Social internet of things using big data analytics and security aspects – a review

S. Deva Arul* and
Meenakshisundaram Iyapparaja

School of Information Technology and Engineering,
VIT University,
Vellore, Tamil Nadu, India
Email: devadevvit@gmail.com
Email: iyapparaja.m@vit.ac.in
*Corresponding author

Abstract: The rapid development of technologies in today's world has become interesting that made millions of people to utilise the major advantages in it. Two main technologies that were emerging in modern society are big data and the social internet of things. Several researchers have studied and developed a major concept of using big data with SIoT and the security development of maintain a large amount of data. In this paper, deep survey regarding the concepts behind the big data analytics with the social internet of things (SIoT) was studied and analysed. Furthermore, the machine learning techniques that were used in previous works were analysed and comparisons of various methods are discussed. The performance comparison of various classifiers on different datasets is shown and SVM has more than 90% of accuracy when compared with other algorithms. KNN has 64% of accuracy which is lowest of any classifier than NB and NN.

Keywords: big data; social internet of things; SIoT; frequent itemset mining; FIM; machine learning.

Reference to this paper should be made as follows: Arul, S.D. and Iyapparaja, M. (2020) 'Social internet of things using big data analytics and security aspects – a review', *Electronic Government*, Vol. 16, Nos. 1/2, pp.137–154.

Biographical notes: S. Deva Arul received his BCA degree from VIT University Vellore in 2014 and MCA degree from VIT University Vellore in 2016. He is currently pursuing his MTech by Research from School of Information Technology and Engineering, VIT University, Vellore. His research interest includes big data analytics and social network analysis.

Meenakshisundaram Iyapparaja received his PhD in Anna University, Chennai, BE degree from Anna University, Chennai and ME degree from Anna University of Technology, Coimbatore. Presently, he is an Associate Professor in School of Information Technology and Engineering, VIT, Vellore. He has 11 years of experience in teaching and big data, software testing and software engineering field. He received University Rank holder award for his ME degree. His research interests include software testing, software engineering, big data, networking and agile testing. He is a life time member of ISTE.

1 Introduction

Internet of things (IoT) is another major technology which became the most prominent usage in daily life by millions of people in this world. IoT is the future era that makes a new way of seeing this world and making each and every thing to interact with technological aspects. One such way is social networks that millions of people nowadays using for communication with others to anywhere in this world. The IoT and social networks jointly form the social internet of things (SIoT) which is a latest and developing technology to make a social relationship between others. Another technology is big data that helps in processing huge amount of data that collected from SIoT (Atzori et al., 2012; Rho and Chen, 2018). Security and privacy are the two basic needs for every technology that has developed in today's world. For big data and SIoT, both have the same impact regarding this possible ways to improve security and privacy over user's information. Many researchers have come across the possibility of developing privacy in big data analytics and secure trust in social networks that are based on IoT.

1.1 Big data

Recent development in the research environment which becomes a tremendous technology is known to be big data. This technology has proven to be the best for handling huge amount of data which are complex in nature. Three main characteristics that are the building blocks of big data are velocity, volume and variety. As millions of people, nowadays, prone to various technology and the huge amount of data are processed continuously and in a different manner which could be handled by this well known big data technology (Oussous et al., 2017). Figure 1 shows the overview of big data analytics that connected with several applications.

In today's modern world of new applications and its development, big data is the only methodology that can solve millions of large complex problems. Several big data applications like smart grid case, e-health, IoT, public utilities, transportation, logistics, etc., were obtaining the concept for the processing of their resources (Raguseo, 2018; Ahmad et al., 2016).

1.2 Big data challenges

The potential of big data have to be fully realised about its challenges even it has substantial and factual benefits over several applications. The functions of the characteristics of big data have some challenges like analysing models and methods that are existing and limitations of data processing system that are currently used. Some extent challenges of big data are to understand the notion are difficult, decision-making, issues of privacy, and constant learning of new approach for building a viable solution to huge data. Infrastructure is one of the major challenges faced by big data is due to its cost efficiency and hardware equipment which is also highly expensive. Another big challenge for big data is to have the expertise to facilitate the huge amount of data's by the humans. Three main categories can be grouped to analyse about challenges of big data which are based on management, process, and data which is shown in Figure 2 (Ahmad et al., 2016; Ahmed et al., 2017). From Figure 2, three main challenges are concerned for big data and these are data challenges, process challenges and management challenges. In data challenge, the characteristics of data are taken into concern regarding

its volume, variety, variability, velocity, visualisation, veracity and value. In process challenge, processing of data is considered for data acquisition and warehousing, data cleansing and mining, data aggregation and integration, analysis and modelling, and data interpretation. In management challenge, the privacy, security, data ownership, data governance and data sharing are concerned.

Figure 1 Overview of big data analytics

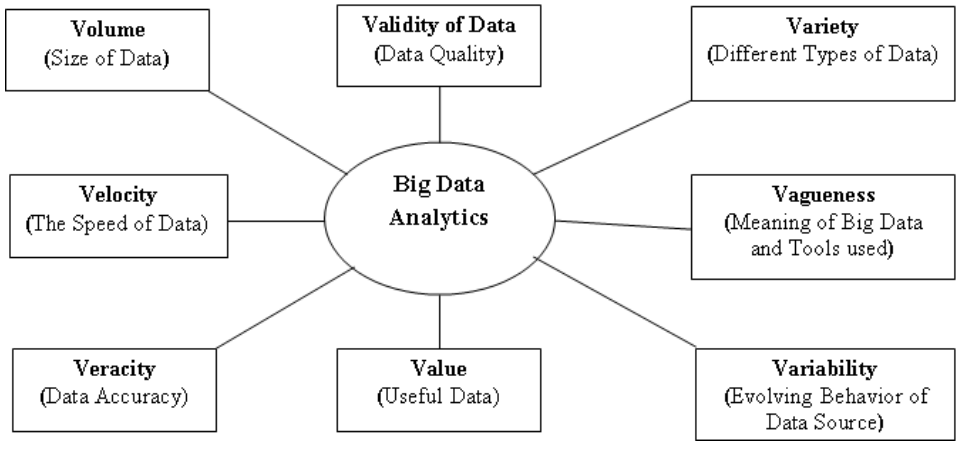
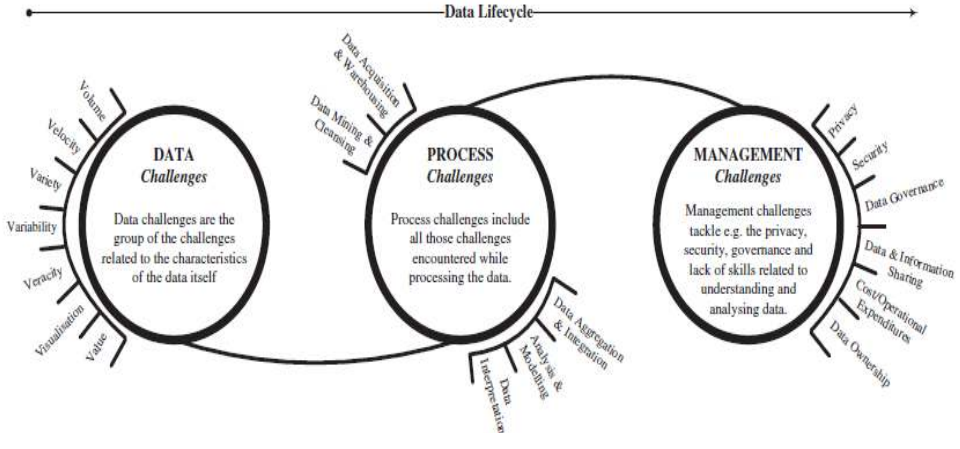


