subsequence is evaluated. In the second step, each average is discretized to a symbol, from a predetermined set of symbols, using equal probability intervals approach. Figure 4(a) illustrates these concepts using three symbols: *a, b, and c.*
- STREND: For each i, we calculate the difference $x_t' = x_i(t+1) - x_i(t), t \in [1 \ldots\ldots n-1]$ and define:

$$S_{i,j}(t) = \begin{cases} 1 & if\ x_i'\ (t).x_j'\ (t) > 0 \\ -1 & if\ x_i'\ (t).x_j'\ (t) < 0 \\ 0 & otherwise \end{cases}$$

  Clearly, $S_{i,j}(t)$ indicates whether or not $x_i(t)$ and $x_j(t)$ change in the same direction at time t. The aggregate measure, over the entire length, n, of the time series is evaluated as,

$$dist\ (i,j) = 1 - \sum_{t\ \in [1 \ldots n-1]} S_{i,j}(t)/n - 1 \quad (1)$$

- DIFFSTD is the standard deviation of differences between two time series, i.e., if $\delta_{i,j}(t) = \|x_i(t) - x_j(t)\|$, and $\mu_{i,j} = \sum_t \frac{\delta_{i,j}(t)}{n}$, then the new distance is defined as $dist\left(x_i, x_j\right) = \sqrt{(\sum_t(\delta_{i,j}(t) - \mu_{i,j}\ (t))^2/n}$. This measure is widely used in *pairs trading* in the financial field, which monitors the performance of correlated securities.

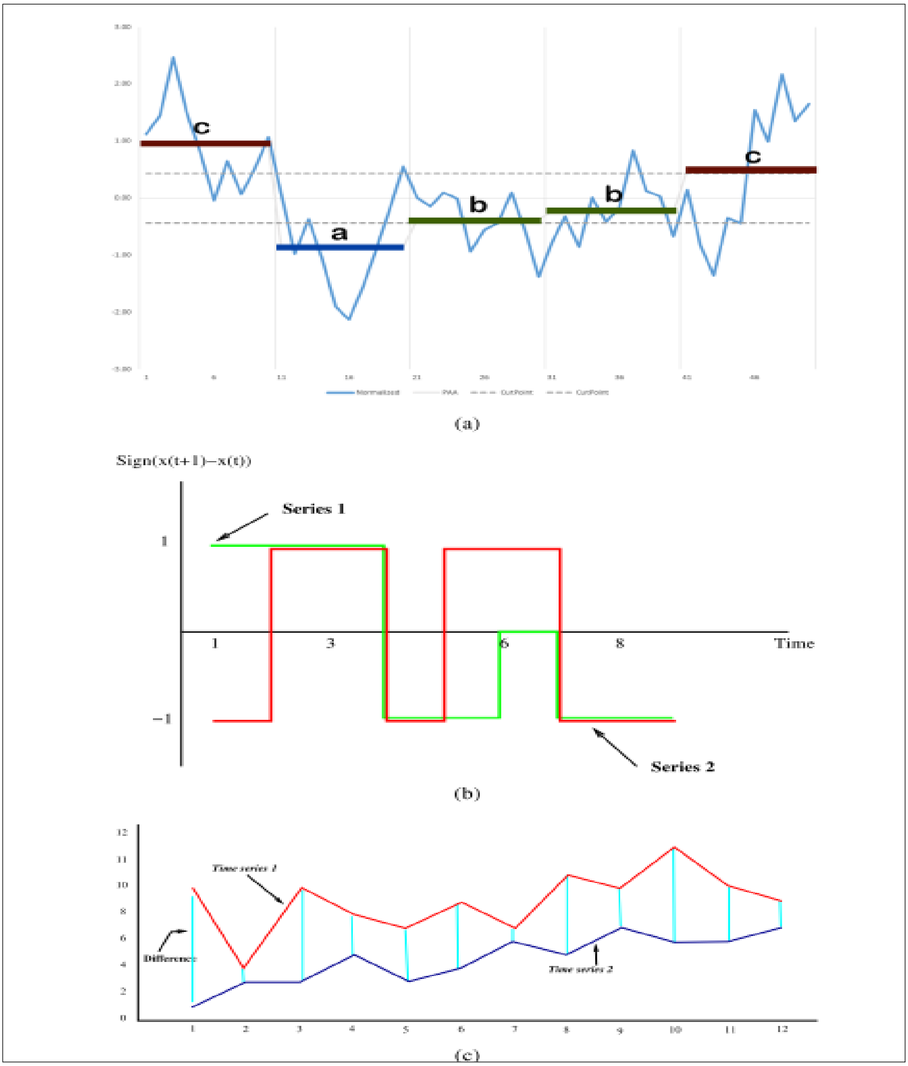


Fig. 4 Illustrations for three key measures (a) Illustrates how a SAX word, in a sliding window, is generated. SAXBAG counts the frequencies of such words in the word-sequence; (b) illustrates STREND (table under the figure shows $S_{i,j}(t)$'s); and (c) illustrates DIFFSTD (vertical lines show the differences between two time series x_1 *and* x_2).

Measures	**Pros**	**Cons**
EUC[29] CCF [10] DIFFSTD	Easy to implement; Computationally efficient	Lock-step measure; Normal series with time lagging cause problems
Dynamic Time Warping [44]	Elastic measure; Comparison with time lagging	Small deviations may not be detected
Discrete Fourier Transform [29] (DFT)	Good in detecting anomalies in frequency domain; Normal series with time lagging do not cause problem	Small deviations may not be detected; Cannot detect anomalies with time lagging
Discrete Wavelet Transform [16] (DWT)	Good in detecting anomalies in frequency domain	Small deviations may not be detected; Sensitive to time lagging
SAX with sliding window [46] (SAXSL)	Tolerates noise, as long as its standard deviation is small	May not detect abnormal subsequence of shorter length than feature window size; Normal series with time lagging can result in false positives
SAX without sliding window [46] (SAXNW)	Tolerates noise, as long as its standard deviation is small	May not detect abnormal subsequence of shorter length than feature window size; Small deviations may not be detected Normal series with time lagging can result in false positives
SAX with bag-of-pattern[51] (SAXBAG)	Tolerates noise, as long as its standard deviation is small; Normal series with time lagging do not cause problem	Cannot detect anomalies with time lagging; Cannot detect anomalous series with similar frequencies but different shapes

Best results were obtained by using the following refinements:

- Elements of each time series are first normalized to have mean 0 and standard deviation 1.
- Normalization was performed by dividing observations by trimmed mean (excluding 5% on either end).
- Higher weights are assigned to the more effective measures, determined using RBDA.
- Deciding whether a point is anomalous is finally performed by thresholding the combined anomaly score, or selecting the highest values.

Sometimes data arrives incrementally, and online anomaly detection algorithms are required. Some distance measures can be updated incrementally, e.g.,

- **DIFFST**: The variance of differences between series *i* and *j* at time *n* can be calculated as:

$$\text{dist}_f\ (i, j)= \frac{n \times ssq(i,j) - (sqs(i,j))^2}{n \times (n-1)}, \qquad (2)$$

where $ssq(i,j) = \sum_{t=1}^{n}(x_i(t) - x_j(t))^2$ and $sqs(i,j) = \sum_{t=1}^{n}(x_i(t) - x_j(t))$ (3)

The numerator in Equation 2 can be updated for the $(n + 1)$th observations by adding $(x_i(n + 1) - x_j(n + 1))^2$ and $(x_i(n + 1) - x_j(n + 1))$ to $ssq(i, j)$ and $sqs(i, j)$ respectively.

- **STrend**: Let $x_i'(n) = x_i(n) - x_i(n-1)$. Then, by definition,

$$S_{i,j}(n) = \begin{cases} 1 \ \ if \ x_i'(n).x_j'(n) > 0 \ \ or \ \ x_i'(n) = x_j'(n) = 0 \\ 0 \ otherwise \end{cases}$$

Consequently,

$$\text{dist}_t(i,j) = \frac{\sum_{t=2}^{n} S_{i,j}}{n-1} \quad (4)$$

Therefore, to update this value using the $(n + 1)$th observation, we modify the numerator by adding the last trend $S_{i,j}(n + 1)$, and accordingly modify the denominator as well.

- **SAXBAG**: converts the data segment in the sliding window of size w to a single SAX word, and then counts the frequency f_i of each word. When data at time $n + 1$ is observed, a new SAX word will be generated based on the sequence

  $x_i(n + 2 - w), x_i(n + 3 - w), x_i(n + 4 - w), \ldots\ldots\ldots, x_i(n + 1).$

  The stored data set can be updated to account for the new SAX word.

- **Normalization and Assignment of Weights:** We normalize $dist_f(i, j)$, $dist(i,j)$, and $dist_s(i,j)$ to $dist_f'(i,j)$, $dist_t'(i,j)$, and $dist_s'(i,j)$, respectively. Then $\text{weight}_l'(i)$ values are also calculated.

- Finally, the anomaly score for the i^{th} series is calculated as follows:

$$A_i = \sqrt{\sum_{l\in\{s,t,f\}}(dist_l'(i))^2 \times \text{weight}_l'(i)}.$$

A simple weighted distance can be used for fast online detection, without considering rank:

$$dist_{msm}(i,j) = \sqrt{\frac{\sum_{l\in\{s,t,f\}}(dist_l'(i,j))^2}{3}} \quad (5)$$

Then, the anomaly score for i^{th} series, $A_i(k)$, can defined as the average distance to its k nearest neighbors:

$$A_i(k) = \frac{\sum_{j\in N_k(i)} dist_{msm}(i,j)}{|N_k(i)|} \quad (6)$$

Now, select any k neighbors of i, and let $A'_i\ (k)$ denote the average $dist_{msm}(i, j)$ over them. Then the average distance of k-nearest-neighbors of i must be *less* or equal to average distance of any k neighbors of i, so:

$$A_i\ (k) \leq A'_i\ (k) \qquad (7)$$

Huang *et al*. [39] propose OMUDIM, a faster version of MUDIM, whose key steps are as follows:

1. Find a threshold λ such that $A_i(k) < \lambda$ implies i is not an anomalous series; then any $A'_j\ (k) < \lambda$ also implies j is not an anomalous series either; thus most of the non-anomalous series can be excluded from anomaly score calculations. To find an estimate of the threshold, λ, we apply the following:
2. Sampling procedure: We calculate $A_i(k)$'s for $i \in S$, where S contains a small fraction (α) of the elements in D, randomly selected. Then λ is chosen to equal the value of $A_i(k)$ which is at the top ($\beta \times 100$)th percentile in descending order.
3. For $x_i \in D - S$, maintain a binary max heap consisting of $dist_{msm}(i, j)$ where j's are selected k neighbors of i. If the average of these k neighbors is less than λ, series i is declared as non-anomalous. Else $dist_{msm}(i, j)$ is calculated for next selected value of j, and the heap is updated by keeping only the smallest k values of $dist_{msm}$. The anomalousness of series i is tested using the above criterion. This process stops if at any stage, the series is found to be non-anomalous or no j is left.
4. Calculate the anomaly scores of all potential anomalous series (found in Step 2) and find the anomalous series, if any.
5. The above steps are repeated once new observations arrive.

This algorithm outperformed three other online detection algorithms based on (a) Euclidean distance, (b) Dynamic Time Warping (DTW), and (c) Autoregressive (AR) approach, proposed by [17], [46] and [32] respectively. The first two of these methods calculate a measure of anomalousness of a time series by (i) finding the k nearest neighbors of the series, and (ii) using the average distance of these k neighbors. The third method constructs a global AR model for all series and then measures the anomaly score at time t as the gap between the observed value and the predicted value.

The MUDIM approach is efficient and detects anomalous series as soon as it begins to drift away from the other (non-anomalous) series, a substantial advantage over other anomaly detection algorithms for time series. This approach can handle data uncertainty very well, and its online version does not require any training data sets. Compared with other methods, it requires less domain knowledge.

5 Anomaly detection with categorical data

Categorical (or nominal) data pose special problems for the formulation of suitable anomaly detection algorithms. The biggest issue is that numerical distance measures cannot easily apply; the simplest distance measure for a single categorical attribute is binary-valued: the distance is 1 if the two values of an attribute are different, and 0 if identical. Transforming a *d*-valued categorical attribute into *d* binary attributes does not make any substantial difference to this problem.

Many practical problems involve such data, in which some of the variables are nominal or categorical, i.e., not numeric. For example, a credit application for a bank may state an occupation which is relevant to decision-making; numeric encoding of the same can only confuse data analysis procedures, since an occupation value of “7” may be as different from “8” as it is from “1”. Hence, anomaly detection algorithms that rely on distance computations must not attempt to reduce categorical variables to numerical values.

One approach for anomaly detection with categorical data is based on data mining approaches that focus on frequent item-set identification; the terminology comes from the grocery shopping analogy, in which shoppers make transactions, and each transaction involves the purchase of a certain set of items. Data mining analysis can help determine which sets of items tend to be purchased together, with actionable implications for the merchant or grocery store. Well-known algorithms [4, 3] have been developed to identify frequent item-sets from large numbers of transactions.

From the anomaly detection perspective, we are interested in transactions that contain infrequent item-sets, i.e., items that are not usually purchased together, and are often purchased separately. An anomaly score can readily be constructed, based on this heuristic, high if $f(\boldsymbol{x},\boldsymbol{y})/f(\boldsymbol{x})f(\boldsymbol{y})$ is small, where *f* indicates the frequency of purchase of that item [REF]. This principle also applies to other contexts in which some variables are categorical. Another perspective is to evaluate the normalized conditional (marginal) probabilities of item-sets.

Another well-founded approach is based on an information compression perspective, relying on a minimum description length (MDL) methodology [34]. First, each transaction can be considered to be covered by a collection of item-sets, of varying frequency in the set of transactions. A coding approach is utilized which assigns codewords (symbol sequences) to various item-sets, with frequent item-sets represented as the shortest codewords, as in the KRIMP algorithm [65]. A transaction that consists only of very frequent item-sets will hence have a much shorter codeword representation than a transaction that consists of randomly assorted items. Thus, the anomaly score of a transaction can be considered to be proportional to the length of its representation using item-set codewords. A small number of items missing from the item-sets can be tolerated; Smets and Vreeken propose a – fault-tolerant cover computation algorithm, which attempts to use as few item-sets as possible to describe a transaction [66].

For example, consider a database with three categorical variables and six transactions (data points in the three-dimensional space): *abx, abx, abx, abx, acx, acy*, where each letter {*a, b, c, x, y*} indicates a different value for one of the three categorical variables. The code assigned to the most frequent item-set, *abx*, is 0, of length 1 bit. The code assigned to the next most frequent item-set, *ac,* is 10, of length 2 bits. The code assigned to *x* is 110, of length 3 bits, and the code assigned to *y* is 111, also of length 3 bits. Now, we can determine the length of the minimal cover for a given transaction using these codes. For example, *acx* requires 2 bits (for *ac*) + 3 bits (for *x*) = 5 bits, indicating that it is far more anomalous than the transaction *abx* which requires only 1 bit.

This approach was further extended in the parameter-free CompreX algorithm [6] that uses multiple dictionaries (compression tables with codewords), and identifies transactions with high compression cost as anomalies.

6 Ensemble methods for anomaly detection

Several anomaly detection algorithms have been developed, and have been fine-tuned to work well in specific data distributions. Ensemble approaches attempt to combine multiple individual algorithms in order to obtain better results over a larger class of problems and data distributions. The two main classes of ensemble approaches [2] are:

1. Sequential: The results obtained by one algorithm are refined by the application of another algorithm.

2. Parallel (or Independent): Each algorithm is applied to the data set, and the results obtained using multiple algorithms are combined. The MUDIM approach discussed in the previous section is an example of this approach, for time series problems.

These approaches work best if different algorithms being utilized are orthogonal to each other, e.g., one catches the anomalies that the other does not.

6.1 Sequential ensembles

The principle of successive refinement is applied by setting algorithm parameters to permit the first algorithm to act as a "coarse sieve" that minimizes false negatives but permits a greater number of false positives, whereas each successive algorithm reduces the number of false positives. Zhiruo *et al.* [72] have developed new sequential ensemble algorithms based on the concept that anomaly detection performs better on a subsample of the dataset, and propose to use the algorithms which have higher diversity among themselves as the base algorithms for a better combination result. Iterative sequential learning over base algorithms processes data in multiple passes, with each round of execution providing a better understanding of the dataset. As a result, the final result will provide a more refined result than obtainable using a single iteration.

Boosting is a well-known example of iterative sequential learning. In particular, the *AdaBoost* algorithm [30] has gained substantial popularity for classification problems. But this approach has not been explored much in unsupervised anomaly detection, since labeled training data are unavailable. Zhiruo *et al*. [72] propose a novel adaptive learning algorithm for unsupervised outlier detection, which uses the score output of the base algorithm to determine the hard-to-detect examples, and iteratively resamples more points.

Random forests constitute ensemble decision-making with multiple randomly generated decision trees, improving performance over single decision trees and other approaches for classification problems. Guha *et al.*[35] have shown that random trees can be used for anomaly detection in a semi-supervised context. However, parameters must be chosen based on empirical learning, and performance suffers for when the problem dimensionality increases. Zhiruo *et al*. [73] have analyzed the impact of parameters used in random trees from both empirical and theoretical points of view, and proposed new algorithms to solve the problem of anomaly detection over high-dimensional data, partitioning the feature space into similar clusters, then building random trees separately.

Future work must address how to extend ensemble methods to streaming data, especially in the presence of concept drift. Since high computational costs are required for iterative computations, it is also important to develop algorithms that reduce computational effort.

6.2 Independent ensembles

The literature on information fusion, e.g., from multiple heterogeneous sensors [70, 15], provides a contrast between data-level, feature-level, and decision-level fusion, providing useful analogies for ensemble methods for anomaly detection.

- Data-level: A single algorithm is applied to data obtained from multiple sensors. This is not an ensemble approach.
- Feature-level: Preliminary analysis identifies important features that are then analyzed together, in the data fusion context. Analogously, we may consider the anomaly scores, obtained from different individual anomaly detection algorithms, as features that need to be combined by averaging the scores, or considering the maximum of the scores, which can then be ranked.
- Decision-level: The final results from various algorithms are combined, in this approach. For anomaly detection, the *ranks* obtained for each point using different anomaly detection algorithms may be considered to be their final decisions; formally, if $\alpha_i(x_j)$ is the anomaly score for the data point x_j using the i^{th} algorithm (applied to the data set D), then the corresponding *rank* is defined as follows:

$$r_i(x_j) = |D| - \left|\{x_k | \alpha_i(x_k) < \alpha_i(x_j)\}\right|.$$

For example, if the data set contains a thousand elements, and all their anomaly scores are different, then the element ranked 1 is the one such that 999 elements have lower anomaly scores, the element ranked 2 is the one such that 998 elements have lower anomaly scores, etc. This definition implies that the higher value is used if elements have the same rank, e.g., if two elements have the same highest anomaly score, then both would have to be ranked 2. In the decision-level approach, these ranks have to be combined; this may be done using one of the following approaches:

1. Min-rank, defined as $r_{min}(x_j) = \min_i r_i(x_j)$. Thus, all the items ranked 1 by any algorithm receive the final min-rank of 1.
2. Aggregate-rank, defined as $r_{sum}(x_j) = \sum_i r_i(x_j)$. This is equivalent to an averaging of the ranks.

