Truth Discovery in Big Data Social Media Sensing Applications

Article *in* International Journal of Innovative Technology and Exploring Engineering · June 2020

DOI: 10.35940/ijitee.H6311.069820

CITATIONS
7

READS
434

2 authors, including:

Sangeeta Gupta
Chaitanya Bharathi Institute of Technology
36 PUBLICATIONS **130** CITATIONS

SEE PROFILE

Truth Discovery in Big Data Social Media Sensing Applications

Omar Ahmed, Sangeeta Gupta, Mohammed Hasibuddin

Abstract**: The detection of truthful information amid data provided by online social media platforms (e.g., Twitter, Facebook, Instagram) is a critical task in the trend of big data. Truth Discovery is nothing but the extraction of true information or facts from unwanted and raw data, which has become a difficult task nowadays in today's day and age due to the rampant spread of rumors and false information. Before posting anything on the social media platform, people do not consider fact-checking and the source authenticity and frantically spread them by re-posting them which has made the detection of truthful claims more difficult than ever. So, this problem needs to be addressed soon since the impact of false information and misunderstanding can be very powerful and misleading. This mission, truth discovery, is targeted at establishing the authenticity of the sources and therefore the truthfulness of the statements that they create without knowing whether it is true or not. We propose a Big Data Truth Discovery Scheme (BDTD) to overcome the major problems. We have three major problems, the main one being "False information spread" where a large number of sources lead to false or fake statements, making it difficult to distinguish true statements, now this problem is solved by our scheme by studying the various behaviors of sources. On Twitter for example rumormongering is common. The second problem is "lack of claims" where most outlets contribute only a tiny small number of claims, giving very few pieces of evidence and making it not sufficient to analyze the trustworthiness of such sources, this problem is addressed by our scheme where it uses an algorithm that evaluates the claim's truthfulness and historic contributions of the source regarding the claim. Thirdly the scalability challenge, due to the clustered design of their existing truth discovery algorithms, many existing approaches don't apply to Big-scale social media sensing cases so this challenge is managed by our scheme by making use of frameworks HTCondor and Work Queue. This scheme computes both the reliability of the sources and, ultimately, the legitimacy of statements using a novel approach. A distributed structure is also developed for the implementation of the proposed scheme by making use of the Work Queue (platform) in the HTCondor method (maybe distributed). Findings of the test on a real-world dataset indicate that the BDTD system greatly outperforms the existing methods of Discovery of Truth both in terms of performance and efficiency.

***Keywords**: Big Data, Rumors, Scalable, Social Media, Truth Discovery, Twitter.*

Revised Manuscript Received on June 01, 2020.

Omar Ahmed, Computer Science Engineering Student, Chaitanya Bharathi Institute of Technology, Hyderabad, India.
Email: omarahmed18071998@gmail.com
Dr Sangeeta Gupta, Associate Professor in Computer Science Engineering Department, Chaitanya Bharathi Institute of Technology, Hyderabad, India.
Email: sangeetagupta_cse@cbit.ac.in
Mohammed Hasibuddin, Computer Science Engineering Student, Chaitanya Bharathi Institute of Technology, Hyderabad, India.
Email: ahmedhasib89@gmail.com

