International Journal of Advanced Science and Technology
Vol. 29, No. 2, (2020), pp. 4484 – 4494

Truth Discovery Algorithms Models and Their Applications in World Wide Web- A Survey

Adilakshmi Vadavalli and Subhashini R
School of computing, Sathyabama Institute of science and Technology, Chennai, 600119, Tamil Nadu, India
Emails: Adilakshmi.it@sathyabama.ac.in, subhaagopi@gmail.com

Abstract

A major challenge in societal recognizing applications lies in determining the perfection of described opinions and trustworthiness of information sources devoid of preceding awareness on both of them. This problem is denoted as truth discovery. Truth discovery is the issue of identifying values that are true from the differing information delivered by several sources on the similar data items. This truth discovery approach in turn incorporates the multi-source noisy data on assessing the consistency of each source, has appeared as an important field. A number of truth discovery methods have been suggested for several consequences, and they have been effectively functional in varied application areas. The world-wide web has become an essential portion of our lives, and might have turn out to be the most significant data source for most individuals. Everyday people repossess all types of data from the web. In this survey, a comprehensive review was made on the truth discovery algorithms. Also, different models its applications in the World Wide Web were discussed.

Index Terms— *Truth discovery, world-wide web, multi-source noisy data.*

I. INTRODUCTION

By the growth of information technology, the Internet has entered into all angle of social common life. The information on internet have gathered abruptly, and these data have been unified into an information deeply. Some of the significant structures of this data is multiplicity, consequently for anything, various reports can be established on internet from several sources. Social media is one of the leading information thresholds, which is used to form consensus among the public's all over the world. When related to other social media comprising Facebook, Instagram, WhatsApp, etc., twitter has achieved a significant consideration in modern times. It delivers an easy and quick network admittance for the consumers to share their data. The irregularity or struggle of these varied descriptions causes excessive misperception for us to recognize [1] true data from each other. Consequently, recognizing the correct and wide-ranging data from incompatible accounts is the important issue for data incorporation. Thus the [2-6] truth discovery problem should be identified. The problem to discover out the actuality from unreliable data is well-defined as Truth Discovery. The principle of truth discovery is to evaluate the quality of source. As a result the computing device of information source will massively disturb the outcome and development of truth discovery. On the other hand the high-tech processes don't consider how source quality is affected when null is provided by source. In several claims, the data concerning to the similar object can be composed from numerous sources. On the other hand, these data that are multi-source are not described dependably. In the bright of this encounter, truth discovery is appeared to recognize truth for each object from multi-source data. Furthermost prevailing truth discovery approaches adopt that ground actualities are entirely unidentified, and they emphasis on the examination of unsubstantiated methodologies to cooperatively evaluate object truths and source consistencies. Conversely, in several real world presentations, a set of ground realities might be moderately accessible. The significant issue in obtaining the results is to detect the trustworthiness of the information obtained from the crowdsourced

ISSN: 2005-4238 IJAST

environment. Hence it is necessary to determine the truth discovery among the conflicts in a crowd sourced environment