Each of these may be sorted, resulting in a linear ordering, from which a final rank may be extracted; we use ρ to denote such a function, so that $(x_i) \in \{1 \ldots \ldots . |D|\}$ when the data set is D, and $(x_i) > (x_j)$ iff either $\alpha(x_i) < \alpha(x_j)$ for a single or composite anomaly score function , or $r(x_i) > r(x_j)$ if r is a single or composite rank function.

Example: Consider a dataset with a thousand points $D = \{x_1, \ldots, x_{1000}\}$, to which three anomaly detection algorithms are applied with the following respective anomaly scores for some of the points as shown in Table 2; each column (other than the first) corresponds to the data points $\{x_1, \ldots, x_5\}$, and we assume that all $\alpha_i$values equal 0 for the remaining points in *D*, i.e., they are not considered to be anomalous even to the least degree by any algorithm. We observe from Table 2 that somewhat different results are obtained (shown in the ρ values) by different ensemble combination methods, although they usually agree in which elements are considered most (x_2)or (x_5) least anomalous.

7 Mathematical perspectives

Many algorithms for anomaly detection have been implemented, as discussed above. We now focus on simple mathematical formulations of the problems being addressed, providing a basis to evaluate competing algorithms that may not directly address a mathematical goal; some earlier definitions are repeated for clarity.

7.1 Prediction

Let *f* be a mathematical model for an unknown process that generates values for an unknown variable $\mathcal{Y}$, given a collection of known variables $\boldsymbol{x}$. Then $| f(\boldsymbol{x}) - \mathcal{Y} |$ measures the amount of deviation between the predicted and actual values of $\mathcal{Y}$.

Table 2 Example of anomaly scores and ranks for five points using three anomaly detection algorithms; the ρ values give the final results of applying the ensemble operations in various ways.

	x_1	x_2	x_3	x_4	x_5
α_1	0.9	0.8	0.7	0.6	0.5
α_2	0.9	0.9	0.9	0.4	0.4
α_3	0.2	0.9	0.5	0.7	0.3
α_{max}	0.9	0.9	0.9	0.7	0.5
$\rho_{\alpha max}$	3	3	3	4	5
α_{sum}	2.0	2.6	2.1	1.7	1.2
$\rho_{\alpha sum}$	3	1	2	4	5
r_1	1	2	3	4	5
r_2	3	3	3	5	5
r_3	5	1	3	2	4
r_{min}	1	1	3	2	4
ρ_{rmin}	2	2	4	3	5
r_{sum}	9	6	9	11	14
ρ_{rsum}	3	1	3	4	5

This leads to defining anomalies as those observed points ($\boldsymbol{x}, \mathcal{Y}$) for which | $f(\boldsymbol{x}) - \mathcal{Y}$| is maximized.

7.2 Classification

Let X_1 and X_0 be two sets of data points, of which the former are known to be "normal" and the latter are known to be anomalous.
A classifier C is first to be trained on this data set, minimizing the number of classification errors, i.e., minimizing

$$|\{\boldsymbol{x} : \boldsymbol{x} \in X_1 \text{ and } C(\boldsymbol{x}) = 0\}| + |\{\boldsymbol{x} : \boldsymbol{x} \in X_0 \text{ and } C(\boldsymbol{x}) = 1\}|$$

To facilitate learning using algorithms such as error back-propagation for feedforward neural networks, often it is preferred to minimize the mean squared error, proportional to,

$$\sum_{x_i \in |\mathrm{X}_1|} (1 - f(\boldsymbol{x}_\mathrm{i}))^2 + \sum_{x_i \in |\mathrm{X}_0|} (f(\boldsymbol{x}_\mathrm{i}))^2$$

where f describes the function implemented by the model after learning, presumed to be constrained such that $0 \le f(\boldsymbol{x}_\mathrm{i}) \le 1$ for all data points $\boldsymbol{x}_\mathrm{i} \in X_1 \cup X_0$ **.** To avoid penalizing points whose values are close to the target values (0 or 1), a minor variation of the above is the task of minimizing

$$\sum_{x_i \in |\mathrm{X}_1| \,\&\, f(x_\mathrm{i}) < 1 - \epsilon} (1 - \epsilon - f(\boldsymbol{x}_\mathrm{i}))^2 + \sum_{x_i \in |\mathrm{X}_0| \& f(x_\mathrm{i}) > \epsilon} (f(\boldsymbol{x}_\mathrm{i}) - \epsilon)^2$$

where $0 < \epsilon < 1$. Since many different models may be constructed for a given dataset, the trade-off between model complexity C_M and error magnitude is achieved by introducing a regularization term, i.e., minimizing,

$$E + \lambda C_M$$

where E is the (misclassification or mean squared or other) error term described above, C_M may be measured as the sum of the number of trainable parameters in the model or the sum of the magnitudes of the model parameter values, and the regularization parameter $\lambda > 0$ describes the permitted tradeoff.

7.3 Clustering

For the purposes of anomaly detection, the goal of clustering is to group data points together, which is quite different from that of partitioning the data space into disjoint regions. We need to address two distinct considerations, discussed below:

1. When does a data point belong to a cluster?
   The goal is to associate each data point with a cluster which is at the least distance. If an *a priori* decision is made to cluster the data set $\boldsymbol{D}$ into k clusters, then the goal is to determine a function C such that $C(\boldsymbol{x}_\mathrm{i}) \in \{1, \ldots, k\}$, minimizing

$$d(C, D) = \sum_{x_i, x_j : C(x_i)=C(x_j)} d(\boldsymbol{x}_i, \boldsymbol{x}_j).$$

However, k (the number of clusters) is not given to us in an anomaly detection task, requiring us to address the following question:

2. How many (k) clusters should we use?
   This question may be answered in many ways.

- A maximum threshold θ may be specified on the distance between a point and the cluster to which it is assigned; we must then determine a function C such that $C(x_i) \in \{1, 2, 3, \ldots\}$, minimizing $d(C, D)$, but subject to the constraint that $C(\boldsymbol{x}_i) = C(\boldsymbol{x}_j)$ implies $d(\boldsymbol{x}_i, \boldsymbol{x}_j) \leq \theta$

- A threshold may instead depend on the point $\boldsymbol{x}_i$ in question, e.g., the local density in a region of the data space containing $\boldsymbol{x}_i$, so that we replace θ in the previous constraint by a function of the distances between points in a neighborhood $N(\boldsymbol{x}_i)$ of $\boldsymbol{x}_i$, such as,

$$\theta(\boldsymbol{x}_i) = \eta \frac{\sum_{x_j \in N(x_i)} \sum_{x_k \in N(x_j)} \mathbf{d}(\mathbf{x}_j, \mathbf{x}_k)}{|N(\mathbf{x}_j)|\,|N(\mathbf{x}_i)|}$$

  where $0 < \eta < 1$. The definition of the neighborhood may consist of a fixed number of nearest points, e.g., such that $|N(\boldsymbol{x}_i)| = n$, a predetermined size, and $\boldsymbol{x}_j \in N(\boldsymbol{x}_i) \,\&\, \boldsymbol{x}_k \notin N(\boldsymbol{x}_i)$ implies $\boldsymbol{d}(\boldsymbol{x}_i, \boldsymbol{x}_k) \geq \boldsymbol{d}(\boldsymbol{x}_i, \boldsymbol{x}_j)$.

- Instead, we may impose a criterion based on inter-cluster distances, e.g., minimizing the ratio of the mean intra-cluster to the inter-cluster distance, i.e.,

$$R = \frac{\frac{1}{k}\sum_{i=1}^{k} \boldsymbol{d}(C_m)}{\frac{1}{k(k-1)}\sum_{i \neq j} \boldsymbol{d}(C_i, C_j)}$$

where the intra-cluster distance for the mth cluster is defined as

$$\boldsymbol{d}(C_m) = \frac{\sum_{x_i, x_j : C(x_i)=C(x_j)} \boldsymbol{d}(\boldsymbol{x}_i, \boldsymbol{x}_j)}{\left|\{\boldsymbol{x}_j : C(\boldsymbol{x}_j) = m\}\right|}$$

and the inter-cluster distance between the i^{th} and j^{th} clusters could be defined in different ways, e.g.,

$$\boldsymbol{d}(C_i, C_j) = \min_{x \in C_i, y \in C_j} \boldsymbol{d}(\boldsymbol{x}, \boldsymbol{y})$$

or

$$\boldsymbol{d}(C_i, C_j) = \frac{\sum_{x \in C_i, y \in C_j} \boldsymbol{d}(\boldsymbol{x}, \boldsymbol{y})}{|C_i| . |C_j|}$$

In an incremental approach, e.g., where clusters are successively merged, a greedy algorithm is usually followed, i.e., continuing to merge clusters until the ratio R begins to increase, or until R exceeds a pre-specified threshold.

7.4 Decision-theoretic optimization

Although minimization or maximization of some measure is inherent in the formulations given above (viz., prediction, classification, or clustering), we may directly pose the anomaly detection task as an optimization task, and address the same using a global optimization algorithm, obtaining potentially "better" solutions, with the caveat that we then expect greater computational expense. Each such formulation provides a standard against which the quality of results obtained by a specific algorithm can be evaluated.

We now consider the decision-theoretic task, where false positive cost $C^+(x_i)$ is associated with mislabeling a non-anomalous data point x_i, and false negative cost $C^-(x_i)$ is associated with mislabeling an anomalous data point x_i. The outputs of an anomaly detection algorithm are interpreted to suggest that only the points in $A \subset D$ are anomalous. Ground truth may be available suggesting that a point x_i is an anomaly with probability $p(x_i)$.

Then the goal is to minimize the total cost of misclassification for the algorithm, i.e.,

$$\sum_{x_i \in A} C^+(x_i)\left(1 - p(x_i)\right) + \sum_{x_i \in D \setminus A} C^-(x_i)\ p(x_i)$$

If the anomaly detection algorithm provides anomaly scores $\alpha(x_i)$ in the [0,1] interval, then the choice of a threshold θ (above which a point is to be considered an anomaly) is obtained by minimizing the total cost obtained by the choice of θ, i.e.,

$$\underset{\theta}{argmin} \sum_{x_i:\, \alpha(x_i) \geq \theta} C^+(x_i)\left(1 - p(x_i)\right) + \sum_{x_i:\, \alpha(x_i) < \theta} C^-(x_i)\ p(x_i)$$

If, on the other hand, the goal is to infer probabilities from the anomaly scores, then a calibration function f_a must be learned using the data points X for which such probabilities are available (from the ground truth), perhaps minimizing mean squared error $\sum_{x_i \in X}(f(\alpha(x_i)) - p(x_i))^2/|X|$ using a neural network or a support vector machine.

8 Concluding remarks

Anomaly detection problems arise in many applications and fields of study, and have been addressed by researchers using traditional statistical tools, data mining approaches, and problem-specific methodologies. Recent years have seen the development of many different algorithms for anomaly detection, and there is considerable potential in exploring their applications in banking and finance. This paper has conducted an overview of various classes of anomaly detection problems and algorithms, focusing on those that are relevant to the context of banking, such as customer and employee behavior tracking and various cases of fraudulent activities.

Traditional unsupervised anomaly detection algorithms have been based on nearest neighbor and clustering procedures, applied to multidimensional numerical data; the choice of distance measures is then critical. Heterogeneous density distributions and asymmetric cluster shapes hinder their applicability, but have been successfully addressed by more recently developed algorithms, which focus on local neighborhoods and relative proximity relationships between nearest neighbors. In some applications, ground truth data may be available, permitting the application of supervised or semi-supervised learning algorithms which use the available data at least to characterize normal data, against which anomalous data may be contrasted and identified.

An entirely different set of problems arises with time-stamped data streams that arrive over time. We then consider the identification of anomalies within a single stream, as well as compare one data stream against others to identify stream-level anomalies. Within a stream, we may distinguish among various special kinds of anomalies such as point anomalies, discords, rate anomalies, contextual anomalies, and others. Between multiple data streams, we may compare individual data streams against subsets of the entire set of streams for which data is available. Algorithms that accomplish these tasks include procedural algorithms such as Regression and Hidden Markov Models, as well as approaches based on transforming data to a different space that is more amenable to the application of anomaly detection algorithms.

These methodologies need substantial modification when the data is categorical or nominal, i.e., data attributes are not numeric. Data mining approaches may then be applicable, identifying frequent item-sets, with anomalies being defined as the infrequent ones, subject to some normalization procedures to account for the possibility that the items (within the infrequent item-sets) themselves occur rarely. Information compression methodologies based on the minimum description length principle have also been explored; transactions with high compression cost are then considered to be anomalies.

We then discussed different ways of combining multiple anomaly detection algorithms, so that we can identify anomalies that may be detected by some (but not all) individual algorithms.

Finally, we presented mathematical formulations of the anomaly detection problems, providing standards for solution quality against which specific anomaly detection algorithms can be evaluated. In addition to prediction, clustering, and classification perspectives, we also introduce decision-theoretic considerations that take asymmetric misclassification costs into consideration.

Although considerable work has been accomplished in all the areas discussed above, extensive empirical results exploring the application of anomaly detection algorithms to banking applications are few. This is a promising area for future study, using real datasets against which different algorithms can be evaluated. The need for such work is critical especially due to the increasing proliferation of cyber-crimes directed at the banking and financial industry, affecting millions of customers, and threatening to drastically diminish the trust in existing financial infrastructure, systems and platforms.

References

1. Abfalg, J., Kriegel, H. P., Kröger, P., Kunath, P., Pryakhin, A., & Renz, M.: Similarity search on time series based on threshold queries. In: Proceedings of the EDBT. 276-294 (2006)
2. Aggarwal, C. C.: Outlier analysis. In Data mining. Springer, 237-263 (2015)
3. Agrawal, R., Mannila, H., Srikant, R., Toivonen, H., Verkamo, A.I.: Fast discovery of association rules. Advances in Knowledge Discovery and Data Mining. **12**(1), 307–328 (1996)
4. Agrawal, R., Srikant, R., et al.: Fast algorithms for mining association rules. In: Proc. of the 20th Int. Conf. on Very Large Data Bases. VLDB. **1215**, 487–499 (1994)
5. Agyemang, M., Barker, K., Alhajj., R.: A comprehensive survey of numeric and symbolic outlier mining techniques. Intelligent Data Analysis. **10**(6), 521–538 (2006)

6. Akoglu, L., Tong, H., Vreeken, J., Faloutsos., C.: Fast and reliable anomaly detection in categorical data. In: Proceedings of the 21st ACM International Conference on Information and Knowledge Management. 415–424, ACM (2012)
7. Amari., S.: Field theory of self-organizing neural nets. IEEE Transactions on Systems, Man, and Cybernetics, **(5)**, 741–748 (1983)
8. Amer, M., Goldstein, M., and Abdennadher, S.: Enhancing one class support vector machines for unsupervised anomaly detection. In: Proceedings of the ACM SIGKDD Workshop on Outlier Detection and Description. 8–15, ACM (2013)
9. Donald, J., Berndt, Clifford, J.: Using dynamic time warping to find patterns in time series. In: AAAI Working Notes of the Knowledge Discovery in Databases Workshop, 359–370 (1994)
10. Bonanno, G., Lillo, F., Mantegna., R.N.: High-frequency cross-correlation in a set of stocks. Quantitative Finance. 96–104 (2001)
11. Boser, B.E., Guyon, I.M., Vapnik, V.N.: A training algorithm for optimal margin classifiers. In: Proceedings of the Fifth Annual Workshop on Computational Learning Theory, 144–152. ACM (1992)
12. Breunig, M.M., Kriegel, H., Ng, R.T., Lof, J.S.: Identifying density-based local outliers. In ACM sigmod record, **29**, 93–104, ACM (2000)
13. Bu, Y., Leung, T., Fu, A.W., Keogh, E., Pei, J., Wat, S.M.: Finding top-k discords in time series database. In: Proceedings of the 2007 SIAM International Conference on Data Mining. 449–454 (2007)
14. Carpenter, G.A., Grossberg, S.: Adaptive resonance theory. In Encyclopaedia of Machine Learning and Data Mining, 1–17, Springer (2016)
15. Chair, Z. and Varshney, P.K.: Optimal data fusion in multiple sensor detection systems. IEEE Transactions on Aerospace and Electronic Systems, **(1)**, 98–101(1986)
16. Chan, K.P. and Fu, A.W.C.: Efficient time series matching by wavelets. In: Proceeding of the 15th International Conference on Data Engineering, Sydney, Australia, 126-127 (1999)
17. Chandola, V., Banerjee, A., Kumar, V.: Anomaly detection. ACM Computing Surveys, **41**(3), 1–58 (2009)
18. Chandola, V, Banerjee, A., Kumar, V.: Anomaly detection for discrete sequences: A survey. IEEE Transactions on Knowledge and Data Engineering, **24**(5), 823–839 (2012)
19. Chang, W., Chang, J.: Using clustering techniques to analyze fraudulent behavior changes in online auctions. In: Proceedings of the International Conference on Networking and Information Technology, 34–38, IEEE (2010)
20. Cheboli, D.: Anomaly detection of time series. PhD thesis, University of Minnesota (2010)
21. Chen, L, Raymond, Ng.: On the marriage of lP-norms and edit distance. In: Proceedings of the Thirtieth International Conference on Very Large Data Bases. 792–803 (2004)
22. Chen, Y., Nascimento, M.A., Chin, B., Anthony, O., Tung, K. H.: Spade: On shape-based pattern detection in streaming time series. In: Proceedings of the 23rd International Conference on Data Engineering, Istanbul, Turkey. 786–795 (2007)
23. Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T.: A fast and elitist multiobjective genetic algorithm: Nsga-ii. IEEE transactions on evolutionary computation. **6**(2),182–197 (2002)
24. Deng, Q., Mei, G.: Combining self-organizing map and k-means clustering for detecting fraudulent financial statements. In: Proceedings of the International Conference on Granular Computing, 126–131. IEEE (2009)
25. Ding, H., Trajcevski, G., Scheuermann, P., Wang, X., Keogh, E.: Querying and mining of time series data: experimental comparison of representations and distance measures. In: Proceedings of the VLDB Endowment, 1542–1552 (2008)
26. Du, K., Swamy, MNS.: Radial basis function networks. In Neural Networks and Statistical Learning. 299–335, Springer (2014)
27. Elkan, C., Noto, K.: Learning classifiers from only positive and unlabeled data. In: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery And Data Mining. 213–220, ACM (2008)
28. Eskin, E., Arnold, A., Prerau, M., Portnoy, L., Stolfo, S.: A geometric framework for unsupervised anomaly detection. In: Applications of Data Mining In Computer Security, 77–101, Springer (2002)
29. Faloutsos, C., Ranganathan, M., Manolopoulos, Y.: Fast subsequence matching in time-series databases. In Proceedings of the ACM SIGMOD International Conference on Management of data, New York, NY, USA, 419 – 429, ACM (1994)
30. Freund, Y., Schapire, R.E.: A decision-theoretic generalization of online learning and an application to boosting. In: Proceedings of the European Conference On Computational Learning Theory, 23–37, Springer (1995)
31. Fritzke, B.: Growing cell structures a self-organizing network for unsupervised and supervised learning. Neural networks. **7**(9), 1441–1460 (1994)