I. INTRODUCTION

THIS paper is presented to address the truth discovery problem in big data social media sensing applications (for example Twitter, Facebook, and Instagram) where many people make claims about their surroundings and make the claim public by posting it on these social media platforms and people believe it without checking the authenticity or the credibility of the claim. This model is inspired by the increasing proliferation of portable data collection devices (e.g., smartphones) and the vast opportunities for data sharing created by social media. Types of social media sensing include real-time situation emergency response; smart transportation system applications use social network apps based on the venue. A critical challenge in social media is the detection of truth where the motive is to differentiate credible sources and true statements from huge social media data that has unextracted and unfiltered data. So, we agreed to introduce this project because it is a problem that requires an efficient solution. It is commonly observed in real-world applications such as social media applications like Twitter, Instagram, Facebook, etc. to determine whether the claim made by a person has credibility or not i.e. if it is truthful or not. There are many claims on social media that have no credibility. In this project, there are three major challenges. The first one is "False information spread" here a small number of outlets lead to false claims, which makes it impossible to differentiate true statements. Rumors, scams, and manipulation bots, such as on Twitter, are common two examples of colluding media, spreading disinformation, and obscuring the facts. From real-life situations we can say that the widely spread false information appears much more believing than the true information, and thus makes truth discovery a difficult job. The test findings on a real-world case tell us that existing approaches for finding reality do not work well in recognizing facts when disinformation is widely spread. The second problem is "lack of claims" here mostly all the sources have very few claims, leading to a shortage of evidence to assess the trustworthiness of those sources, but many existing algorithms are highly dependent on accurate estimating of source reliability, which usually involves a relatively dense data collection. For example, due to spontaneous nature of social media sensing, sources lack the motivation and alternatively they can choose to disregard non-interested events and cases and only produce data in interesting topics or events Third, because of the centralized nature of their truth discovery algorithms, that is the current solutions did not have deep knowledge on the scalability dimension of the truth discovery problem and, many present solutions are not compatible to large scale social events.

Applications for social sensing also produce large quantities of data during critical events (e.g. accidents, sports, distress) where there is a high chance of fake news and rumors spreading. In this paper, we build up a Big Data Truth Discovery (BDTD) scheme to discourse the three challenges that are False information spread, lack of claims, and scalability challenges in big data social media applications. To tackle the false information spread challenge, this scheme categorizes various behaviors that sources show such as copying/ forwarding, self-correction, and spamming. To tackle data sparsity, which is the second challenge, the scheme employs an algorithm that estimates claim truthfulness by computing the source data and the historical contributions made to the claim by source. To overcome the scalability challenge, we develop a lightweight distributed framework developed by the combination of Work Queue and HTCondor [4], making the system scalable and effective. When our scheme is evaluated by comparing the state-of-the-art baselines on a real-world dataset collected from Twitter during the event Charlie Hebdo Attack in 2015. From the results, we see that BDTD wins and outperforms the baseline scheme by its effective features of detecting and extracting the true information from the vast mixed and unfiltered data with high efficiency.

II. RELATED WORK

2.1 Social Media Sensing

Social media sensing has become a trend these days, where people and especially the youth use the platforms to tell about their opinions, perceptions, and observations around the globe. When social media analyzing strategies and social media sensing is combined it leads us to various good applications such as social event summarization and Emergency responses.

This work aims on identifying the trustful information where the motive is to extract truthful information by computing the truthfulness of claims and reliability of social media users. The solution to this problem can benefit in extracting information from unfiltered data set.

2.2 Truth Discovery

Truth Discovery is the process of finding the actual true value or information or fact when many sources provide conflicting information on it. Truth discovery has come into eyes from few years. Yen et al was the first who defined Truth Discovery problem which led to proposal of Heuristic Algorithm Truth Finder, later many other solutions like AvgLog were developed. Dong et al considered that source dependency plays an important role in truth discovery problems [2]. There were many other machine learning schemes such as maximum likelihood, semi-supervised graph learning and many others. However, the current existing solutions of truth discovery are not able to compute in identifying truthful claims among noisy data which is both a challenging and critical task. In our work we come up with a new Big Data truth discovery system that is strong against false information spread and capable to find truthful claims even if most of the sources are providing conflicting and false information.