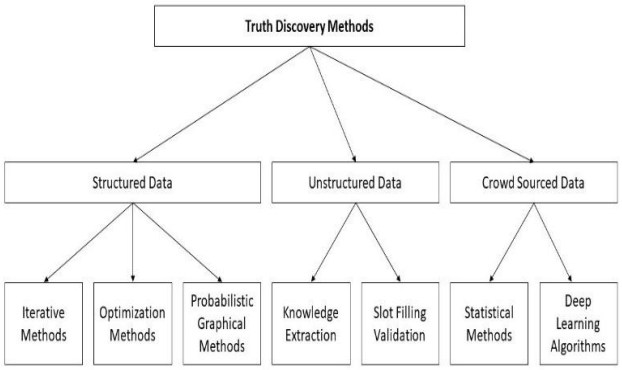

Fig 1. Truth Discovery methods

In figure 1 the major truth discovery methods with respect to different kinds of data has been shown. These methods can be applied on structured data, unstructured data and more recently the crowd sourced data as well and better insights can be drawn. The primitive approach for truth discovery is the Majority Voting which chooses the majority answers that are obtained from all the sources. But this approach fails to consider the reliability levels of various sources which may further causes poor performance. To overcome this failure, several techniques for multi-source aggregation has been introduced which considers the estimation of source reliability. But this approach also has the most common disadvantage of estimating one reliability degree for each source. This does not reflect the variations in the reliability properly in real world problems. Hence the process of data fusion is carried out which integrates the several data sources to form an accurate results than the results that are obtained by individual data sources [7-12]. Also the process of finding the true triples from multiple data sources are carried out using knowledge fusion. Several existing techniques are used for determining the truth discovery over crowd sourcing. But they had some challenging drawbacks like inaccurate results, less efficiency and reduces the overall performance of the system. In the existing method, the true values are determined by using batch truth algorithm [13-17]. However, that method works efficiently only for the small datasets and the results are also not scalable. The batch truth discovery algorithm has two main drawbacks when dealing with big data as mentioned below: It requires to iterate over the entire dataset numerous times until convergence, which results in time complexities. It is essential to load and process the entire dataset, which may cause an "out of memory" issue.

II. SURVEY ON DATA MINING

The advancement in information technology leads to the storage of large volume of data in several formats such as documents, files, images, sound recordings, scientific data, and the videos and a several new formats of data. For the better decision making, the data collected from the different applications will require the [18] appropriate mechanism for the extraction of information or knowledge from the huge sources. The area of data mining has been posed into the new fields of human life with the advancement and integration of areas such as databases, [19] statistics, pattern recognition, machine learning, computer capabilities, artificial intelligence and so on. The major intention of using this survey was to study security threats in the data mining using the cloud and to increase the security level in the cloud system by the creation of user access policies. The detection of recurrent patterns, correlation relationships, and association rules between vast quantities of data are valuable to the professional intellect. A distinctive model of frequent item set mining is the marketplace basket investigation. This procedure examines [20] consumer obtaining behaviors by verdict relations among the altered items that consumers place in their

"shopping baskets". The finding of such relations can aid traders to improve promoting approaches by means of fast awareness into which items are regularly acquired together by consumers. As there was more integration of data to mine, the application of data fusion increased the value of data mining. The data that are health-care related are of explosive still there were some challenges for the management of data, processing, and storage. The main intention of IoT is to alter [21, 22] the objects that are outmoded to make them smart on exploiting the extensive range of technologies that are more advanced. The prospective impact on economic of the IoT is suspected to make several business opportunities thereby accelerating the financial raise of IoT dependent services. However, the machine learning may have some adverse effects on workforce, and job as various parts of jobs are more preferable for the application of machine learning which in turn increases the demand of ML approaches and their products. The report [23] declares that improvements in ML systems, like neural networks and deep learning, are the chief enablers of information work automation.

III. TRUTH DISCOVERY ALGORITHMS

This section provides the survey of various approaches regarding the truth discovery model. [24] Elucidated the different truth discovery approaches to tackle the problem of collecting the data from similar objects that makes conflicts between the composed multi source information. In order to overcome this issue, the multi-source noisy information were integrated in the truth discovery through the estimation of each source reliability. It was appeared as one of the hot matter. In various application domain, numerous trust discovery methods were proposed for the different scenario. An inclusive overview of truth discovery was provided and also summarized them in the diverse aspects. Further research is made on the truth discovery research. It was possible for us to understand better on the truth discovery and it also recommends a procedures for applying these methods over the application domain.