32. Fujimaki, R., Yairi, T., Machida, K.: An anomaly detection method for spacecraft using relevance vector learning. Advances in Knowledge Discovery and Data Mining, 785–790 (2005)
33. Ghosh, S., Reilly, D.L.: Credit card fraud detection with a neural-network. In: Proceedings of the Twenty-Seventh Hawaii International Conference on System Sciences, **3**, 621–630, IEEE (1994)
34. Grünwald, P.D.: The minimum description length principle. MIT Press (2007)
35. Guha, S., Mishra, N., Roy, G., Schrijvers, O.: Robust random cut forest based anomaly detection on streams. In: proceedings of the International Conference on Machine Learning, 2712–2721 (2016)
36. Gupta, M., Gao, J., Aggarwal, C.C., Han, J.: Outlier detection for temporal data: A survey. IEEE Transactions on Knowledge and Data Engineering, **26**(9), 2250–2267, (2014)
37. Hodge, V.J., Austin, J.: A survey of outlier detection methodologies. Artificial Intelligence Review, **22**(2), 85–126 (2004)
38. Hsu, C., Chang, C., Lin, C. et al.: A practical guide to support vector classification. (2003)
39. Huang, H., Mehrotra, K., Mohan, C.K.: An online anomalous time series detection algorithm for univariate data streams. In: Proceedings of the International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems. 151–160. Springer (2013)
40. Huang, H., Mehrotra, K., Mohan, C.K.: Rank-based outlier detection. Journal of Statistical Computation and Simulation, **83**(3) ,518–531(2013)
41. Huang, H., Mehrotra, K., Mohan, C.K.: Detection of anomalous time series based on multiple distance measures. In: Proceedings of the 28th International Conference on Computers and Their Applications, Honolulu, Hawaii, USA, (2013)
42. Issa, H., Vasarhelyi, M.A., Application of anomaly detection techniques to identify fraudulent refunds, (2011)
43. Jin, W., Tung, A. KH., Han, J., Wang, W.: Ranking outliers using symmetric neighbourhood relationship. Advances in Knowledge Discovery and Data Mining, 577–593, Springer (2006)
44. Keogh, E., Ratanamahatana, C.A.: Exact indexing of dynamic time warping. Knowl. Inf. Syst. **3**, 358–386 (2005)
45. Keogh, E., Folias, T.: The ucr time series data mining archive, (2002)
46. Keogh, E., Lin, J.: Hot SAX: Efficiently finding the most unusual time series subsequence. In: Proceedings of the 5th IEEE International Conf. on Data Mining, Houston, Texas, 226 – 233 (2005)
47. Keogh, E., Lin, J., Fu, A.: Hot sax: Efficiently finding the most unusual time series subsequence. In: Proceedings of the Fifth IEEE International Conference on Data Mining, pp. 8, IEEE (2005)
48. Keogh, E., Lin, J., Lee, S., Herle, H.V.: Finding the most unusual time series subsequence: Algorithms and applications. Knowledge and Information Systems, **11**(1), 1–27 (2007)
49. Kohonen, T: The self-organizing map. In: Proceedings of the IEEE, **78**(9),1464–1480 (1990)
50. Kohonen, T.: Essentials of the self-organizing map. Neural Networks, **37**, 52–65 (2013)
51. Lin, J., Keogh, E., Li, W., Lonardi, S.: Experiencing sax: A novel symbolic representation of time series. Data Mining and knowledge discovery, **15**(2), 107 (2007)
52. Ma, J., Theiler, J., Perkins, S.: Accurate on-line support vector regression. Neural computation. **15**(11), 2683–2703 (2003)
53. Markou, M., Singh, S. Novelty detection: a review part 1: Statistical approaches. Signal processing. **83** (12), 2481–2497 (2003)
54. Markou, M., Singh, S.: Novelty detection: a review part 2: Neural network based approaches. Signal Processing. **83**(12), 2499–2521 (2003)
55. Mehrotra, K., Mohan, C.K., Ranka, S.: Elements of artificial neural networks. MIT Press (1997)
56. Moayedi, H.Z., Masnadi-Shirazi, MA.: Arima model for network traffic prediction and anomaly detection. In: Proceedings of the International Symposium on Information Technology, ITSim.**4**, 1–6. IEEE (2008)
57. Park, J., Sandberg, I.W.: Universal approximation using radial basis-function networks. Neural computation. **3**(2), 246–257 (1991)
58. Patcha, A., Park, J: An overview of anomaly detection techniques: Existing solutions and latest technological trends. Computer networks. **51**(12), 3448–3470 (2007)
59. Pelleg, D., Moore, A.: X-means: Extending k-means with efficient estimation of the number of clusters. In: Proceedings of the 17th International Conf. on Machine Learning, 727–734, Morgan Kaufmann (2000)
60. Marco, AF., Pimentel, Clifton, DA., Clifton, L., Tarassenko, L.: A review of novelty detection. Signal Processing. **99**, 215–249 (2014)
61. Pincombe, B.: Anomaly detection in time series of graphs using arma processes. Asor Bulletin. **24** (4), 2 (2005)

62. Protopapas, P., Giammarco, JM., Faccioli, L., Struble, MF., Dave, R., and Alcock, C.: Finding outlier light curves in catalogues of periodic variable stars. Monthly Notices of the Royal Astronomical Society. **369** (2), 677–696 (2006)
63. Rakthanmanon, T., Campana, B., Mueen, A., Batista, G., Westover, b., Zhu,Q., Zakaria, J., and Keogh, E. : Searching and mining trillions of time series subsequences under dynamic time warping. In: Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. 262–270 (2012)
64. Sabau, A.S.: Survey of clustering based financial fraud detection research. Informatica Economica. **16** (1), 110 (2012)
65. Siebes, A., Vreeken, J., Leeuwen, M.V.: Item sets that compress. In: Proceedings of the SIAM International Conference on Data Mining. 395–406 (2006)
66. Smets, K., Vreeken, J.: The odd one out: Identifying and characterising anomalies. In: Proceedings of the SIAM international conference on data mining. 804–815 (2011)
67. Smola, A.J., Schölkopf, B.: A tutorial on support vector regression. Statistics and Computing. **14**(3), 199–222 (2004)
68. Tang, J., Chen, Z., Fu, A.W., Cheung, D.: A robust outlier detection scheme for large data sets. In: Proceedings of the 6th Pacific-Asia Conf. on Knowledge Discovery and Data Mining. Citeseer (2001)
69. Thiprungsri, S., Vasarhelyi, M.A.: Cluster analysis for anomaly detection in accounting data: An audit approach (2011)
70. Varshney, P.K.: Distributed detection and data fusion. Springer Science & Business Media (2012)
71. Li, W., Nitin, K., Lolla, V.N., Keogh, E.J., Lonardi, S., Ratanamahatana, C.: Assumption-free anomaly detection in time series. In: SSDBM, **5**, 237–242 (2005)
72. Zhao, Z., Mehrotra, K.G., Mohan, C.K.: Ensemble algorithms for unsupervised anomaly detection. In: Proceedings of the International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems. 514–525, Springer (2015)
73. Zhao, Z., Mehrotra, K.G., Mohan, C.K.: Card. Evolution of space-partitioning forest for anomaly detection. In: Proceedings of the Workshop on Genetic Programming Theory and Practice. (2017)
74. Zhu, X. Semi-supervised learning literature survey (2005)

IJBT (2017) 01: 49-65
RESEARCH ARTICLE

Artificial neural networks for artificial intelligence

Nikola Kasabov[1]

Abstract Artificial neural networks now have a long history as major techniques in computational intelligence with a wide range of applications for learning from data. There are many methods developed and applied so far, from multiplayer perceptrons (MLP) to the recent ones being deep neural networks and deep learning machines based on spiking neural networks. The paper addresses a main question for researchers and practitioners: Having data and a problem in hand, which method would be most suitable to create a model from the data and to efficiently solve the problem? In order to answer this question, the paper reviews the main features of the most popular neural network methods and then lists examples of applications already published and referenced. The methods include: simple MLP; hybrid systems; neuro-fuzzy systems; deep neural networks; spiking neural networks; quantum inspired evolutionary computation methods for network parameter optimization; deep learning neural networks and brain-like deep learning machines. The paper covers both methods and their numerous applications for data modelling, predictive systems, data mining, pattern recognition, across application areas of engineering, health, robotics, security, finances, etc. It concludes with recommendations on which method would be more suitable to use, depending on the data and the problems in hand, in order to create efficient information technologies across application domains.

Keywords: Artificial intelligence (AI) · Artificial neural networks · Evolving Connectionist Systems (ECOS) · Neuro-fuzzy systems · Spiking Neural Networks (SNN) · Evolving spiking neural networks · NeuCube · Quantum inspired neural networks · Spatio-temporal pattern recognition · Data mining

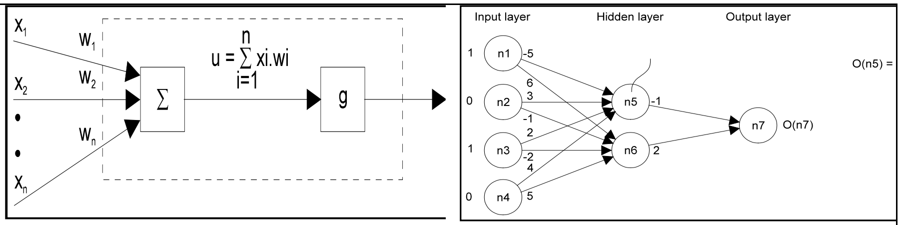
1 Artificial neural networks and hybrid systems

Artificial Neural Networks (ANNs) are computational models that mimic the nervous system in its main function of adaptive learning and generalization. ANNs are universal computational models. One of the most popular artificial neuron models is the McCulloch and Pitts neuron developed in 1943 (Figure 1a). It was used in early ANNs such as Perceptron [1] and multilayer perceptron [2-5] – a simple example is given in Figure 1b.

✉ Nikola Kasabov
nkasabov@aut.ac.nz, www.kedri.aut.ac.nz

1 Fellow IEEE, Director, Knowledge Engineering and Discovery Research Institute – KEDRI
Auckland University of Technology, New Zealand

These ANN are suitable when trained on a small scale static (vector-based) data, but are not adaptive to new data and in most cases they are ‘black boxes’ – they do not reveal internal structures in the data to be used to extract new knowledge. Optimal structures of ANNs are difficult to design.



(a) A simple artificial neuron. *(b) An example of a simple ANN.*

Fig. 1 Examples of simple artificial neuron models

In order to incorporate human knowledge into an intelligent system, an ANN module can be combined with a rule-based module in the same system. The rules can be fuzzy rules as a partial case [6, 7]. An exemplar system is shown in Figure 2, where, at a lower level, an ANN module predicts the next day value of a stock index and, at a higher level, a fuzzy reasoning module combines the predicted values with some macro-economic variables, using the following types of fuzzy rules [8]:

IF <the module stock predicted by the ANN is high> AND <the economic situation is good> (1)
THEN <buy stock>

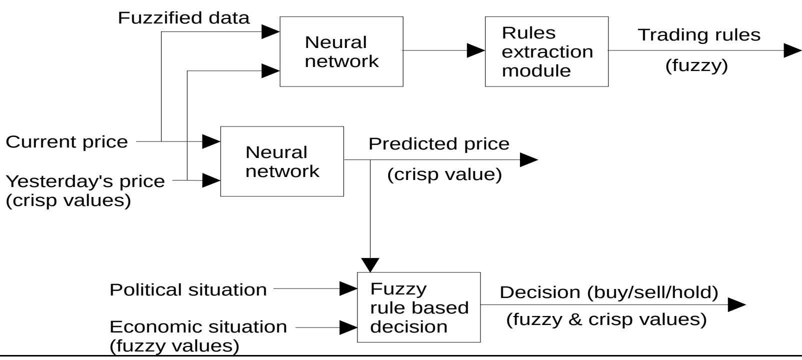


Fig. 2 A hybrid ANN-fuzzy rule-based expert system for financial decision support [8]

Hybrid systems can also use crisp propositional rules, along with fuzzy rules [9]. The hybrid systems from Figure 2 are suitable to use when decision rules are available to integrate with data.

Another group of ANN methods can be used not only to learn from data, but to extract rules from a trained ANN and/or insert rules into an ANN as initialization procedure. These are the neuro-fuzzy systems as discussed in the next section on the case of the evolving connectionist systems (ECOS).

2 Neuro-fuzzy and evolving connectionist systems

2.1 The principles of neuro-fuzzy and evolving connectionist systems

The integration of neural networks and fuzzy systems into one ANN attracted many researchers. The integration of fuzzy rules into a single neuron model and then into larger neural network structures, tightly coupling learning and fuzzy reasoning rules into connectionists structures, was initiated by Professor Takeshi Yamakawa and other Japanese scientists [10]. Many models of fuzzy neural networks were developed based on these principles [8, 11, 12].

In the evolving connectionist systems (ECOS) these ideas were developed further, where instead of training a fixed connectionist structure, the structure and its functionality are evolving from incoming data, often in an online, one-pass learning mode [12-16].

ECOS are modular connectionist based systems that evolve their structure and functionality in a continuous, self-organized, online, adaptive, interactive way from incoming information [13]. They can process both data and knowledge in a supervised and/or unsupervised way. ECOS learn local models from data through clustering of the data and associating a local output function for each cluster represented in a connectionist structure. They can learn incrementally single data items or chunks of data and also incrementally change their input features [15, 17]. Elements of ECOS have been proposed as part of the classical neural network models, such as Self-Organizing Maps, Radial Basis Functions, Fuzzy ARTMap, growing neural gas, neuro-fuzzy systems, Resource Allocation Network (for a review see [17]). Other ECOS models, along with their applications, have been reported in [18] and [19].

The principle of ECOS is based on *local learning* – neurons are allocated as centers of data clusters and the system creates local models in these clusters. Fuzzy clustering, as a means to create local knowledge-based systems, was stimulated by the pioneering work of Bezdek, Yager, and Filev [20, 21].

To summarize, the following are the main principles of ECOS as stated in [13]:

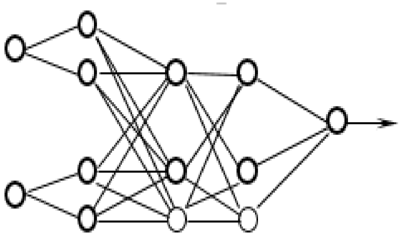
(1) Fast learning from large amount of data, e.g. using "one-pass" training, starting with little prior knowledge;
(2) Adaptation in real-time and in an on-line mode where new data is accommodated as it comes based on local learning;
(3) "Open", evolving structure, where new input variables (relevant to the task), new outputs (e.g. classes), new connections and neurons are added/evolved "on the fly";
(4) Both data learning and knowledge representation is facilitated in a comprehensive and flexible way, e.g., supervised learning, unsupervised learning, evolving clustering, "sleep" learning, forgetting/pruning, fuzzy rule insertion and extraction;
(5) Active interaction with other ECOSs and with the environment in a multi-modal fashion;
(6) Representing both space and time in their different scales, e.g., clusters of data, short- and long-term memory, age of data, forgetting, etc.;

(7) System's self-evaluation in terms of behavior, global error and success, and related knowledge representation.

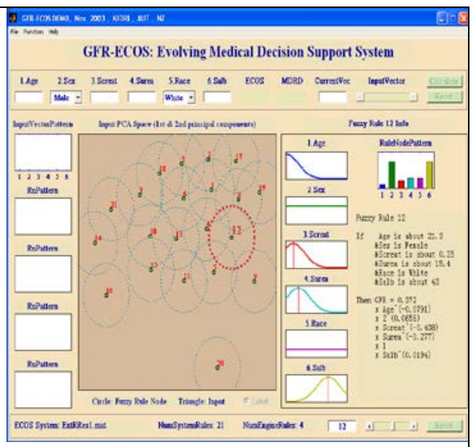
Here the concept of ECOS is illustrated on two implementations: the evolving fuzzy neural network (EFuNN) [14] and the dynamic evolving neuro-fuzzy inference system (DENFIS) [16]. Examples of EFuNN and DENFIS are shown in Figure 3a and Figure 3b, respectively. In ECOS, clusters of data are created based on similarity between data samples either in the input space (this is the case in some of the ECOS models, e.g., DENFIS), or in both the input and output space (this is the case, e.g., in the EFuNN models). Samples (examples) that have a distance to an existing node (cluster center, rule node) less than a certain threshold are allocated to the same cluster. Samples that do not fit into existing clusters form new clusters. Cluster centers are continuously adjusted according to new data samples, and new clusters are created incrementally. ECOS learn from data and automatically create or update a local fuzzy model/function, e.g.:

$$IF < data\ is\ in\ a\ fuzzy\ cluster\ C_i > THEN < the\ model\ is\ F_i > \quad (2)$$

where F_i can be a fuzzy value, a logistic or linear regression function (Figure 3b) or ANN model [16, 17].



a) An example of EFuNN model (from [14])



b) An example of DENFIS model (from [15] and [17])

Fig. 3 Example implementations of ECOSs

The ECOS methods are realized as software modules as part of the free development system NeuCom (www.theneucom.com).

A special development of ECOS is *transductive reasoning and personalized modelling*. Instead of building a set of local models (i.e., prototypes) to cover the whole problem space and then use these models to classify/predict any new input vector, in transductive modelling for every new input vector, a new model is created based on selected nearest neighbor vectors from the available data. Such ECOS models are the neuro-fuzzy inference system (NFI) and the transductive weighted NFI (TWNFI) [22]. In TWNFI, for every new input vector the neighborhood of closets data vectors is optimized using both the distance between the new vector and the neighboring ones and the weighted importance of the input variables, so that the error of the model is minimized in the neighborhood area [23].

