Scientific Journal of Impact Factor (SJIF): 5.71

e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 5, Issue 04, April -2018

AN ADVANCE MINING APPROACH FOR SENTIMENT ANALYSIS OVER ENRICH DATABASE

PRANJALI MALVIYA, PROF. DR. DEEPAK GOUR

**Department of Computer Science Engineering, RGPV, Bhopal, TIT, Bhopal, Madhya Pradesh India*

ABSTRACT:*Web mining is important when it comes to real world application. Sentiment analysis is a field where many people and their data is get analysis using the various library. There are many approaches were introduced by the previous author. Algorithm such as pattern based, signature oriented makes use of fixed format of data and use them for sentiment determination. Sentiment analysis help in different application area such as e-commerce, medical, politics and in every sector where opinion poll is involved. Previous algorithm does not includes enrich database for the sentiment analysis. A limited pattern, data is involved in analysis of sentiment over large dataset. A problem still persist in the terms of accuracy in detection. Data such as emoji and other relevant information doesn't included for analysis. In this paper an advance approach used for efficient sentiment analysis over different range of values and working with the enhance parameters. An experiment is performed using the real time social media dataset and tagging approach. Discussed finding result is compared with the traditional sentiment analysis approach. Executed result shows the efficient of our approach over existing available approach.*

Keywords:*Web Mining, Sentiment Analysis, Emoji Sentiment Extraction, Web Extraction, Pattern recognition.*

INTRODUCTION

Sentiment analysis is a type of natural language processing for tracking the mood of the public about a particular product or topic. Sentiment analysis, which is also called opinion mining, involves in building a system to collect and examine opinions about the product made in blog post, comments, reviews or tweets. There are several challenges in sentiment Analysis. The first is an opinion word that is considered to be positive in one situation may be considers negative in another situation. A second challenge is that people don't always express opinions in the same way. Most traditional text processing relies on the fact that small differences between two pieces of text don't change the meaning very much [1].

The Semantic Web enhances the traditional web by adding a semantic layer on top of the well-known web data formats to make the web machine readable. It serves as a standard language to define the terms in vocabularies and the relationships between those terms. Opposed to databases, ontologies serve as conceptual structures to describe the entire application domain, instead of just describing one specific application.

In the recent years, a huge amount of data is being gathered and stored in databases everywhere across the globe, which is mainly coming from information industry and social networking sites**.** There is a need to extract and classify useful information and knowledge from such a data collected [2]. Data mining is an interdisciplinary field of computer science and is referred to as extracting or mining knowledge from large databases. It is the process of performing automated extraction and generating the predictive information from a large database. It is actually the process of finding the hidden information or patterns from the repositories .The fields that use Data mining techniques include medical research, marketing, telecommunication, and stock markets, health care and so on[3].

LITERATURE REVIEW

This section discuss about the security approach and data accessing approach over the computing scenario. Discussed algorithm shows the procedure of sentiment analysis and accessing approach discussed by previous authors.
In this paper author research [4] on the WTM word based translation model which exploit the relation between the extracted data reviews and comments from the dataset they have considered such as restaurant, mp3 and other. They have presented capture opinion relations more precisely, especially for long-span relations.

They compared the algorithm with the existing such as syntax based technique and other. By using the technique they removed noises from parsing errors when dealing with informal texts and reviews extracted. Also they uses graph based approach which actual opinion targets are extracted in a global process, which can effectively alleviate the problem of error propagation in traditional bootstrap-based methods, such as Double Propagation as it present effective result than single approach in any of the tradition relation mining approach. In future they tend to involve

some model such as discriminative model, syntactic approach with their proposed model and monitor the effectiveness of the approach.

In this paper author [5] discuss about syntactic based approach for the word extraction and analysis, such model behaves with the help of data dictionary and applies in the scenario. Syntax based methods usually exploited syntactic patterns to extract opinion targets, which were however prone to suffer from parsing errors when dealing with online informal texts.
In this work author [6] described their work model on extraction data set and analyze them author defined some rules to identify text which says about different entity need to consider. They have fixed dependency relationship R in between words and then analyze the relation words. They have also focused on the word style such as personal style , short words and grammar, some short form , abbreviation is been consider to actual precision calculation on considering all the entities from the extracted dataset. Finally they have considered experiment using three dataset and outperform their presented technique best while compare with existing technique.