[25] proposed a recent IoT research through investigating the literature, telling contests that impend IoT diffusion, classifying current trends, bestowing open research questions and upcoming directions, and accumulating an inclusive reference list for assisting researchers. In this work, classification scheme covering six categories has been proposed and it was specified as technology, challenges, applications, business models, future directions and impression/survey. After categorizing, based on this classification scheme, literature pool was classified. From that reviews it was analyzed that the IoT favored for the improvement of peoples live in the automation and augmentation. And also IoT was capable of saving people and the time of the organization. It was also improved the decision making for application with wider range. The question raised concerning the IoT was whether it holds better pattern for another paradigm or not. Thus the question answered that if the current research was combined with IoT, there was a chance for acquiring the better potential on IoT for reshaping it.

[26] elucidated a novel approach to achieve multi source data aggregation with the source reliability which is critical thing. Several ways are proposed in the multi-source data aggregation for estimating source reliability. In the existing approach, a naïve adaptation split up the data on the basis of topics, then the aggregation methods were applied for each and every group well-defined through separate topic. However it faced a challenge due to inadequate data for supporting the better estimation of reliability of source. A novel probabilistic Bayesian model was introduced for addressing the contest of fine grained source reliability. Through the collective content of question and answers, the topic of questions, topic-specific expertise of sources, and the true answers were learnt in this approach. The two real crowd sourced datasets proved with the experimental results that the Fait Crowd model was effective in terms of skilled source detecting the true answers in the corresponding topics though the fewer answers are present in the answer set.

[27] Elucidated a new approach in the framework of truth discovery for crowd sensing of correlated entities. On the crowd sensing systems, among the set of correlated entities, users create their

ISSN: 2005-4238 IJAST

observations over some objects and the observations were reported towards central server through smart phones. In this approach, the user's reliability and truths between correlated entities were inferred. The task was framed as an optimization problem where the user reliability and truths were unidentified variables, and as a regularization terms the correlations were modeled. The optimization was difficult for regularization terms because of correlations between unknown variables. In order to overcome this issue, the variables were partitioned toward separate self-determining sets, and the block coordinate descent was conducted for updating the truths and user reliabilities iteratively. The Hadoop cluster will be used far along for the huge scale data. The foremost advantage of this method was its efficiency and it was proved with the experimental results.

[28] Elucidated efficient approach for event discovery and tracking from large micro-blog streams. Using Symbolic Aggregate Approximation (SAX) was employed for converting a word temporal series into string of symbols. In this context, SAX, regular expression learning and hierarchical clustering were combined together to create instinctive framework in the event detection of micro blogs. Initially, a temporal series of terms were discredited into fewer set of levels that leads to each term string. After this, to distinguish event-like terms from non-event terms, a regular expression (regex) was well-read from a group of known events. At last, identified events generated the clusters after the clustering algorithm was applied over the upper part of the event-like strings. While comparing it with other traditional approach, this method has yielded better computational complexity which was more critical to attain in large micro blog streams. In prescribed way, the complexity was computed as a function of system strictures with the dimension of the vocabulary and temporal granularity. The performance of the method was analyzed systematically underneath diverse settings of the model parameters for gaining the vision toward influence over the type and event quality that reliably detected. Also to evaluate a detected events, an objective way was proposed that includes searching for associated Google news on similar temporal slot.

[29] Presented a community detection over networks which was similar to cluster searching in self-governing vector data. It was common for evaluating the performance of community detection algorithm through their capabilities for finding ground truth communities. This was well suited for synthetic networks along with planted communities due to the explicit formation of network link on the basis of recognized communities but in real world networks, there were no planted communities. As an alternative, it was a normal practice for treating some detected discrete-valued node attributes, or metadata, as ground truth It was shown that the metadata was not similar to ground truth and handling them persuades simple theoretical and practical problems. The community detection was not solved exclusively, and general No Free Lunch theorem for community detection was showed that suggests no algorithm acquiring best for every likely community detection tasks. Though, community detection leftovers influential tool and there was a value in node metadata, consequently a cautious survey of their network structure relationship produced visions of honest worth. This point was illustrated through the familiarizing two statistical methods which enumerate the association amongst community structure and metadata for a models with comprehensive class. Both synthetic and real-world networks were utilized to establish those techniques and multiple kinds of metadata and community structure.