Figure 2 Classification of big data challenges



1.3 Methods in big data analytics

The main motivation of this analytical method is to uncover the insight of a large amount of data and the hidden patterns and correlations related to it. Speed and efficiency are the two major benefits that brought by big data analytics. New opportunities are identified by harnessing the data using this analytic approach. The organisational output and decision making is increased by enhancing big data for extracting sense from the data and these are (Sivarajah et al., 2017):

- Descriptive analytics – Defines the current situation of the business by scrutinising information and data in such a way that patterns, developments, and exceptions can become apparent, which produce standard reports and alerts.
- Inquisitive analytics – Is certified or reject the business proportions by probing the data, factor analysis and statistical analysis are two major parts of drilling the data.
- Predictive analytics – Is to obtain future possibilities by concerning statistical modelling and forecasting.
- Prescriptive analytics – Is the randomised testing and optimisation for accessing how enhancement of service levels of business while decreasing the expenses.
- Pre-emptive analytics – The precautionary actions can be taken on any events in which the organisational performance is influenced.

Figure 3 Classification of big data analytical solutions for IoT systems (see online version for colours)

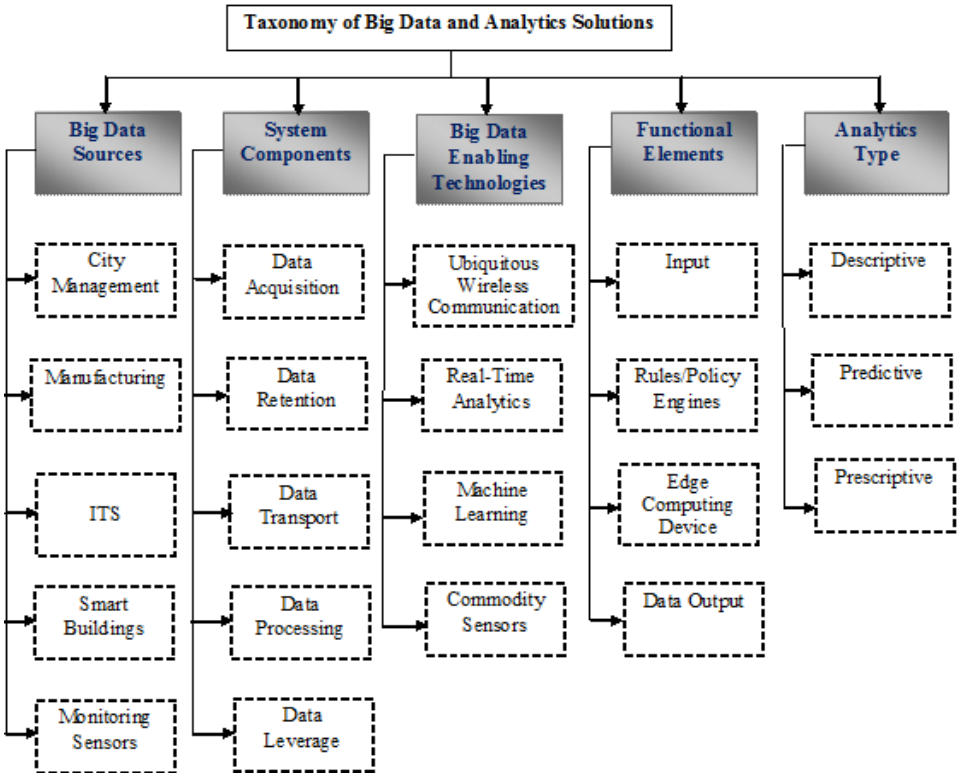


Figure 3 shows the classification of big data analytical solutions that are used for the IoT. The solutions are mainly categorised using the following attributes and these are big data sources, system components, and big data enabling technologies, functional elements and analytics type (Ahmad et al., 2016; Ahmed et al., 2017; Iyapparaja and Sivakumar, 2017; Iyapparaja and Sharma, 2017).

1.4 Security aspects in big data analytics

Enormous values are created by big data for technical innovation and economic growth that are aware by many researchers about large amount of data leads to many privacy concerns. For balancing the big data benefits which should be identified as intensely as possible and privacy preservation of an individual (Lu et al., 2014; Iyapparaja et al., 2012):

- Requirements for privacy in big data collection – Eavesdropping is possible when extensive taken place for big data collection which data could be leaked incidentally. Physical protection methods are resorted when collected data is sensitive and personal; the data privacy also ensured using information security techniques before storing securely.
- Requirements for big data storage – Storage of big data can be compromised which could be harmful when comparing collection phase an individual data could be eavesdropping. Personal information of individual can be disclosed once it is successful. Henceforth, the confidentiality should be ensured for stored data in cyber and physical ways.
- Security aspects in big data processing – It is the key component for the analytics of big data and new knowledge is developed for the technical innovation or economic growth. Efficiency and individual privacy should be ensured and cannot be sacrifice at any cost. Privacy preserving algorithms are desirable to design in an efficient manner for bid data processing (Iyapparaja and Tiwari, 2017).

