Truth Detection Algorithm in Social Media Tweets Using Similarity Measures

*Vishesh Gupta[1], G.Vadivu [1]
[1]SRM Institute of Science and Technology
[1*]vs4084@srmist.edu.in
[2]SRM Institute of Science and Technology
[2]vadivug@srmist.edu.in

Abstract. In the social media environment, many tweets are posted by users, it is amongst the way of living with social media culture in this era. Internet and Online Media are an important part of communication these days. Sometimes fake news can create a lot of issues that may not be expected by the users. Therefore, it is regulated by cyber-crime authorities to regulate the news and general rules are created by social media websites. This makes it necessary for a proper truth detection algorithm. In this paper, a java implementation of a modified Jaccard Algorithm is applied which then classifies news from the dataset as true or false. The dataset is initially classified on sentiments which are done through the Stanford CoreNLP library in Java. The results show good accuracy for truth detection. This is used in big data truth discovery algorithm to prevent misinformation spread.

Keywords: Social media, Truth detection, Tweet sentiment, Parallel credibility, NLP, Data analytics, Jaccard similarity.

1 Introduction

Different kinds of online media detecting incorporate ongoing circumstance crisis reaction; advanced transportation framework applications utilize informal community applications dependent on the setting. A basic test in web-based media is the discovery of truth where the rationale is to separate dependable sources and genuine proclamations from colossal web-based media information that has unextracted and unfiltered information [20]. To tackle reality revelation issue, a rich arrangement of principled methodologies has been proposed in AI, information mining, and organization detecting networks [1].

Online media detecting has arisen as another huge information application worldview to gather perceptions and cases about the deliberate factors in actual climate from regular residents. [2] The ability to judge the evolving truth of cases and the unwavering accuracy of information sources without understanding all of them deduced is a major problem in online internet detection applications, Successful truth discovery is a term used to describe this process [3].

In the digital world, truth discovery is very important. While we need more accurate data than ever before, inconsistency is inevitable due to big data's "variety" feature. The advancement of truth discovery has a wide range of implications in a variety of fields where important decisions must be taken based on accurate data obtained from different sources [26]. Data extraction, medical services, crowd/social sensing, information diagram development, and so on are some examples. This and other examples show how truth analysis has a wider effect on multi-source data integration. Although considerable advancement have been conducted in solving the truth discovery challenge, several significant problems would be remained. Firstly, current truth discovery solutions failed to [4], [5] address the complex truth discovery challenge, in which the ground truth of statements varies with time. Secondly, since their truth discovery algorithms are clustered, many existing implementations are not scalable to large-scale social sensing cases. Thirdly, the heterogeneity and unconventionality of the social detecting information traffic represent extra difficulties to the resource distribution [12] and framework responsiveness.

Truth disclosure strategies have been proposed to together gauge occasion realities and specialist reliabilities by collecting uproarious data from numerous specialists [27]. For instance, organizations may accumulate item appraisals from online media to gauge the prevalence of their items and administrative offices may utilize participatory social detecting to decide whether certain occasions like gridlock have happened by permitting the general population to report such occasions to them [4].

Truth disclosure has gotten a lot of consideration as of late, and past examinations have created different models to address this significant test Truth revelation issue has been concentrated to determine the contention among sources. [5] The fundamental thought is by fusing the source quality, data from great sources is more dependable, and ought to gauges more in truth assessment. [6] The problem formulation of the truth discovery algorithm is investigated using a modified Jaccard Algorithm technique in this paper.

In terms of detecting fake news from tweets and retweets, the proposed system is more efficient. The credibility score, which includes Independent, Attitude, and Uncertainty scores, was implemented to solve the issue of false news from tweets [21]. These scores encompass specific levels of tweets that assist in determining the reliability of scores using the SRTD method. The model uses an approach of similar tweet analysis method to solve the issue of fake news from retweets, which includes special factors such as wordnet, dictionary, and stopword list, these factors assist in the identification of tweets of the same meaning, allowing us to determine how many retweets are identical, as well as how many retweeted tweets use the same meaning words to spread misinformation [25]. The reliability is again predicted on this new parameter of similar tweet analysis using SRTD which provides help in determining more efficiency in detecting the trueness from the retweets. In this part, the Work Queue framework is utilized to provide an appropriate deployment of Jaccard Similarity in the scalable and robust truth discovery (SRTD) scheme. It helps in analyzing the credibility score for tweets in a shorter period of time.