2.2 A survey of neuro-fuzzy and ECOS-based methods

The following is a survey list of selected methods that use the ECOS principles (full publications and details are available from www.ieeexplore.ieee.org, Google Scholar and Scopus):

- Evolving Self-Organized Maps (ESOM) [24];
- Evolving Clustering Method (ECM) [25];
- Incremental feature learning in ECOS [26];
- Online ECOS optimization [27];
- Assessment of EFuNN accuracy for pattern recognition using data with different statistical distributions [28];
- Recursive clustering based on a Gustafson–Kessel algorithm [29];
- Using a map-based encoding to evolve plastic neural networks [30];
- Evolving Takagi–Sugeno fuzzy model based on switching to neighboring models [31];
- A soft computing based approach for modeling of chaotic time series [32];
- Uni-norm based evolving neural networks and approximation capabilities [33];
- Global, local and personalised modelling and profile discovery in Bioinformatics: An integrated approach [34];
- FLEXFIS: a robust incremental learning approach for evolving Takagi–Sugeno fuzzy models [35];
- Evolving fuzzy classifiers using different model architectures [36];
- RSPOP: Rough Set–Based Pseudo Outer-Product Fuzzy Rule Identification Algorithm [37];
- SOFMLS: Online self-organizing fuzzy modified least-squares network [38];
- On-Line Sequential Extreme Learning Machine [39];
- Finding features for real-time premature ventricular contraction detection using a fuzzy neural network system [40];
- Evolving fuzzy rule-based classifiers [41];
- A novel generic Hebbian ordering-based fuzzy rule base reduction approach to Mamdani neuro-fuzzy system [42];
- Implementation of fuzzy cognitive maps based on fuzzy neural network and application in prediction of time series [43];
- Backpropagation to train an evolving radial basis function neural network [44];
- Smooth transition autoregressive models and fuzzy rule-based systems: Functional equivalence and consequences [45];
- Development of an adaptive neuro-fuzzy classifier using linguistic hedges [M46];

- A meta-cognitive sequential learning algorithm for neuro-fuzzy inference system [47];
- Meta-cognitive RBF network and its projection based learning algorithm for classification problems [48];
- SaFIN: A self-adaptive fuzzy inference network [49];
- A sequential learning algorithm for meta-cognitive neuro-fuzzy inference system for classification problems [50];
- Architecture for development of adaptive online prediction models [51];
- Clustering and co-evolution to construct neural network ensembles: An experimental study [52];
- Algorithms for real-time clustering and generation of rules from data [53];
- SAKM: Self-adaptive kernel machine – A kernel-based algorithm for online clustering [54];
- A BCM theory of meta-plasticity for online self-reorganizing fuzzy-associative learning [55];
- Evolutionary strategies and genetic algorithms for dynamic parameter optimization of evolving fuzzy neural networks [56];
- Incremental leaning and model selection for radial basis function network through sleep learning [57];
- Interval-based evolving modeling [58];
- Evolving granular classification neural networks [59];
- Stability analysis for an online evolving neuro-fuzzy recurrent network [60];
- A TSK fuzzy inference algorithm for online identification [61];
- Design of experiments in neuro-fuzzy systems [62];
- EFuNNs ensembles construction using a clustering method and a co-evolutionary genetic algorithm [63];
- eT2FIS: An evolving type-2 neural fuzzy inference system [64];
- Designing radial basis function networks for classification using differential evolution [65];
- A meta-cognitive neuro-fuzzy inference system (McFIS) for sequential classification problems [66];
- An evolving fuzzy neural network based on the mapping of similarities [67];
- Incremental learning by heterogeneous bagging ensemble [68];
- Fuzzy associative conjuncted maps network [69];
- EFuNN ensembles construction using CONE with multi-objective GA [70].

2.3 Neuro-fuzzy and ECOS applications for AI

Based on the ECOS concepts and methods, sustained engineering applications have been developed, such as:

- Risk analysis and discovery of evolving economic clusters in Europe [71];
- Adaptive time series prediction for financial applications [72];
- Adaptive speech recognition [73];
- and others [17].

While the ECOS methods presented above use the McCulloch and Pitts model of a neuron (Figure 1a) and have been efficiently used for vector-based data, for rule extraction from data and for classification and prediction purposes, the further developed spiking neural networks (SNN) and evolving SNN (eSNN) architectures use a spiking neuron model and spike information representation. Spike information representation accounts for time in the data and for changes in the data over time. This is where SNN can be chosen as preferred methods and used efficiently.

3 Spiking Neural Networks and the brain-like AI

3.1 Main principles, methods and examples of SNN and evolving SNN (eSNN)

A spiking neuron model receives input information represented as trains of spikes over time. When sufficient input information is accumulated in the membrane of the neuron, the neuron's post synaptic potential exceeds a threshold and the neuron emits a spike at its axon (Figure 4).

Some of the-state-of-the-art models of a spiking neuron include: early models by Hodgkin and Huxley [74]; more recent models by Maas, Gerstner, Kistler, Izhikevich and others, e.g., Spike Response Models (SRM); Integrate-and-Fire Models (IFM) (Figure 4); Izhikevich models; adaptive IFM; probabilistic IFM [75, 76].

Based on the ECOS principles, an evolving spiking neural network architecture (eSNN) was proposed [17]. It was initially designed as a visual pattern recognition system. The first eSNNs were based on Thorpe's neural model [77], in which the importance of early spikes (after the onset of a certain stimulus) is boosted, called rank-order coding and learning. Synaptic plasticity is employed by a fast supervised one-pass learning algorithm. Different eSNN models were developed, including:

- Reservoir-based eSNN for spatio- and spectro-temporal pattern recognition (Figure 5) [78];
- Dynamic eSNN (deSNN) [79] – a model that uses both rank-order and time-based spike-time dependent plasticity (STDP) learning rules [80] to account for spatio-temporal data.

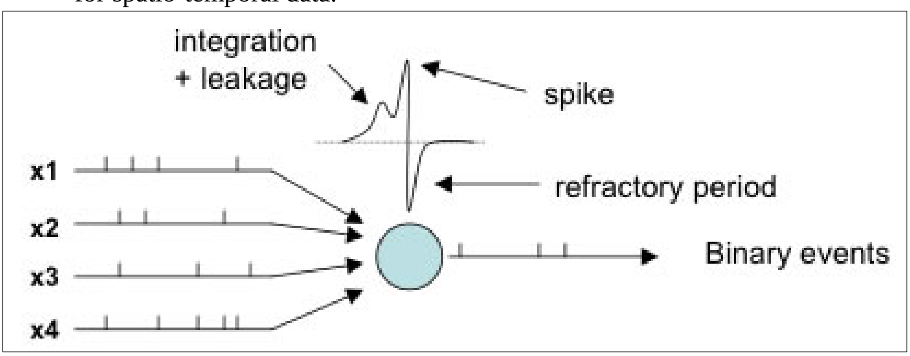


Fig. 4 The structure of the LIFM of a spiking neuron

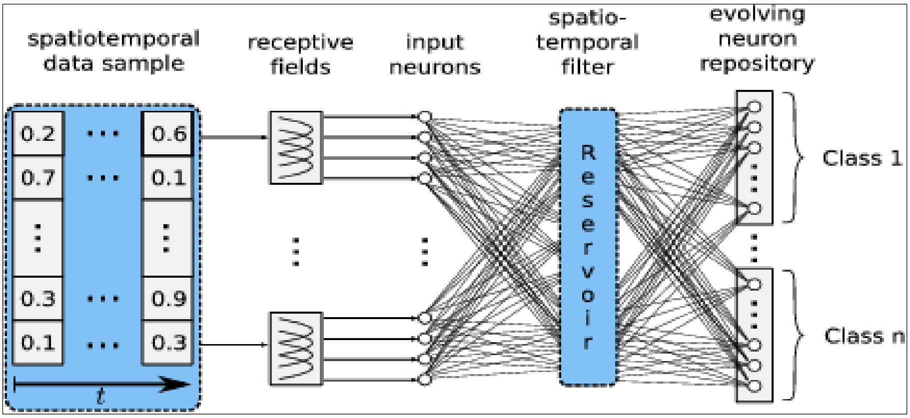
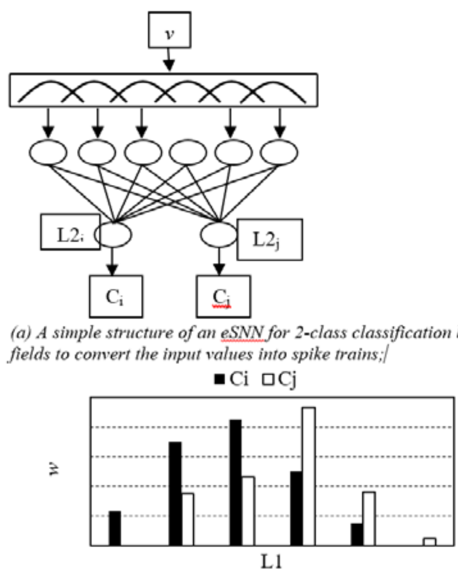


Fig. 5 A reservoir-based eSNN for spatio-temporal data classification

Extracting fuzzy rules from an eSNN would make the eSNN not only efficient learning models, but also knowledge-based models. A method was proposed [81] and illustrated in Figure 6a and Figure 6b. Based on the connection weights (W) between the receptive field layer (L1) and the class output neuron layer (L2), the following fuzzy rules can be extracted:

$$\textit{IF (input variable v is SMALL) THEN class } C_i; \quad \textit{IF(v is LARGE) THEN class } C_j \qquad (3)$$



(a) A simple structure of an eSNN for 2-class classification based on one input variable using 6 receptive fields to convert the input values into spike trains;

(b) The connection weights of the connections to class C_i and C_j output neurons respectively are interpreted as fuzzy rules.

Fig. 6 Knowledge extraction from evolving spiking neural networks

The eSNN use spike information representation, spiking neuron models, and spike learning and encoding rules, and the structure is evolving to capture spatio-temporal relationship from data.

3.2 Applications of eSNN for AI

Numerous applications based on different eSNN models have been reported, among them:

- Advanced spiking neural network technologies for neurorehabilitation [82];
- Object movement recognition [83];
- Multimodal audio and visual information processing [84];
- Ecological data modelling and prediction of the establishment of invasive species [85];
- Integrated brain data analysis [86];
- Predictive modelling method and case study on personalized stroke occurrence prediction [87].

3.3 Quantum inspired optimization of eSNN

eSNN have several parameters that need to be optimized for an optimal performance. Several successful methods have been proposed for this purpose, among them are: Quantum-inspired evolutionary algorithm (QiEA) [88], and Quantum-inspired particle swarm optimization method (QiPSO) [89].

Quantum inspired optimization methods use the principle of superposition of states to represent and optimize features (input variables) and parameters of the eSNN [17]. Features and parameters are represented as qubits, which are in a superposition of 1 (selected) with a probability α, and 0 (not selected) with a probability β. When the model has to be calculated, the quantum bits "collapse" into a value of 1 or 0.

3.4 Neuromorphic implementations of SNN

Using SNN neuromorphic computational models can be developed. Opposite to the traditional von Neumann computational architecture, where memory, control and ALU are separated, in neuromorphic models all these modules are integrated together as they are in the brain.

To make the implementation of SNN models more efficient, specialized neuromorphic hardware has been developed, including:

- A hardware model of an integrate-and-fire neuron [90];
- A silicon retina [91];
- INI Zürich SNN chips [92, 93];
- IBM True North [94]. The system enables parallel processing of 1mln spiking neurons and 1 billion synapses;
- DVS and silicon cochlea (ETH, Zurich);
- Stanford NeuroGrid [95]. The system has 1 million neurons on a board, 6 3 billion connections, and is realized as hybrid analogue/digital circuits;
- SpiNNaker [96]. The system is a general-purpose, scalable, multichip multicore platform for real-time massively parallel simulations of large scale SNN.

The neuromorphic platforms are characterized by massive parallelism, high speed and low power consumption. For their efficient application, they require the development of SNN computational models for learning from data.

4 Deep learning neural networks and brain-like AI machines. NeuCube

4.1 Deep learning neural networks (DNN)

Deep learning neural networks (DNN) are ANN that have several layers of neurons and connections in their structures (rather than 3 as shown in Figure 1b). A class of DNN is the convolutional DNN, where neurons at the first layer learn features only within a small subsection of the input vector data (e.g., a small square of pixels from an image). These neurons are connected to the next layer where features are combined, until the output classification layer, where output classes are determined. An example is shown in Figure 7.

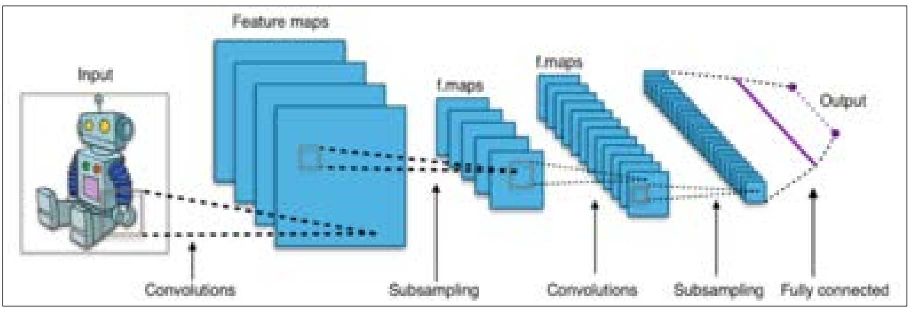


Fig.7 An example of a convolutional DNN (from https://en.wikipedia.org/wiki/Convolutional_neural_network)

DNNs are excellent for vector- or frame-based data, but not much for temporal (or spatio-/spectro-temporal data). There is no *time of asynchronous events* learned in the model. They are difficult to adapt to new data and the structures are not flexible.

4.2 Brain-like AI machines. NeuCube.

Inspired by the deep learning in the brain, a deep learning machine was developed, named NeuCube [86]. It was initially designed for spatio-temporal brain data modelling, but then it was also used for climate data modelling and stroke occurrence prediction and other applications [87].

The NeuCube framework is depicted in Figure 8. It consists of the following modules:

- Input information encoding module;
- 3D SNN reservoir module (SNNr);
- Output (e.g. classification) module;
- Gene regulatory network (GRN) module (optional);
- Optimization module (optional).

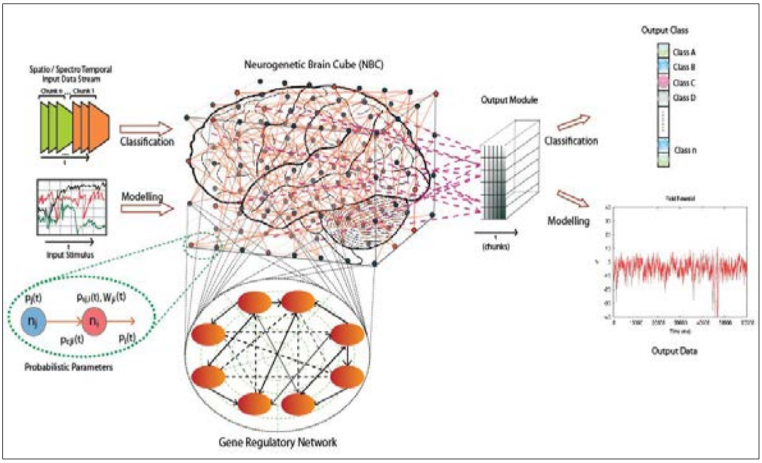


Fig. 8 A block diagram of the NeuCube deep learning machine (from [86])

The input module transforms input data into trains of spikes. Spatio-temporal data (such as EEG, climate, cybersecurity, financial, etc.) is entered after the encoding into the main module – the 3D SNN reservoir (SNNr). Input data is entered into *pre-designated spatially distributed* areas of the SNNr that correspond to the spatial location in the origin where data was collected (if there is such).

Learning in the SNN is performed in two stages:

- Unsupervised training, where spatio-temporal data is entered into relevant areas of the SNNr over time. Unsupervised learning is performed to modify the initially set connection weights. The SNNr will learn to activate the same groups of spiking neurons when similar input stimuli are presented, also known as a *polychronization* effect [76].
- Supervised training of the spiking neurons in the output classification module, where the same data that was used for unsupervised training is now propagated again through the trained SNN and the output neurons are trained to classify the spatio-temporal spiking pattern of the SNNr into pre-defined classes (or output spike sequences). As a special case, all neurons from the SNN are connected to every output neuron. Feedback connections from output neurons to neurons in the SNN can be created for reinforcement learning. Different SNN methods can be used to learn and classify spiking patterns from the SNNr, including the deSNN [79] and SPAN models [97]. The latter is suitable for generating motor control spike trains in response to certain patterns of activity of the SNNr.

Memory in the NeuCube architecture is represented as a combination of the three types of memory described below, which are mutually interacting:

- Short-term memory, represented as changes of the PSP and temporary changes of synaptic efficacy;
- Long-term memory, represented as a stable establishment of synaptic efficacy – long-term potentiation (LTP) and long-term depression (LTD);
- Genetic memory, represented as a genetic code.

In NeuCube, similar activation patterns (called “polychronous waves”) can be generated in the SNNr with recurrent connections to represent short term memory. When using STDP learning, connection weights change to form LTP or LTD, which constitute long-term memory.

Results of the use of the NeuCube suggest that the NeuCube architecture can be explored for learning long (spatio-) temporal patterns and to be used as associative memory. Once data is learned, the SNNr retains the connections as a long-term memory. Since the SNNr learns functional pathways of spiking activities represented as structural pathways of connections, when only a small initial part of input data is entered the SNNr will “synfire” and “chain-fire” *learned connection pathways* to reproduce *learned functional pathways.* Thus, a NeuCube can be used as an associative memory and as a predictive system with a wide scope of applications.

4.3 Applications of NeuCube for AI

For a classification or a regression problem based on temporal- or spatio/spectro-temporal data, a NeuCube based system can be designed and implemented following the steps below:

(a) Input data transformation into spike sequences;
(b) Mapping input variables into spiking neurons;
(c) Unsupervised learning spatio-temporal spike sequences in a scalable 3D SNN reservoir;
(d) On-going learning and classification of data over time;
(d) Dynamic parameter optimization;
(e) Evaluating the time for predictive modelling;
(f) Adaptation on new data, possibly in an on-line/real time mode;
(g) Model visualization and interpretation for a better understanding of the data and the processes that generated it;
(h) Implementation of the SNN model as both software and a neuromorphic hardware system (if necessary).

A NeuCube development system, that allows for the above steps to be explored for a final design of an efficient application system, is available from: http://www.kedri.aut.ac.nz/neucube/. Several applications of NeuCube are described in [98]. In [99], a method for using SNN for efficient data compression is introduced, with wide range of applications in telecommunication. In [100] a survey of applications using NeuCube SNN machine can be found.

5 Conclusion

This paper presents briefly methods of artificial neural networks (ANN) and directs users which method to use depending on the data and the problem in hand. According to the type of the data, these applications can be classified as:

- Static (vector-based) data;
- Temporal data (e.g. climate, financial);
- Spatio-temporal data with fixed spatial location of variables (e.g., cybersecurity, brain data);

- Spatio-temporal data with changing locations of the spatial variables (e.g., moving objects);
- Spectro-temporal data (e.g. radio-astronomy, audio).

According to the characteristics of the data, the applications can be:

- Sparse features/low frequency (e.g. climate data, ecological data, multisensory data);
- Sparse features/high frequency (e.g. EEG brain signals, seismic data related to earthquakes);
- Dense features/low frequency (e.g. financial data);
- Dense features/high frequency (e.g. radio-astronomy data).

The paper presents first principles of various ANN and then surveys their use in practical systems. The paper specifically addresses the problem of designing systems for adaptive learning and knowledge discovery from complex data, emphasizing on one particular direction named ECOS. Different types of ANN and ECOS methods in particular are suitable for different applications as discussed in the paper. The paper advocates that integrating principles derived from neural networks, fuzzy systems, evolutionary computation, quantum computing and brain information processing, could lead to more efficient information technologies [100].

Further use of SNN is anticipated in several directions:

- Spatio-temporal data compression [101];
- Security systems [102, 103];
- Personal assistance [104,105]
- Automated financial, banking and trading systems [106].

Acknowledgements The work on this paper is supported by the Knowledge Engineering and Discovery Research Institute (KEDRI, http://www.kedri.aut.ac.nz) under a SRIFT AUT grant. I would like to thank Prof. Ravi and the editors of the journal for inviting me to submit this paper.