1.5 Social internet of things

Recent research activities have integrated social networking concepts into the field of IoT which probably taken into a new paradigm and this concept is now known to be SIoT. Furthermore, this technology has the potential for supporting networking services and novel applications of IoT in an efficient manner. The basic concept behind this SIoT is to establish the relationship of social activities with other objects which is also defined as IoT (Geetha, 2016; Jara et al., 2014). The social networking websites are the major concern towards this development of SIoT for humans to connect together with Facebook, Twitter, LinkedIn, etc. Figure 4 shows the basic architecture of SIoT and Figure 5 gives the collaborative service between IoT and social networking which represent the ecosystem which allows interaction between the smart objects with people within a social structure of relationship (Nitti et al., 2013).

SIoT became the most popular future application by many state-of-the-arts, and rapid developing technologies and these are IP-enabled embedded devices, long-range and short-range communication technologies, smart objects, analysis, visualisation tools, data collection, and processing from big market giants and its different advantages are scalability, network navigability, trustworthiness, object discovery, behaviour prediction and classification, service composition, and evaluation of objects (Ning and Wang, 2011; Kranz et al., 2010; Murugan and Devi, 2018). A variety of research is discussed based on these attributes regarding a designing platform for constructing SIoT service environment. The relationships of SIoT are summarised as (Atzori et al., 2011):

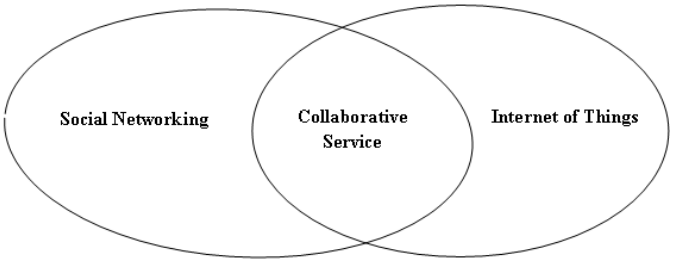
- 'Parental object relationship' (PRO) – Is established among the objects that are belong to same production batch, i.e., by the same period and manufacturer the homogeneous objects usually originated (Atzori et al., 2011).
- 'Co-location object relationship' (C-LOR) – It is established either homogenous or heterogeneous object that are used in the same place when it comes to sensor, augmented objects, and actuators that are used in smart city or at the home. A common goal is achieved when cooperate with each other by establishing C-LORs between objects. With 'short links', it is useful for filling the network (Saleem et al., 2016).
- 'Ownership object relationship' (OOR) – Is belonging to the same users like music players, mobile phones, game consoles, etc., that are established among heterogeneous objects (Farris et al., 2015).
- 'Co-work object relationship' (C-WOR) – A common IoT application is provided by collaborating objects by establishing this relationship. It is used in emergency response that is come in touch by the objects and also telemedicine is another aspect of this work (Atzori et al., 2011).
- 'Social object relationship' (SOR) – Continuous or sporadically the objects are contacted together when establishing this relationship due to the owners have live touch with objects of each other (e.g., sensors and some devices that are usually belong to classmates, relative, friends, colleagues. etc.) (Saleem et al., 2016).

Figure 4 Basic architecture of SIoT (see online version for colours)

It has been witnessed about SIoT that a massive growth in data which has been generated by the social objects and things that are surrounded by people. This kind of growth can be seen in terms of velocity, volume, and veracity of the data that are generated by some common medium and that is known to be internet. As per the Cisco analysis, 2.5 quintillion bytes of data are used per day by several social devices. Due to these circumstances, the need of big data applications is very much useful in the field of SIoT (Saleem et al., 2016).

For example, 'like-art' is one the mentioned part in which the 'like-it' that comes from Facebook is popular and potential of social networks and the physical interaction is combined by moving the physical-organs of the human body. Therefore, the likes and dislikes can be easily understand by the humans and favourite places to visit medical assistance which commonly known to be wireless body area network that are present in the cyber world and enormous data is generated by interacting with physical world, which require large storage databases (Farris et al., 2015). Moreover, retort to probes the archived data is run over and stated as relational databases and one such example is credit card transactions. Precisely, the data collections are maintained by developing different data processing techniques in an efficient and fault-tolerant way that data can be processed at any time and the issues regarding the users query also analysed.

Figure 5 Collaboration service of SIoT



Data aggregation is processed by providing relation database vendors by using big data analytics in various platforms. In today's modern society, several big data applications were developed and used in SIoT into a real time streaming of data (Ortiz et al., 2014; Nitti et al., 2015). However, several new challenges were posted by state of the art when a huge amount of datasets are processed for analysis into a single machine for defining human behaviours. Streaming of large data and storage are distinct challenges that are faced today which several researchers are finding new solutions for the defining data that are in streams within a minimum time span for predicting human behaviour.

Hence, data processing techniques of SIoT that were used in the traditional way should be changed for real time streaming of data for human dynamics for processing. For this reason, new system architecture is needed for analysing and processing large-volumes of data when it arrives as streams. This advancement can help big data in SIoT to overcome the challenges that faced earlier in the fields of human dynamics. Hence, we can say it clearly that "SIoT and big data are the collection of social networks and internet of things (IoT) for defining the human dynamics" (Singh et al., 2015; Murugan and Devi, 2018).

1.6 Major concerns in big data SIoT

Though the development and research about the possibilities and usefulness of the big data in SIoT is deliberate but still some concerns and issues related to this technology is being overwhelmed and some of these factors are as follows (Singh et al., 2015; Suresh et al., 2014):

- Security and privacy – Personal life moments are reflected when most of the user generated content (UGC) on social networks processed. Huge amount of data cannot be secured in a proper manner as of continuous processing all over the world.
- Explosive growth rate – More UGC are generated when growth in IoT and internet. From all social networks, the semi-structured and unstructured data of new sources should accommodate by the infrastructure for processing, capture and analysis, and storing. Apache Hive and Apache Hadoop are used by the Facebook for storing of data due to high scalability and for log collection Scribe is strategy is used.
- Remove noise from valid UGC – Natural language processing (NLP), data mining (DM), and text mining (TM) techniques are needed for optimising to find valid data.
- Real time analytics – Break up occurs when stream processing (SP) is done. Incoming events are checked and analysed by developing models through batch processing (BP) in real time.
- Software and analytical tools – Expectation of high visualisation and low latency from big data SIoT tools and software are still a major concern. Sophisticated algorithms and software tools are currently done as a research project in Lincoln Laboratory for generating networks from semi-structured/unstructured data.