2 Literature Review

The inconsistency or battle of these fluctuated depictions makes inordinate misperception for us to perceive [1] genuine information from one another. Thusly, perceiving the right and wide-going information from contradictory records is a significant issue for information fuse. Along these lines, reality disclosure issues ought to be recognized. The issue to find out the reality from problematic information is very much characterized as Truth Discovery. The standard of truth disclosure is to assess the nature of the source. [5], [6] therefore, the figuring gadget of data source will hugely upset the result and advancement of truth revelation. Then again, the innovative cycles don't consider what source quality is meant for when an invalid is given by the source. In a few cases, the information worried to the comparable item can be formed from various sources. Then again, this information that are multi-source are not portrayed reliably. [7] In the brilliant of this experience, truth disclosure is seemed to perceive truth for each article from multi-source information. Furthermost winning truth disclosure approaches receive that ground facts are completely unidentified, and they accentuate on the assessment of unconfirmed systems to helpfully assess object certainties and source textures [8], [11]. In the work of S Bhuta [23], the users of social networking internet sites have increased dramatically, people use such websites to communicate their thoughts and ideas on a wide range of issues. Sentiment analysis of the data including people's opinions is critical for estimating public opinion on a specific issue or phenomenon. This work examines a variety of methodologies for text sentiment analysis, including both lexicon-based and learning-based approaches [22]. A number of concerns and obstacles must be overcome in order to accommodate these approaches for sentiment analysis of data obtained from one of the social media platforms [13], Twitter, as outlined in this study. The proposed framework, as of now, misuses Twitter for information revelation and other online media, the strategies are space explicit. Various motivations for sensing mood behaviors have been explored in the paper of D Wang [4], [14], [17], this research is driven by the need to offer users trustworthy information recommendations in the field of social sensing. Individuals are used as sensors in social sensing, which allows them to observe and report occurrences in the physical environment. This research proposes a novel principled technique for solving a mood-sensitive truth discovery issue in social sensing applications for reliable recommendation systems. In truth-finding solutions, the new technique considers the mood-sensitive elements of both sources and claims. Using expectation-maximization techniques, the proposed method evaluates the mood sensitivity and dependability of sources, as well as the mood proportionality and accuracy of assertions. Unfortunately, it presents remedies which may be entirely skewed in terms of the mood of individual sources, resulting in useless or even false advice. Different algorithms based on the number of tweets and other factors have also been used to make predictions. In terms of using truth detection algorithms for tweets and retweets prediction, our study varies from previous studies. Alternately, in a few genuine introductions, a bunch of ground real factors may be decently open. [9], [10] By literature review, few papers were studied and it is seen that mainly the truth discovery algorithms are divided into categories as given in the figure. 2.