References

1. Rosenblatt, F.: The perceptron: A probabilistic model for information storage and organization in the brain. Psychology Review. **65,** 386-402 (1958)
2. Amari, S.: A theory of adaptive pattern classifiers. IEEE Transactions on Electronic Computers. **3** (vol.EC-16), pp. 299-307 (1967)
3. Amari, S.: Mathematical foundations of neurocomputing. Proceedings of the IEEE. **9** (vol.78), pp. 1443-1463 (1990)
4. Rumelhart, D., McLelland, J. (eds): Parallel and distributed processing. MIT Press, 1986
5. Werbos, P. Backpropagation through time: Proc. of the IEEE. 87:10 (1990)
6. Zadeh, L.A.: Fuzzy sets. Information and Control. **3** (vol.8), pp. 338-353 (1965)
7. Zadeh, L. A.: Fuzzy logic. IEEE Computer, **4** (vol.21), pp. 83-93 (1988)
8. Kasabov, N. K.: Foundations of neural networks, fuzzy systems and knowledge engineering. MIT Press, Cambridge, Massachusetts, USA (1996)
9. Kasabov, N. K., Shishkov, S. I.: A connectionist production system with partial match and its use for approximate reasoning. Connection Science. **3-4,** 5, pp. 275-305 (1993)

10. Yamakawa, T., Uchino, E., Miki, T., Kusanagi, H.: A neo fuzzy neuron and its application to system identification and prediction of the system behavior. Proc. 2nd Int. Conf. on Fuzzy Logic and Neural Networks. Japan, pp. 477-483 (July 1992)
11. Furuhashi, T., Hasegawa, T., Horikawa, S., Uchikawa, Y.: An adaptive fuzzy controller using fuzzy neural networks. Proc. of 5th Int'l Fuzzy System Association World Congress. Korea, pp.769-772 (July 1993)
12. Kasabov, N. K., Kim, J. S., Watts, M. J., Gray, A. R.: FuNN/2 – A fuzzy neural network architecture for adaptive learning and knowledge acquisition. Information Sciences – Applications. **3-4** (vol.101), pp. 155-175 (1997)
13. Kasabov, N. K.: Evolving fuzzy neural networks – algorithms, applications and biological motivation. In: Yamakawa, T. and Matsumoto, G. (Eds.). Methodologies for the Conception, Design and Application of Soft Computing, World Scientific, pp 271 – 274 (1998)
14. Kasabov, N. K.: Evolving fuzzy neural networks for on-line supervised/unsupervised. Knowledge–Based Learning. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, **6** (vol.31), 2001, pp. 902-918
15. Kasabov, N. K.: Evolving connectionist systems: methods and applications in bioinformatics, brain study and intelligent machines. Springer-Verlag, London, England, 2002
16. Kasabov, N. K. Song, Q.: DENFIS: Dynamic evolving neural-fuzzy inference system and its application for time-series prediction. IEEE Transactions on Fuzzy Systems, **2** (vol.10), 2002, pp. 144-154
17. Kasabov, N. K.: Evolving connectionist systems: the knowledge engineering approach (2nd ed.), Springer-Verlag, London, England (2007)
18. Futschik, M. E., Kasabov, N. K.: Fuzzy clustering in gene expression data analysis. Proc. of the IEEE Int'l Conf. on Fuzzy Systems, USA. pp. 414-419 (2002)
19. Watts, M. J.: A decade of kasabov's evolving connectionist systems: a review. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews. 3 (vol.39), 2009, pp. 253-269
20. Bezdek, J. C.: Analysis of fuzzy information. CRC Press, Boca Raton, Florida, USA (1987)
21. Yager, R. R., Filev, D. P.: Generation of fuzzy rules by mountain clustering. Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology. **3** (vol.2), pp. 209-219 (1994)
22. Song, Q., Kasabov, N. K.: TWNFI – A transductive neuro-fuzzy inference system with weighted data normalization for personalized modelling. Neural Networks, 10 (vol.19), 2006, pp. 1591-1596
23. Kasabov, N. K., Hu, Y.: Integrated optimisation method for personalised modelling and case study applications. Int'l Journal of Functional Informatics and Personalised Medicine. 3 (vol.3), pp. 236-256 (2010)
24. Deng, D., Kasabov N. K.: On-line pattern analysis by evolving self-organizing maps. Neurocomputing. (vol.51), pp. 87-103 (2003)
25. Song Q. and N. Kasabov, NFI: A neuro-fuzzy inference method for transductive reasoning. IEEE Trans. on Fuzzy Systems. **13,** 6, 799-808 (2005)
26. Ozawa, S., S.Too, S.Abe, S. Pang and N. Kasabov: Incremental learning of feature space and classifier for online face recognition. Neural Networks, 575-584 (August 2005)
27. Chan Z.and N.Kasabov: Evolutionary computation for on-line and off-line parameter tuning of evolving fuzzy neural networks. Int. J. of Computational Intelligence and Applications. Imperial College Press, **4**, 3, 309-319 (2004)
28. de Moraes, R. M.: Assessment of EFuNN accuracy for pattern recognition using data with different statistical distributions. Proc. of the 2nd Brazilian Congress on Fuzzy Systems. Brazil, pp. 672-685 (November 2012)
29. Dovžan, D., Škrjanc, I.: Recursive clustering based on a gustafson–kessel algorithm. Evolving Systems. **1** (vol.2), pp. 15-24 (2011)
30. Tonelli, P., Mouret, J.-B.: Using a map-based encoding to evolve plastic neural networks. Proc. of the IEEE Workshop on Evolving and Adaptive Intelligent Systems. France, pp. 9-16 (April 2011)
31. Kalhor, A., Araabi, B. N., Lucas, C.: Evolving takagi–sugeno fuzzy model based on switching to neighboring models. Applied Soft Computing, **2** (vol.13), pp. 939-946 (2013)
32. Vajpai, J., Arun, J. B.: A soft computing based approach for modeling of chaotic time series. Proc. of the 13th Int'l Conf. on Neural Information Processing. China. pp. 505-512 (October 2006)
33. Bordignon, F., Gomide, F.: Uninorm based evolving neural networks and approximation capabilities. Neurocomputing, (vol.127), pp. 13-20 (2014)
34. Kasabov, N.: Global, local and personalised modelling and profile discovery in bioinformatics: An integrated approach. Pattern Recognition Letters. **28**, 6, 673-685 (2007)
35. Lughofer, E. D.: FLEXFIS: A robust incremental learning approach for evolving takagi–sugeno fuzzy models. IEEE Transactions on Fuzzy Systems. **6** (vol.16), pp. 1393-1410 (2008)

36. Angelov, P. P., Lughofer, E. D., Zhou, X.: Evolving fuzzy classifiers using different model architectures. Fuzzy Sets and Systems. **23** (vol.159), pp. 3160-3182 (2008)
37. Ang, K. K., Quek, C.: RSPOP: Rough set-based pseudo outer-product fuzzy rule identification algorithm. Neural Computation. **1** (vol.17), pp. 205-243 (2005)
38. de Jesús Rubio, J.: SOFMLS: Online self-organizing fuzzy modified least-squares network. IEEE Transactions on Fuzzy Systems, **6** (vol.17), pp. 1296-1309 (2009)
39. Huang, G. B., Liang, N. Y., Rong, H. J.: On-line sequential extreme learning machine. Proc. of the IASTED Int'l Conf. on Computational Intelligence. Canada, pp. 232-237 (July 2005)
40. Lim, J. S.: Finding features for real-time premature ventricular contraction detection using a fuzzy neural network system. IEEE Transactions on Neural Networks. **3** (vol.20), pp. 522-527 (2009)
41. Angelov, P. P., Zhou, X., Klawonn, F.: Evolving fuzzy rule-based classifiers. Proc. of the IEEE Symposium on Computational Intelligence in Image and Signal Processing. USA, pp. 220-225 (April 2007)
42. Liu, F., Quek, C., Ng, G. S.: A novel generic hebbian ordering-based fuzzy rule base reduction approach to mamdani neuro-fuzzy system. Neural Computation. **6** (vol.19), pp. 1656-1680 (2007)
43. Song, H., Miao, C., Roel, W. Shen, Z., Catthoor, F.: Implementation of fuzzy cognitive maps based on fuzzy neural network and application in prediction of time series. IEEE Transactions on Fuzzy Systems. **2** (vol.18), pp. 233-250 (2010)
44. de Jesús Rubio, J., Vázquez, D. M., Pacheco, J.: Backpropagation to train an evolving radial basis function neural network. evolving systems. **3** (vol.1), pp. 173-180 (2010)
45. Aznarte, J. L., Benítez, J. M., Castro, J. L.: Smooth transition autoregressive models and fuzzy rule-based systems: functional equivalence and consequences. fuzzy sets and systems. **24** (vol.158), pp. 2734-2745 (2007)
46. Cetisli, B. Development of an Adaptive Neuro-Fuzzy Classifier Using Linguistic Hedges. Expert Systems with Applications. 8 (vol.37), pp. 6093-6101 (2010)
47. Subramanian, K., Suresh, S.: A meta-cognitive sequential learning algorithm for neuro-fuzzy inference system. applied soft computing. 11 (vol.12), pp. 3603-3614 (2012)
48. Babu, G. S., Suresh, S. Meta-cognitive RBF network and its projection based learning algorithm for classification problems. Applied Soft Computing. **1** (vol.13), pp. 654-666 (2013)
49. Tung, S. W., Quek, C., Guan, C. SaFIN: A self-adaptive fuzzy inference network. IEEE Transactions on Neural Networks. **12** (vol.22), pp. 1928-1940 (2011)
50. Suresh, S., Subramanian, K.: A sequential learning algorithm for meta-cognitive neuro-fuzzy inference system for classification problems. Proc. of the Int'l Joint Conf. on Neural Networks. USA, pp. 2507-2512 (August 2011)
51. Kadlec, P., Gabrys, B.: Architecture for development of adaptive on-line prediction models. memetic computing. **4** (vol.1), pp. 241-269 (2009)
52. Minku, F. L., Ludemir, T. B.: Clustering and co-evolution to construct neural network ensembles: an experimental study. Neural Networks. **9** (vol.21), pp. 1363-1379 (2008)
53. Filev, D. P., Angelov, P. P.: Algorithms for real-time clustering and generation of rules from data. In: Valente di Oliveira, J., and Pedrycz, W. (Eds.). Advances in Fuzzy Clustering and its Applications. John Wiley & Sons, Chichester, UK, (2007)
54. Amadou Boubacar, H., Lecoeuche, S., Maouche, S.: SAKM: Self-adaptive kernel machine: a kernel-based algorithm for online clustering. Neural Networks. **9** (vol.21), pp. 1287-1301 (2008)
55. Tan, J., Quek, C.: A BCM theory of meta-plasticity for online self-reorganizing fuzzy-associative learning. ieee transactions on neural networks. **6** (vol.21), pp. 985-1003 (2010)
56. Minku, F. L., Ludermir, T. B.: Evolutionary strategies and genetic algorithms for dynamic parameter optimization of evolving fuzzy neural networks. Proc. of the IEEE Congress on Evolutionary Computation. Scotland, pp. 1951-1958 (2005)
57. Yamauchi, K., Hayami, J.: Incremental leaning and model selection for radial basis function network through sleep. IEICE Transactions on Information and Systems. **4** (vol.e90-d), pp. 722-735 (2007)
58. Leite, D. F., Costa, P., Gomide, F.: Interval-based evolving modeling. Proc. of the IEEE Workshop on Evolving and Self-Developing Intelligent Systems. USA, pp. 1-8 (March 2009)
59. Leite, D. F., Costa, P., Gomide, F.: Evolving granular neural networks from fuzzy data streams. Neural Networks. (vol.38), pp. 1-16 (2013)
60. de Jesús Rubio, J.: Stability analysis for an online evolving neuro-fuzzy recurrent network. In: Angelov, P. P., Filev, D. P., Kasabov, N. K. (Eds.). Evolving Intelligent Systems: Methodology and Applications. John Wiley & Sons, Hoboken, New Jersey, USA (2010)
61. Kim, K., Whang, E. J., Park, C. W., Kim, E., Park, M.: A TSK Fuzzy inference algorithm for online identification. Proc. of the 2nd Int'l Conf. on Fuzzy Systems and Knowledge Discovery. China, pp. 179-188 (August 2005)

62. Zanchettin, C., Minku, L. L., Ludermir, T. B.: Design of experiments in neuro-fuzzy systems. international journal of computational intelligence and applications. **2** (vol.9), pp. 137-152 (2010)
63. Minku, F. L., Ludermir, T. B.: EFuNNs ensembles construction using a clustering method and a coevolutionary genetic algorithm. Proc. of the IEEE Congress on Evolutionary Computation. Canada, pp. 1399-1406 (July 2006)
64. Tung, S. W., Quek, C., Guan, C. eT2FIS: An evolving type-2 neural fuzzy inference system. Information Sciences. (vol.220), pp. 124-148 (2013)
65. O'Hara, B., Perera, J., Brabazon A.: Designing radial basis function networks for classification using differential evolution. Proc. of the Int'l Joint Conf. on Neural Networks. Canada, pp. 2932-2937 (July 2006)
66. Subramanian, K., Sundaram, S., Sundararajan, N.: A metacognitive neuro-fuzzy inference system (McFIS) for sequential classification problems. IEEE Transactions on Fuzzy Systems. **6** (vol.21), pp. 1080-1095 (2013)
67. Hernández, J. A. M., Castañeda, F. G., Cadenas, J. A. M.: An evolving fuzzy neural network based on the mapping of similarities. IEEE Transactions on Fuzzy Systems. **6** (vol.17), pp. 1379-1396 (2009)
68. Zhao, Q. L., Jiang, Y. H., Xu, M.: Incremental learning by heterogeneous bagging ensemble. Proc. of the Int'l Conf. on Advanced Data Mining and Applications. China, pp. 1-12 (November 2010)
69. Goh, H., Lim, J. H., Quek, C.: Fuzzy associative conjuncted maps network. IEEE Transactions on Neural Networks. **8** (vol.20), pp. 1302-1319 (2009)
70. Minku, F. L., Ludermir, T. B.: EFuNN ensembles construction using cone with multi-objective GA. Proc. of the 9th Brazilian Symposium on Neural Networks. Brazil, pp. 48-53 (October 2006)
71. Kasabov, N.: Adaptation and interaction in dynamical systems: modelling and rule discovery through evolving connectionist systems. Applied Soft Computing. **6,** 3, 307-322 (2006)
72. H.Widiputra, R. Pears, N. Kasabov: Dynamic interaction network versus localized trends model for multiple time-series prediction. Cybernetics and Systems. **42,** 2, 100-123 (2011)
73. Ghobakhlou A., M. Watts and N. Kasabov: Adaptive speech recognition with evolving connectionist systems. Information Sciences. **156,** 71-83 (2003)
74. Hodgkin, A. L., Huxley, A. F.: A Quantitative Description of Membrane Current and Its Application to Conduction and Excitation in Nerve. Journal of Physiology. **4** (vol.117), pp. 500-544 (1952)
75. Hopfield, J. J.: Pattern recognition computation using action potential timing for stimulus representation. Nature, **6535** (vol.376), pp. 33-36 (1995)
76. Izhikevich, E. M.: Which model to use for cortical spiking neurons? IEEE Transactions on Neural Networks. **5** (vol.15), pp. 1063-1070 (2004)
77. Thorpe, S., Delorme, A., van Rullen, R.: Spike-based strategies for rapid processing. Neural Networks. **6-7** (vol.14), pp. 715-725 (2001)
78. Verstraeten, D., Schrauwen, B., D'Haene, M., Stroobandt, D.: An experimental unification of reservoir computing methods. Neural Networks. **3** (vol.20), pp. 391-403 (2007)
79. Kasabov, N. K., Dhoble, K., Nuntalid, N., Indiveri, G.: Dynamic evolving spiking neural networks for on-line spatio- and spectro-temporal pattern recognition. Neural Networks. (vol.41), pp. 188-201 (2013)
80. Song, S., Miller, K. D., Abbott, L. F.: Competitive hebbian learning through spike-timing-dependent synaptic plasticity. Nature Neuroscience. **9** (vol.3), pp. 919–926 (2000)
81. Soltic, S., Kasabov, N. K.: Knowledge extraction from evolving spiking neural networks with rank order population coding. International Journal of Neural Systems. 6 (vol.20), pp. 437-445 (2010)
82. Chen, Y., Hu, J., Kasabov, N. K., Hou, Z., Cheng, L.: NeuroCubeRehab: A pilot study for eeg classification in rehabilitation practice based on spiking neural networks. Proc. of the 20th Int'l Conf. on Neural Information Processing. South Korea, pp.70-77 a(November 2013)
83. EU FP7 Marie Curie EvoSpike Project. INI/ETH/UZH, http://ncs.ethz.ch/projects/evospike (2011-2012)
84. Wysoski, S. G., Benuskova, L., Kasabov, N. K.: Evolving spiking neural networks for audiovisual information processing. Neural Networks. **7** (vol.23), pp. 819-835 (2010)
85. Schliebs, S., Defoin-Platel, M., Worner, S., Kasabov, N. K.: Integrated feature and parameter optimization for evolving spiking neural networks: exploring heterogeneous probabilistic models. Neural Networks. **5-6** (vol.22), pp. 623-632 (2009)
86. Kasabov, N. K.: NeuCube: A Spiking neural network architecture for mapping, learning and understanding of spatio-temporal brain data. neural networks. (vol.52), pp. 62-76 (2014)
87. Kasabov, N. K., Feigin, V. L., Hou, Z.-G., Chen, Y., Liang, L., Krishnamurthi, R., Othman, M., Parmar, P.: Evolving spiking neural networks for personalized modelling of spatio-temporal data and early prediction of events: a case study on stroke. neurocomputing. (vol.134), pp. 269-279 (2014)