Table 1 shows the description the tools and software that are used nowadays. There is several considerations needed when selecting the proper tool and these factors are ease of use, cost effectiveness, intended goal, operating platform, etc. 'Visualisation' capability is one of the major and vital role that should be considered out of all these. In this survey, we deeply discuss about the big data analytics and its literature survey using SIoT (Sagiroglu and Sinanc, 2013; Katal et al., 2013).

Table 1 Description of big data tools and software for SIoT

Software and tools	*Representation*
NetLogo (Kiran et al., 2015)	Platform independency is supported as it is free software. Visualising dynamics is supported by this software in network formation. This tool is used for studying network behaviour.
Pajek (Batagelj and Mrvar, 2014)	It runs only on windows platform and it is also free software. Network formation, information diffusion, dynamics and many other features are inbuilt.
Gephi (Berzinji et al., 2012)	An open source license is needed and it is platform independent software. It is good tool for the relationship for visualising networks.
iGraph (Nepusz and Csárdi, 2006)	Heavy calculations are performed by this tool and it is free software.
NodeXL (Smith et al., 2009)	It is new to the market which can integrate SNA with Excel. Available only on windows platform and free of cost.
UCINet (Borgatti et al., 2002)	Supported only by windows platform and it is a commercial software
NetworkX (Hagberg et al., 2008)	It is used for program perspective. Fortral libraries and C is used for developing this software. Large matrices are scaled by optimising this tool.

1.7 Hybrid techniques

Many researchers have proposed several hybrid models to improve the efficiency and secure big data system and SIoT. The hybrid techniques are much more important and useful in analysing and processing large amount of data using the IoT and especially for the social networks as millions of people interact daily. These hybrid algorithms not only aggregate data smoothly but also enhance the speed of the system to process the result. The main advantages of the hybrid models is can be used more than one algorithms at once and can add as much to improve the performance with the help of another algorithm. Several hybrid techniques that were used by several researchers like machine learning algorithms combine with some of the evolutionary models and also authentication techniques to provide a better security models and such examples are, mixed encryption algorithm for improving security efficiency, particle swarm optimisation (PSO) with genetic algorithm (GA), etc. (Alturki et al., 2017; Xin, 2015).

1.8 MapReduce

The Hadoop ecosystem is processed by this MapReduce which is a core component as it provides the processing logic. The applications are written by the MapReduce which is software framework that can process large amount of datasets using the parallel and distributed algorithms inside the Hadoop environment. Two main functions are used in MapReduce and these are: Map() that performs actions like grouping, filtering, and sorting whereas Reduce() summarises and aggregates result that produced by the Map() (Rajendran et al., 2016).

1.9 Security aspects of SIoT

The major security aspect of SIoT is 'trustworthiness' which at least depend on two entities that known to be a trustee and trustor. Network system like SIoT has an important feature which is defined as trust management. Some of the trust properties are presented for SIoT environment and these are:

- Direct trust: Which experiences, interactions or observations are taken place directly between the trustee and trustor.
- Indirect trust: Is defined to be opinions are given by the recommendation of available nodes for the trust which basically not having any past interactions or experiences.
- Local trust: Depends on couple trustee/trustor which are considered to be and different from each other.
- Global trust: Is also known as reputation means in which a unique trust value is consisted by every node in the network that are known to other nodes.
- Subjective trust: Various evidence or factors are given by a personal opinion which is inherent and carry more weight (Nitti et al., 2014; Yan and Holtmanns, 2008).

The review of this paper has following sections, in Section 2, the literature review of the previous works are discussed. In Section 3, the simulation results and comparison are analysed. In Section 4, the conclusion of this survey is expressed.

2 Related works

Nowadays, many researchers are doing experiment that is based on latest generation of objects. In human society, these objects play a fundamental role in all aspects. Hence, the object having its own nature which is intended to be intelligent and smart that helps in interacting with people. In this literature, the notion that the social networks and IoT are not too far as people may think. The 'ubiquitous IoT' is the term that is considered for the prospect of internet which is very specific. The social organisation framework model is very much closer to this ubiquitous IoT. The social network basic theme is not comprehended with IoT by this concept as per the knowledge that discovered. In last decade, various methods and architecture were analysed and developed by many researchers for describing the concepts of IoT and social network that are append with big data analytics. Sequential and parallel algorithms of DM is already available for finding frequent itemsets but there is a need to widely work on using MapReduce, and distributed platform due to the issues regarding the performance and scalability in the era of big data. Table 2 shows the comparison of existing frequent mining techniques and its analysis.

Ahmad et al. (2016) proposed a method using big data analytics for social internet of things paradigm for analysing human behaviour. Three operational domains are comprised in their proposed architecture. Ecosystem is analysed in this research which is created by big data and smart cities. The future work of this research is to use collaborative filtering techniques for analysing human behaviour accurately. Arora et al. (2016) proposed a model for analysing network-enabled devices using big data analytical techniques. Four machine learning algorithms are used in their research and these are support vector machine (SVM), random forest (RF), naïve Bayes (NB) and K-nearest neighbour (KNN). The simulation results from their research work shows that highest accuracy is obtained by RF and lowest for NB. Furthermore, KNN and SVM are obtaining accuracy with that of close to RF.

Yen et al. (2015) analysed the potential of composition techniques and service discovery for solving real-world problems that are based on IoT generated data. Various technologies like artificial intelligence and data analytics are examined for the situational facts that are derived from smart world and the required actions accordingly. A gaming-based crowd sourcing platform is proposed for certain control task to complete successfully by using human intelligence. Rathore et al. (2016) analysed problems faced in smart city environment like enabling objects, cost minimising for collection data, and insights into data by proposing four-tier architecture. Data collection and generation is done in bottom tier, communication among sensors are enabled in intermediate tier 1, data processing and management is analysed in the intermediate tire 2 by using Hadoop framework, and data analysis techniques are applied in the top tier. Their simulation results obtained efficient and scalable throughput and processing time than other proposed works.