2.3 Data Sparsity

Data sparsity or lack of claims is one of the most important challenge in big data research areas. The problem of not having enough data for modelling has always been an issue. However, only few solutions have considered the problem of Data sparsity abundant. A scheme called Confidence-Aware Truth Discovery (CATD) was proposed where the reliability of sources is not reliable if sources contribute very less claims. This method derived a confidence interval to calculate the accuracy of source reliability. It was further extended to consider the confidence interval of truthfulness of claims. It was argued that when a claim has few sources contributing it, then the estimation score for the truthfulness of the claims become less meaningful. Then a new truth discovery scheme was proposed called Estimating Truth and Confidence Interval via Bootstrapping (ETCIBooT) which was capable for constructing claims, confidence intervals as well as identifying Truth [3]. Although both the works have taken lack of claims or data sparsity into consideration, but they evaluated the performance by excluding the detection of widespread false information. And moreover, the results have indicated that the solutions are not robust against false information.

2.4 Distributed System for Social Media Sensing

The scheme we proposed also resembles to few other distributed system executions used for social media sensing. For instance, a Parallel algorithm which used MapReduce framework was developed for handling streaming data. A cloud serving medium system was developed to combine social and sensor data to deal with massive and continuous data streams. Another cloud computing using Hadoop framework was introduced for large scale social network analysis. One of the defects of Hadoop is that it cannot compute with time critical applications which require faster responses time, since Hadoop is designed for only Big Data and thus it cannot manage quicker response for small data sets. In this paper, we develop a lightweight distributed framework which is the combination of HTCondor and Work Queue to improve the efficiency. And this developed framework is perfect for quick response-based systems because:

i) HTCondor is an efficient distributed computing system that lets thousands of tasks to be computed in a parallel manner, thus making the overall operating time reduced.

ii) There is no priority scheduling followed, it is flexible and adaptive enough to allow important tasks to be processed faster to meet the deadlines.

iii) The initializing time is less in HTCondor tasks when compared to Hadoop, making it ideal for handling data streaming.

III. REVIEW CRITERIA

The aim of the task is to compute the truthfulness of each claim and the reliability of each source, which can be elaborated as follows:

DEFINITION 1. Truthfulness for a claim: The possibility of a claim to be true.

DEFINITION 2. Reliability for a source: The score tells us how much the source can be trusted, the more the reliability the more chances of source having credible and trustworthy data.

As sources are often not thoroughly examined on social media platform and can report fake claims and not truthful sometimes, we need to examine the reality of sources. However, it is very difficult to evaluate the source reliability when data is not sufficient for experimenting. Fortunately, the reports contain evidence, hints and useful information to conclude the truth of a claim. For Example, in Twitter we have Geo Tags, Pictures, links, Video, Locations having in a Tweet to be considered as extra evidence to the report. To use such evidence in our scheme, we come up with credibility score for each report to tell us how much the report helps in gaining the truthfulness of a claim.

We first define the following terms related to the credibility:

DEFINITION 3. **Attitude Score:** This score has a range of (-1,1). When a source believes the claim is true and supports the claim then we provide a score of 1.

If the source does not believe the claim as to be true and instead believes it to be false and opposes the claim, then we provide a score of -1.

If the source does not provide any report then we simply give a score of 0.

DEFINITION 4. **Uncertainty Score:** A score in the range of (0,1) that measures the ambiguity or uncertainty of a report. It sees with how much confidence has been made on the claim. The higher the uncertainty the higher the score.

DEFINITION 5. **Independent Score:** This score has a range of (0,1) it measures the dependability of the score. If the report has been made independently without being copied from other sources, then a score of 1 is awarded. If the source is directly copied from other source without any edits or modification, then a score of 0 is awarded.

When combing the above scores, we come up with Credibility score. The presumptions which are made, depend on namely these three scores (Attitude, uncertainty and Independent). Now we can clearly distinguish the reports on a claim in the following behaviors:

i) a report that supports or opposes the claim.

ii) a report made with high or low uncertainty on the claim.

iii) an original, independent, copied, edited, modified, or forwarded report on the claim.

All the factors are considered important in detecting truthful claims from vast false information. Our system also takes into consideration the source's historical reports on the same claim. For example, on Twitter a spammer keeps posting the same exact tweet over and over which are unedited and unmodified, which in most cases contain unrelated, misleading, or inappropriate claims. Whereas a reliable source such as Government department or news outlet may proactively correct its previous reports that contain false information. Therefore, a time-series matrix has been defined which is used for computing historical contributions of a source on its claims.

