iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL
FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 Issue: VII Month of publication: July 2021

DOI: https://doi.org/10.22214/ijraset.2021.37122

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VII July 2021- Available at www.ijraset.com

Data Mining for the Internet of Things: Literature Review and Challenges

Dr. Mohd Zuber
Assistant Professor of Girdher group of institutions

Abstract: The huge data generate by the Internet of Things (IOT) are measured of high business worth, and data mining algorithms can be applied to IOT to take out hidden information from data. In this paper, we give a methodical way to review data mining in knowledge, technique and application view, together with classification, clustering, association analysis and time series analysis, outlier analysis. And the latest application luggage is also surveyed. As more and more devices connected to IOT, huge volume of data should be analyzed, the latest algorithms should be customized to apply to big data. We reviewed these algorithms and discussed challenges and open research issues. At last a suggested big data mining system is proposed.
Keyword: Internet of things, Classification, clustering, Association analysis.

I. INTRODUCTION

The Internet of Things (IoT) and its related technologies can effortlessly mix classical networks with networked devices. IoT has been live an vital role ever since it appear which covers from conventional tools to general household objects [1] and has been attracting the notice of researchers from academic world, engineering and government in current years. There is a great vision that all things can be easily controlled and monitored can be recognized robotically [2]. Lots of analysis technologies are introduced into IoT one of the most precious technologies is data mining.

Data mining discover narrative, motivating, and potentially useful patterns from large data sets and applying algorithms to the taking out of unseen information. Many other conditions are used for data mining for example, knowledge discovery databases (KDD), knowledge taking out, data analysis, data archeology, data dredging, and information harvesting [3]. The purpose of any data mining procedure is to make an well-organized predictive or descriptive model of a huge amount of data that not only best explains it but is also able to simplify to new data [4]. Data mining is the procedure of discovering motivating knowledge from huge amounts of data stored in databases, data warehouses or other information repositories. On the base of the description of data mining and the definition of data mining functions, a typical data mining process includes the following steps (see Figure 1).

1) Data preparation: organize the data for mining. It includes 3 sub steps:(a) integrate data in a variety of data sources and clean the noise from data; (b) extract some parts of data into data mining system; (c) preprocess the data to facilitate the data mining.
2) Data mining: apply algorithms to the data to find the patterns and evaluate patterns of discovered knowledge.
3) Data presentation: visualize the data and represent mined knowledge to the user.

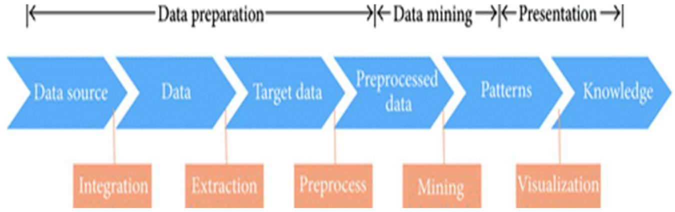

Figure 1 The data mining general idea.

A variety of researches focus on knowledge view, technique view, and application view can be originated in the literature. However, no preceding effort has been made to review the different views of data mining in a systematic way, especially in nowadays big data [5–7]; mobile internet and Internet of Things [8–10] produce rapidly and some data mining researchers shift their attention from data mining to big data. There are lots of data that can be mine like a relational database, data warehouse, data stream, time series, sequence, text and web, multimedia [11], graphs, the World Wide Web, Internet of Things data [12–14]. In this paper, we attempt to make a complete survey of the important fresh developments of data mining research. This survey focuses on knowledge view, utilized techniques view, and application view of data mining. Our main input in this paper is that we chosen some well-known algorithms and studied their strengths and limitations.

The involvement of this paper includes 3 parts: the primary part is that we suggest a narrative way to review data mining in knowledge view, technique view, and application view; the next part is that we talk about the new characteristics of big data and analyze the challenges. Another important involvement is that we propose a suggested big data mining system. It is precious for readers if they want to build a big data mining system with open source technologies.

The rest of the paper is planned as follows. In Section second we survey the main data mining functions from knowledge view and technology view with classification, clustering, association analysis, and outlier analysis, and bring in which techniques can hold up these functions. In Section three we review the data mining applications in e-commerce, industry, health care, and public service and talk about which knowledge and technology can be useful to these applications. In Section four IoT and big data are discussed new technologies to mine big data for IoT are surveyed, the challenges in big data age are overviewed and a new big data mining system architecture for IoT is proposed. In Section five we give a conclusion.

II. DATA MINING FUNCTIONALITIES

Data mining functionalities include classification, clustering, association analysis, time series analysis,
Outlier analysis describes and models regularities or trends for objects whose performance changes over time.

1) Classification is the method of finding a set of models or functions that explain and differentiate data classes or concepts for the reason of predicting the class of objects whose class label is unidentified.
2) Clustering analyzes data objects without consult a known class model.
3) Association analysis is the discovery of association rules displaying attribute-value situation that regularly occur together in a given set of data.
4) Time series analysis comprises methods and techniques for analyzing time series data in order to extract significant statistics and other characteristics of the data.

A. *Classification*

Classification is significant for management of decision making. Given an object, transmission it to one of predefined target categories or classes is called classification. For example, a classification model could be used to identify loan applicants as low, medium, or high credit risks [15].

There are many methods to classify the data with decision tree induction, frame-based or rule-based expert systems, hierarchical classification, neural networks, Bayesian network, and support vector machines (see Figure 2).

1) A decision tree is a flow-chart-like tree structure where each internal node is denoted by rectangles and leaf nodes are denoted by ovals. All internal nodes have two or more child nodes. All internal nodes contain splits which test the value of a look of the attributes. Arcs from an internal node to its children are labeled with separate outcomes of the test. Each leaf node has a class label linked with it. Classification and Regression Trees (CART) is a nonparametric decision tree algorithm. It produces either classification or regression trees, based on whether the response variable is categorical or continuous.
2) The KNN (*K*-Nearest Neighbor) algorithm is introduced by the Nearest Neighbor algorithm which is intended to find the nearest point of the experiential object. The main idea of the KNN algorithm is to discover the *K*-nearest points [16]. There are a lot of dissimilar improvements for the traditional KNN algorithm. such as the Wavelet Based *K*-Nearest Neighbor Partial Distance Search (WKPDS) algorithm [17], Equal-Average Nearest Neighbor Search (ENNS) algorithm [18], Equal-Average Equal-Norm Nearest Neighbor code word Search (EENNS) algorithm [19], the Equal-Average Equal-Variance Equal-Norm Nearest Neighbor Search (EEENNS) algorithm [20], and other improvements [21].
3) Bayesian networks are directed acyclic graphs whose nodes represent random variables in the Bayesian sense. Edges stand for conditional dependencies; nodes which are not connected stand for variables which are conditionally independent of each other. The research includes naïve Bayes [22, 23], selective naïve Bayes [24], seminaïve Bayes [25], one-dependence Bayesian classifiers [26, 27], *K*-dependence Bayesian classifiers [28], Bayesian network-augmented naïve Bayes [29], unrestricted Bayesian classifiers [30], and Bayesian multinets [31].
4) Support Vector Machines algorithm is supervised learning model with associated learning algorithms that examine data and be familiar with patterns which is based on statistical learning theory. SVM is widely used in text classification [23, 32], marketing, pattern recognition, and medical diagnosis [33]. A lot of further research is done, GSVM (granular support vector machines) [34–36], FSVM (fuzzy support vector machines) [37–39], TWSVMs (twin support vector machines) [40–42], VaR-SVM (value-at-risk support vector machines) [43], and RSVM (ranking support vector machines) [44].