88. Defoin-Platel, M., Schliebs, S., Kasabov, N. K.: Quantum-inspired evolutionary algorithm: a multimodel eda. ieee transactions on evolutionary computation. **6** (vol.13), pp. 1218-1232 (2009)
89. Hamed, H. N. A., Kasabov, N. K., Shamsuddin, S.: Probabilistic evolving spiking neural network optimization using dynamic quantum inspired particle swarm optimization. Australian Journal of Intelligent Information Processing Systems, **1** (vol.11), pp. 23-28 (2010)
90. Mead, C.: Analog VLSI and neural systems: Addison-Wesley Longman Publishing Co, Boston, Massachusetts, USA (1989)
91. *Mahowald, M., Mead, C.: The silicon retina. Scientific American.* ***264*** *(5), 76–82 (1991)*
92. Corradi, F. and Indiveri, G.: A neuromorphic event-based neural recording system for smart brain-machine-interfaces, biomedical circuits and systems. IEEE Transactions. **9:(5)** 699 - 709 (2015)
93. Binas, J. and Rutishauser, U. and Indiveri, G. and Pfeiffer, M.: Learning and stabilization of winner-take-all dynamics through interacting excitatory and inhibitory plasticity. Frontiers in Computational Neuroscience. **8:(68)** (2014)
94. Sawada, J., et al.: Truenorth ecosystem for brain-inspired computing: scalable systems, software, and applications. proc. of the int'l conf. for high performance computing, Networking, Storage and Analysis. USA, pp. 130-141 (November 2016)
95. Benjamin, B. V., et al.: Neurogrid: A mixed-analog-digital multichip system for large-scale neural simulations. Proceedings of the IEEE. 5 (vol.102), pp. 699-716 (2014)
96. Furber, S. B., Lester, D. R., Plana, L. A., Garside, J. D., Painkras, E., Temple, S., Brown, A. D.: Overview of the spinnaker system architecture. IEEE Transactions on Computers. 12 (vol. 62), pp. 2454-2467 (2012)
97. Mohemmed, A., Schliebs, S., Matsuda, S., Kasabov, N. K.: Training spiking neural networks to associate spatio-temporal input–output spike patterns. Neurocomputing. (vol.107), pp. 3-10 (2013)
98. Kasabov, N. K., Scott, N., Tu, E., Marks, S., Sengupta, N., Capecci, E., Othman, M., Doborjeh, M., Murli, N., Hartono, R., Espinosa-Ramos, J. I., Zhou, L., Alvi, F., Wang, G., Taylor, D., Feigin, V. L., Gulyaev, S., Mahmoudh, M., Hou, Z.-G., Yang, J.: Design methodology and selected applications of evolving spatio-temporal data machines in the neucube neuromorphic framework. Neural Networks. (vol.78), pp. 1-14 (2016)
99. Sengupta, N., Kasabov, N. K.: Spike-time encoding as a data compression technique for pattern recognition of temporal data. Information Sciences. (vol.406–407), pp. 133-145 (2017)
100. Kasabov, N. (ed): The Springer Handbook of Bio- and Neuroinformatics. Springer pp. 1230 (2014)
101. N. Sengupta, N. Kasabov: Spike-time encoding as a data compression technique for pattern recognition of temporal data. Information Sciences. 406–407, pp. 133–145 (2017)
102. Alvi,F., R. Pears, N. Kasabov: An evolving spatio-temporal approach for gender and age group classification with spiking neural networks, evolving systems, Springer (2017)
103. Kasabov, N.: From multilayer perceptrons and neuro-fuzzy systems to deep learning machines: which method to use? – A survey. Int. Journal on Information Technologies and Security. vol.**9**, No. 20, pp. 3-24 (2017)
104. Kasabov, N., M. Doborjeh, Z. Doborjeh: Mapping, learning, visualisation, classification and understanding of fMRI data in the neucube spatio temporal data machine. IEEE Transactions of Neural Networks and Learning Systems. vol. **28**,4, 887-899, DOI: 10.1109/TNNLS.2016.2612890 (2017)
105. Ge, C., N Kasabov, Z Liu, J Yang: A spiking neural network model for obstacle avoidance in simulated prosthetic vision, Information Sciences **399**, 30-42 (2017)
106. Khansama, R., V. Ravi, N. Sengupta, A.R. Gollahalli and N. Kasabov: Stock market movement prediction using evolving spiking neural networks, IEEE TNNLS, submitted 2017

IJBT (2017) 01: 66-84
RESEARCH ARTICLE

Upcoming research challenges in the financial services industry: a technology perspective

Raghu Krishnapuram[1] · Anirban Mondal[2]

Abstract Several key transformations in the macro environment, coupled with recent advances in technology, have opened up tremendous innovation opportunities in the financial services industry. However, many research challenges need to be addressed for realizing the full potential of innovation in financial services. Examples of such challenges include context-aware analytics over uncertain and imprecise data, design of user-friendly interfaces for improved expressiveness in querying financial service providers, and personalization based on fine-grained user preferences especially in the presence of sparse data. In this paper, we arrive at these research challenges based on an analysis of the macro environment and technology trends, and provide our vision and perspective on the same.

Keywords: Financial services industry · Macro environmental trends · Financial technology · Intelligent digital mesh · Innovations

1 Introduction

Over the past decade, there have been several key transformations across many aspects of the macro environment in which the financial services industry operates. Such aspects include demographics and social behaviour (e.g., prevalence of social media, significant increases in life expectancy, and emergence of the millennial generation), globalization (e.g., effect of natural and man-made disasters on global supply chains, protectionism, war/terrorism, and business process outsourcing), economics (e.g., recession markets, unstable labour markets, and higher unemployment rates), legal (e.g., stricter financial reporting and compliance requirements from Securities Exchange Commissions of different countries, and accounting policy shifts from GAAP to IFRS), and technology (e.g., the ever-increasing popularity of mobile devices/applications, ubiquitous network connectivity, and improved availability of cost-efficient platforms for analyzing big data). Such a transformation in the macro environment, coupled with data about such aspects (which was previously unavailable, especially in real-time), has opened up avenues for creating a major paradigm shift in terms of innovation in financial services as a key value proposition to end-users (both individuals as well as firms).

✉ Raghu Krishnapuram, Anirban Mondal
raghuk@msrit.edu, anirban.mondal@snu.edu.in

[1] Fellow, IEEE and Former Program Manager, Xerox Research Center, India
[2] Former Senior Research Scientist, Xerox Research Center, India

The macro environmental factors mentioned above have resulted in a significant amount of activity in the financial products and services space. “Financial technology” (abbreviated as “fintech”) refers to any technology-based innovation in the financial industry [1-2]. Innovations concerning fintech essentially cover the entire gamut of financial services ranging from loan/mortgage management to investment banking to virtual currencies (e.g., Bitcoin [3-4]). Notably, much of this activity is emanating from technology start-up firms that are primarily outside of the traditional financial sector. In order to keep up with the disruptive cross-boundary innovation plays by technology start-ups, traditional financial firms have started pumping funding into fintech companies. To put things into perspective, fintech companies across the world obtained funding to the tune of about $36 billion in 2016 [5]. For example, Wells Fargo provides mentoring for start-ups, while Bank of America has an annual conference for fintechs in Silicon Valley [6]. Furthermore, BBVA invested $13.5M on renovating an 80,000 square-foot operation centre in Birmingham for housing technology teams that create fintech applications [6].

In this paper, we review the macro environmental factors and technology trends, and then point out some of the key research challenges that need to be addressed for realizing the full potential of innovation in the financial services industry. We also summarize our perspective on the innovation opportunities that they open up. The remainder of this paper is organized as follows. Section 2 discusses key trends in the financial services industry as well as their ramifications based on an analysis of macro-environmental factors. Section 3 presents recent developments from a technology perspective. Section 4 identifies the key research challenges related to the innovation opportunities in financial services arising out of these factors and trends. It also provides our perspective on how these research challenges, if addressed, could benefit the industry. Finally, Section 5 provides the summary and conclusions.

2 Analysis of macro environmental trends impacting the financial services industry

This section provides an analysis of the macro environment trends in the financial services industry. The trends are summarized in Figure 1 based on the factors of the macro-environment. We now examine how these macro-environmental factors impact the financial services industry.

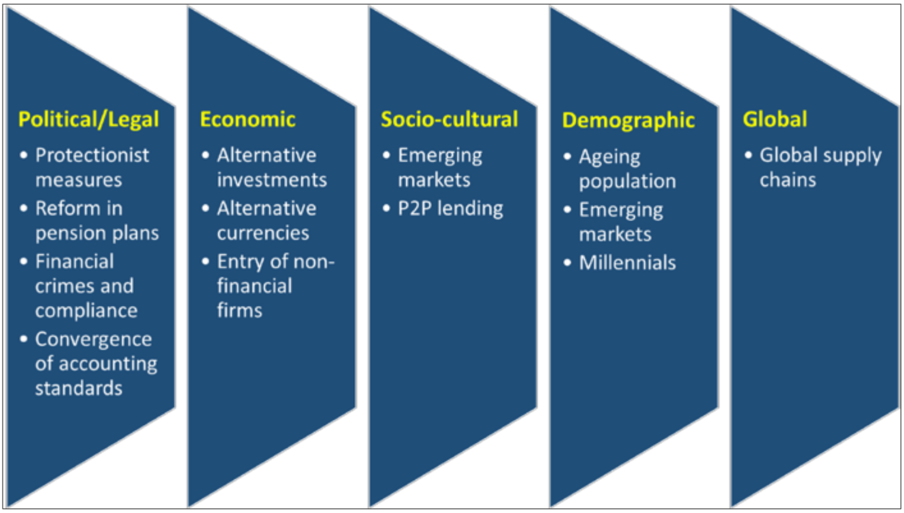


Fig. 1 Macro-environmental analysis of trends in the financial services industry

2.1 Political/legal factors

- **Protectionist measures:** In the recent years, governments have been becoming increasingly protectionist in an effort to safeguard the economic sustainability of their respective countries. Examples of protectionist-oriented measures include the creation of more trade barriers, increased import duties, and stricter policies on work visas. A recent example relates to the US administration's H1 visa restrictions [7]. Although some may argue that such protectionist measures may go against the principles of free market economics, the global financial crisis of 2008, the reality of recession markets, as well as the high rates of unemployment have compelled several governments across the world to adopt such measures. The protectionist trend is likely to continue well into the future as well. The implication of protectionism is that product/service differentiation would become a key factor for any given firm to enter the market of a particular country, given that advantages due to offering goods/services at a lower cost would be practically offset by protectionist measures.

  Interestingly, protectionist measures open up significant opportunities for research concerning the improvement of financial services and products. For example, consider the case of visa restrictions by any given country. In such a scenario, new tools and technologies can be developed for automated problem-solving and for remote troubleshooting. Furthermore, "intelligent" (see Section 3.1) software could facilitate remote monitoring of the systems for preventative maintenance, in order to reduce the chances of service failures and emergencies.

- **Reform in pension plans:** Over the past few years, there has been a trend in terms of a gradual shift from defined benefit pension plans towards defined contribution pension plans [8]. In case of defined benefit plans, employees typically receive a fixed set of retirement benefits that are not dependent upon market conditions. On the other hand, retirement benefits arising from defined contribution plans depend upon market conditions, thereby creating a significant amount of uncertainty for employees [9]. Given the current economic climate of recession and uncertainty in market conditions, where long-term employment at most firms is fast fading away, we believe that this trend is likely to continue. It is primarily because this trend shifts the financial risk to the employees as opposed to shifting it to the firms. It is to be noted that this shift increases the complexity of long-term investment risk management for retirement asset management firms, especially given that almost the entire portfolio of a given customer entails a high degree of volatility and uncertainty, i.e., no retirement benefits can be taken for granted anymore.

Given such complexity, automation of the analysis of long-term investment risk planning becomes key for effective investment decision-making. Consequently, this trend in pension plans opens up research opportunities towards creating novel financial services that incorporate a considerable amount of automated investment risk planning and decision-making. The challenge here is to design automated financial risk modelling and risk management approaches that are capable of simulating the way in which human financial experts make decisions based on the macro environmental context.

- **Financial crimes and compliance:** Over the last couple of decades, there has been an increase in financial crimes such as financial accounting fraud, securities fraud, credit card fraud, insurance fraud, Internet-based fraud and money laundering. Financial accounting fraud [10-11] perpetrated by companies such as Enron [12], WorldCom [13] and Madoff [14] have shaken public trust in stock markets, thereby resulting in increasingly stringent compliance requirements by regulatory authorities. An example of such compliance regulations is the Sarbanes-Oxley Act (SOX) [12, 15]. Ensuring compliance with respect to such stringent governmental regulations is time-consuming and expensive for firms. It is especially problematic for small to medium-sized businesses. Here, automated fraud detection technologies as well as technologies for creating automated financial reports and statements can play a prominent role in keeping the costs down for all firms. Building such automation technologies raises several interesting research challenges such as heuristics for detecting fraud, intelligent approaches for identifying cases of creative accounting and techniques for working with huge amounts of uncertain data from different modalities and identifying cross-connections and correlations.

- **Convergence of accounting standards:** Over the past decade, there have been several initiatives to combine the two widely used accounting standards, namely the US Generally Accepted Accounting Principles (GAAP) [16] and the International Financial Reporting Standards (IFRS) [17-18]. The goal is to create a unified framework of accounting standards to make the process of financial reporting easier across countries, especially given the importance of globalization [19-22]. Notably, these two accounting standards differ significantly in terms of their philosophy, thereby creating challenges for accountants. In other words, an accountant, who has been trained in US GAAP accounting standard, would find it challenging to understand financial statements written using the IFRS accounting standard, and vice versa. This opens up new research avenues for the creation of new technologies and software that are capable of automating the process of translating US GAAP-based financial reports and statements to IFRS and vice versa. The challenges here are around interpreting principles expressed in natural language and converting them to algorithms. Furthermore, GAAP and IFRS are philosophically different in that GAAP provides clear-cut quantifiable guidelines for financial accounting, while IFRS is more about principles. The implication is that translation of US GAAP-based financial reports and statements to IFRS (and vice versa) would often be subjective based on the financial experts' human interpretation. Modelling and incorporating such subjectivity into the translation algorithms would open up further opportunities for research.

2.2 Economic factors

- **Impact of alternative investment markets:** Alternative investments refer to investments that are not a part of the standard asset classes (e.g., stocks, bonds and cash). Examples of alternative investments include real estate, futures, derivative contracts and hedge funds [23]. The growth of alternative investment markets [24-25] have been fuelled by investors' increasing lack of trust in the stock markets due to events such as the global financial crisis of 2008 and multiple large-scale scams such as Enron [12], WorldCom [13] and Madoff [14]. Alternative investments can be used for diversifying a given portfolio to reduce risk and volatility, especially because they have little direct correlation with the standard asset classes. For example, Gravitas [26] serves as an outsourcing platform provider for the alternative investment industry. Their thought leadership event in 2015 was aimed at understanding the trends in the alternative investment industry. The areas of focus identified by Gravitas were (a) big data (b) branding (c) cyber security and (d) innovation. While these issues impact the alternative investment industry, they also impact other aspects of financial services, e.g., general (traditional) investment banking services. Hence, addressing these issues would be beneficial to other financial services as well.

- **Emergence of alternative currencies:** The emergence of alternative currencies has been becoming a major trend for the financial services industry. An example of alternative currency is Bitcoin [3-4], which is becoming increasingly popular. While we certainly share the optimism about alternative currencies, we also believe that regulatory frameworks are not yet adequate to deal with alternative currencies. In particular, we believe that the rate of adoption of alternative currencies would depend significantly upon how the regulatory authorities and large financial institutions respond to alternative currencies. This is because the opinions concerning alternative currencies vary across governments. Incorporating alternative currencies into the mainstream entails research challenges such as traceability, security, trust, privacy, non-repudiation protocols and distributed management of the transactional log data. On the other hand, the rewards for developing technologies that can overcome these problems are huge. As a result, many financial firms, start-ups, as well as traditional IT companies are investing heavily in this area. We will come back to this issue again in Section 3.2.

- **Entry of non-financial firms:** Over the past decade, the financial services industry has been witnessing the entry of firms that are not traditionally financial firms. These firms have been entering the financial services industry with their own expertise and they have been instrumental in continuously redefining the scope of financial services. Examples of new technologies introduced by such firms include Google wallet [27] for payments via mobile devices, Amazon's store card [28-29] for purchase financing purposes and NFC-enabled checkout terminals [30]. Furthermore, companies, such as AT&T and Verizon, are using NFC technologies for creating digital wallets in mobile phones [31]. Moreover, Google has been trying to enter the funds management business, while Apple with its huge cash reserves is also making forays into financial markets.

  Additionally, companies, such as Google, Apple and Amazon, have a significant amount of data about their customers. These companies have been increasingly using this data for analytics and bundling their core services with the financial services e.g., for increasingly personalized financial product and service offerings. Given the prevalence of mobile devices and mobile technologies, we believe that mobile services in finance will continue to grow and increase in importance with the key contributions and innovation most likely to come from non-financial firms.

  In essence, financial services may no longer remain the sole domain for banks and other traditional financial institutions, as the non-financial firms keep increasing their market share. This would lead to a large-scale integration of financial services with other domains such as retail and travel, thereby bringing in research challenges around the integration and analysis of data from multiple and disparate data sources, including social media.

2.3 Socio-cultural factors

- **Impact of emerging markets:** In emerging markets, two key trends include micro-financing/financial inclusion [32-33], and correspondent banking models [34]. Notably, in emerging markets, a significant percentage of the population are completely excluded from the financial and banking systems and they have little or no access to loans. Micro-financing concerns the small loans that are provided by an organization or private investors to small business owners for enabling financial inclusion and facilitating community development. More recently, the concept of micro-financing has become applicable to start-ups. Consequently, several technological innovations have started to occur in the micro-financing space. As an example, "Kiva" (www.kiva.org) is an online platform that essentially provides a match-making service between private investors and poor entrepreneurs in emerging markets.

Additionally, firms such as Allianz are focusing on innovations in the micro-insurance market.

Over the next decade, we believe that innovations due to technologies such as smartphones and improved credit scoring software (which rely on big data) will create significant positive market disruptions in the micro-financing space. This is also supported by the increasing drive towards entrepreneurship in the startup ecosystems. As more sophisticated data analytics techniques are developed, loan-default analysis would be likely to become more accurate, thereby increasing the trust of investors in the micro-financing space. Furthermore, since micro-financing has the potential to contribute significantly to community development, governments and regulatory agencies have an incentive to provide considerable tax benefits to investors in the micro-financing markets. Here, the role of electronic transactions (e.g., by using mobile devices) would become more prominent because electronic transactions can typically be tracked, thereby alleviating the concerns of governments about money laundering.

Incidentally, the correspondent banking model [34] is becoming increasingly popular towards facilitating financial inclusion for low-income consumers living in remote areas. Correspondents are representatives of a given bank, who are authorized to act on behalf of the bank e.g., in executing various kinds of transactions. Correspondents connect to the banking services requested by the consumers by means of mobile devices. Thus, the correspondent banking model essentially tries to provide consumers with access to banking services without necessitating their physical presence. Given that traditional banks cannot open branches in remote areas due to cost issues, the correspondent banking model has become a necessity to provide financial inclusion to consumers in these remote areas. Furthermore, such areas lack electricity and Internet connectivity, thereby precluding the possibility of installing ATMs.

We see a significant amount of trust issues with Correspondent banking models. For example, a correspondent could be fraudulent and dupe illiterate consumers of their hard-earned savings. We believe that a combination of mobile technologies, security and authentication technologies, as well as stronger government regulations would significantly facilitate in the success of Correspondent banking models over the next decade. Given the recent developments in “intelligent things” (See Section 3.1), it is not unthinkable that some of these issues might find a solution in autonomous vehicles and drones, which could replace correspondents.

- **Popularity of P2P lending:** Peer-to-Peer (P2P) lending [35-36] involves lending money among strangers as opposed to lending from a bank. Based on the study in [37] by Transparency Market Research, the size of the global P2P lending market will be $897.85 billion by the year 2024, i.e., significantly higher than its market size of $26.16 billion in 2015. The findings in [37] also revealed that the size of the global P2P lending market is expected to grow at a CAGR of 48.2% between 2016 and 2024. Websites specializing in P2P lending use automated technologies for activities such as credit checking and loan default predictions. In particular, web-based P2P lending firms typically incur considerably lower costs for providing lending services to consumers as compared to that of traditional banks primarily because of the automated technologies that they use. However, P2P loans are essentially unsecured. In case of any default on a given P2P loan, the lender has no recourse to any legal action.

Interestingly, P2P lending can also be seen as supporting entrepreneurship and innovation in emerging markets (and in some cases, in developed markets as well). In particular, crowd-funding-oriented websites allow potential entrepreneurs to pitch their ideas for novel products and services; then interested individuals can invest money on ideas, which they judge to be of high potential value. We believe that automated technologies for loan default predictions w.r.t. the overall macro-environmental context will play a key role in the near-term as well as long-term future towards further increasing the popularity of P2P lending.