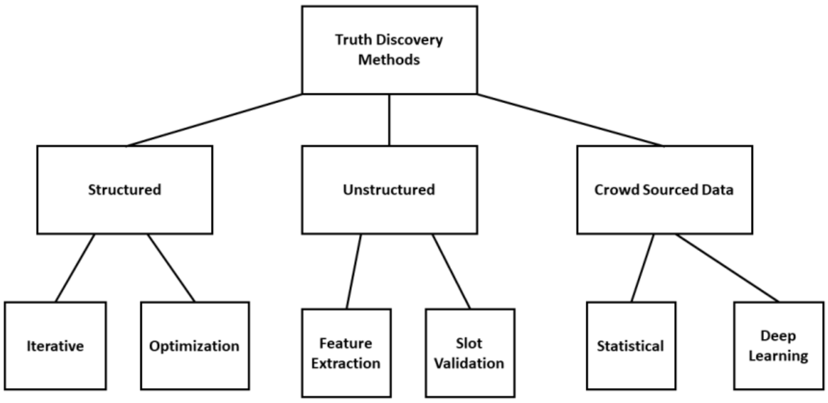

Fig. 1. Types of Truth Discovery Data

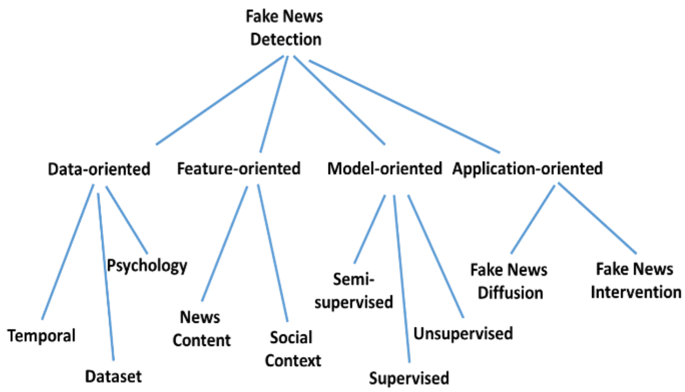

Fig. 2. Truth Discovery Methods [9]

2 Truth Discovery

In the implementation structure, firstly the dataset is loaded which taken from https://apollo.cse.nd.edu/datasets.html. The information is related to the July 7, 2016, Dallas Shooting. An intensely equipped armed force veteran trapped a gathering of cops in Dallas, in which five officials were killed, and nine others were injured. The firing was widely regarded as the worst incident affecting US law enforcement since the 9/11 terrorist attacks. However, the sources in the information are regularly unvetted and may not generally report truthful claims. Our key challenge is to make current truth discovery solutions in applications that detect social media more efficient in addressing the three essential factors. The first is the spread of misinformation, which involves a large number of sources disseminating fake information on social media. The second issue is data sparsity or a lack of information from a massive dataset. The third point is that existing truth discovery approaches do not sufficiently address the problem's robustness. This dataset is first loaded through java coding, then Stanford CoreNLP library is applied to analyzed the no. of retweets counts and for the sentiment analysis. To address the first two causes more efficiently we have performed certain approaches on this i.e., calculating Retweet/uncertainty score, independent score, Sentiment/Attitude score. To determine the Attitude Score, we use Sentiment analysis with Stanford CoreNLP. Text data analysis [18] is simple and effective by using Stanford's CoreNLP. CoreNLP is capable of extracting a wide range of text properties, such as part-of-speech tagging or named-entity recognition [19]. CoreNLP is written in Java, which means you'll need to have Java installed on your computer, which I'll be using in this demonstration. Sentiment values are measured on a scale of zero to four. 0 and 1 means that the sentence is very negative while 3 and 4 mean it's extremely positive. For the neutral value, it is equal to 2. These attitudes scores are represented by 1, -1, and 0 as stated in Table 2. The Sentiment and retweet count is an important parameter for consideration with Jaccard Algorithm [16], [24]. In Table 1, the output of some of the tweets is shown. The Jaccard algorithm is used to calculate the other two score which is independent score and the No. of retweets. After uploading the dataset, the Jaccard index is calculated for each tweet and comparing every tweet with the Jaccard similarity of source and target. To determine the Uncertainty Score, we can compare two tweets to see whether they are identical. If they are, then the tweets appear to be true and we will give a high score of 1; if they are not, we will allocate a low score of -1. The Jaccard distance will be used to determine similarity, and all identical tweets will be placed in the same entity. To compute the Independent Score, if a tweet is merely copied and retweeted, it will be considered a dependent tweet, implying that the user has not added something to it. It will be considered untruthful and will be given a score of -1, while all independent tweets will be given a score of 1 because they were created by the user by inserting text and claiming it. Now, Jaccard Algorithm is applied to analyze the Jaccard Similarity index. On the basis of this, a score matrix is made which is shown in Table 1.

Table 1. Dataset Analysis of Dallas Shooting Tweets

Tweet Text	No. of Retweets	Sentiment Analysis	Jaccard Algorithm
Top story: Four officers killed in Dallas protests against police shootings -… https://t.co/zMe2F0aZ6Q,	0	Negative	0.168269231