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VII July 2021- Available at www.ijraset.com

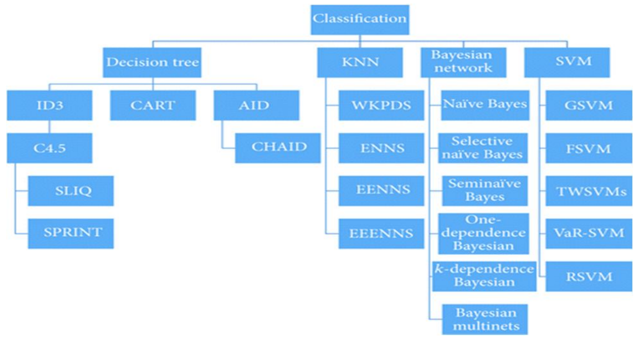

Figure 2 The research structure of classification.

B. Clustering

Clustering algorithms [45] divide data into significant groups (see Figure 3) so that patterns in the same group are similar in some sense and patterns in different group are dissimilar in the same sense. Searching for clusters involves unsupervised learning [46].for example the search engine clusters billions of web pages into different groups such as news, reviews, videos, and audios. One straightforward example of clustering problem is to divide points into different groups [15].

1) Hierarchical clustering technique combines data objects into subgroups; those subgroups combine into larger and high level groups form a hierarchy tree. Hierarchical clustering methods have two classifications, agglomerative (bottom-up) and divisive (top-down) approaches. The agglomerative clustering starts with one-point clusters and recursively merge two or more of the clusters. The divisive clustering in contrast is a top-down strategy.
2) Partitioning algorithms find out clusters either by iteratively relocating points between subsets or by identifying areas heavily populated with data. Density-based partitioning methods effort to discover low-dimensional data which is dense-connected known as spatial data. Grid based partitioning algorithms use hierarchical agglomeration as one stage of processing and perform space segmentation and then aggregate appropriate segments.
3) In arrange to handle definite data researchers change data clustering to pre clustering of items or categorical attribute values typical research includes ROCK [47].
4) Scalable clustering investigates faces scalability problems for computing time and memory requirements including DIGNET [48] and BIRCH [49].
5) High dimensionality data clustering methods are designed to handle data with hundreds of attributes including DFT [50] and MAFIA [51].

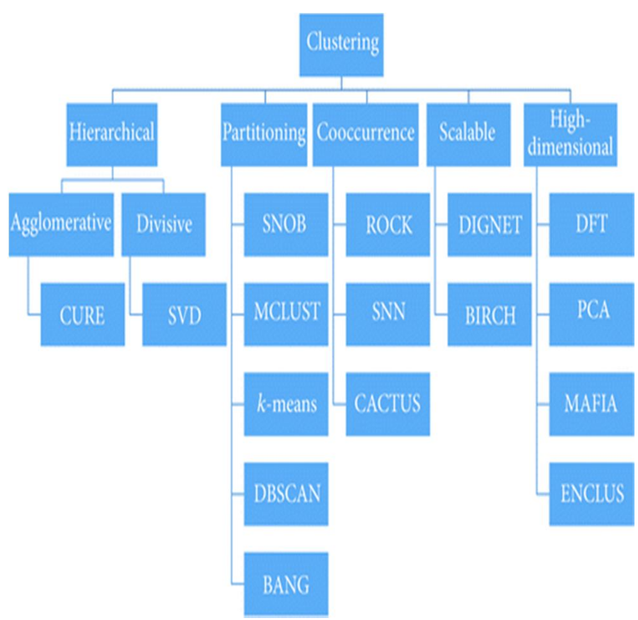

Figure 3 The research structure of clustering.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VII July 2021- Available at www.ijraset.com

C. Association Analysis

Association rule mining [52] focuses on the market basket analysis or transaction data analysis and it targets finding of rules showing attribute-value associations that occur regularly and also help in the generation of all-purpose and qualitative knowledge which in turn helps in decision making [53]. The research structure of association analysis is shown in Figure 4.

1) For the primary list of association analysis algorithms the data will be process sequentially. The a priori based algorithms have been used to find out intra transaction associations and then discover associations there are lots of extension algorithms. According to the data record format it clusters into 2 types: Horizontal Database Format Algorithms and Vertical Database Format Algorithms. Pattern growth algorithm is more complex but can be faster to calculate given large volumes of data. The typical algorithm is FP-Growth algorithm [54].
2) In some region the data would be a flow of actions and therefore the problem would be to discover event patterns that occur frequently together. It divides into 2 parts: event-based algorithms and event-oriented algorithms.

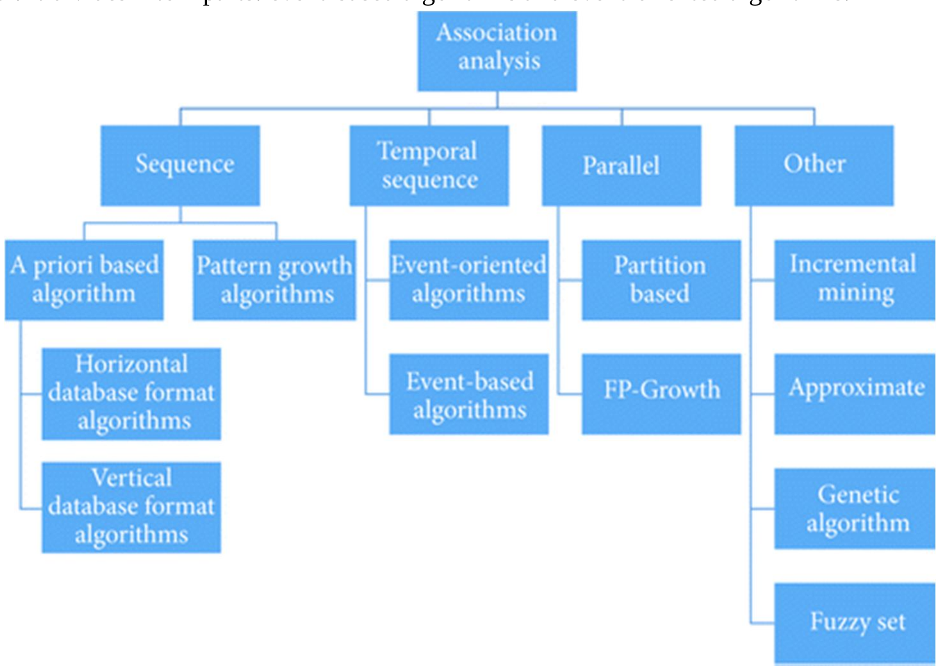

Figure 4 The research structure of association analysis.