2.4 Demographic factors

- **Impact of ageing population:** Based on the study in [38], the percentage of the world's population over the age of 60 would become almost double in the next 25 years. Ageing of the population has different ramifications for developed and emerging markets. In the case of developed markets, ageing-themed investing has been dramatically increasing in popularity [39]. Examples of ageing-themed investing include investments in healthcare, insurance companies and travel (e.g., cruises). On the other hand, for emerging markets, a huge number of elderly people tend to work with generally unpredictable incomes, thereby implying that reducing risk in their investments and assets becomes a necessity. The financial services industry could play a major role in managing the volatility of individual consumers' investments and assets (e.g., real estate) as well as the volatility of large-sized funds that contain the investments of a large number of firms and consumers. This requires a comprehensive contextual understanding of the environment in which these individual consumers' investments and assets, and large-sized funds operate. It also requires a what-if analysis of different kinds of scenario analysis and reasoning about how the investments would play out under a wide gamut of potential circumstances and events. Hence, in this space, we see significant research opportunities for contextual analysis and fine-grained understanding of financial data.

- **Impact of emerging markets**: In emerging markets such as India and China, the dramatic pace of economic growth has contributed to the middle class expanding at a rapid pace. This trend would imply an increased demand for services towards financing home and car loans as well as investment recommendations for small amounts of money. Notably, such services were originally designed for a relatively small percentage of the population. However, the huge and rapidly expanding size of the middle class in emerging economies, such as India and China, implies that these services need to be provided to a much larger population size than ever before. The ability to provide these services at scale entails some key research challenges related to cost considerations. Here, automated technologies need to be developed to facilitate economy of scale in providing these services to a large number of consumers. For example, it would be practically unviable to provide individualized/personalized customer relationship managers for everyone in a large population. However, the automated technologies would make it feasible for large population sizes to leverage value from these financial services primarily due to economies of scale. This also requires a significant amount of research in predictive analytics in financial services for understanding the huge amounts of data that are generated from multiple and disparate sources.

There has been a rapid growth in the number of millionaires in Asia. Based on the information in [40], at the end of the year 2015, the net worth of Asian millionaires was $17.39 trillion as compared to North America's $16.61 trillion. This market segment of Asian millionaires will increase the demand for personalized advice concerning investments and services associated with the search, sales and financing of high-end products and services. Furthermore, financing of loans for high-end homes would also need to be provided as a service to this market segment. As this market segment becomes larger in size, automation of these financial services would become a necessity to offer the services in a personalized manner. This would create new research challenges and opportunities again with respect to technologies related to digital personal assistants, natural language interfaces, scalability, data collection and analytics, which can be incorporated into “intelligent things” (see Section 3.1).

- **Emergence of the millennials:** According to a Goldman report [41-42], Millennials will inherit $30 trillion in wealth. This marks a paradigm shift for investment management firms. This is because multiple studies have demonstrated that Millennials tend to rely on information from social networks and the results of their own online research for making financial investment decisions. Notably, this is in contrast with other generations (e.g., Generation X), which typically use the services of financial investment specialists for making their investment decisions. In fact, the survey in [43] indicates that 67% of affluent Millennials would find it acceptable to obtain services from non-financial firms as compared to 45% for affluent Gen X consumers. Furthermore, 91% of affluent Millennials would seek information about financial investments from social networks, while only about 53% of affluent Gen X consumers would be willing to make investments based on information obtained from social networks. Hence, we believe that the emergence of Millennials and their huge $30 trillion wealth inheritance would lead to innovations in financial investment services. Such innovations would be associated with the collection and analysis of huge amounts of data from social media.

2.5 Global factors

- **Impact of global supply chains:** Trade in supply chains accounted for about $20 trillion i.e., 60% of the global trade and approximately 30% of the world’s GDP in 2013 [44]. Given the significant amount of globalization in today’s world, global supply chains are becoming more common and are typically associated with multiple countries across the world [45]. Given that global supply chains span across geographies, events in any given part of the world can impact the entire global supply chain in an adverse manner. Examples of such events include earthquakes, tsunamis, and regional conflicts. Given that firms typically have marketable securities (investments made by a given firm in other firms) in their respective investment portfolios, the complexity of risk management planning increases dramatically for financial investment firms, especially given the wide gamut of potential exposure due to most major global/local events. This trend again opens up avenues for research in predictive analytics and reasoning under uncertainty for determining the probability and consequences of events that may have considerable impact on global supply chains. Some initial work in this direction may be found in [46-48].

3. Recent Developments in Information Technology

Recent developments in information technology can be summed up in three terms: “intelligent,” “digital,” and “mesh”, i.e., intelligent digital mesh is the way the technology is moving. See for example [49-52].

3.1 Intelligent

- **The trend towards “intelligent things”:** This means that almost any device or artifact that we use in our daily life or professional life could become intelligent, capable of learning and adaptation.

  The implication is that artificial intelligence and machine learning would become ubiquitous and will be part of our daily lives. The advent of sophisticated machine architectures and associated learning algorithms, combined with computing power to deal with massive amounts of data as well as the availability of a variety of data (including real-time) from a variety of sources, have the potential to provide huge business value. Access to cheaper and deeper machine learning and advanced chip manufacturing technology (such as the 10-nanometer technology) [50] means that intelligence can be embedded in massive-scale systems as well as in very small end devices. Intelligent apps such as personal assistants that can handle natural language or other immersive interfaces, as well as analytics-enabled features that are integrated into many processes such as ERP, marketing and end-user services will enhance efficiencies of and improve customer experience. “Intelligent things” such as safe robots, drones, autonomous vehicles, and even air taxis [50] are expected to dominate as businesses become more digital. Internet-of-Things (IoT) devices will evolve into “intelligent things” with powerful capabilities.

- **Implications of “intelligent” to financial services:** Financial organizations can be expected to move out of traditional machine learning niche areas such as automated high-frequency trading and venture into broader areas such as fraud and abuse detection, compliance (such as Know Your Customer, or KYC), as well as risk data aggregation and risk modeling. These are some of the areas that can benefit immensely from the “intelligent” trend. These applications will need to analyze large amounts of diverse data to extract hidden information and to identify complex patterns. The issues related to fraud were eloquently illustrated by the 2016 Wells Fargo account scandal [53]. The idea of “intelligent things” that can send and receive information from the ‘edge’ via a variety of sources such as ATMs, mobile phones, trade centres and financial markets can lead to new types of application scenarios such as real-time risk computation and monitoring. Intelligent assistants are potential game changers that could replace relationship managers and financial advisers. We may also imagine other intelligent things, such as autonomous vehicles and drones, providing banking services in remote areas. However, there are several research as well as regulatory challenges related to privacy and security that will limit some of these applications. We will discuss this again in Section 4.

3.2 Digital

- **The developments towards “digital”:** This means that digital and physical worlds will become increasingly intertwined and eventually seamlessly integrated. Virtual and augmented reality which allows people to simulate and experience situations, will create new business models in the future in terms of providing innovative services. Research challenges in this area are related to freeing up the user from the requirement of wearing appendages such as glasses and capturing the user input (or intent) on a device without a keyboard or a touchscreen (See for example [54]). Digital twins, namely, dynamic software models or physical entities including humans (or organs of humans), will allow us to model real-world situations accurately, leading to a better understanding of the world, including analysis, monitoring and control.

- **Blockchain technology:** Blockchain technology, also known as distributed ledger technology, is also part of “digital”. The technology has its origins in cryptocurrencies (Bitcoin). However, it has now become mainstream, and has the potential to make a substantial impact. The technology works by maintaining identical copies of a digital, distributed transaction ledger in the computers of all participating members. It functions by grouping transactions into blocks and then chaining the blocks. Historical transactions cannot be altered or tampered with, because the links between blocks and their content are protected by cryptography.

- **Implications of “digital” to financial services:** Digital twins will provide financial organizations the capability to understand their customers much better, and thus increase customer satisfaction levels through personalization. Similarly, virtual and augmented technologies have the potential to transform employee training and customer experience. Blockchain technology, which maintains identical copies of transaction history, can bring more trust, better access and transparency to all sorts of value-exchange transactions [55-56].

3.3 Mesh

- **Evolution of the mesh:** The evolution of the digital mesh (which provides dynamic connectivity across people, processes and things), will enable many more applications involving real-time responses. This will be accelerated by technologies such as 5G [57], which will increase the speed of connectivity by an order of magnitude. Conversational systems that use a variety of modalities for interaction (voice, sight, text, touch, gesture) will enable new types of interactions and services. The development of flexible (e.g. fold-up) interfaces will facilitate easy adoption of such technologies. Multi-channel solution architectures that are built on cloud and serverless computing and use APIs and services can be used to deliver dynamic and flexible solutions. Digital technology platforms and services will help build the digital business, streamlining many aspects such as business ecosystems and customer experience.

- **Implications of "mesh" to financial services:** Multi-channel solution architectures, combined with multi-modality interaction, can produce powerful solutions for automating many aspects of financial services such as loan underwriting, reconciliation, risk model development, real-time monitoring, automated portfolio management, and customer relationship management.

4 Research challenges

In this section, we summarize the top research challenges that we believe need to be addressed in order to realize the full potential of the innovation opportunities in the financial services industry. We also point out the typical applications that will be enabled by finding solutions to these challenges:

1. **Blockchain technology:** Blockchain technology is projected to cause the next big transformation in financial services. However, in order to realize its vision, many open problems need to be addressed. Some critical issues are: scalability with respect to performance and latency requirements, extension to scenarios other than cryptocurrency (such as contracts and licensing), and security. If the research challenges mentioned above are addressed, blockchain technology has the potential to reduce the costs of transfer-of-value transactions (particularly across borders), and also fraud. They can also speed up the settlement process in share trading. By pushing identity management to the blockchain, banks can provide a wider range of quick and reliable options to customers for validating identities by eliminating intermediaries. Because blockchains can record anything of value, including contracts, they can also facilitate smarter (intelligent) contracts, where the contracting process as well as the enforcement of the obligations of the contract can be automated. Many banks are already projecting billions of dollars in savings by moving to blockchain technology.

2. **Analytics and machine learning on big data:** Big data is now ubiquitous and financial services industry is no exception. However, in order to derive the full benefits out of big data, we need to address several problems in advanced analytics and machine learning. These include how to reduce requirements for labelled data (especially in the case of deep learning approaches), how to learn structural as well as complex (e.g. 3D) invariances, and how to perform automated complex reasoning. Deep learning approaches are still plagued by problems such as providing theoretical guarantees, quantifying model complexity, tuning hyper-parameters, and designing architectures to perform multiple tasks. Issues related to analyzing large amounts of financial data include storing personal and private information (e.g., facial images or other biometric data), capture, transferring and storing big data. Other challenges include how to guarantee accuracy or correctness in an automated system. While it is possible to divert specific cases to a human when they are too complex, it is not always possible to identify cases to be sent to human operators automatically. For example, if a machine learning algorithm classifies a certain input with high confidence, we may assume that the algorithm provides the correct answer, but there is no guarantee, due to false positives. Much of financial data is stored in disparate locations in a variety of formats, based on the application and the process in which it is used. This leads to challenges in data storage, transfer, as well as filtering by interestingness or representativeness to reduce computational burden. In the case of analytics, low-complexity, adaptive, sequential, and multi-scale versions of algorithms that are amenable to running on scalable parallel architectures are needed to handle high-velocity streaming data [58-59]. In the case of integration with question answering systems, other analytics problems to be solved are goal-oriented knowledge discovery and context-aware analytics.

   Solving the research problems mentioned above will help address many issues in financial services. For example, analytics over huge amounts of disparate data is essential for automatic detection of crime and compliance. Similarly, detection of fraud and identifying cases of creative accounting will involve working with huge amounts of uncertain and imprecise data from different modalities and discovering cross-connections and correlations. Also, as mentioned in Section 2.2, entry of firms from non-financial domains (such as retail and travel) into the financial domain brings in a variety of challenges around integration and analysis of data from multiple and disparate data sources, including social media. Big data analytics can also facilitate remote monitoring and preventative maintenance of systems, thus reducing the cost of services. This is particularly useful in the era of protectionism. Big data analytics and machine learning provide tools for understanding trends in alternative investment markets, and can potentially help design services that provide automated financial investment advice to specific demographic groups such as the Millennials.

3. **Cyber security and privacy:** Security and privacy issues continue to be of paramount importance in financial applications that run in an intelligent digital mesh environment. In spite of a large amount of effort put into addressing many problems in this area, several challenges remain. Security is not just

limited to storage and transmission of data. Solutions need to integrate security features at every stage. Layered security architectures that use entity behaviour analytics will need to be designed and developed to monitor suspicious activities. Also, owing to many regulatory challenges, much more progress in privacy-preserving analytics will be needed if we were to automate services such as financial advice. Solving problems related to security will also help in problems related to emerging markets. For example, a combination of mobile technologies, security and authentication technologies, as well as stronger government regulations would significantly facilitate the success of the correspondent banking models.

4. **Intelligent things:** Intelligent assistants are potential game changers that could replace relationship managers and financial advisers. One of the challenges in this area is to combine natural language understanding and conversational technologies to design effective dialog systems that are able to query databases or initiate query-specific analytics (e.g. goal-oriented knowledge discovery and context-aware analysis) that could help synthesize the answer for a user query. In many cases, the answer could also be in the form of a plot or a graph. Other difficult problem to solve is how to generate recommendations on the fly that make sense given the time, geography and other contextual information about the user. Other intelligent things such as self-driving vehicles and drones could have potential applications in delivering services that need physical contact. This could be particularly useful in branchless banking and in reaching out to remote customers in emerging markets. We believe that a combination of mobile technologies, security and authentication technologies, as well as stronger government regulations would significantly facilitate in the success of correspondent banking models over the next decade.

5. **Natural language understanding and conversational interfaces:** Natural language understanding has made a lot of progress, and has been deployed in scenarios such as question answering in restricted domains. Much of current work is based on statistical methods that requires large amounts of data. On the other hand, many applications in financial services, such as translation of US GAAP-based financial reports and statements to IFRS (and vice versa) will require much deeper understanding of the text (including many nuances) and in the context in which the text needs to be interpreted. The context may include information about the entity about which the report is written, as well as other rules, policies and laws that may apply. The translation process is often subjective and is based on the expert's interpretation. Natural language understanding algorithms that incorporating domain knowledge, context, and other information into would open up further applications in the financial domain. Challenges in conversational technologies include paraphrase detection, deducing the context from the earlier conversations, and discovering user intent.

6. **Modeling and reasoning in the context of complex dependencies:** Many problems in the financial services industry involve analyzing highly intricate relationships between various financial parameters, market variables, events,

and entities. Changes in the macro environment ripple through the network of such complex relationships, and if not properly modelled and analyzed, can manifest themselves in unexpected ways. Examples of such problems are automated investment risk planning and decision-making in the context of managing pension plans (see Section 2.1), managing individual consumers' investments in the context of the ageing population, and assessing impact of events on global supply chains. All of these problems require comprehensive models that are based on contextual understanding of the environment. The models should be capable of performing predictions and reasoning under uncertainty, because, very often, the information about changes in the macro environment is not precise and reliable.

7. **Personalization technologies:** Personalization based on fine-grained user preferences, especially in the presence of sparse data, is still a technical challenge. High levels of personalization will need fine-grained segmentation of the customer base. Other dimensions such as personas and geography will further complicate the issue. We also need to create customized content for each segment. Effective matching of the content to the user without storing (or caching) multiple copies of the content will bring in issues related to scalability. Identifying the segment to which a new (or anonymous) user may belong is also a non-trivial problem. Ideally, a "digital twin" needs to be created to understand and model a customer, so that appropriate recommendations can be made in at the right place, right time, and the right context. However, given the restrictions imposed by privacy laws, there are limits to what kind of data may be collected to build such twins. Solving the personalization problem is crucial for several applications such as intelligent agents for personalized financial advice.

8. **Automated technologies for loan default predictions:** Estimating reliable creditworthiness scores and predicting loan delinquency or default based on peer-to-peer interactions, transaction history, other online behavior (including social media data, when available), and limited demographic data is still a largely unsolved problem. The solution to this problem will play a key role in the near-term as well as the long-term future towards making P2P lending more attractive to investors and participants.

9. **Virtual and augmented reality technologies for customer interactions, training and remote support:** In many situations, physical presence of an agent or an employee of a financial organization to address a customer problem may not be feasible. This can be due to protectionist measures, cost of travel, availability of personnel, cost of training agents, etc. The challenge is to develop technologies based on virtual/augmented reality and intelligent agents to handle such situations. For example, new tools based on these technologies can be developed for providing self-help to customers for problem-solving and trouble shooting. Similarly, "virtual banks" can provide a scalable solution to the problem of branchless banking. Design of user-friendly interfaces (other than key pads) for collecting user input is critical for the success of solutions based on virtual and augmented reality.

5 Conclusion

We are living in exciting times of dramatically fast-paced innovation in the financial services industry in the form of novel financial products and services. Given this background, we have reviewed the macro-environmental factors and technology trends that are giving rise to this innovation, and identified some of the key research (technical) challenges that need to be overcome if we want to realize the full potential of the innovation opportunities. We envisage a future in which such financial products and services would be offered increasingly by a wide gamut of stakeholders (including start-ups) outside of the traditional financial institutions to meet the high expectations and preferences of newer market segments that are emerging, (e.g., the Millennials and the ageing population). The day may not be far when a person in a remote location might “walk into” a virtual bank to order a credit card that may be delivered by a drone, or a personal assistant might help manage the investments of a Millennial.

We believe that the solutions developed in response to the research challenges faced by the financial services industry would also be broadly applicable (albeit possibly with some modifications) to other industry domains such as retail, travel & tourism, healthcare, and transportation. As a single instance, the research advances in natural language processing and contextualization of textual data may very well be applied to facilitate users in determining appropriate healthcare vendors that are aligned with their preferences or to automate insurance claims. We hope that researchers from academia, industry and other government organizations will come together to realize these innovations in the financial services space.