Zaidi et al. (2015) proposed an algorithm for specifying depth of the mode for deep broad learning and it is tuned for achieving better accuracy in the performance when using huge amount of data when comparing with other state of the art research. Alsheikh et al. (2016) collected a large amount of data from mobile devices and addressed the challenges that rose from it. They have proposed a method for mobile big data analytics using deep learning technique. They used Apache Spark and presented an improvement for learning framework. The simulations results show the better performance in terms of

speed for learning process which compared with other algorithms. Chen and Lin (2014) discussed about the deep learning methods for handling the data that are large in size.

Bing and Chan (2014) analysed the public opinion that are extracted from the social networks in which amount of data is huge and used fuzzy logic for this application and obtained a better performance. Duggal et al. (2015) performed big data analytics using MapReduce and matching algorithms-based fuzzy logic for clinical decision support. Zadeh (1996) proposed fuzzy logic method for analysing uncertain and imprecise data for approximate reasoning and qualitative data is modelled and maintain adaptive control.

Parpinelli and Lopes (2011) used optimisation algorithms for complex real-world data and process. The experimental data is learned the features are extracted using artificial neural network (ANN). This method resembles the speaking of neurons that exists in our brains which identify the interconnection assembly of basic elements. These are connected using the values of the weights that analysed by this technique. Wang et al. (2011) applied fuzzy sets to various areas like patter recognition, machine learning and control systems. The distinct levels of information granularity is represented and processed by using this fuzzy set. Furthermore, it plays an important role in the field of big data value chain like uncertainties of raw data are handled, secondly annotating the data, and finally granular representation of data is prepared specifically for artificial intelligent algorithms. Zhang et al. (2009) proposed a model for big stream data with concept drift using ensemble and incremental learning. Various difficulties like limited resources, addressing data availability are tackled. User profiling and stock trend prediction applications are adapted by this method for producing faster classification and forecasting times are calculated for the receiving new data.

Razavi et al. (2015) analysed a model using evolutionary algorithms for proving machine learning algorithms can be used to solve problems faced today by big data like clustering, feature selection, etc. Naruchitparames et al. (2011) proposed a model using GA for friend recommendation system in social networks. In this work, the authors presented two main categories of research and first one is to tell how formation of links in the social networks generated and second one is to present why the links are forming. The results show a higher performance for listing quality friends that are appropriate and relevant.

Shvachko et al. (2010) proposed a model for Hadoop file system by using the framework of Hadoop MapReduce that can be run over Hadoop-distributed file server (HDFS). Its main advantage is to provide better data-locality using the local disks on computer node. Suguna and Thanushkodi (2011) proposed ABC algorithm along with rough set approach for reducing the dimensions of various datasets related to medical. Also, neural network (NN) is used in this algorithm for feature selection. The major drawback of this proposed approach is cost and computational efficiency. Dean and Ghemawat (2008) analysed programming paradigm for MapReduce which used in large datasets that leads to extensive problems over real-world tasks. The input data is divided using MapReduce into small independent chunks that are completely in a parallel manner. The maps output which are classified using MapReduce are sent to the reduced jobs. The file system will have the input and output tasks. Three main tasks are performed simultaneously as MapReduce is parallel programming model and these are load balancing, simplicity and fault tolerance. Large database applications are needed efficient and reliable data storage which is give by Google file system (GFS) that normally dealt with MapReduce concept.

Table 2 Comparison of different frequent itemset mining (FIM) techniques

Author	*Technique used*	*MapReduce will work/not work*	*Characteristics*	*Advantages*	*Disadvantages*
Han et al. (2000)	FP-growth	No	FP-tree data structure is employed, recursive approach	Focus on small database search	Performance is poor
Agrawal and Srikant (1994)	Apriori	No	Monotonicity property, level wise search and easy implementation	Association rules and frequent itemsets are generated	Scalability
Zaki et al. (1997)	Eclat	No	Intersection of tid_list and works on vertical database, depth first search	Few scans are required to database and enhance locality	Transaction is in large number which degrade performance
Zaki and Gouda (2003)	dEclat	No	Diffsets and vertical databases are used over tidset	Improvement in significant performance	Advantage over tidset is lost when using sparse database diffsets
Hammoud (2011)	MRApriori	Yes	Both vertical and horizontal of hybrid data structure and data in original format are scanned in single	Good and efficient performance over large datasets	Significant reduction is not possible over processing time
Li et al. (2012)	Parallel FP growth	Yes	Grouping of items and independent mining of FP-tree	Linear scalability	Memory and speed are not efficient
Riondato et al. (2012)	PARMA	Yes	Random sampling technique is used	Scaling linearly, data replication minimised, runs faster	Approximate frequent itemset collection is found
Moens et al. (2013)	Dist-Eclat	Yes	Eclats distribution version is used	Speed	Scalability
Zhou et al. (2010)	Balanced FP-growth	Yes	The groups of PFP is balanced by using frequencies of frequent items and FP-growth improvement	Singletons is used for faster execution with balanced distribution	Single items usage for search space partition is not efficient

Bifet and Gavalda (2007) proposed a concept that is named as adaptive sliding window (ADWIN) to solve the existing problem of huge amount of evolving data or concept drift. The window splitting criterion is enabled using the core idea of divide-and-conquer for estimating mean value of variations that observed from change rate of sliding window. Subsequently, when automatic shrinking or expanding in window size occurs, then it becomes adaptive which depends on whether the data could be stationary or not. Within a sliding window, all data points are helpful in deriving the performance of ADWIN which rely on excessive parameter selection. Wong and Kerkez (2016) analysed IoT sensor network data using an adaptive sampling pre-processing technique for hydrology applications. Gallo and Haerri (2017) analysed automatic sampling measurements that compared with previous work which process all data at any time receiving with high consumption. The times of desired events and sampling frequencies are optimised by this technique for detecting appearance. Two major parts are included in this adaptive sampling mechanism and these are rule-based optimisation procedure and another one is prequential forecast phase.

Chen et al. (2016) proposed a scalable and adaptive protocol of trust management for SOA-based IoT systems. Trust feedback is selected by using distributed collaborative filtering technique and sharing similar social interests form owners of IoT nodes and considered three relationships and these are friendship, social contact and community of interest. The direct and indirect trust is combined when employing this adaptive filtering technique for overall and minimising convergence time and bias of trust evaluation.