D. Time Series Analysis

A time series is a group of temporal data objects the characteristics of time series data comprise large data size, high dimensionality, and updating incessantly. Commonly, time series task relies on 3 parts of mechanism, including representation, similarity measures, and indexing (see Figure 5) [55, 56].

1) One of the main reasons for time series representation is to decrease the dimension, and it divides into three categories: model based representation, non-data-adaptive representation and data adaptive representation. The model based representations desire to find parameters of fundamental model for a representation. Important research works include ARMA [57] and the time series bitmaps research [58]. In non-data-adaptive representations the parameters of the alteration remain the same for every time series despite of its nature related research including DFT [59], wavelet functions related topic [60], and PAA [50]. In data adaptive representations, the parameters of a alteration will change according to the data available and related works including representations version of DFT [61]/PAA [62] and indexable PLA [63].
2) The similarity calculates of time series analysis is characteristically carried out in an estimated manner the research information include subsequence matching [64] and full sequence matching [65].
3) The indexing of time series analysis is intimately associated with representation and similarity measure part the research topic includes SAMs (Spatial Access Methods) and TS-Tree [66].

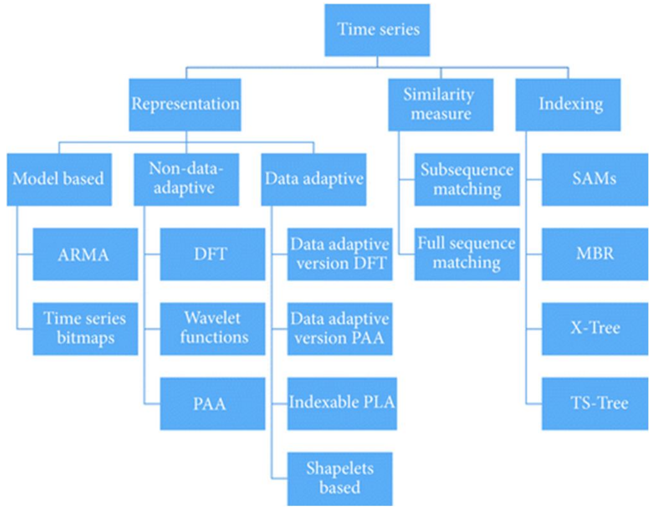

Figure 5 The research structure of time series analysis.

III. DATA MINING APPLICATIONS

A. Data Mining in e-Commerce

e-commerce is one of the most possible domains for data mining because data records, customer data, product data, users' action log data, are overflowing IT team has enriched data mining ability and return on investment can be calculated. Researchers influence association analysis and clustering to give the insight of what product combinations were purchased it encourages clients to purchase related products that they may have been missed or unnoticed. Users' behaviors are monitored and analyzed to find similarities and patterns in Web surfing behavior so that the Web can be more successful in meeting user needs [67].

B. Data Mining in Industry

Data mining can highly advantage industries such as retail, banking, and telecommunications; classification and clustering can be applied to this region [68].Retailers gather customer information, related transactions information, and product information to considerably improve exactness of product demand forecasting, variety optimization, product recommendation, and ranking across retailers and manufacturers [69, 70]. Researchers leverage SVM [71], support vector regression [72], or Bass model [73] to forecast the products' demand.

C. Data Mining in Health Care

In health care, data mining is becoming increasingly popular, if not increasingly essential [74–79]. Heterogeneous medical data have been generated in various health care organizations, including payers, medicine providers, pharmaceuticals information, prescription information, doctor's notes, or clinical records produced day by day. These quantitative data can be used to do clinical text mining, predictive modeling [80], survival analysis, patient similarity analysis [81], and clustering, to improve care treatment [82] and reduce waste. In health care area, association analysis, clustering, and outlier analysis can be applied [83, 83].

Treatment record data can be mined to explore ways to cut costs and deliver better medicine [85, 86]. Data mining also can be used to identify and understand high-cost patients [87] and applied to mass of data generated by millions of prescriptions, operations, and treatment courses to identify unusual patterns and uncover fraud [88, 89].

D. Data Mining in City Governance

In public service area, data mining can be used to find out public wants and improve service performance, decision making with computerized systems to reduce risks, classification, clustering, and time series analysis which can be developed to solve this area difficulty's-government improve excellence of government service, cost savings, wider political participation, and more effectual policies and programs [90, 91], and it has also been proposed as a solution for rising citizen communication with government agencies and political trust [92]. City event information management system can put together data mining methods to provide a complete appraisal of the impact of usual disasters on the agricultural production and rank disaster affected areas impartially and assist governments in disaster preparation and resource allocation [93].

By using data analytics, researchers can forecast which residents are likely to move away from the city [94], and it helps to suppose which factors of city life and city services lead to a resident's choice to leave the city [95].
Also data mining can be used to notice criminal individuality deceptions by analyzing people information such as name, address, date of birth, and social-security number [96] and to uncover previously unknown structural patterns from criminal networks [139].In transport system, data mining can be used for map modification according to GPS traces [97–100], and based on numerous users' GPS trajectories researchers discover the interesting locations and classical travel sequences for location recommendation and travel recommendation [101].

E. Summary

The data mining application and nearly everyone popular data mining functionalities can be summarized in Table 1.

Click to view table

Table 1 The data mining application and most popular data mining functionalities.

Table 1 The data mining application and most popular data mining functionalities.

TABLE 1: The data mining application and most popular data mining functionalities.

Application	Classification	Clustering	Association analysis	Time series analysis	Outlier analysis
e-commerce		√	√		
Industry	√	√	√		
Health care		√	√		√
City governance	√	√	√	√	

View larger version

IV. CHALLENGES AND UNLOCK RESEARCH ISSUES IN IOT AND BIG DATA AGE

The fast development of IoT, big data, and cloud computing the most basic challenge is to discover the huge volumes of data and take out useful information or knowledge for expectations actions [102]. The key characteristics of the data in IoT age can be measured as big data they are as follows.

1) Huge volumes of data to read and write: the quantity of data can be TB (terabytes), even PB (petabytes) and ZB (zettabyte) so we require exploring fast and effectual mechanisms.
2) Heterogeneous data sources and data types to integrate in big data age the data sources are miscellaneous for example we need to put together sensors [103–105], cameras, social media all these data are different in format, byte, binary and string, number. We require to converse with different types of devices and also need to extract data from web pages.