References

1. Investopedia article on Fintech. http://www.investopedia.com/terms/f/fintech.asp
2. Sironi, P. (ed) Bibliography, in FinTech Innovation: From Robo-Advisors to Goal Based Investing and Gamification. John Wiley & Sons, Ltd., Chichester, UK. doi: 10.1002/9781119227205.biblio (2016)
3. Clay A, Camfield J.: Big ideas for a new economy: Social currency unleashed. http://www.fastcoexist.com/1679259/big-ideas-for-a-new-economy-social-currency-unleashed. (2012)
4. Blundell-Wignall A.: The bitcoin question: currency versus trust-less transfer technology. OECD Working Papers on Finance, Insurance and Private Pensions, No. 37, OECD Publishing. http://dx.doi.org/10.1787/5jz2pwjd9t20-en (2014)
5. Mesropyan E.: Global fintech funding reached $36 bn in 2016 with payments companies securing 40% of total funds. https://letstalkpayments.com/global-fintech-funding-36-bn-2016/. (2017)
6. Crosman P.: BBVA compass takes emulating startups to another level. http://www.americanbanker.com/news/bank-technology/bbva-compass-takes-emulating-startups-to-another-level-1074384-1.html. (2015)
7. H1-B visa rules: Trump tightens the noose on US firms; entry-level Indian techies to bear the brunt. Business Today.http://www.businesstoday.in/current/economy-politics/h1-b-visa-rules-donald-trump-us-firms-indian-techies/story/249335.html. (2017)
8. Broadbent J., Palumbo M., Woodman E.: The shift from defined benefit to defined contribution pension plans - implications for asset allocation and risk management. Working Group on Institutional Investors, Global Savings and Asset Allocation established by the Committee on the Global Financial System. https://www.bis.org/publ/wgpapers/cgfs27broadbent3.pdf (2006)
9. McWhinney JE.: The demise of the defined-benefit plan. http://www.investopedia.com/articles/retirement/06/demiseofdbplan.asp. Accessed 20 June 2017

10. The 10 Worst Corporate Accounting Scandals of All Time. http://www.accounting-degree.org/scandals/. Accessed 20 June 2017
11. Brody RG, Melendy SR, Perri FS (2012) Commentary from the American Accounting Association's 2011 Annual Meeting Panel on Emerging Issues in Fraud Research. Accounting Horizons: September 2012, 26(3): 513-531. https://doi.org/10.2308/acch-50175
12. Asthana S, Balsam S, Kim S (2009) The effect of Enron, Andersen, and Sarbanes-Oxley on the US market for audit services. Accounting Research Journal 22(1): 4-26. https://doi.org/10.1108/10309610910975306
13. Kuhn JR Jr., Sutton SG.: Learning from worldcom: implications for fraud detection through continuous assurance. Journal of Emerging Technologies in Accounting (3): 61–80, (2006)
14. Tomasic R.: The financial crisis and the haphazard pursuit of financial crime. Journal of Financial Crime, (18)1: 7-31, https://doi.org/10.1108/13590791111098771 (2011)
15. The laws that govern the securities industry. US Securities and Exchange Commission. http://www.sec.gov/about/laws.shtml. Accessed 20 June 2017
16. Kothari SP., Ramanna K., Skinner DJ.: Implications for GAAP from an analysis of positive research in accounting. Journal of Accounting and Economics. Elsevier 50(2-3): 246-286 (2010)
17. Nobes C.: The survival of international differences under IFRS: towards a research agenda. Accounting and Business Research 36(3): 233-245 (2006)
18. Bradshaw M., Callahan C., Ciesielski J., Gordon EA., Hodder L., Hopkins PE., Kohlbeck M., Laux R, McVay S., Stober T., Stocken P., Yohn TL.: Response to the SEC's proposed rule-roadmap for the potential use of financial statements prepared in accordance with international financial reporting standards (IFRS) by U.S. Issuers. Accounting Horizons (24)1: 117-128. (2010)
19. Pologeorgis N.: The impact of combining the U.S. GAAP and IFRS. http://www.investopedia.com /articles/economics/12/impact-gaap-ifrs-convergence.asp. Accessed 20 June 2017
20. FASB. Progress report on commitment to convergence of accounting standards and a single set of high quality global accounting standards. http://www.fasb.org/cs/ContentServer?site=FASB&c= Document_C&pagename=FASB%2FDocument_C%2FDocumentPage&cid=1176156953931 (2010)
21. Jeffers AE., Mengyu W., Askew S.: The switch from US GAAP to IFRS. Proceedings of the Northeast Business & Economics Association: 48–54 (2010)
22. Arens AA., Loebbecke JK.: Auditing, an integrated approach. Ed.7. Prentice-Hall Series in Accounting. (1997)
23. Alternative investment. http://www.investopedia.com/terms/a/alternative_investment.asp Accessed 23 June 2017
24. Jeffrey M.: A decade of alternative investment market (AIM) – but little impact on the UK apparel sector. Journal of Fashion Marketing and Management: An International Journal 11(3): 441-444. https://doi.org/10.1108/13612020710763164 (2007)
25. Fuglie K et al.: Agricultural research and development: public and private investments under alternative markets and institutions. Agricultural Economic Report/U.S. Department of Agriculture (no. 735), Publisher: [Washington]: U.S. Department of Agriculture, 1996
26. Turner C.: Alternative investment trends for 2015 and outlook for 2014. http://www.hedgethink.com/resources/alternative-investment-trends-for-2015-and-outlook-for-2014/. (2014)
27. Chandler N (2012) What is Google Wallet? http://electronics.howstuffworks.com/google-wallet1.htm. Accessed 23 June 2017
28. Amazon Store Cardholder. https://www.amazon.com/iss/credit/storecardmember/ref=storecard_van. Accessed 23 June 2017
29. James K.: The new amazon store card gives prime users 5% cash back. http://twocents.lifehacker.com/the-new-amazon-store-card-gives-prime-users-5-cash-bac-1694419150. (2015)
30. Profis S.: Everything you need to know about NFC and mobile payments. http://www.cnet.com/how-to/how-nfc-works-and-mobile-payments/. (2014)
31. Gross G.: Verizon, AT&T, T-Mobile form joint digital wallet venture. http://www.computerworld.com/article/2514394/enterprise-applications/verizon--at-t--t-mobile-form-joint-digital-wallet-venture.html. (2010)
32. Peachey S., Roe A.: Access to finance: what does it mean and how do savings banks foster access. World Savings and Retail Banking Institute (WSBI) (2006)
33. Tyson J.: FINCA international CEO on the future of microfinance. https://www.devex.com/news/finca-international-ceo-on-the-future-of-microfinance-85615. (2015)

34. Kishore A.: Business correspondent model boosts financial inclusion in india, https://www.minneapolisfed.org/publications/community-dividend/business-correspondent-model-boosts-financial-inclusion-in-india. (2012)
35. Wikipedia article. http://en.wikipedia.org/wiki/Peer-to-peer_lending. Accessed 23 June 2017
36. Uster S.: P2P lending will be a game-changing evolution if we don't curb innovation. http://www.theglobeandmail.com/report-on-business/rob-commentary/p2p-lending-will-be-a-game-changing-evolution-if-we-dont-curb-innovation/article24715236/. (2015)
37. Bajpai P.: The Rise of Peer-To-Peer (P2P) Lending. http://www.nasdaq.com/article/the-rise-of-peertopeer-p2p-lending-cm685513. (2016)
38. The world's growing aging population presents financial challenges and market opportunities for public and private sectors. Center for Financial Inclusion at Accion. http://www.prnewswire.com/news-releases/the-worlds-growing-aging-population-presents-financial-challenges-and-market-opportunities-for-public-and-private-sectors-300033331.html. (2015)
39. Dyson R.: How to make money from the world's ageing population. http://www.telegraph.co.uk/finance/personalfinance/investing/10191182/How-to-make-money-from-the-worlds-ageing-population.html. (2013)
40. McCarthy N.: Asian millionaires now control more wealth than those in north america. https://www.forbes.com/sites/niallmccarthy/2016/06/28/asian-millionaires-now-control-more-wealth-than-those-in-north-america-infographic/#2076154941d3. (2016)
41. Data Story Millennials. Goldman Sachs Report. http://www.goldmansachs.com/our-thinking/pages/millennials/index.html. (2017)
42. The $30 trillion challenge: what will millennials do with their forthcoming $30 trillion inheritance? Morgan Stanley Wealth Management Report: https://www.morganstanley.com/ ideas/30-trillion-challenge. (2015)
43. Shin L.: How millennials' money habits could shake up the financial services industry. http://www.forbes.com/sites/laurashin/2015/05/07/how-millennials-money-habits-could-shake-up-the-financial-services-industry/. (2015)
44. Global value chains: investment and trade for development. World Investment Report. UNCTAD. http://unctad.org/en/PublicationsLibrary/wir2013_en.pdf. (2013)
45. Seifert RW, Bochukova P.: Global value chains under threat. http://www.imd.org/research/challenges/TC029-14-global-value-chains-under-threat-seifert-bochukova.cfm. (2014)
46. Kenkre S, Kondapally R, Pandit V.: Algorithmic Aspects of Planning under Uncertainty for Service Delivery Organizations. Proceedings of the International Conference on Service Oriented Computing (ICSOC): 615-622 (2012)
47. Karthik S, Kenkre S, Narayanam K, Pandit V.: Resiliency Analytics Framework for Service Delivery Organizations. SRII Global Conference. (2012)
48. Karthik S, Kenkre S, Narayanam K, Pandit V.: Recourse aware resource allocation for contingency planning in distributed service delivery, IEEE Conference on Service Operations, Logistics and Informatics (SOLI) (2012)
49. Panetta K.: Gartner's Top 10 Strategic Technology Trends for 2017. http://www.gartner.com/smarterwithgartner/gartners-top-10-technology-trends-2017/. (2016)
50. Here are some technologies you'll be reading about this year. IEEE Spectrum, 21-51. (2017)
51. Press Release: Gartner Identifies the Top 10 Strategic Technology Trends for 2016. October 2015, http://www.gartner.com/newsroom/id/3143521. (2015)
52. Special Report: 2016 Top Tech to Watch. IEEE Spectrum, (2016)
53. Egan M.: 5,300 Wells Fargo employees fired over 2 million phony accounts. http://money.cnn.com/2016/09/08/investing/wells-fargo-created-phony-accounts-bank-fees/index.html#sthash.Y80TJ32O.dpuf. (2016)
54. Hesselmann T., Boll S., Heuten W. SCIVA: designing applications for surface computers. Proceedings of the 3rd ACM SIGCHI Symposium on Engineering Interactive Computing Systems: 191-196. (2011)
55. Tapscott A., Tapscott D.: How blockchain is changing finance. Harvard Business Review. https://hbr.org/2017/03/how-blockchain-is-changing-finance (2017)
56. Yli-Huumo J., Deokyoon K., Choi S., Park S., Smolander K.: Where is current research on blockchain technology?—a systematic review," Plos One, http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0163477 (2016)
57. Wikipedia article on "5G". https://en.wikipedia.org/wiki/5G. (2017)

58. Najafabadi MM, Villanustre F, Khoshgoftaar TM, Seliya N, Wald R, Muharemagic E: Deep learning applications and challenges in big data analytics. Journal of Big Data. 2015.https://journalofbigdata.springeropen.com/articles/10.1186/s40537-014-0007-7(2015)
59. Isah H: Research opportunities and challenges in big data. BigDat2015 Lessons, University of Bradford, http://www.computing.brad.ac.uk/~hisah/docs/ACM_BigData_Talk.pdf (2015)

IJBT (2017) 01: 85-88
PRACTITIONER'S PERSPECTIVE

Disruptive innovation in the financial sector

Duvvuri Subbarao[1]

Abstract Disruptive innovations have been happening in the field of finance as a consequence of a magical combination of technology and financial engineering. The Indian financial sector too has seen disruptive innovation not just in payment systems but also in the segments of deposits, credit and insurance. This article discusses the key drivers behind this disruptive innovation in Indian financial sector including internet penetration, usage of smartphones, improved online presence and policy initiatives. Further, the article elaborates the disruptive innovation that is presently caused by FinTech companies in accelerating the pace of change and reshaping the financial services industry radically.

Keywords: Disruptive innovation · FinTech · Indian Finance · Technologies · Demonetisation

1 Introduction

When we look at disruptive innovation today, we can't but acknowledge the big elephant in the room – demonetisation in which 86% of the currency was delegalised overnight [1] in what is arguably one of the most cash-intensive economies in the world.

While the objectives, costs and benefits of demonetisation are a subject matter of contentious debate, what is not contentious is that this has possibly been the most disruptive policy innovations in India since the 1991 reforms.

We owe to the celebrated economist Joseph Schumpeter who gave us the phrase – creative destruction [2], which he called – a process or mutation that revolutionises the economic structure from within – destroying the old one and creating a new one.

The smartphone, for example, is a creative destruction because it killed the market not just for regular cell phones, but also for PDAs, point and shoot cameras, wrist watches, calculators, voice recorders, planners and even music systems.

Demonetisation is in that sense creative destruction too, but it is a unique type of creative destruction because what it aims to destroy is itself a destructive creation. So, demonetisation is a creative destruction of a destructive creation.

✉ Duvvuri Subbarao
subbarao@gmail.com
[1] Former Governor, Reserve Bank of India

2 FinTech revolution

The world of finance has witnessed rapid disruptive innovation over the last decade as a consequence of a magical combination of technology and financial engineering. Tasks once handled by paper money and bulky computers are now being accomplished entirely on digital interfaces.

Even as late as 2008, one of the problems was the duration taken to clear outstation cheques. Now, a generation of millennials is growing up without knowing what a cheque is. Large number of young enthusiastic entrepreneurs are taking aim at the heart of financial sector and are revolutionising how financial transactions are carried out. Borrowing money through peer-to-peer lending club, transferring money through independent payment systems rather than through a bank and investing based on online customised advice tailored to individual needs are all made possible by start-ups [3]. Fighting cybercrime with help of financial services company that checks customer identities based on biometrics, behavioural profiling, push notifications and analytics across all channels – from mobile phones to physical branches - is a great illustration of the emergence of FinTechs [4].

One could get into a taxi, go to the destination and walk away – the billing is automatically taken care of. One could go into a restaurant, wave the phone and touch the thumb for biometrics and walk away. This is how fintech is revolutionising financial services.

FinTechs will change financial sector in many ways. They will cut costs and improve the quality of financial services, unburdened by legacy IT systems, large branch networks and traditional regulation. These companies would explore new ways of assessing risk through social media review and usage of logistic firms. For instance, motor insurance risk gets assessed not just by how much one drives and but also by how one drives. They will be able to diversify risk without the burden of geography.

3 Disruptive innovation beyond finance

Disruptive innovation extends beyond Fintech. Think about the concept of microfinance which was in existence for a long time but took centre stage in 2006 when Dr. Muhammad Yunus, founder of Grameen Bank, received the Noble Prize [5]. The world took notice of this amazing experiment of lending in small amounts to closely-linked communities.

Similarly, self-help groups in India, which are closely linked small communities based on mutual guarantee, are taking advantage of the community norm that default is a social taboo.

Another example is carbon credit, which in simple terms is a tradable certificate that permits a right to emit one ton of carbon-dioxide or other greenhouse gas of this equivalence. The concept was created as part of Kyoto Protocol to discourage countries/companies that emit carbon-dioxide, and that they should pay to countries/companies that don’t emit such toxic gases by using clean technology.

4 Disruptive innovation

So what exactly is disruptive innovation? [6] The Harvard Business School's Clayton Christensen calls it "wild and unexpected changes that radically restructure markets typically by harnessing new technologies."

Examples of disruptive innovation include mobile phone which has nearly killed fixed line phones; digital photography which sent sales of camera film plummet and made Kodak change its business model; online retailing which is bruising traditional retailing; online ticket booking which has made travel agents reinvent their business models and MOOCS, which are a paradigm shift in the business model of higher education.

But disruptive innovation is not necessarily new technologies. It can as well be using of an existing technology with a new business model. For instance, Netflix is moving away from sending DVDs to streaming content online on demand.

Disruptive innovations usually find their first customers at the bottom of the market. They move up the value chain as successive refinements improve them to the point that they start to steal customers. They may end up reshaping entire industry as in the following cases:

- Craigslist has transformed classified ads
- Skype has changed long distance calls
- iTunes has changed record stores
- Uber has changed the business model of taxis. [7]

In future, driverless cars, 3D printing, shipping via drones and Internet of Things are expected to have huge impact.

5 Are all innovations disruptive?

Not all innovations are disruptive even if they are revolutionary. The first automobiles in the late 19th century were not a disruptive innovation because those automobiles remained expensive luxury items that did not disrupt the market for the horse drawn carriages. It was the low-priced Ford Model which came 30 years later, that was a disruptive innovation because it was cheap, mass produced and changed the rules of the transportation market.

Similarly, the first mainframe computers which used to occupy acres of space were not a disruptive innovation. It was the introduction of the personal computer that disrupted the market.

The Indian financial sector is witnessing disruptive innovation not just in payment systems but in the segments of deposits, credit, and insurance. The key drivers behind this disruptive innovation include: [8]

Internet penetration: The National Optic Fibre Initiative under Digital India will link every village with a broadband connection. As the fastest growing internet economy in the world, India is slated to add 300 million internet users by 2020.
Usage of smartphones: India currently ranks #2 in the world with over 1 billion mobile subscriptions. Of this, approximately 240 million consumers use smartphones and this base is projected to increase to over 520 million by 2020.
Improved O2O (Online to Offline) presence: Online shopping, bill payment and mobile recharge are the foremost reasons for using a digital mode of payment. Sensing these opportunities, merger of e-commerce and e-wallet firms is adding synergy.

Policy incentives: The Unified Payment Interface (UPI) which introduced a virtual payment address, Payment and Small Finance Banks, Aadhaar based Jan Dhan Yojana accounts, Tax rebates for digitisation, and Start-up India initiative are some of the recent important policy initiatives.

6 Conclusion

FinTechs, Payment Banks and Small Finance Banks will challenge the business model of banks and can potentially take away some of the lucrative segments of the banking business by offering tailor-made products and better service.

So, Indian banks have to find strategies to face competition from the new entities. Banks may need to collaborate, build alliances with some of them and even consider taking over a few activities. Ceding the last mile to them is also an option, but what is not an option is not changing their attitude and not becoming more responsive.

The regulators are caught between two competing objectives of letting a *laissez-faire* regime run by encouraging innovation and regulating in the interest of financial stability and consumer protection. Managing the right balance is the challenge for regulators.

Regulation of financial sector is different from regulating other sectors because the financial sector is interconnected. If it doesn’t self-correct in good time, risk builds up, and it’s difficult to know where the implosion will occur. It has to be recognized that a collection of healthy financial institutions does not make a healthy financial system.

The most important thing for regulators is to inspire trust that they will do the right thing. They must ensure that what is regulated is prudently regulated and what is left unregulated is because of a conscious decision.

References

1. Gupta Anupam: Note demonetisation. 86% of Indian currency has been frozen overnight. https://scroll.in/article/821073/note-demonetisation-86-of-indian-currency-and-12-of-gdp-has-been-frozen-overnight (November 2016)
2. Schumpeter Joseph: Theory of economic development. Routledge (2017)
3. The fintech revolution. https://www.economist.com/news/leaders/21650546-wave-startups-changing-financefor-better-fintech-revolution (May 2015)
4. Jenkins Patrick, Arnold Martin, McLannahan Ben, Gardner Laura and Mitevska Tatjana: The 10 financial technology companies to watch. Financial Times, https://www.ft.com/content/797a4924-9553-11e6-a1dc-bdf38d484582 (Nov. 2016)
5. Bhatt Suyash: Disruptive innovation in finance. Forbes India, http://www.forbesindia.com/article/weschool/disruptive-innovation-in-finance/44203/1 (September 2016)
6. Christensen Clayton M., Raynor Michael E. and McDonald Rory: What is disruptive innovation? https://hbr.org/2015/12/what-is-disruptive-innovation (December 2015)
7. What disruptive innovation means. The Economist, https://www.economist. com/blogs/economist-explains/2015/01/economist-explains-15 (January 2015)
8. Kashyap Karan: India’s fintech startups will cause a cashless ‘payment revolution’ By 2023. Forbes, https://www.forbes.com/sites/krnkashyap/2016/08/18/indias-fintech-startups-will-cause-a-cashless-payment-revolution-by-2023/#40e6426c2b42 (August 2016)

IDRBT Journal of Banking Technology

Volume 1 | Number 1 | Jul-Dec 2017

Printed and Published by Sahadevan Rashmi Dev on behalf of Institute of Development and Research in Banking Technology and printed at M/s Karshak Art Printers, 40-APHB, OU Road, Vidyanagar -500044 at Hyderabad. Editor: Dr. Allamraju Subrahmanya Ramasastri