[30] elucidated a novel approach in truth discovery to handle numerous kinds of uncertainties and automatic learning effectively. The purpose of this review was to acquire effective truth discovery for that two unsupervised probabilistic models to solve the participants' mobility and reliability that was indeterminate. The location popularity, location visit indicators, truths of events, and three-way participant reliability in an integrated framework were modelled in TSE. The personal location was modeled in PTSE. The location tracking was not required for these models. For TSE and PTSE, a batch and online model algorithms were developed. An extensive experiments were conducted for evaluating the performance of proposed and conventional methods.

ISSN: 2005-4238 IJAST

Tuarob, et al [31] hired an collaborative heterogeneous classification methodology to discover the health related knowledge over social media. In this review, various types of classification techniques were measured, that includes Random Forest (RF), Support Vector Machine (SVM), Multinomial Naïve Bayes (MNB), and Bernouli Naïve Bayes (BNB). Similarly, the sensitivity parameter was exploited for assessing the performance of the feature extraction techniques to attain the best features from separate feature type.

Wang, et al [32] molded an estimation theoretic approach to determine the theme relevant truth on twitter. To provide the suitable solutions for the truth discovery problem, this analytical model measured the theme relevance feature. Furthermore, through estimating the correctness and theme relevance of claims a bi-dimensional estimation problem was solved. Formerly, the analytical model could be exploited for deriving the solutions that were reliable with the detected twitter data.

Shen, et al [33] developed a comprehensive social spammer detection framework through integrating multiple view information and social regularization. Similarly, an empirical in-depth analysis was accompanied on a real world twitter dataset that determined the feature distributions amongst the spammers and genuine users. The advantage of this method was, it measured multiple view information for detecting spammers on the basis of single view methods, optimization methods, and combination methods. Through implementing simple strategy to compliment the missing values, the performance of the spammer detection can be still improved.

Shi, et al [34] applied a user interest model based event evolution strategy for determining the events in a social data streams. For accessing the correlation among the events, a cosine measure based event similarity detection method was introduced. In this method, a set of tweets were categorized into different classes like positive, negative or neutral. But, it required to increase the accuracy of classification through extracting the utmost relevant features.

Schulz, et al [35] organized a semantic abstraction method to improve the generalization in tweet classification. In this method, the location and temporal mentions features were extracted from the accompanied open data. At that moment, the incident related tweets were separated concerning various incident types and neutral class. Furthermore, using Heidel time framework the temporal expression extraction and replacement were performed in tweets.

Ji, et al [36] established a twitter sentiment classification method to address the concerns of public health. In this work, a two-step classification model related to mixtures of clue based search and machine learning was established for categorizing the tweet sentiments. At this time, the sentiment timeline and news timeline were correlated in a quantitative and qualitative manner. But even, it required to rise the efficiency of discovery by applying better classification technique, which was the limitation of this model.

[37] described a novel approach which helped to find the true values from the conflicting information from large number of sources. Among this huge source some might be copied form the others. In this work, a case study was presented to prove the intended algorithm could improve the accuracy of the truth discovery. Also it was more scalable even with the large number of data sources. [38] surveyed the methods that were used for finding the true information from the conflicted data. The basic principle of the truth discovery approaches were provided in this review to find the deviation from the conflicted data that were gathered from several sources. These methods were compared with each other for choosing the suitable approach based on the data types such as categorical, numerical and continuous data.

[39] presented an organized picture about the aggregation of truth discovery and crowdsourcing. The main purpose of this combination of truth discovery and crowdsourcing was to resolve the conflicts in the information and also it helped to achieve high quality data. Also this combination could be compared

based on both the theoretical as well as application levels. The theoretical analysis required more attention from both the truth discovery as well as crowdsourcing.