see more https://t.co/C7abbVO0ib			
RT @Hope_Fl0ats: Shooting someone for being a cop, is no better than a cop shooting someone for their skin color. Both are equally disgusti…	198	Negative	0.085585586
RT @PrisonPlanet: Time for the left to dial back its racist anti-white rhetoric in the aftermath of #Dallas.	186	Negative	0.121359223
RT @MattSmethurst: Let the record show. #Dallas https://t.co/5r2oAH4MD8	9951	Neutral	0.149068323
RT @ChrisMurphy67: Its hard to feel good about anything these days. Mostly b/c many will not acknowledge legit concerns of others. Until it…	1	Positive	0.103305785
RT @dcexaminer: "The suspect stated he wants to kill white people." https://t.co/BvzeUJczBa https://t.co/4UUdlaA5sS	48	Neutral	0.143564356
RT @cristinalaila1: Police officers are ready to lay down their lives to protect us‼ SHAME on u leftist TERRORIST COP KILLERS! #Dallas htt…	272	Negative	0.127659574
RT @Libreriamo: "Una delle armi più potenti è il dialogo..." (Proverbio africano) #Dallas https://t.co/dBAhZ4PywE	37	Negative	0.135
Investigador @SergioAndCab, ahora en @UTSA Texas, entrevistado en	0	Negative	0.079295154

@ExpositoCOPE a partir de 16:04h sobre tiroteo en Dallas.			
RT @KarenAttiah: First question for #PhilandoCastile 's mother from CNN was about her reaction to Dallas. "My son died 48 hours ago"	945	Negative	1
Restaurant Roundup: Chili's launches new marketing campaign; adds burgers https://t.co/B3MYEO56La #IHLive	0	Negative	0.103773585

In table 2, the computation of each tweet parameter is reevaluated by -1 and 1 by the use of parallel processing. Parallel processing is known to improve the time execution of large datasets [15]. Hence, the queue is modified for the parallel processing of several tweets. In table 2, different scores are evaluated as shown.

Table 2. Computing by Parallel Processing

Sentiment Value	Jaccard Algorithm	No. of Counts
-1	1	1
-1	-1	1
-1	1	1
-1	1	-1
-1	-1	1
-1	1	1
1	1	1
1	1	-1
1	1	1
-1	1	1

The word "parallel processing" is used to denote a broad class of simultaneous data-processing operations for the purpose of improving the computing speed of a computer system [28]. In comparison, a parallel processing system can process data in parallel, resulting in quicker execution times than normal processing of credibility scores. Combining the scores mentioned above in table 2, the truth analysis is done, by evaluating the weightage of each score and then adding it. Hence, the final evaluation result is shown in Table 3.

Table 3. Truth Analysis

Tweet No.	Weightage of Reliability	Result
0	0.333333333	FALSE
1	0.333333333	FALSE
2	0.333333333	FALSE
3	0.333333333	FALSE
4	0.333333333	FALSE
5	0.333333333	FALSE
6	1	TRUE
7	0.333333333	FALSE
8	1	TRUE
9	0.333333333	FALSE
10	0.333333333	FALSE
11	1	TRUE

The Scalable and Robust Truth Discovery (SRTD) algorithm is an optimization technique that directly incorporates the source' contribution scores when calculating source reliability and claim truthfulness. We start the model with a set of sources which provides us uniform reliability scores of tweets and claim truthfulness. In every iteration, the updating of each source's reliability score first takes place using the truthfulness scores of claims provided by the given source and also the source's contribution score.

The approach is to apply the SRTD Algorithm on the combination of independent score, sentiment score, and uncertainty score and the evaluation of this results the reliability score for each tweet. Certain parameters have been set for the reliability score; if the reliability score for a given tweet is greater than 0.75, the claim's reliability is True, and if the score is less than 0.75, the claim's reliability is false. Hence, we calculate the source reliability and statement truthfulness scores independently before combining them into a matrix. The SRTD algorithm produces a simple result, as seen in Table 3.