IV. SOLUTION

Here, we define the Big Data Truth Discovery (BDTD) scheme to solve the truth discovery problem in big data social media platforms formulated in the previous section. A few observations have been observed relevant to our model. Then the design is discussed and BDTD scheme is presented.

4.1 Observations

The following observations has been noticed:

Observation 1: False information is usually spread by sources by simply forwarding, copying from other sources without modification.

Observation 2: False claims are often controversial and sources tend to disagree with each other and have intensive debates on those claims.

Observation 3: If a previous claim is exposed by a new claim made by the source then they are more chances that the old claim was false.

4.2 Algorithm Design

Before getting into the details of the proposed BDTD scheme, we review the current scenario of the truth discovery solutions in social media sensing. The current truth discovery solutions can be mainly classified into two categories:

(i) principled solutions: In this solution specific optimization techniques are used to find the intersecting points at the global best of functions (e.g., MLE).

ii) data-driven solutions: In this solution Machine learning techniques are used to compute some practical data driven challenges for instance, insufficient data, these problems were not able to compute by Principled solution since they work great on relatively dense datasets but fail in sparse data scenarios this is because Principled solutions results mainly depend on accuracy of large set of parameters which are density sensitive. However, data driven solutions are heuristic in nature and explore the content of data to manage the insufficient data problem. Our BDTD scheme comes under data-driven solutions type. It follows the mechanism of our previous work where the semantics of tweets are found to be useful and important in evaluating the truthfulness of claim when source reliability is hard to compute since the data is insufficient. When we compared our scheme with a few principled driven schemes and found that BDTD beats those previous models when the data is sparse or insufficient.

V. IMPLEMENTATION

In this section, we discuss the lightweight distributed model of BDTD scheme which is the combination of HTCondor and Work Queue will be presented first. Then, BDTD implementation will be presented which is focused on memory allocation and distributed truth discovery task control.

5.1 HTCondor and Work Queue

5.1.1 HTCondor

We have used HTCondor as the distributed system for implementation of our BDTD scheme. HTCondor consists of more than 1850 machines and over 14700 crores at writing time. HTCondor is used by many organizations including government and industries. HTCondor was used in public desktops and system allocates the tasks on ideal machines connected to the system when it is running.

5.1.2 Work Queue

It is a lightweight framework for computing distributed systems on larger scale. This framework prioritizes the master process to initialize a set of tasks and submit the tasks to the queue and wait for completion.

Work Queue has a flexible worker pool that scales on required number of workers by application. A worker is a process which operates functions when ordered by tasks. Once they are running, each worker calls back to the master process and executes the tasks, we use Work Queue with HTCondor because of its Master process dynamic allocation.

VI. EXPERIMENT ON REAL WORLD DATA

6.1 Baseline Methods

Nine truth discovery solutions were chosen to be considered as the baselines in our model evaluation: AvgLog, Invest, TruthFinder, 2-Estimates, CATD, ETBoot, EM-SD, EM-Conflict, and RTD.

AvgLog: It is a basic truth discovery scheme where it estimates the reliability of source and true claim by making use of basic average-log function.

Invest: This scheme obtains the targets by using a nonlinear function.

TruthFinder: It uses the interdependence between source trustworthiness and fact confidence to get its solution result.

2-Estimates: It estimates the two defined parameters in their models related to source reliability and claim truthfulness.

CATD, ETBoot: These schemes provide interval estimators for source reliability in a insufficient dataset where most sources make one or two claims.

EM-SD, EM-Conflict: These two schemes both use a maximum likelihood estimation approach of Machine Learning for obtaining results. EM-SD openly considers the retweeting behavior of Twitter users and uses that to build a source dependent model.

RTD: RTD is a robust truth discovery solution. It leverages the semantic scores of tweets to identify the false information spread in social media platform.