[40] provided a wide review and highlighted the growth which was made on truth discovery from information extraction, knowledge and data fusion and modeling of misinformation dynamics in social networks. The existing techniques, algorithms and models were reviewed in detail. The goal of these contributions was to estimate the veracity of the data in the changing environment. Also this review helped to bridge the theory and practical. This challenges of truth discovery in big data were addressed by introducing current work from different disciplines to the database

IV. APPLICATION OF TRUTH DISCOVERY IN WORLD WIDE WEB

A truth discovery approach can be applicable in several fields like health care, social or mobile sensing, crowd sourced aggregation of data, information retrieval and data fusion, wireless sensor networks or IoT, Question answering system, toxic content classification, management of diasaster and so on.

[41] The World-Wide Web grants survey assistants with an exceptional means for the group of information. The overheads in relations of both money and time for reproducing a survey on the web are small compared with costs accompanying with conservative surveying approaches. The data admittance phase is abolished for the survey manager, and software can safeguard that the information assimilated from contributors is permitted from collective access faults. Prominently, web reviews can interactively deliver applicants with modified response. These structures instigated at a cost of conservation that appropriately transcribed software accomplishes the information gathering manner. Even though the prospective for lost data, improper reactions, replacement submissions, and web manipulation happen, events can be occupied when generating the review software to diminish the occurrence and undesirable significances of such occurrences.

[42] World Wide Web or Web is the popular and biggest source of data accessible, within reach and available at low rate delivers quick answer to the consumers and decreases problem on the handlers of physical activities. The information on the Web is piercing. The noise originates from two foremost sources. Initially, an illustrative Web page covers several portions of data, e.g., the foremost routing links, content of the page, copyright notices, privacy policies, ads, etc. Second, because of the circumstance that the Web does not consume feature control of data, i.e., some can compose nearly everything that one adores, a huge extent of facts on the Web is of small eminence, inaccurate, or even misleading. Retrieving of the essential web page on the web, competently and efficiently, is becoming a challenging.

[43] This paper progresses a different upright background for manipulating time-sensitive data to progress the actuality discovery correctness in social identifying applications. This work is driven by the appearance of social recognizing as a new standard of accumulating interpretations around the physical atmosphere from persons or strategies on their behalf. These interpretations possibly false or true, and henceforth are observed as dual assertions. An important problem in social sensing uses lies in determining the precision of assertions and the dependability of data sources. We mention to this problematic as truth discovery. Time is a serious measurement that wants to be prudently demoralized in the truth discovery resolutions. In this broadsheet, we progress a novel time-sensitive truth discovery system that obviously includes the source receptiveness and the assertion lifetime into a severe logical background. The new truth discovery system resolves a determined probability approximation delinquent to regulate mutually the assertion exactness and the source consistency.

V. TRUTH DISCOVERY METHODS IN VARIOUS SCENARIOS

As discussed there are various scenarios in real time where these methods can be applied like in a crowd sourced system when multiple data values are submitted by crowd users we need a system to ascertain the

true values of the data and the reliability of the users, Also, in a wireless sensor network based environment when multiple sensors are providing conflicting values there is a greater need to do data aggregation/fusion inorder to arrive at accurate values. Also in a question Answering system when multiple users submit answers to different questions we can evaluate the answers whether they are true or false and simultaneously evaluate the reliability of the users. In any social media network like facebook, twitter or linked in we could track genuine users from fake ones and eliminate spam by performing certain feature extraction techniques and truth discovery methods.

[44] The web is a vast source of valued data. Conversely, in recent periods, there is an accumulative development on the way to incorrect assertions in social media, further web-sources, and even in news. Therefore, fact-checking websites have turn out to be progressively prevalent to recognize such misrepresentation created on physical examination. Recent investigation projected approaches to evaluate the trustworthiness of assertions repeatedly. Conversely, there are foremost restrictions: most mechanisms accept rights to be in an organized method, and a limited arrangement with word-based privileges but necessitate that causes of data or counter-evidence are effortlessly repossessed from the web.