3 Results

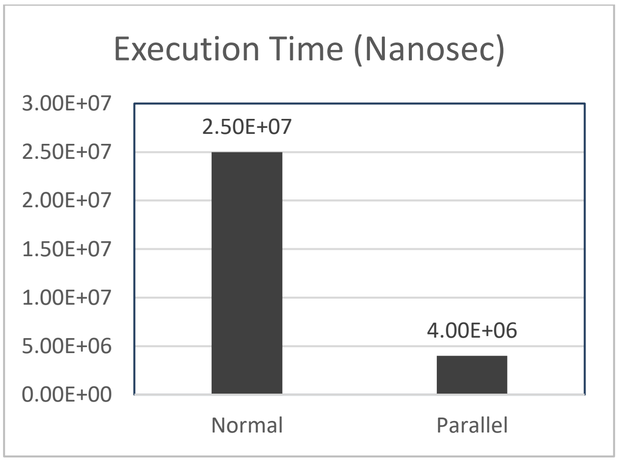

Fig. 3. Execution Time of Credibility Scores

In this segment, the outcomes are introduced. We measure the execution times of all compared schemes to determine performance. We use SRTD on computing credibility score and in parallel computing credibility score as shown in Table 3. The outcome shows that our parallel computing credibility score beats any remaining processing by completing the truth discovery task in a shorter period. We also note that as the size of the data grows larger, the efficiency gather obtained by SRTD becomes more important, demonstrating the scalability of our scheme in social sensing applications on large data traces. The execution time is evaluated. The result for execution time is shown in the figure. 3.

Table 4. Similar Tweet Analysis

Tweet	Word	Retweet Modification using Dictionary
rt prison-planet time for the left to dial back its racist antiwhite rhetoric in the aftermath of dallas	Dallas	[Dallas, urban-centre, municipality, urban area, City, geographical region, physical object, metropolis, district, territory, territorial-dominion, ...]
top story four officers killed in dallas protests against police shootings	Top	[top, region, part, location, vertex, physical object, physical entity, entity.]
rt hopeflats shooting someone for being a cop is no better than a cop shooting someone for their skin color both are equally disgusting	Cop	[patrol-man, officer of the law, cop, trooper, constable, deputy, policeman, sheriff, police officer, lawman, law_officer, peace_officer, defender, guardian, physical _entity, entity, causal_agent, cause, causal_agency.]

Many tweets are really similar to one another, which may be attributed to people tweeting on the same subject, which creates an amount of fuss and misinformation throughout the data. As a result, in order to reduce the misinformation, it would be useful to find out the same meaning words in the tweets, and on the basis of Jaccard similarity, we can identify the number of same meaning tweets. In order to capture the attributes of the retweet text, a similar tweet analysis is used. Some keywords, such as dallas, problem, shooting, violence, and others, appear more frequently in retweeted content than in regular tweets, thus spreading misinformation. This identification can be done using the Wordnet library, Dictionary, Stopword list, and this approach is known as similar tweet analysis. Three distinctive factors are discussed in this methodology to get higher performance using wordnet library and dictionary namely keyword, tweet text, and words with the same meaning. A larger dictionary will usually improve detection accuracy, but it will also raise the risk of overfitting. As a result, each word in the tweet is compared to words that have the same meaning.

However, after estimating the claim's accuracy, multiple truth discovery approaches sometimes struggle to have the right accuracy, resulting in the dissemination of misinformation. Because of the usage of common words/synonyms, the number of fake tweets is on the rise. In this strategy, fake tweets are more likely to be perceived as truth. In order to gain the effectiveness of the approach, the similarity of the tweet has been used to check the same meaning words using the wordnet dictionary. The similarity is always calculated using the similarity index which is thus circulated by the Jaccard algorithm itself. If the similarity of the particular tweet is seen, the similarity index will return 1, otherwise, it will return -1. In other words, same meaning word/dictionary concept has been used in this problem to redefine all the retweets. After the evaluation of similarity of tweets, again Scalable and robust truth discovery approach has been used on the fourth parameter (similarity index of tweet) to obtain the highest accuracy reliability score as shown in Table 4.

Table 4. Similar Index Truth Analysis

Tweet ID	Reliability Score	Reliability
0	0.42	FALSE
1	0.25	FALSE
2	0.35	FALSE
3	0.19	FALSE
4	0.24	FALSE
5	0.25	FALSE
6	0.74	TRUE
7	0.25	FALSE
8	0.82	TRUE
9	0.25	FALSE
10	0.25	FALSE
11	0.71	TRUE

The proposed model calculates the reliability of tweets using the three parameters i.e., independent score, attitude score, uncertainty score. To get more efficiency on the retweets we have taken the new parameter (Retweet modification using the word-net and dictionary) on the similarity index to get a better reliability score.