6.2 Data Collection

In this paper, we assess our proposed scheme on a real-world data find collected from Twitter in the aftermath of recent emergency and disaster events. The trace is the Paris Charlie Hebdo Attack which took place on Jan. 1, 2015. Offices of a French news magazine were attacked by several gunmen, they killed 12 people where Employees and two police officers were also included.

A data crawler was developed which was based on Twitter search API, the job of data crawler was to collect the data traces by presenting query terms and locations of the events. The statistics of the data trace is summarized in Table 1.

Data Trace	Charlie Hebdo Attack
Start Date	January 1 2015
Time Duration	3 days
Location	Paris, France
Search Keywords	Paris, Shooting, Charlie Hebdo
# of Tweets (Original Set)	253,536
# of Tweets (Evaluation Set)	60,559
# of Tweets per User	1.09
% of Tweets Related to Misinformation	22.86%

Table 1: Data Trace Statistics

It was observed that the data set was very insufficient for evaluating. In the Charlie Hebdo dataset, 90.8% of sources were found to have a single claim and only 2.3% of sources provided more than two claims.

Our evaluation results have indicated that most of the tweets are related to false information. We have found the proportions of false information and it is 22.16%. And out of that 79.15% are simply retweets.

6.3 Data Preprocessing

Data preprocessing steps were conducted to prepare the datasets for the experiment:

(i) like tweets will be clustered into same cluster for generating claims.

(ii) semantic link scores will be produced.

(iii) TSC Matrix will be generated.

(iv) ground truth labels will be generated.

The details of these steps are summarized below.

Clustering: Firstly, similar tweets are grouped into same cluster using K-means algorithm which is specialized in large datasets and can handle streaming Twitter data. And Jaccard distance is used to calculate the similarity between tweets, which is the distance between tweets in the cluster. Jaccard distance is used for calculating dissimilarity and multi-dimensional scaling. We chose a statement as a claim for each generated cluster and Twitter user are considered as source.

Computing Credibility Score: To find this score we begin by calculating Attitude score of a source by using Sentiment analysis and keyword matching. We have also performed Polarity analysis to detect tweets that express strong negative sentiment as “disagree” using NLTK 3. Words like “fake”, “false”, “debunked” etc. A score of 1 is given if the tweet is agreeing and a score of -1 is awarded if the tweet is not agreeing. Then uncertainty score is calculated using text classifier by using skit-learn with data. Now to calculate the independent score, a tweet is labeled “dependent” if the below conditions match:

i) If it is a retweet

ii) If it is mostly like other tweets (J distance < 0.1)

To check the effectiveness of these methods, 200 label tweets were picked randomly, and their derived scores were checked manually. So, the accuracy of Attitude score was 82.1%, for Uncertainty score it was 78.6% and for Independent score it was 90.3%.

Generating the Time-series Source-Claim Matrix: The TSC matrix is generated in this manner: for source, all the reports were recorded. The tweets were sorted in sequential order and credibility score was computed for each tweet.

Labeling Ground Truth: Firstly, it was checked that whether the claim was Truthful or unconfirmed based on the following:

Truthful claims: These are the claims that are the statements Of physical events related to the selected topic and noticed by multiple independent observers.

Unconfirmed Claims: These are the claims that do not meet the above claim’s criteria. This includes tweets that represent feelings, quotes, shout-outs etc.

After the label processing is done, the unconfirmed claims are removed since they are not verified. Now Manually truthfulness of Factual claims based on historic facts are verified.

6.4 Evaluation Metrics

When the labeling is done, we see that our datasets are not balanced: true claims were more than false claims. For example, only 14% in Charlie Hebdo Attack data trace are false. But it does not indicate that our problem is basic since the false claims also share many similar features as that of truth claims. If a user misidentifies a false claim as true then it can lead to High negative impact especially in critical situations like disasters, accidents, virus infections etc.