Table 1: Truth discovery algorithms and approaches:

Baseline Methods	**Input Data**	**Model**	**Output data**	**Performance Measure**
CRH	Categorical, Continuous	Optimization model	Single truth	Error rate, MNAD
Mean	Continuous	Mean of all observations of an object	Single truth	MNAD
Median	Continuous	Median of all observations of an object	Single truth	MNAD
GTM	Continuous	Bayesian Probabilistic Model	Single truth	MAE, RMSE
LTM	Categorical	Probabilistic Graphical model	Multiple Truth	False Positive, False Negative, Recall Precision
Voting	Categorical, Continuous	Majority voting/Averaging	Single truth	Error rate, MNAD, RMSE
Investment	Categorical	Object relation considered, **nonlinear** function, Source Invests reliability among claimed values	Unknown Truths	MAE, RMSE
Pooled Investment	Categorical	Object relation considered	Unknown Truths	MAE, RMSE
2-estimate	Categorical	Adopts Complementary vote	Single truth	MAE, RMSE

ISSN: 2005-4238 IJAST

3-estimate	Categorical	Object difficulty considered	Single truth	MAE, RMSE
Truthfinder	Categorical, Continuous	Bayesian Analysis	Single truth	MAE, RMSE
AccuSim	Categorical, Continuous	Bayesian Analysis	Single truth	MAE, RMSE
SSTF	Categorical, Continuous	Semi Supervised learning Model	Labeled Truth	MAE, RMSE
AccuCopy	Categorical, Continuous	Source dependence Considered	Single truth	MAE, RMSE

Table 2: Comprehensive survey of truth discovery approaches:

Type	Model\algorithm or task	Description	Usage examples in business
Supervised	Neural network [45]	Computations are structured in terms of interconnected groups, much like the neurons in a brain. Neural networks are used to model complex relationships between inputs and outputs to find patterns in data or to capture a statistical structure among variables with known and unknown relationships. They may also be used to discover unknown input.	Predicting financial results Fraud detection
Supervised	Classification and\or regression [46]	Computations are structured in terms of categorized outputs or observations based on defined classifications. Classification models are used to predict new outputs based on classification rules. Regression models are generally used to predict outputs from training data	Spam filtering Fraud detection
Supervised	Decision tree [47]	Computations are particular representationsof possible solutions to a decision based oncertain conditions. Decision trees are great for building classification models becausethey can decompose datasets into smaller,more manageable subsets	Risk assessment Threat management Systems in which Any optimization problem where an

			exhaustive search is not feasible
unsupervised	Cluster analysis [48]	Computations are structured in terms of groups of input data (clusters) based on howsimilar they are to one another. Cluster analysis is heavily used to solve exploratory challenges where little is known about the data.	Financial transactions Streaming analytics in IoT Underwriting in insurance
unsupervised	Pattern recognition [49]	Computations are used to provide a description or label to input data, such as in classification. Each input is evaluated and matched based on a pattern identified. Pattern recognition can be used for supervised learning as well	Spam detection Biometrics Identity management
unsupervised	Association rule learning [50]	Computations are rule-based in order to determine the relationship between different types of input or variables and to make Predictions.	Security and intrusion Detection Bioinformatics Manufacturing and assembly

Not any of these mechanisms can manage with recently developing statements, and no preceding technique can offer user-interpretable clarifications for its decision on the statement's trustworthiness.

However, there were some challenges in the truth discovery techniques like the handling of unstructured data, source reliability initialization, model selection, Efficiency for parallel, streaming and large data sets, Theoretical Analysis for convergence, and the performance evaluation with the limited labeled ground truths.