The four parameters are determined to determine the tweets' reliability, which determines whether or not the claims are true, so by executing the SRTD algorithm we got the reliability score for particular tweets, and the score which is having the reliability less than 0.7 are considered in False category similarly the reliability which greater than 0.7 are considered in the truth category. Then, a view matrix has been formed as shown in Table 4 which contains all the reliability score for the tweets and a truth-count chart has been performed on the basis of the view matrix has shown in figure 4.

In figure. 4, the truth count detected in the execution result for the dataset taken is presented.

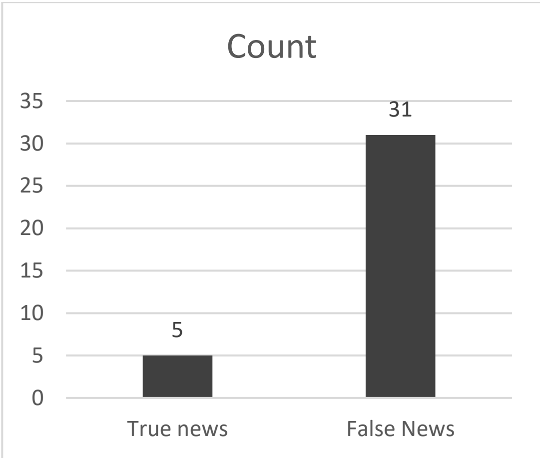

Fig. 4. Truth Count of Tweets

3.1 Evaluation of Result

We introduce a distributed implementation of the Jaccard Similarity in the SRTD scheme using the Work Queue framework in this section. The HTCondor scheme and Work Queue framework are first introduced. We have combined the SRTD model with an efficient workload distribution method to dynamically assign resources (such as Retweet, Sentiment score, and Independent Score) to truth discovering approaches depending on their computing requirements. The execution of the SRTD scheme is then presented, with an emphasis on distributed truth discovery activity management, and allocation.

The final tweet id graph is generated using the similarity tweet table as shown in Table 4. It shows how many tweets have used the same or similar meaning terms or synonyms. By using the Dictionary and similar text [Wordnet Library] we calculated the total number of similar words used on the tweets in the Dallas dataset and thus getting the highest 195 counts on the tweet number 5 as shown in the figure. 5.

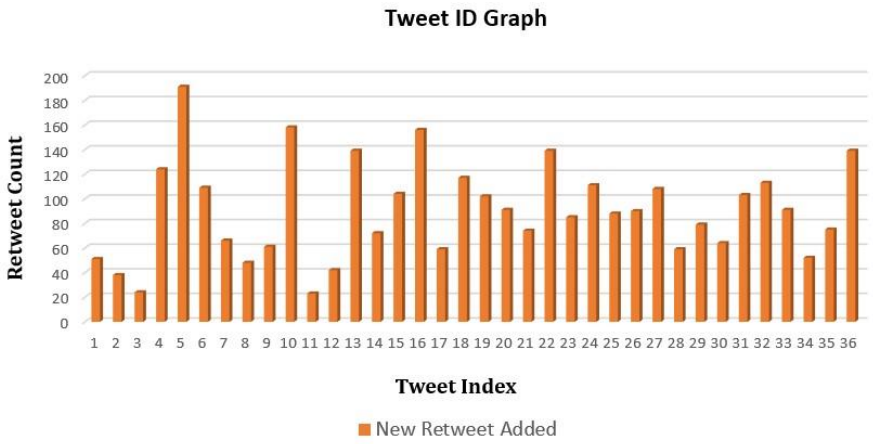

Fig. 5. Tweet ID Graph

4 Conclusion

Misinformation may lead to negative consequences and should be prevented. The method should be reliable for contributing to society in a positive way. The truth discovery combining the evaluation of reliability score with Jaccard Algorithm gives a better view of accuracy measurement for the existing truth discovery algorithms. Hence, it is seen in this paper, the truth discovery was done efficiently, the results easily indicate the truth and false counts. It can also be implemented on further platforms with natural language processing easily and can be applied on any kind of dataset for twitter. The implementation was tested on GUI based on Java programming language. A model like this may be useful to social media users by enhancing and reinforcing their own credibility judgments on each tweet, which would be a huge help in social media sensing. These findings might be useful in determining if propaganda on social media tweets is true or fake and whether they follow similar behaviors. For better efficiency techniques based on the truth discovery algorithm, we used only the popular truth detection algorithm i.e., SRTD. However, there exist other types of truth detection techniques based on NLP. The analysis for determining the number of synonyms of retweets might be addressed at a depth and more detailed level as a way to improve efficiency in Truth Detection, taken into account domains like as government, health, climate, municipal services, and so on (i.e., identifying tweets based on their contents or issues in terms of hashtags, as well as examining certain Twitter Security broadcasts). This might be done to see whether more specific measurements and models can be identified in comparison to the ones examined in this study, perhaps leading to greater accuracy results. Comparing the accuracy of the techniques like NLP (retweet-based techniques) with that of Similar tweet analysis are important topics for future work.