A. Challenges

There are plenty of challenges when IoT and big data approach the amount of data is big but the superiority is low and the data are various from different data sources heterogeneous, as-structured, semi structured, and even entirely unstructured. We analyze the challenges in data extracting, data mining algorithms, and data mining system region. Challenges are summarized under.

1) The first challenge is to access, extracting big data from different data storage locations. We need to contract with the diversity, heterogeneity and noise of the data and it is a large challenge to find the error and even harder to correct the data. In data mining algorithms region how to adapt traditional algorithms to big data atmosphere is a big challenge.
2) Second challenge is how to mine unsure and unfinished data for big data applications. In data mining system an effectual and safety solution to share data between different applications and systems is one of the mainly vital challenges since responsive information such as banking dealings and medical minutes should be a matter of concern.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VII July 2021- Available at www.ijraset.com

B. Unlock Research Issues

In big data age there are some unlock research issues counting data inspection, parallel programming model, and big data mining structure.

1) Lots of researches on finding errors unseen in data such as [106] cleaning, filtering, and reduction mechanisms.
2) Parallel programming model is introduced to data mining and some algorithms are adopted to be practical in it. Researchers have expanded obtainable data mining methods in many ways. For example parallel association rule mining [107, 108] and parallel *k*-means algorithm based on Hadoop stage are good practice. But there are still some algorithms which are not adapted to parallel platform.
3) The most significant work for big data mining system is to expand a well-organized framework to support big data mining. We need to consider security, privacy, data sharing mechanism and growth of data size. A well calculated data mining framework for big data is a very important way and a big challenge.

C. Current Works of Big Data mining system for IoT

In data mining system area many great companies as Facebook, Yahoo, and Twitter advantage and contribute works to open source projects. Big data mining infrastructure includes the following.

1) Apache Mahout Project equipment a broad range of machine learning and data mining algorithms [109].
2) R Project is a programming language and software environment planned for statistical computing and visualization [110].
3) MOA project performs data mining in real time [111] and SAMOA [112] project integrates MOA with Strom and S4.
4) Pegasus is a pet scale graph mining records for the Hadoop platform [113].

Some researchers from IoT region also projected big data mining system architectures for IoT and these systems center on the combination with devices and data mining technologies [114]. Figure 6 shows architecture maintains of social network and cloud computing in IoT. They included the big data and KDD into the extraction, management and mining, and interpretation layers. The extraction layer maps onto the insight layer.

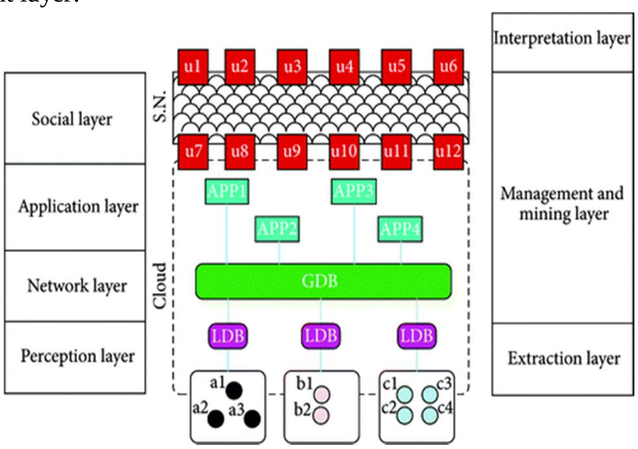

Figure 6 Big data mining system for IoT.

D. Recommended System Architecture for IoT

In this system it includes 5 layers as shown in Figure 7.

1) Devices: lots of IoT devices such as sensors, RFID, cameras, and other devices, can be incorporated into this system to apperceive the world and produce data continuously.
2) Raw data: in the big data mining system structured, semi structured, and unstructured data can be integrated.
3) Data collect: real-time data and batch data can be supported and all data can be parsed, analyze, and compound.
4) Data processing: lots of open source solutions are integrated, including Hadoop, HDFS, Storm, and Oozie.
5) Service: data mining functions will be providing as service.
6) Security/privacy/standard: security, privacy, and standard are very significant to big data mining system. Security and privacy defend the data from illegal access and privacy revelation.

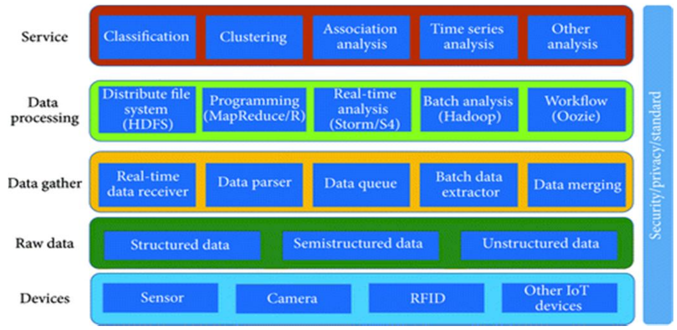

Figure 7 The suggested big data mining system.

V. CONCLUSIONS

The Internet of Things idea arises from the need to supervise, computerize, and explore all devices, instruments, and sensors in the world. In arrange to make intelligent decisions both for community and for the things in IoT, data mining technologies are included with IoT technologies for decision making maintain and system optimization. Data mining involves discovering narrative, motivating, and potentially practical patterns from data and applying algorithms to the taking out of hidden information. In this paper, we survey the data mining in three different views: knowledge view, technique view, and application view. In knowledge view we analysis classification, clustering, association, time series and outlier analysis. In application view we analysis the typical data mining application, including e-commerce, industry, health care, and public service. The technique view is discussed with knowledge view and application view. At the present time, big data is a hot topic for data mining and IoT. We also converse about the new characteristics of big data and examine the challenges in data extracting, data mining algorithms. Based on the survey of the present research a optional big data mining system is proposed.