PROBLEM DESCRIPTION

There are several problems which occurs due to Semantic identity deception or Semantic**s**' in a website or social media platform .Semantic identity sometime it pretend as attack with multiple identities for malicious intent named after the famous multiple personality disorder patients "Semantic". This particular attack has been used by spammers to create multiple websites or ids with identical domain names with junk and duplicate content [7,8]. These pages have no quality content and are created just with the intention to create spam and drive traffic.

The problems associated with technique:

1. No proper information based technique is introduced.
2. Anti-Spamming technique which can refine Semantic based on object and identify based on its log basis is still required.
3. Previous approaches doesn't involve Meta information, user log information and further relation between their inputs.
4. Semantic identity deception and matching is not applied with real time implemented scenario.
5. Limitation with the accuracy, its computation time, computation cost were observed in solution given by previous technique.

Thus, this is the requirement of research to work toward a proper approach, which can work with identity deception and prevention over the algorithm analysis[9].

PROPOSED METHODOLOGY

In the proposed given approach enhancement is done with inclusion of Meta data and some other component for better improvement. Inclusion of email id, Meta information, NLP data extraction and analysis steps are involved in high end identity extraction, deception process. The below process are taken for identity deception in proposed approach.
Steps Taken for the work performance:
As per the proposed work given by our work, steps are explained here which is implemented using Java framework and web API.

1. An extraction of user's data at the time loading.
2. Further an analysis over the comments, post performed by the particular user or user of different group is performed.
3. Performing NLP (Natural language process) library over the comments.
4. Finding relevant verb, keyword, noun, pronoun, adjective over the given comments is analysed.
5. An identification of the document, sentence and comment behaviour is performed.
6. Applying the POS (part of speech approach) along with the NLP [8] to find the same meaning and usage of sentences.
7. Finding the comment search, finding the positive or negative comment count, their frequency and percentage of their effectiveness either positive or negative.
8. Finding group relevance and score finding for the identity deception from the group utility algorithm process.

CHESA(Computing based Hybrid Emoji Sentiment Analysis) User End.
Input: Multiple social identities -sit, Meta information of identity –mit, score=0, identity matching=0;
Output: Number of identity match, score.
Steps:
Begin [
Loading Social media platform API ();
Build social platform execution ();

Login(){
Matching with Comment ();
Post and other information ();
}
Execution failed ();
}
Finding Optimal parameter()
Finding Matching identity deception();
Return result parameters();
}
}
] End;

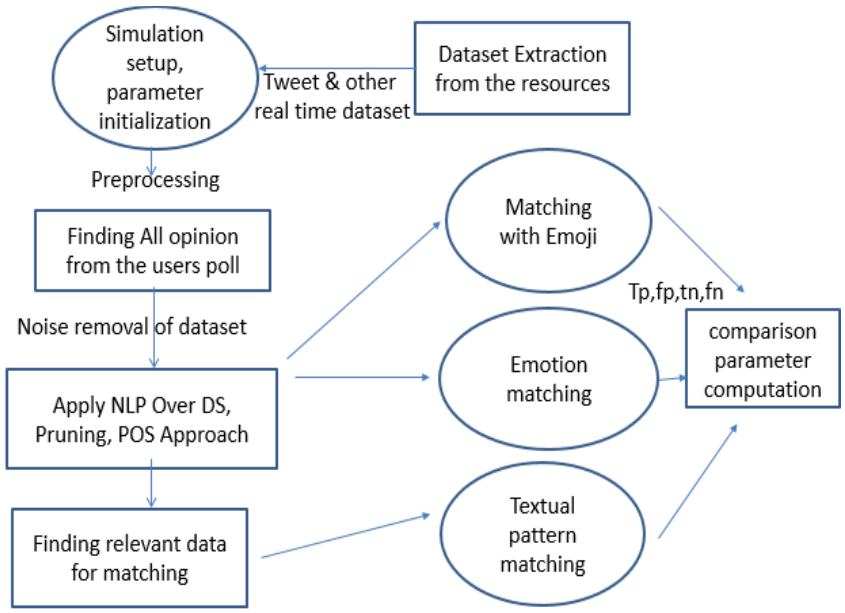

Figure 1: Detail architecture flow of proposed approach

The above pseudo code and flow diagram in figure 1, shows the complete process execution of proposed architecture.

EXPERIMENT & RESULT EVALUATION

A comparison analysis between the previously obtained technique which is compressive sensing and data processing approach is implemented. Further the proposed approach in two phases which is data storage and then query processing is given by the output interface [10].