3 Results

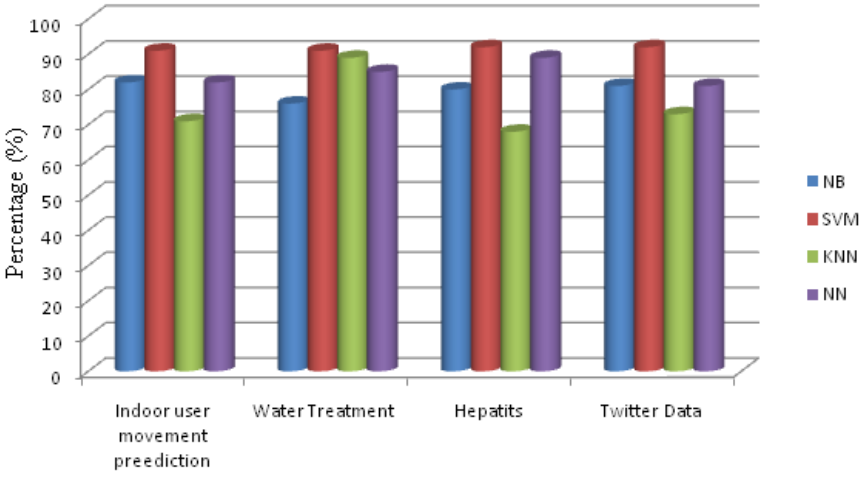
The results for the comparison of previous works are described in Figure 6. The comparison of four classifiers namely NB, SVM, NNs, and KNN are used for the classification of the four datasets. Here, we used indoor movement prediction data for classifying the movements of person, the water treatment data is used for analysing the quality as good or bad. The hepatitis data is used for analysing the patient for the treatment and the Twitter data is used for the analysis of sentiment classification of Arabic words. The overall performance of the classifiers are shown in Figure 6 in which the SVM has more than 90% of accuracy when comparing with the other classifiers and KNN has the least accuracy of 64% than NB and KNN. For different dataset, the classifiers showing the various performances due to its features that includes in it. The PC configurations used for the implementation process is Intel Core2Duo Process with 4 GB RAM and having 500 GB of storage space.

4 Conclusions

In this research, the survey of big data analytics using SIoT was discussed deeply. The characteristics and challenges faced today by big data were studied and scope of social IoT is at peek that is categorised. In addition, the major concerns towards the big data SIoT are also discussed and various factors are studied. Furthermore, the representation of software and tools of big data are shortlisted in Table 1 and related works shows the different algorithms and deep learning methods that are used in big data and as well for

social IoT. Figure 6 shows the comparison of various classifiers performance and SVM shows the better improvement on different datasets. The SVM performance obtains classification accuracy of about 90% when compared with other classifiers like NB, KNN and NN. Although, various techniques were used by researchers to analyse the huge amount of data using big data analytical techniques for SIoT but still the problem exist for streaming of data that continuously sent through social networks as number of users increasing. Security is also a major concern towards the huge amount of data that processing and privacy or trustworthiness is basic fulfilment required in protecting SIoT. For this reason, new hybrid techniques should be considered for improving the performance and reduce the time efficiency which could be the near future research.

Figure 6 Performance comparison of different classifiers (see online version for colours)



References

Agrawal, R. and Srikant, R. (1994) 'Fast algorithms for mining association rules', *in Proc. 20th Int. Conf. Very Large Data Bases, VLDB*, September, Vol. 1215, pp.487–499.

Ahmad, A., Rathore, M.M., Paul, A. and Rho, S. (2016) 'Defining human behaviors using big data analytics in social internet of things', *in 2016 IEEE 30th International Conference on Advanced Information Networking and Applications (AINA)*, IEEE, March, pp.1101–1107.

Ahmed, E., Yaqoob, I., Hashem, I.A.T., Khan, I., Ahmed, A.I.A., Imran, M. and Vasilakos, A.V. (2017) 'The role of big data analytics in internet of things', *Computer Networks*, DOI: 10.1016/j.comnet.2017.06.013.

Alsheikh, M.A., Niyato, D., Lin, S., Tan, H.P. and Han, Z. (2016) 'Mobile big data analytics using deep learning and apache spark', *IEEE Network*, Vol. 30, No. 3, pp.22–29.

Alturki, B., Reiff-Marganiec, S. and Perera, C. (2017) 'A hybrid approach for data analytics for internet of things', *in Proceedings of the Seventh International Conference on the Internet of Things*, ACM, October, p.7.

Arora, D., Li, K.F. and Loffler, A. (2016) 'Big data analytics for classification of network enabled devices', *in 2016 30th International Conference on Advanced Information Networking and Applications Workshops (WAINA)*, IEEE, March, pp.708–713.

Atzori, L., Iera, A. and Morabito, G. (2011) 'SIoT: giving a social structure to the internet of things', *IEEE Communications Letters*, Vol. 15, No. 11, pp.1193–1195.

Atzori, L., Iera, A., Morabito, G. and Nitti, M. (2012) 'The social internet of things (SIoT) – when social networks meet the internet of things: concept, architecture and network characterization', *Computer Networks*, Vol. 56, No. 16, pp.3594–3608.

Batagelj, V. and Mrvar, A. (2014) 'Pajek', in *Encyclopedia of Social Network Analysis and Mining*, pp.1245–1256, Springer, New York, NY.

Berzinji, A., Kaati, L. and Rezine, A. (2012) 'Detecting key players in terrorist networks', *in 2012 European Intelligence and Security Informatics Conference (EISIC)*, IEEE, August, pp.297–302.

Bifet, A. and Gavalda, R. (2007) 'Learning from time-changing data with adaptive windowing', *in Proceedings of the 2007 SIAM International Conference on Data Mining*, Society for Industrial and Applied Mathematics, April, pp.443–448.

Bing, L. and Chan, K.C. (2014) 'A fuzzy logic approach for opinion mining on large scale Twitter data', *in Proceedings of the 2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing*, IEEE Computer Society, December, pp.652–657.

Borgatti, S.P., Everett, M.G. and Freeman, L.C. (2002) *UCInet for Windows: Software for Social Network Analysis*, Analytic Technologies, Harvard, MA.

Chen, R., Guo, J. and Bao, F. (2016) 'Trust management for SOA-based IoT and its application to service composition', *IEEE Transactions on Services Computing*, Vol. 9, No. 3, pp.482–495.

Chen, X. and Lin, X. (2014) 'Big data deep learning: challenges and perspectives', in *IEEE Access*, Vol. 2, pp.514–525, DOI: 10.1109/ACCESS.2014.2325029.