VI. CONCLUSION

This survey summarizes different trust discovery methods considering several approaches and its challenges. World Wide Web is the data source for accessing information. Most of the data are retrieved from the web for all major data analysis tasks. As such ascertaining the truth of the data and reliability of

the users is a major challenge in this information era. The truth discovery problem is to identify the true values from different information delivered by several sources on the similar data items. In order to effectively ascertain truth values more number of truth discovery methods has been recommended and were utilized in different application domains. Enormous research has been made to analyze the best truth discovery methods. At each level source consistency must be accessed in this truth discovery field by incorporating multi-source noisy data also. Accordingly this paper provides a comprehensive on the truth discovery algorithms. Similarly, its applications in the World Wide Web were also discussed. Finally, truth discovery methods can be seen as the need of the hour as the data veracity can be estimated based on these and that better insights can be generated from dark data. By making these dynamic and integrating them with optimization techniques the web search goes to the next level.

REFERENCES

1. XuG et al. 2017, "Achieving efficient and privacy-preserving truth discovery in crowd sensing systems,", pp.114-126.
2. Nguyen QVH, 2017, "Argument discovery via crowd sourcing". *Th e VLDB Journal.* 2017, pp.511-535.
3. Bianchi FM, 2017, "An agent-based algorithm exploiting multiple local dissimilarities for clusters mining and knowledge discovery",pp.1347-1369.
4. Cao J, 2016, "Web video topics discovery and structuralization with social network, ",pp. 53-63.
5. Yang S, 2017, "On designing data quality-aware truth estimation and surplus sharing method for mobile crowd sensing", pp.832-847.
6. RuanY, 2015, "Community discovery: Simple and scalable approaches". *User Community Discover,* .pp.23-54.
7. Dong XL, 2014, "From data fusion to knowledge fusion. *Proceedings of the VLDB En- dowment",pp.*881-892.
8. SivaraksH, Ratanamahatana CA, 2015, "Robust and Accurate Anomaly Detection in ECG Artifacts Using Time Series Motif Discovery",pp.1-20.
9. Zelst SV, "Filter techniques for region-based process discovery". *BPMreports.* 2015;1504.
10. Zhu Y, 2016, "Matrix profile ii: Exploiting a novel algorithm and gpus to break the one hundred million barrier for time series motifs and joins". *IEEE 16th International Conference on.*pp.739-748.
11. Huang C, Wang D, 2017, "Critical source selection in social sensing applications",pp.53-60.
12. Tuarob S, Tucker CS, "Automated Discovery of Lead Users and Latent Product Features by Mining Large Scale Social Media Networks" . *Journal of Mechanical Design.*2015;137 (7):71402-71402.
13. RW Ouyang, 2016,"Parallel and streaming truth discovery in large-scale quantitative crowd sourcing" ,*pp.*2984-2997.
14. Pendyala VS, S Figueira, 2015, "Towards a truthful world wide web from a humanitarian perspective" *Global Humanitarian Technology Conference(GHTC),pp.*137- 143.
15. SokolovaE, 2017, "Handling hybrid and missing data in constraint-based causal discover to study the etiology of ADHD",*pp.*105-119.
16. Ding W., 2017, "Node embedding via word embedding for network community discovery", *pp.*539-552.
17. ErNAS.Truthfulnessofcandidatesinsetoft-uplesexpansion.*International Con- ference on Database and Expert Systems Applications.*2017:314-323.
18. YangY.OntheDiscoveryofContinuousTruth:ASemi-supervisedApproachwith Partial Ground Truths. *International Conference on Web Information Systems En- gineering.*2018:424-438.
19. Ranjan R, "Streaming big data processing in datacenter clouds". *IEEE Cloud Com- puting.*2014,pp.78-83.
20. Yi X, 2015, " Privacy-preserving association rule mining in cloud computing", pp.439-450.
21. Mohammadi M, 2018, "Deep Learning for IoT Big Data and Streaming Analytics: A Survey". *IEEE Communications Surveys & Tutorials*.
22. Zakaria J, 2016, "Accelerating the discovery of unsupervised-shapelets". *Data mining and knowledge discovery,* 243-281.
23. Mohammadi M, Al-Fuqaha A, 2018, "Enabling Cognitive Smart Cities Using Big Data and Machine Learning: Approaches and Challenges",pp.94-101.
24. LiY., 2016, "Asurvey on truth discovery",pp.1-16.
25. Whitmore A, 2015, "The Internet of Things-A survey of topics and trends". *Information Systems Frontiers,pp.*261-274.
26. Ma F. Faitcrowd, 2015,"Fine grained truth discovery for crowd sourced data aggregation",*pp.*745-754.
27. Meng C, "Truth discovery on crowd sensing of correlated entities", 2015,pp.169- 182.
28. StiloG, Velardi P, 2016, "Efficient temporal mining of micro-blog texts and its application to event discovery". *Data Mining and*