REFERENCES

[1] Jing, Q., Vasilakos, A. V., Wan, J., Lu, J., Qiu, D.Security of the internet of things: perspectives and challengesWireless Networks2014208248125010.1007/s11276-014-0761-7Google Scholar | Crossref | ISI

[2] Tsai, C.-W., Lai, C.-F., Vasilakos, A. V.Future internet of things: open issues and challengesWireless Networks2014208220121710.1007/s11276-014-0731-02-s2.0-84901572822Google Scholar | Crossref | ISI

[3] Jiawei, H., Kamber, M.Data Mining: Concepts and Techniques2011Morgan KaufmannGoogle Scholar

[4] Mukhopadhyay, A., Maulik, U., Bandyopadhyay, S., Coello, C. A. C.A survey of multiobjective evolutionary algorithms for data mining: part IIEEE Transactions on Evolutionary Computation201418141910.1109/tevc.2013.22900862-s2.0-84893833215Google Scholar | Crossref | ISI

[5] Zhang, Y., Chen, M., Mao, S., Hu, L., Leung, V.CAP: crowd activity prediction based on big data analysisIEEE Network201428452571 0.1109/mnet.2014.6863132Google Scholar | Crossref | ISI

[6] Chen, M., Mao, S., Liu, Y.Big data: a surveyMobile Networks and Applications20141921712091 0.1007/s11036-013-0489-02-s2.0-84898796363Google Scholar | Crossref | ISI

[7] Chen, M., Mao, S., Zhang, Y., Leung, V.Big Data: Related Technologies, Challenges and Future Prospects2014SpringerSpringerBriefs in Computer ScienceGoogle Scholar | Crossref

[8] Wan, J., Zhang, D., Sun, Y., Lin, K., Zou, C., Cai, H.VCMIA: a novel architecture for integrating vehicular cyber-physical systems and mobile cloud computingMobile Networks and Applications201419215316010.1007/s11036-014-0499-62-s2.0-84898828128Google Scholar | Crossref | ISI

[9] Rong, X. H., Chen, F., Deng, P., Ma, S. L.A large-scale device collaboration mechanismJournal of Computer Research and Development2011489158915962-s2.0-80053997391Google Scholar

[10] Chen, F., Rong, X.-H., Deng, P., Ma, S.-L.A survey of device collaboration technology and system softwareActa Electronica Sinica20113924404472-s2.0-79955052781Google Scholar

[11] Zhou, L., Chen, M., Zheng, B., Cui, J.Green multimedia communications over Internet of ThingsProceedings of the IEEE International Conference on Communications (ICC '12)June 2012Ottawa, Canada194819521 0.1109/icc.2012.63639092-s2.0-84871967365Google Scholar | Crossref

[12] Deng, P., Zhang, J. W., Rong, X. H., Chen, F.A model of large-scale Device Collaboration system based on PI-Calculus for green communicationTelecommunication Systems201352213131326 10.1007/s11235-011-9643-92-s2.0-84879603230Google Scholar | Crossref | ISI

[13] Deng, P., Zhang, J. W., Rong, X. H., Chen, F.Modeling the large-scale device control system based on PI-CalculusAdvanced Science Letters201146-72374237910.1166/asl.2011.13982-s2.0-80051542301Google Scholar | Crossref

[14] Zhang, J., Deng, P., Wan, J., Yan, B., Rong, X., Chen, F.A novel multimedia device ability matching technique for ubiquitous computing environmentsEURASIP Journal on Wireless Communications and Networking201320131, article 181121 0.1186/1687-1499-2013-1812-s2.0-84894120909Google Scholar | Crossref | ISI

[15] Sog, S. Analysis and acceleration of data mining algorithms on high performance reconfigurable computing platforms [Ph.D. thesis]2011Iowa State UniversityGoogle Scholar

[16] Larose, D. T. k-nearest neighbor algorithmDiscovering Knowledge in Data: An Introduction to Data Mining2005John Wiley & Sons90106 Google Scholar | Crossref

[17] Hwang, W.-J., Wen, K.-W.Fast kNN classification algorithm based on partial distance searchElectronics Letters19983421206220610.1049/el:199814272-s2.0-3743053629
Google Scholar | Crossref | ISI
[18] Jeng-Shyang, P., Yu-Long, Q., Sheng-He, S.Fast k-nearest neighbors classification algorithmIEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences2004874961963Google Scholar
[19] Pan, J.-S., Lu, Z.-M., Sun, S.-H.An efficient encoding algorithm for vector quantization based on subvector techniqueIEEE Transactions on Image Processing20031232652701 0.1109/TIP.2003.8105872-s2.0-0038192328Google Scholar | Crossref | Medline | ISI
[20] Lu, Z.-M., Sun, S.-H.Equal-average equal-variance equal-norm nearest neighbor search algorithm for vector quantizationIEICE Transactions on Information and Systems20038636606632-s2.0-0038719254Google Scholar
[21] Tang, L. L., Pan, J. S., Guo, X., Chu, S. C., Roddick, J. F.A novel approach on behavior of sleepy lizards based on K-nearest neighbor algorithmSocial Networks: A Framework of Computational Intelligence2014526Cham, SwitzerlandSpringer287311Studies in Computational Intelligence10.1007/978-3-319-02993-1_13 Google Scholar | Crossref
[22] Maron, M. E., Kuhns, J. L.On relevance, probabilistic indexing and information retrievalJournal of the ACM19607321624410.1145/321033.321035Google Scholar | Crossref | ISI
[23] Minsky, M. Steps toward artificial intelligenceProceedings of the IRE1961491830MR0134428Google Scholar | Crossref
[24] Langley, P., Sage, S.Induction of selective Bayesian classifiersProceedings of the 10th International Conference on Uncertainty in Artificial Intelligence1994399406Google Scholar
[25] Kononenko, I. Semi-naive Bayesian classifierMachine Learning—EWSL-911991482Berlin, GermanySpringer206219Lecture Notes in Artificial Intelligence10.1007/BFb0017015MR1101397Google Scholar | Crossref
[26] Zheng, F., Webb, G. I.Tree Augmented Naive Bayes2010Berlin, GermanySpringerGoogle Scholar
[27] Jiang, L., Zhang, H., Cai, Z., Su, J.Learning tree augmented naive bayes for rankingProceedings of the 10th International Conference on Database Systems for Advanced Applications (DASFAA '05)2005688698Google Scholar
[28] Sahami, M. Learning limited dependence Bayesian classifiersProceedings of the 2nd International Conference on Knowledge Discovery and Data MiningAugust 1996Portland, Ore, USA335338Google Scholar
[29] Friedman, N. Learning belief networks in the presence of missing values and hidden variablesProceedings of the 14th International Conference on Machine Learning1997125133Google Scholar
[30] Lei, Y., Ding, X. Q., Wang, S. J.Visual tracker using sequential Bayesian learning: discriminative, generative, and hybridIEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics200838615781591 10.1109/tsmcb.2008.9282262-s2.0-57049142065Google Scholar | Crossref | Medline
[31] Geiger, D., Heckerman, D.Knowledge representation and inference in similarity networks and Bayesian multinetsArtificial Intelligence1996821-245741 0.1016/0004-3702(95)00014-3MR13910562-s2.0-0030125397Google Scholar | Crossref | ISI
[32] Joachims, T. Text categorization with support vector machines: learning with many relevant featuresMachine Learning: ECML-9819981398Berlin, GermanySpringer13714210.1007/bfb0026683
Google Scholar | Crossref
[33] Yingxin, L., Xiaogang, R.Feature selection for cancer classification based on support vector machineJournal of Computer Research and Development2005421017961801Google Scholar | Crossref
[34] Tang, Y., Jin, B., Sun, Y., Zhang, Y.-Q.Granular support vector machines for medical binary classification problemsProceedings of the IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB '04)October 200473782-s2.0-17044427050Google Scholar
[35] Guo, H.-S., Wang, W.-J., Men, C.-Q.A novel learning model-kernel granular support vector machineProceedings of the International Conference on Machine Learning and CyberneticsJuly 200993093510.1109/icmlc.2009.52124132-s2.0-70350716755
Google Scholar | Crossref
[36] Lian, K., Huang, J., Wang, H., Long, B.Study on a GA-based SVM decision-tree multi-classification strategyActa Electronica Sinica200836815021507Google Scholar
[37] Lin, C.-F., Wang, S.-D.Fuzzy support vector machinesIEEE Transactions on Neural Networks200213246447110.1109/72.9914322-s2.0-0036505650
Google Scholar | Crossref | Medline
[38] Huang, H.-P., Liu, Y.-H.Fuzzy support vector machines for pattern recognition and data miningInternational Journal of Fuzzy Systems200243826835MR1933593Google Scholar
[39] Yan, W.-Y., He, Q.Multi-class fuzzy support vector machine based on dismissing marginProceedings of the International Conference on Machine Learning and CyberneticsJuly 20091139114410.1109/icmlc.2009.52123682-s2.0-70350708096
Google Scholar | Crossref
[40] Qi, Z., Tian, Y., Shi, Y.Robust twin support vector machine for pattern classificationPattern Recognition20134613053161 0.1016/j.patcog.2012.06.0192-s2.0-84866023118Google Scholar | Crossref | ISI
[41] Khemchandani, R., Chandra, S.Twin support vector machines for pattern classificationIEEE Transactions on Pattern Analysis and Machine Intelligence200729590591010.1109/TPAMI.2007.10682-s2.0-34047225880Google Scholar | Crossref | Medline | ISI
[42] Qi, Z., Tian, Y., Shi, Y.Structural twin support vector machine for classificationKnowledge-Based Systems20134374811 0.1016/j.knosys.2013.01.0082-s2.0-84875271568Google Scholar | Crossref | ISI
[43] Tsyurmasto, P., Zabarankin, M., Uryasev, S.Value-at-risk support vector machine: stability to outliersJournal of Combinatorial Optimization20142812182321 0.1007/s10878-013-9678-9MR32151082-s2.0-84888793792Google Scholar | Crossref | ISI
[44] Herbrich, R., Graepel, T., Obermayer, K.Large margin rank boundaries for ordinal regressionAdvances in Neural Information Processing Systems1999MIT Press115132Google Scholar
[45] Jain, A. K., Dubes, R. C.Algorithms for Clustering Data1988Englewood Cliffs, NJ, USAPrentice HallGoogle Scholar
[46] Ansari, S., Chetlur, S., Prabhu, S., Kini, G. N., Hegde, G., Hyder, Y.An overview of clustering analysis techniques used in data miningInternational Journal of Emerging Technology and Advanced Engineering2013312284286Google Scholar