Dean, J. and Ghemawat, S. (2008) 'MapReduce: simplified data processing on large clusters', *Communications of the ACM*, Vol. 51, No. 1, pp.107–113.

Duggal, R., Khatri, S.K. and Shukla, B. (2015) 'Improving patient matching: single patient view for clinical decision support using big data analytics', *in 2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions)*, IEEE, September, pp.1–6.

Farris, I., Girau, R., Nitti, M., Atzori, L., Bruschi, R., Iera, A. and Morabito, G. (2015) 'Taking the SIoT down from the cloud: integrating the social internet of things in the INPUT architecture', *in 2015 IEEE 2nd World Forum on Internet of Things (WF-IoT)*, IEEE, December, pp.35–39.

Gallo, L. and Haerri, J. (2017) 'Unsupervised long-term evolution device-to-device: a case study for safety-critical V2X communications', *IEEE Vehicular Technology Magazine*, Vol. 12, No. 2, pp.69–77.

Geetha, S. (2016) 'Social internet of things', *World Scientific News*, Vol. 41, p.76.

Hagberg, A., Swart, P. and Schult, D. (2008) *Exploring Network Structure, Dynamics, and Function Using NetworkX*, No. LA-UR-08-05495, LA-UR-08-5495, Los Alamos National Lab. (LANL), Los Alamos, NM, USA.

Hammoud, S. (2011) *MapReduce Network Enabled Algorithms for Classification Based on Association Rules*, Doctoral dissertation, Brunel University School of Engineering and Design PhD Theses.

Han, J., Pei, J. and Yin, Y. (2000) 'Mining frequent patterns without candidate generation', *ACM Sigmod Record*, May, Vol. 29, No. 2, pp.1–12, ACM.

Iyapparaja, M. and Sharma, B. (2017) 'Augmenting SCA project management and automation framework', in *Materials Science and Engineering Conference Series*, November, Vol. 263, No. 4, p.042018, DOI: 10.1088/1757-899X/263/4/042018.

Iyapparaja, M. and Sivakumar, P. (2017) 'Metrics based evaluation for disease affection in distinct cities', *Research J. Pharm. and Tech.*, Vol. 10, No. 8, pp.2487–2491.

Iyapparaja, M. and Tiwari, M. (2017) 'Security policy speculation of user uploaded images on content sharing sites', in *Materials Science and Engineering Conference Series*, November, Vol. 263, No. 4, p.042019, DOI: 10.1088/1757-899X/263/4/042019.

Iyapparaja, M. et al. (2012) 'Coupling and cohesion metrics in java for adaptive reusability risk reduction', *IET Chennai 3rd International Conference on Sustainable Energy and Intelligent Systems (SEISCON 2012)*, pp.52–57.

Jara, A., Bocchi, Y. and Genoud, D. (2014) 'Social internet of things', *in Proceedings of International Conference on Intelligent Networking and Collaborative Systems 2014*, 10–12 September, No. CONFERENCE-2015-144.

Katal, A., Wazid, M. and Goudar, R.H. (2013) 'Big data: issues, challenges, tools and good practices', *in 2013 Sixth International Conference on Contemporary Computing (IC3)*, IEEE, August, pp.404–409.

Kiran, M., Murphy, P., Monga, I., Dugan, J. and Baveja, S.S. (2015) 'Lambda architecture for cost-effective batch and speed big data processing', *in 2015 IEEE International Conference on Big Data (Big Data)*, IEEE, October, pp.2785–2792.

Kranz, M., Roalter, L. and Michahelles, F. (2010) 'Things that twitter: social networks and the internet of things', *in What Can the Internet of Things Do for the Citizen (CIoT) Workshop at the Eighth International Conference on Pervasive Computing (Pervasive 2010)*, May, pp.1–10.

Li, N., Zeng, L., He, Q. and Shi, Z. (2012) 'Parallel implementation of apriori algorithm based on MapReduce', *in 2012 13th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing*, IEEE, August, pp.236–241.

Lu, R., Zhu, H., Liu, X., Liu, J.K. and Shao, J. (2014) 'Toward efficient and privacy-preserving computing in big data era', *IEEE Network*, Vol. 28, No. 4, pp.46–50.

Moens, S., Aksehirli, E. and Goethals, B. (2013) 'Frequent itemset mining for big data', *in BigData Conference*, October, pp.111–118.

Murugan, N.S. and Devi, G.U. (2018) 'Detecting streaming of twitter spam using hybrid method', *Wireless Personal Communications*, Vol. 103, No. 1353, pp.1–22, https://doi.org/10.1007/s11277-018-5513-z .

Murugan, N.S. and Devi, G.U. (2018) 'Feature extraction using LR-PCA hybridization on Twitter data and classification accuracy using machine learning algorithms', *Cluster Computing*, pp.1–10, https://doi.org/10.1007/s10586-018-2158-3.

Naruchitparames, J., Güneş, M.H. and Louis, S.J. (2011) 'Friend recommendations in social networks using genetic algorithms and network topology', *in 2011 IEEE Congress on Evolutionary Computation (CEC)*, IEEE, June, pp.2207–2214.

Nepusz, G.C.A.T. and Csárdi, G. (2006) 'The iGraph software package for complex network research', *Complex Systems*, Vol. 1695, No. 5, pp.1–9.

Ning, H. and Wang, Z. (2011) 'Future internet of things architecture: like mankind neural system or social organization framework?', *IEEE Communications Letters*, Vol. 15, No. 4, pp.461–463.

Nitti, M., Atzori, L. and Cvijikj, I.P. (2015) 'Friendship selection in the social internet of things: challenges and possible strategies', *IEEE Internet of Things Journal*, Vol. 2, No. 3, pp.240–247.

Nitti, M., Girau, R. and Atzori, L. (2013) 'Trustworthiness management in the social internet of things', *IEEE Transactions on Knowledge and Data Engineering*, Vol. 26, No. 5, pp.1253–1266.

Nitti, M., Girau, R. and Atzori, L. (2014) 'Trustworthiness management in the social internet of things', *IEEE Transactions on Knowledge and Data Engineering*, Vol. 26, No. 5, pp.1253–1266.

Ortiz, A.M., Hussein, D., Park, S., Han, S.N. and Crespi, N. (2014) 'The cluster between internet of things and social networks: review and research challenges'. *IEEE Internet of Things Journal*, Vol. 1, No. 3, pp.206–215.