To manage the data imbalance, Specificity (SPC), Mathews correlation Coefficient (MCC), Cohen's Kappa (Kappa) have been chosen.

6.5 Evaluation Results - Truth Discovery Effectiveness

When we evaluate the effectiveness of truth discovery, the results of Charlie Hebdo Attack in figure 1 from the experiments show that BDTD outperforms all baselines. Also, the gain achieved by BDTD compared to the best performing baseline (RTD) on SPC is 9:3%, on MCC it is 2:1% and on Kappa it is 8:1%.

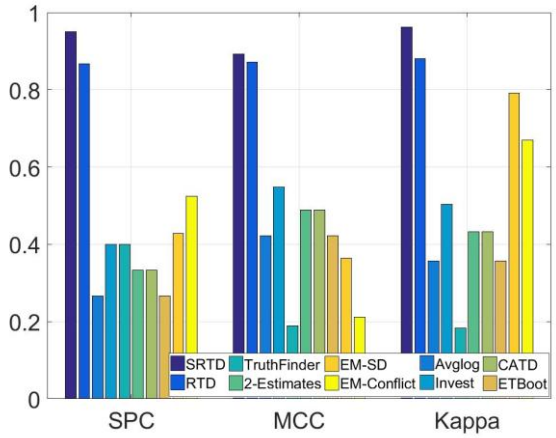

Figure 1: Truth Discovery Effectiveness.

6.6 Evaluation Results - Scalability and Efficiency

Next, we estimate the efficiency of the BDTD scheme.

BDTD is run on HTCondor cluster with range of workers (1,9). All other baselines are also run with same compatibility conditions. This is to check the performance when there are less resources. The execution of all compared schemes is calculated. Since HTCondor has crores of workers but still to match the compatibility we have used only 10 workers. The results show that BDTD scheme outperforms all the other baseline schemes by finishing the execution in a shorter time as shown in table 2.

Method	Charlie Hebdo
SRTD	**4.169**
RTD	11.023
CATD	34.846
TruthFinder	40.213
2-Estimates	152.707
Invest	83.827
AverageLog	17.922
ETBOOT	81.170
EM-SD	52.793
EM-Conflict	56.214

Table 2: Execution Time

The scalability of the BDTD scheme is further evaluated by extending each of the data traces by adding some logical tweets. Specifically, Whole data trace with tweets more than our evaluation set (unconfirmed, irrelevant tweets) have been used. The size of data trace is listed in Table 1 previously. The results are shown in Figure 2. It has been observed that BDTD scheme is the fastest of all compared schemes as data size increases. It is also noticed that performance gain of BDTD gets more effective as the data size increases. This evaluation indicates the scalability of our scheme when large data is taken into consideration.

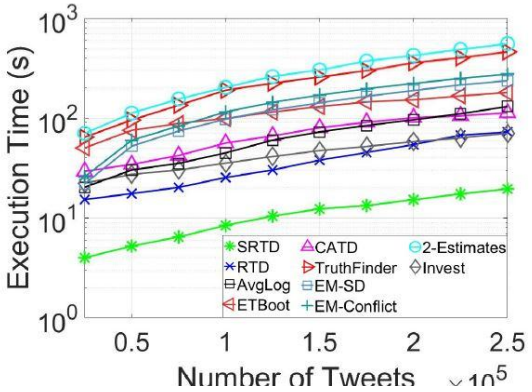

Figure 2: Charlie Hebdo Attack Data Trace

VII. DISCUSSION AND FUTURE WORK

This section is for discussion of limitations that have been detected as well as the future work to overcome this limitation. Firstly, BDTD scheme trusts on heuristically defined scoring functions to solve. And BDTD uses semantic information of reports to manage the data insufficiency problem. And we plan to explore the statistical model of upcoming schemes that can manage data insufficiency problem.

Secondly, our scheme does not take Unconfirmed claims into consideration that do not have ground truth. However, these claims appear very much, so we plan to overcome this limitation by detecting and shortlisting these claims by making use of current sentimental analysis [1]. Our BDTD scheme can also be improved by allowing our types to include these claims also.