[47] Guha, S., Rastogi, R., Shim, K.ROCK: a robust clustering algorithm for categorical attributesProceedings of the 15th International Conference on Data Engineering (ICD ′99)March 19995125212-s2.0-0032652570Google Scholar | Crossref

[48] Thomopoulos, S. C. A., Bougoulias, D. K., Wann, C.-D.Dignet: an unsupervised-learning clustering algorithm for clustering and data fusionIEEE Transactions on Aerospace and Electronic Systems199531121381O.1109/7.3662892-s2.0-0029221045 Google Scholar | Crossref | ISI

[49] Zhang, T., Ramakrishnan, R., Livny, M.BIRCH: a new data clustering algorithm and its applicationsData Mining and Knowledge Discovery19971214118210.1023/a:10097838243282-s2.0-21944442892Google Scholar | Crossref | ISI

[50] Keogh, E., Chakrabarti, K., Pazzani, M., Mehrotra, S.Dimensionality reduction for fast similarity search in large time series databasesKnowledge and Information Systems200133263286 Google Scholar | Crossref

[51] Nagesh, H. S., Goil, S., Choudhary, A. N.Adaptive grids for clustering massive data setsProceedings of the 1st SIAM International Conference on Data Mining (SDM ′01)April 2001Chicago, Ill, USA117Google Scholar

[52] Agrawal, R., Imieliński, T., Swami, A.Mining association rules between sets of items in large databasesProceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD ′93)19932072161O.1145/170035.170072Google Scholar | Crossref

[53] Gosain, A., Bhugra, M.A comprehensive survey of association rules on quantitative data in data miningProceedings of the IEEE Conference on Information & Communication Technologies (ICT ′13)April 2013JeJu Island, Republic of Korea10031008 10.1109/cict.2013.65582442-s2.0-84881620516 Google Scholar | Crossref

[54] Han, J., Pei, J.Mining frequent patterns by pattern-growth: methodology and implicationsACM SIGKDD Explorations Newsletter20002214201O.1145/380995.381002Google Scholar | Crossref

[55] Fu, T.-C. A review on time series data miningEngineering Applications of Artificial Intelligence20112411641811O.1016/j.engappai.2010.09.0072-s2.0-78649672225Google Scholar | Crossref | ISI

[56] Esling, P., Agon, C.Time-series data miningACM Computing Surveys2012451, article 123410.1145/2379776.23797882-s2.0-84871210043Google Scholar | Crossref | ISI

[57] Kalpakis, K., Gada, D., Puttagunta, V.Distance measures for effective clustering of ARIMA time-seriesProceedings of the IEEE International Conference on Data Mining (ICDM ′01)December 2001San Jose, Calif, USA2732802-s2.0-781492994181O.1109/ICDM.2001.989529Google Scholar | Crossref

[58] Kumar, N., Lolla, V. N., Keogh, E., Lonardi, S., Ratanamahatana, C. A., Wei, L.Time-series bitmaps: a practical visualization tool for working with large time series databasesProceedings of the 5th SIAM International Conference on Data Mining (SDM ′05)April 20055315352-s2.0-84880090937Google Scholar

[59] Chan, F. K.-P., Fu, A. W.-C., Yu, C.Haar wavelets for efficient similarity search of time-series: with and without time warpingIEEE Transactions on Knowledge and Data Engineering20031536867051O.1109/tkde.2003.11983992-s2.0-0038294452Google Scholar | Crossref | ISI

[60] Shasha, D. E., Zhu, Y.High Performance Discovery in Time Series: Techniques and Case Studies2004Springer10.1007/978-1-4757-4046-2MR2079948Google Scholar | Crossref

[61] Vlachos, M., Gunopulos, D., Das, G.Indexing time-series under conditions of noiseData Mining in Time Series Databases200457World Scientific67100Series in Machine Perception and Artificial Intelligence10.1142/9789812565402_0004Google Scholar | Crossref