References

1. Zhang, Daniel Yue & Wang, Dong & Vance, Nathan & Zhang, Yang & Mike, Steven. (2018). On Scalable and Robust Truth Discovery in Big Data Social Media Sensing Applications. IEEE Transactions on Big Data. PP. 1-1. 10.1109/TBDATA.2018.2824812.
2. Gupta, Sangeeta & Ahmed, Omar. (2020). Truth Discovery in Big Data Social Media Sensing Applications. 9. 2278-3075. 10.35940/ijitee.H6311.069820..
3. Zhang, Daniel Yue & Badilla, Jose & Zhang, Yang & Wang, Dong. (2018). Towards Reliable Missing Truth Discovery in Online Social Media Sensing Applications. 143-150. 10.1109/ASONAM.2018.8508655.
4. Zhang, Daniel Yue & Han, Rungang & Wang, Dong & Huang, Chao. (2016). On robust truth discovery in sparse social media sensing. 1076-1081. 10.1109/BigData.2016.7840710.
5. Zhang, Daniel Yue & Wang, Dong & Zhang, Yang. (2017). Constraint-aware dynamic truth discovery in big data social media sensing. 57-66. 10.1109/BigData.2017.8257911.
6. Yang, Jielong & Wang, Junshan & Tay, Wee Peng. (2019). Using Social Network Information in Bayesian Truth Discovery. PP. 10.1109/TSIPN.2019.2914610.
7. Shao, Huajie & Sun, Dachun & Yao, Shuochao & Su, Lu & Wang, Zhibo & Liu, Dongxin & Liu, Shengzhong & Kaplan, Lance & Abdelzaher, Tarek. (2020). Truth Discovery with Multi-modal Data in Social Sensing. IEEE Transactions on Computers. PP. 1-1. 10.1109/TC.2020.3008561.
8. Khan, Moin & Jain, Amisha & Chouhan, Rishi & Sheikh, Sakeeb. (2020). Fake News Identification on Social Media. International Journal of Engineering Research and. V9. 10.17577/IJERTV9IS010183.
9. Zhao, Zhou & Cheng, James & Ng, Wilfred. (2014). Truth Discovery in Data Streams: A Single-Pass Probabilistic Approach. CIKM 2014 - Proceedings of the 2014 ACM International Conference on Information and Knowledge Management. 1589-1598. 10.1145/2661829.2661892.
10. Yang, Jielong & Wang, Junshan & Tay, Wee Peng. (2018). Using Social Network Information in Bayesian Truth Discovery.
11. J. Wiebe and E. Riloff, "Finding Mutual Benefit between Subjectivity Analysis and Information Extraction," in IEEE Transactions on Affective Computing, vol. 2, no. 4, pp. 175-191, Oct.-Dec. 2011, doi: 10.1109/T-AFFC.2011.19.
12. C. Banea, R. Mihalcea and J. Wiebe, "Porting Multilingual Subjectivity Resources across Languages," in IEEE Transactions on Affective Computing, vol. 4, no. 2, pp. 211-225, April-June 2013, doi: 10.1109/T-AFFC.2013.1.
13. W. Che, Y. Zhao, H. Guo, Z. Su and T. Liu, "Sentence Compression for Aspect-Based Sentiment Analysis," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 23, no. 12, pp. 2111-2124, Dec. 2015, doi: 10.1109/TASLP.2015.2443982.
14. Marshall, Jermaine, and Dong Wang. "Mood-sensitive truth discovery for reliable recommendation systems in social sensing." In Proceedings of the 10th ACM Conference on Recommender Systems, pp. 167-174. 2016.