Knowledge Discovery.pp.372-402.
29. PeelL, 2017, "The ground truth about meta data and community detection in networks. *Science advances, pp.1602548*-1602548.
30. Ouyang RW, "Truth discovery in crowd sourced detection of spatial events", 2016,pp.1047-1060.
31. TuarobS, RamN, 2014, "An ensemble heterogeneous classification methodology for discovering health-related knowledge in social media messages",pp.255-268.
32. Wang D. Theme-Relevant Truth Discovery on Twitter: An Estimation Theoretic Approach.in *Theme- Relevant Truth Discoveryon Twitter: An Estimation Theoretic Approach* (Icwsm . , ed.):408-4162016.
33. Shen H. Discovering social spammers from multiple views.*Neurocomputing.* 2017;225:49-57.
34. Shi LL, "Event detection and user interest discovering in social media data streams". *IEEE Access.* 2017;5:20953-20964.
35. SchulzA, "Semantic Abstraction for generalization of tweet classification: A new valuation of incident related tweets". *Semantic Web.* 2017;8:353-372.
36. Ji X, "Twitter sentiment classification for measuring public health concerns", *Social Network Analysis and Mining.*2015; 5:13-13.
37. Dong XL, "Data fusion: resolving conflicts from multiple sources". *Handbook of Data Quality.*2013:293-318.
38. ThiyagarajMPB,Aloysius A, "A Survey on Truth Discovery Methods for Big Data". *International Journal of Computational Intelligence Research.* 2017;13:1799- 1810.
39. GaoJ, "Truth discovery and crowd sourcing aggregation: A unified perspective. *Pro- ceedings of the VLDB Endowment.*2015; 8:2048-2049.
40. Berti-Equille L "Scaling up truth discovery". *IEEE 32nd International Conference on.*2016:1418 1419.
41. Schmidt WC, "World-Wide Web Survey Research: Benefits, Potential Problems, andSolutions".
42. JainS, "A survey paper on techniques and applications of web usage mining". *Emerg- ingTrends in Computing and Communication Technologies (ICETCCT), Interna- tional Conference on.*2017:1 6.
43. HuangC," Towards time-sensitive truth discovery in social sensing applications". *Mobile Ad Hoc and Sensor Systems (MASS).* 2015:154-162.
44. PopatK. "Where the truth lies: Explaining the credibility of emerging claims on the web and social media",2017,pp.1003 1012.
45. KraußJ. "Selection and Application of Machine Learning- Algorithms in Production Quality". *Machine Learning for Cyber Physical Systems.*2019:46-57.
46. ZhongguoY., "Choosing Classification Algorithms and Its Optimum Parameters based on Data Set Characteristics". *Journal of Computers.*2017;28:26-38.
47. AlqurashiS, BatarfiO, "A Comparison of Malware Detection Techniques Based on Hidden Markov Model". *Journal of Information Security.*2016;07(03):215-223.
48. Bonab MB. An Efficient Robust Hyper-Heuristic Algorithm to Clustering Prob- lem. *International Conference on Computational Intelligence in Information Sys- tem.*2018:48-60.
49. Borchani H, VarandoG, Bielza C, LarrañagaP. A survey on multi-output regres- sion. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery.* 2015;5(5):216-233.
50. Elsayed S, "Survey of uses of evolutionary computation algorithms and swarm in- telligence for network intrusion detection" ,2015;14, 1550025-1550025.