[62] Megalooikonomou, V., Li, G., Wang, Q.A dimensionality reduction technique for efficient similarity analysis of time series databasesProceedings of the 13th ACM International Conference on Information and Knowledge Management (CIKM ′04)November 2004Washington, DC, USA1601612-s2.0-1874441601410.1145/1031171.1031203Google Scholar | Crossref

[63] Chen, Q., Chen, L., Lian, X., Liu, Y., Yu, J. X.Indexable PLA for efficient similarity searchProceedings of the 33rd International Conference on Very Large Data BasesSeptember 2007Vienna, Austria435446Google Scholar

[64] Dong, X. L., Gu, C. K., Wang, Z. O.Research on shape-based time series similarity measureProceedings of the International Conference on Machine Learning and CyberneticsAugust 2006125312581O.1109/icmlc.2006.2586482-s2.0-33947269937 Google Scholar | Crossref

[65] Megalooikonomou, V., Wang, Q., Li, G., Faloutsos, C.A multiresolution symbolic representation of time seriesProceedings of the 21st International Conference on Data Engineering (ICDE ′05)April 20056686791O.1109/icde.2005.102-s2.0-28444432990 Google Scholar | Crossref

[66] Assent, I., Krieger, R., Afschari, F., Seidl, T.The TS-tree: efficient time series search and retrievalProceedings of the 11th International Conference on Extending Database Technology: Advances in Database Technology (EDBT ′08)2008252263Google Scholar

[67] Heer, J., Chi, E. H.Identification of web user traffic composition using multi-modal clustering and information scentProceedings of the Workshop on Web Mining, SIAM Conference on Data Mining20015158Google Scholar

[68] Elgendy, N., Elragal, A.Big data analytics: a literature review paperAdvances in Data Mining. Applications and Theoretical Aspects20148557Cham, SwitzerlandSpringer214227Lecture Notes in Computer Science10.1007/978-3-319-08976-8_16Google Scholar | Crossref

[69] Liu, Q., Wan, J., Zhou, K.Cloud manufacturing service system for industrial-cluster-oriented applicationJournal of Internet Technology2014153373380Google Scholar | ISI

[70] Maaß, D., Spruit, M., de Waal, P.Improving short-term demand forecasting for short-lifecycle consumer products with data mining techniquesDecision Analytics201411117Google Scholar | Crossref

[71] Du, X. F., Leung, S. C. H., Zhang, J. L., Lai, K. K.Demand forecasting of perishable farm products using support vector machineInternational Journal of Systems Science201344355656710.1080/00207721.2011.6178882-s2.0-84870610658Google Scholar | Crossref | ISI

[72] Lu, C.-J., Wang, Y.-W.Combining independent component analysis and growing hierarchical self-organizing maps with support vector regression in product demand forecastingInternational Journal of Production Economics201012826036110.1016/j.ijpe.2010.07.0042-s2.0-78049295934Google Scholar | Crossref | ISI

[73] Lee, H., Kim, S. G., Park, H.-W., Kang, P.Pre-launch new product demand forecasting using the Bass model: a statistical and machine learning-based approachTechnological Forecasting and Social Change20138649641O.1016/j.techfore.2013.08.0202-s2.0-84883881196Google Scholar | Crossref | ISI

[74] Chen, M., Gonzalez, S., Leung, V., Zhang, Q., Li, M.A 2G-RFID-based e-healthcare systemIEEE Wireless Communications2010171374310.1109/mwc.2010.54163482-s2.0-77649143550Google Scholar | Crossref | ISI

[75] Liu, J., Wan, J., He, S., Zhang, Y.E-healthcare supported by big dataZTE Communications20141234652Google Scholar

[76] Chen, M., Ma, Y., Wang, J., Mau, D. O., Song, E.Enabling comfortable sports therapy for patient: a novel lightweight durable and portable ECG monitoring systemProceedings of the IEEE 15th International Conference on e-Health Networking, Applications and Services (Healthcom '13)October 2013Lisbon, PortugalIEEE2712710.1109/healthcom.2013.67206812-s2.0-84894166007Google Scholar | Crossref

[77] Liu, J., Wang, Q., Wan, J., Xiong, J., Zeng, B.Towards key issues of disaster aid based on Wireless Body Area NetworksKSII Transactions on Internet and Information Systems20137510141035 10.3837/tiis.2013.05.0052-s2.0-84878463160Google Scholar | Crossref | ISI

[78] Chen, M. NDNC-BAN: supporting rich media healthcare services via named data networking in cloud-assisted wireless body area networksInformation Sciences20142841014215610.1016/j.ins.2014.06.023Google Scholar | Crossref | ISI

[79] Chen, M., Mau, D. O., Wang, X., Wang, H.The virtue of sharing: efficient content delivery in wireless body area networks for ubiquitous healthcareProceedings of the IEEE 15th International Conference on e-Health Networking, Applications & Services (Healthcom '13)October 2013Lisbon, Portugal6696710.1109/HealthCom.2013.6720760Google Scholar | Crossref

[80] Wan, J., Zou, C., Ullah, S., Lai, C.-F., Zhou, M., Wang, X.Cloud-enabled wireless body area networks for pervasive healthcareIEEE Network201327556610.1109/MNET.2013.66161162-s2.0-84885589116Google Scholar | Crossref | ISI

[81] Duan, L., Street, W. N., Xu, E.Healthcare information systems: data mining methods in the creation of a clinical recommender systemEnterprise Information Systems20115216918110.1080/17517575.2010.5412872-s2.0-79952445361Google Scholar | Crossref | ISI

[82] Schuerenberg, B. K. An information excavation. Las Vegas payer uses data mining software to improve HEDIS reporting and provider profilingHealth Data Management200311680822-s2.0-0043267751Google Scholar | Medline

[83] Sun, J., Reddy, C. K.Big data analytics for healthcareProceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data MiningAugust 2013Chicago, Ill, USA152510.1145/2487575.2506178Google Scholar | Crossref

[84] Kincade, K. Data mining: digging for healthcare goldInsurance & Technology199823227Google Scholar

[85] Bellazzi, R., Zupan, B.Predictive data mining in clinical medicine: current issues and guidelinesInternational Journal of Medical Informatics20087728197 10.1016/j.ijmedinf.2006.11.0062-s2.0-37249089420Google Scholar | Crossref | Medline | ISI

[86] Liu, J., Pan, J., Wang, Y., Lin, D., Shen, D., Yang, H., Li, X., Luo, M., Cao, X.Component analysis of Chinese medicine and advances in fuming-washing therapy for knee osteoarthritis *via* unsupervised data mining methodsJournal of Traditional Chinese Medicine20133335686691 10.1016/s0254-6272(14)60043-12-s2.0-84887756703Google Scholar | Crossref | ISI

[87] Silver, M., Sakata, T., Su, H. C., Herman, C., Dolins, S. B., O'Shea, M. J.Case study: how to apply data mining techniques in a healthcare data warehouseJournal of Healthcare Information Management20011521551642-s2.0-0035380848Google Scholar | Medline