Thirdly, BDTD is not aware of the dynamic truth problem, where the truth changes time to time. For instance, escape path of a thief, virus infection rate etc. To consider dynamic truth challenge there are two important jobs, first is to catch the conversion of truth sharply. The second job is to be strong to manage unfiltered data that may lead to false identification of the conversion. This task can be very tough since social media is filled with rumors. So, this can be overcome by combining Hidden Markov Model (HMM) with our BDTD scheme.

Fourth, a very popular limitation is that a false claim can spread from one domain to another without changing any information.

Fifth, our scheme is made to assume independence between claims, which might not be correct all the time, since tweets like weather forecast, where the weather of one place maybe supported by depending on other place's weather. So here we can overcome this by having deep knowledge on the domains, and by using location services such as Google Maps.

Finally, one of the limitations is collusion attack by hackers, now we have to improve the robustness of the scheme. For instance, a group of users can deliberately spread false information to mislead the crowd. And unfortunately, this problem is not considered since it is hard to overcome. Honey Pots can be used if the cyber security is compromised.

VIII. CONCLUSION

It is evident that the spread of fake news and misinformation creates havoc among the public, this is because of people reposting and retweeting posts and their opinions without verifying the sources and facts. So, to overcome this rampant spread of false information we have proposed a Scheme whose function is to not only detect the truth but also deal with scalable data. Hence our scheme solves three major problems which were not answered by previous works. The problems were False Information Spread, where claims were made without credibility. Secondly, Sources with lacking evidence to back their claims were considered authentic, and thirdly the scalability problem. These all problems are addressed by our proposed BDTD Scheme in this paper. The scheme was evaluated by using a Real-world Dataset and other baselines approaches were also compared. The results showed us that our scheme outperforms all other schemes with a high margin in terms of execution time, scalability, truth discovery, and effectiveness. In conclusion, the BDTD scheme outperforms other approaches and can be used as an application. Improvements in the scheme can be done, where it caters to other social media platforms other than twitter.

REFERENCES

[1] S. Bhuta and U. Doshi. A review of techniques for sentiment analysis of twitter data. In Proc. Int Issues and Challenges in Intelligent Computing Techniques (ICICT) Conf, pages 583–591, Feb. 2014.
[2] X. L. Dong, L. Berti-Equille, and D. Srivastava. Integrating conflicting data: the role of source dependence. In Proceedings of the VLDB Endowment, pages 550–561, 2009.
[3] X. X. et al. Towards confidence in the truth: A bootstrapping based truth discovery approach. In Proceedings of the 22th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD ’16, 2016.
[4] D. Thain and C. Moretti. Abstractions for cloud computing with condor. 2010.

AUTHORS PROFILE

Omar Ahmed is a Final year student pursuing a Bachelor of Engineering program in Computer Science Engineering at Chaitanya Bharathi Institute of Technology, Hyderabad, India. He has completed his primary and high school education from Delhi Public School, Jeddah, and owing to his interest in Computer Science, enrolled in the BE program in CSE at CBIT. He has a passion for the field of Data Science and Machine Learning. He has a knack for research and is always looking to push the envelope in learning new things and technologies.

Dr Sangeeta Gupta is an Associate Professor in Computer Science and Engineering Department at Chaitanya Bharathi Institute of Technology, Hyderabad, India. She completed her Ph. D in Cloud Computing with NoSQL Databases from JNTU Kakinada in 2017. She is Passionate about doing research and exploring emerging areas.

Mohammed Hasibuddin is a Final year student pursuing Bachelor of Engineering in Computer Science Engineering at Chaitanya Bharathi Institute of Technology, Hyderabad, India. He has completed his primary and high school education from International Indian School, Dammam. His research interest includes in the field of Cloud Computing and IoT Technology. He has keen interest in Big Data and IoT technology.