These comparison is explained in tabular form as following:

Comparison among algorithm on the basis of different computation:

Algorithm computation	Existing Accuracy %	Proposed Accuracy%	Existing technique Precision %	Proposed Technique Precision %
Identity 1	**68.6**	**71.33**	**83.1**	**86.5**
Identity 2	**76**	**73.2**	**81.3**	**88.9**
Identity 3	**87**	**91.2**	**89**	**89.90**

Table 1 Time Comparison among Algorithms

Table 1 shows the comparison among different considered algorithms with respect to completion time. Here results shows that proposed algorithm is taking less time with respect to different number of tweets up to 1 Crore in Analyst simulation environment.

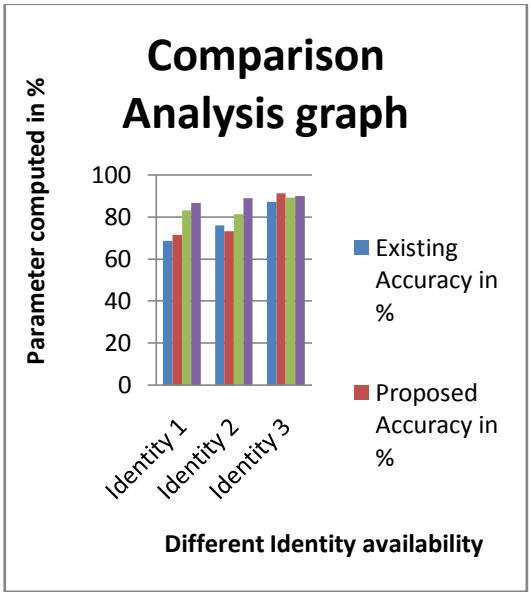

Figure 2 Time based comparison among query processing algorithms

In the figure 2 above, a comparison between the existing and proposed secure accessing approach is performed. The result evaluation shows the efficiency of proposed approach.

As per analysis of result retrieve by existing and proposed approach over cloud data analysis, it is observed that proposed approach exhibit low computation time as well as less execution time to process multiple test.

CONCLUSION

Semantic deception is an important aspect today in various social media and generic platform which talk about the usage of their data. Social media platform help in organic visibility of users content, they help in advertising and reach of correct content to the genuine number of public. It works according to user interest and their application use. Social media platform face a challenge from the different user and organization such as facing a Semantic identity and populating wrong data among the users. It makes use of false review, rating, like and other scenario by the fake id created by Semantic. Semantic can be identifying by identifying proper given text or other data by user. Identifying textual content, identifying multimedia content is important aspect.

REFERENCE

[01] Xiaohui Liang and Xiaodong Lin. "Fully Anonymous Profile Matching in Mobile Social Networks", IEEE areas in communication vol:31 no:9 year 2013.

[02] N. Eagle and A. Pentland, "Social serendipity: mobilizing social software, "IEEE Pervasive Computing, vol. 4, no. 2, pp. 28–34, 2005.

[03] J. Teng, B. Zhang, X. Li, X. Bai, and D. Xuan, "E-shadow: Lubricating social interaction using mobile phones," in ICDCS, 2011, pp. 909–918.

[04] B. Han and A. Srinivasan, "Your friends have more friends than you do: identifying influential mobile users through random walks," in MobiHoc, 2012, pp. 5–14.

[05] Z. Yang, B. Zhang, J. Dai, A. C. Champion, D. Xuan, and D. Li, "E-smalltalker: A distributed mobile system for social networking in physical proximity," in ICDCS, 2010, pp. 468–477.

[06] R. Lu, X. Lin, X. Liang, and X. Shen, "A secure handshake scheme with symptoms-matching for mhealthcare social network," ACM MobileNetworks and Applications (MONET), vol. 16, no. 6, pp. 683–694, 2011.

[07] J. Katz, A. Sahai, and B. Waters, "Predicate encryption supporting disjunctions, polynomial equations, and inner products," in EUROCRYPT, 2008, pp. 146–162.

[08] M. Motani, V. Srinivasan, and P. Nuggehalli, "Peoplenet: engineering a wireless virtual social network," in MobiCom, 2005, pp. 243–257.

[9] M. Brereton, P. Roe, M. Foth, J. M. Bunker, and L. Buys, "Designing participation in agile ridesharing with mobile social software," inOZCHI, 2009, pp. 257–260.

[10] E.Bulut and B.Szymanski, "Exploiting friendship relations for efficient routing in delay tolerant mobile social networks," IEEE Transactions on Parallel and Distributed Systems, vol. 23, no. 12, pp. 2254–2265, 2012.