[88] Koh, H. C., Tan, G.Data mining applications in healthcareJournal of Healthcare Information Management201119265Google Scholar

[89] Thornton, D., Mueller, R. M., Schoutsen, P., van Hillegersberg, J.Predicting healthcare fraud in medicaid: a multidimensional data model and analysis techniques for fraud detectionProcedia Technology2013912521264Google Scholar | Crossref

[90] Helbig, N., Gil-García, J. R., Ferro, E.Understanding the complexity of electronic government: implications from the digital divide literatureGovernment Information Quarterly20092618999 10.1016/j.giq.2008.05.0042-s2.0-56849130203Google Scholar | Crossref | ISI

[91] Wan, J., Li, D., Zou, C., Zhou, K.M2M communications for smart city: an event-based architectureProceedings of the IEEE 12th International Conference on Computer and Information Technology (CIT '12)October 2012Chengdu, China89590010.1109/cit.2012.1882-s2.0-84872352579Google Scholar | Crossref

[92] Chadwick, A., May, C.Interaction between states and citizens in the age of the internet: 'e-government' in the United States, Britain, and the European UnionGovernance200316227130010.1111/1468-0491.002162-s2.0-0037787917Google Scholar | Crossref | ISI

[93] Peng, Y., Zhang, Y., Tang, Y., Li, S.An incident information management framework based on data integration, data mining, and multi-criteria decision makingDecision Support Systems201151231632710.1016/j.dss.2010.11.0252-s2.0-79953753719Google Scholar | Crossref | ISI

[94] Sullivan, B., Mitra, S.Community issues in American metropolitan cities: a data mining case studyJournal of Cases on Information Technology201416123391 0.4018/jcit.2014010103Google Scholar | Crossref

[95] Chen, M. Towards smart city: M2M communications with software agent intelligenceMultimedia Tools and Applications20136711671781 0.1007/s11042-012-1013-42-s2.0-84881146585Google Scholar | Crossref | ISI

[96] Wang, G., Chen, H., Atabakhsh, H.Automatically detecting deceptive criminal identitiesCommunications of the ACM200447370761 0.1145/971617.9716182-s2.0-1942443495Google Scholar | Crossref | ISI

[97] Chen, H., Chung, W., Qin, Y., Chau, M., Xu, J. J., Wang, G., Zheng, R., Atabakhsh, H.Crime data mining: an overview and case studiesProceedings of the Annual National Conference on Digital Government Research200315Google Scholar

[98] Cao, X., Cong, G., Jensen, C. S.Mining significant semantic locations from GPS dataProceedings of the VLDB Endowment201031-210091020Google Scholar | Crossref

[99] Wan, J., Zhang, D., Zhao, S., Yang, L. T., Lloret, J.Context-aware vehicular cyber-physical systems with cloud support: architecture, challenges, and solutionsIEEE Communications Magazine2014528106113 10.1109/mcom.2014.6871677Google Scholar | Crossref | ISI

[100] Schroedl, S., Wagstaff, K., Rogers, S., Langley, P., Wilson, C.Mining GPS traces for map refinementData Mining and Knowledge Discovery200491598710.1023/b:dami.0000026904.74892.89MR20555552-s2.0-3543072870Google Scholar | Crossref | ISI

[101] Zheng, Y., Zhang, L., Xie, X., Ma, W.Mining interesting locations and travel sequences from GPS trajectoriesProceedings of 18th International Conference on World Wide Web2009791800Google Scholar

[102] Hu, T., Chen, H., Huang, L., Zhu, X.A survey of mass data mining based on cloud-computingProceedings of the International Conference on Anti-Counterfeiting, Security and Identification (ASID '12)August 20121410.1109/icasid.2012.63253532-s2.0-84870624640Google Scholar | Crossref

[103] Sun, Y., Han, J., Yan, X., Yu, P. S.Mining knowledge from interconnected data: a heterogeneous information network analysis approachProceedings of the VLDB Endowment201220222023Google Scholar

[104] Chen, M., Yang, L. T., Kwon, T., Zhou, L., Jo, M.Itinerary planning for energy-efficient agent communications in wireless sensor networksIEEE Transactions on Vehicular Technology20116073290329910.1109/TVT.2011.21341162-s2.0-80052854703Google Scholar | Crossref | ISI

[105]Zhang, D., Wan, J., Liu, Q., Guan, X., Liang, X.A taxonomy of agent technologies for ubiquitous computing environmentsKSII Transactions on Internet and Information Systems20126254756510.3837/tiis.2012.02.0062-s2.0-84862173924 Google Scholar | Crossref | ISI

[106]Chen, M., Leung, V. C. M., Mao, S.Directional controlled fusion in wireless sensor networksMobile Networks and Applications20091422202910.1007/s11036-008-0133-62-s2.0-62249146158Google Scholar | Crossref | ISI

[107]Wu, X., Zhang, S.Synthesizing high-frequency rules from different data sourcesIEEE Transactions on Knowledge and Data Engineering20031523533610.1109/TKDE.2003.11858392-s2.0-0037339975Google Scholar | Crossref | ISI

[108]Su, K., Huang, H., Wu, X., Zhang, S.A logical framework for identifying quality knowledge from different data sourcesDecision Support Systems20064231673168310.1016/j.dss.2006.02.0122-s2.0-33750493303Google Scholar | Crossref | ISI

[109]Owen, S., Anil, R., Dunning, T., Friedman, E.Mahout in Action2011ManningGoogle Scholar

[110]R Development Core Team R: A Language, and Environment for Statistical Computing2012Vienna, AustriaR Foundation for Statistical ComputingGoogle Scholar

[111]Bifet, A., Holmes, G., Kirkby, R., Pfahringer, B.Moa: massive online analysisThe Journal of Machine Learning Research201011160116042-s2.0-77953527363Google Scholar | ISI

[112]de Francisci Morales, G. SAMOA: a platform for mining big data streamsProceedings of the 22nd International Conference on World Wide Web (WWW '13)May 20137777782-s2.0-84893053113Google Scholar

[113]Kang, U., Chau, D. H., Faloutsos, C.Pegasus: mining billion-scale graphs in the cloudProceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '12)March 2012Kyoto, Japan5341534410.1109/icassp.2012.62891272-s2.0-84867619547Google Scholar | Crossref

[114]da Silva, W. M., Alvaro, A., Tomas, G. H. R. P., Afonso, R. A., Dias, K. L., Garcia, V. C.Smart cities software architectures: a surveyProceedings of the 28th Annual ACM Symposium on Applied Computing (SAC '13)March 20131722172710.1145/2480362.24806882-s2.0-84877940180Google Scholar | Crossref

I am Dr.Mohd Zuber, Assistant Professor in Girdher Group of institution.