Oussous, A. et al. (2017) 'Big data technologies: a survey', *Journal of King Saud University – Computer and Information Sciences*, Vol. 30, pp.431–448.

Parpinelli, R.S. and Lopes, H.S. (2011) 'New inspirations in swarm intelligence: a survey', *International Journal of Bio-Inspired Computation*, Vol. 3, No. 1, pp.1–16.

Raguseo, E. (2018) 'Big data technologies: an empirical investigation on their adoption, benefits and risks for companies', *International Journal of Information Management*, Vol. 38, No. 1, pp.187–195.

Rajendran, P.K., Asbern, A., Kumar, K.M., Rajesh, M. and Abhilash, R. (2016) 'Implementation and analysis of MapReduce on biomedical big data', *Indian Journal of Science and Technology*, Vol. 9, p.31.

Rathore, M.M., Ahmad, A. and Paul, A. (2016) 'IoT-based smart city development using big data analytical approach', *in IEEE International Conference on Automatica (ICA-ACCA)*, IEEE, October, pp.1–8.

Razavi, S.H., Ebadati, E.O.M., Asadi, S. and Kaur, H. (2015) 'An efficient grouping genetic algorithm for data clustering and big data analysis', in *Computational Intelligence for Big Data Analysis*, pp.119–142, Springer, Cham.

Rho, S. and Chen, Y. (2018) 'Social internet of things: applications, architectures, and protocols', *Future Generation Computer Systems*, Vol. 84, pp.667–668.

Riondato, M., DeBrabant, J.A., Fonseca, R. and Upfal, E. (2012) 'PARMA: a parallel randomized algorithm for approximate association rules mining in MapReduce', *in Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, ACM, October, pp.85–94.

Sagiroglu, S. and Sinanc, D. (2013) 'Big data: a review', *in 2013 International Conference on Collaboration Technologies and Systems (CTS)*, IEEE, May, pp.42–47.

Saleem, Y., Crespi, N., Rehmani, M.H., Copeland, R., Hussein, D. and Bertin, E. (2016) 'Exploitation of social IoT for recommendation services', *in 2016 IEEE 3rd World Forum on Internet of Things (WF-IoT)*, IEEE, December, pp.359–364.

Shvachko, K., Kuang, H., Radia, S. and Chansler, R. (2010) 'The Hadoop distributed file system', *in 2010 IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST)*, IEEE, May, pp.1–10.

Singh, M., Rajan, M.A., Shivraj, V.L. and Balamuralidhar, P. (2015) 'Secure MQTT for internet of things (IoT)', *in 2015 Fifth International Conference on Communication Systems and Network Technologies (CSNT)*, IEEE, April, pp.746–751.

Sivarajah, U., Kamal, M.M., Irani, Z. and Weerakkody, V. (2017) 'Critical analysis of big data challenges and analytical methods', *Journal of Business Research*, Vol. 70, pp.263–286, ISSN: 0148-2963, https://doi.org/10.1016/j.jbusres.2016.08.001.

Smith, M.A., Shneiderman, B., Milic-Frayling, N., Mendes Rodrigues, E., Barash, V., Dunne, C., Gleave, E. et al. (2009) 'Analyzing (social media) networks with NodeXL', *in Proceedings of the Fourth International Conference on Communities and Technologies*, ACM, June, pp.255–264.

Suguna, N. and Thanushkodi, K.G. (2011) 'An independent rough set approach hybrid with artificial bee colony algorithm for dimensionality reduction', *American Journal of Applied Sciences*, Vol. 8, No. 3, p.261.

Suresh, P., Daniel, J.V., Parthasarathy, V. and Aswathy, R.H. (2014) 'A state of the art review on the internet of things (IoT) history, technology and fields of deployment', *in 2014 International Conference on Science Engineering and Management Research (ICSEMR)*, IEEE, November. pp.1–8.

Wang, J., Abid, H., Lee, S., Shu, L. and Xia, F. (2011) *A Secured Health Care Application Architecture for Cyber-physical Systems*, arXiv preprint arXiv: 1201.0213.

Wong, B.P. and Kerkez, B. (2016) 'Real-time environmental sensor data: an application to water quality using web services', *Environmental Modelling & Software*, Vol. 84, pp.505–517, ISSN: 1364-8152, https://doi.org/10.1016/j.envsoft.2016.07.020.

Xin, M. (2015) 'A mixed encryption algorithm used in internet of things security transmission system', *in 2015 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)*, IEEE, September, pp.62–65.

Yan, Z. and Holtmanns, S. (2008) 'Trust modeling and management: from social trust to digital trust', in Subramanian, R. (Ed.): *Computer Security, Privacy and Politics: Current Issues, Challenges and Solutions*, pp.290–323, IGI Global, Hershey, PA, DOI: 10.4018/978-1-59904-804-8.ch013.

Yen, I.L., Zhou, G., Zhu, W., Bastani, F. and Hwang, S.Y. (2015) 'A smart physical world based on service technologies, big data, and game-based crowd sourcing', *in 2015 IEEE International Conference on Web Services (ICWS)*, IEEE, June, pp.765–772.

Zadeh, L.A. (1996) 'On fuzzy algorithms', in *Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems: Selected Papers by Lotfi A Zadeh*, pp.127–147, World Scientific, USA.

Zaidi, N.A., Webb, G.I., Carman, M.J. and Petitjean, F. (2015) *Deep Broad Learning – Big Models for Big Data*, arXiv preprint arXiv: 1509.01346.

Zaki, M.J. and Gouda, K. (2003) 'Fast vertical mining using diffsets', *in Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, August, pp.326–335.

Zaki, M.J., Parthasarathy, S., Ogihara, M. and Li, W. (1997) 'Parallel algorithms for discovery of association rules', *Data Mining and Knowledge Discovery*, Vol. 1, No. 4, pp.343–373.

Zhang, Y., Meratnia, N. and Havinga, P. (2009) 'Adaptive and online one-class support vector machine-based outlier detection techniques for wireless sensor networks', *in International Conference on Advanced Information Networking and Applications Workshops, 2009, WAINA'09*, IEEE, May, pp.990–995.

Zhou, L., Zhong, Z., Chang, J., Li, J., Huang, J.Z. and Feng, S. (2010) 'Balanced parallel FP-growth with MapReduce', *in 2010 IEEE Youth Conference on Information Computing and Telecommunications (YC-ICT)*, IEEE, November, pp.243–246.