15. Z. Jianqiang, G. Xiaolin and Z. Xuejun, "Deep Convolution Neural Networks for Twitter Sentiment Analysis," in IEEE Access, vol. 6, pp. 23253-23260, 2018, doi: 10.1109/ACCESS.2017.2776930.
16. A. Kumar, V. T. Narapareddy, V. Aditya Srikanth, L. B. M. Neti and A. Malapati, "Aspect-Based Sentiment Classification Using Interactive Gated Convolutional Network," in IEEE Access, vol. 8, pp. 22445-22453, 2020, doi: 10.1109/ACCESS.2020.2970030.
17. Z. Zha, J. Yu, J. Tang, M. Wang and T. Chua, "Product Aspect Ranking and Its Applications," in IEEE Transactions on Knowledge and Data Engineering, vol. 26, no. 5, pp. 1211-1224, May 2014, doi: 10.1109/TKDE.2013.136.
18. X. Zhou, X. Wan and J. Xiao, "CMiner: Opinion Extraction and Summarization for Chinese Microblogs," in IEEE Transactions on Knowledge and Data Engineering, vol. 28, no. 7, pp. 1650-1663, 1 July 2016, doi: 10.1109/TKDE.2016.2541148.
19. A. Ghose and P. G. Ipeirotis, "Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics," in IEEE Transactions on Knowledge and Data Engineering, vol. 23, no. 10, pp. 1498-1512, Oct. 2011, doi: 10.1109/TKDE.2010.188.
20. R. Cobos, F. Jurado and A. Blázquez-Herranz, "A Content Analysis System That Supports Sentiment Analysis for Subjectivity and Polarity Detection in Online Courses," in IEEE Revista Iberoamericana de Tecnologias del Aprendizaje, vol. 14, no. 4, pp. 177-187, Nov. 2019, doi: 10.1109/RITA.2019.2952298
21. X. Lei, X. Qian and G. Zhao, "Rating Prediction Based on Social Sentiment From Textual Reviews," in IEEE Transactions on Multimedia, vol. 18, no. 9, pp. 1910-1921, Sept. 2016, doi: 10.1109/TMM.2016.2575738.
22. Almatarneh S, Gamallo P (2018) A lexicon based method to search for extreme opinions. PLoS ONE 13(5): e0197816. https://doi.org/10.1371/journal.pone.0197816
23. Bhuta, Sagar, Avit Doshi, Uehit Doshi, and Meera Narvekar. "A review of techniques for sentiment analysis Of Twitter data." In 2014 International conference on issues and challenges in intelligent computing techniques (ICICT), pp. 583-591. IEEE, 2014.
24. T. Yang, Q. Yin, L. Yang and O. Wu, "Aspect-based Sentiment Analysis with New Target Representation and Dependency Attention," in IEEE Transactions on Affective Computing, doi: 10.1109/TAFFC.2019.2945028.
25. C. Liu, M. Chen and C. Tseng, "IncreSTS: Towards Real-Time Incremental Short Text Summarization on Comment Streams from Social Network Services," in IEEE Transactions on Knowledge and Data Engineering, vol. 27, no. 11, pp. 2986-3000, 1 Nov. 2015, doi: 10.1109/TKDE.2015.2405553.
26. Kim Schouten, Onne van der Weijde, Flavius Frasincar, and Rommert Dekker, “Supervised and Unsupervised Aspect Category Detection for Sentiment Analysis with Co-occurrence Data”, IEEE Transactions on Cybernetics, Vol. 48, No. 4, April 2018.
27. Andreea Onaciu, Anca Nicoleta Marginean, “Ensemble of Artificial Neural Networks for Aspect-Based Sentiment Analysis”, 2018 IEEE 14th International Conference on Intelligent Computer Communication and Processing (ICCP), published in IEEE, 2018.
28. Mitra, Ayushi. "Sentiment Analysis Using Machine Learning Approaches (Lexicon based on movie review dataset)." Journal of Ubiquitous Computing and Communication Technologies (UCCT) 2, no. 03 (2020): 145-152.
29. Chen, Joy Iong Zong, and Kong-Long Lai. "Data Conveyance Maximization in Bilateral Relay System using Optimal Time Assignment." Journal of Ubiquitous Computing and Communication Technologies (UCCT) 2, no. 02 (2020): 109-117