

The Amalgamation Of Lexical, Syntactic And Semantic Analysis Is A Quintessential Approach In Maintaining Sentiment Analysis On Social Networking (Facebook + Twitter)

Rajat Verma, Harshita Mishra

Department of Computer Science & Engineering

Amity School of Engineering & Technology

Amity University, Lucknow

rajatverma310795@gmail.com, mharshu1995@gmail.com

Abstract— we the people are dominated in this modern era of social networking. This is clearly depicted in this paper. The combination of lexical, syntactic, semantic usage in the domain of social networking is also mentioned in this paper. The individual functionalities of lexical, syntactic and semantic analysis is also illustrated in this paper. These all features will clearly increase the dimension of clarity in the field of social networking.

Keywords— *Social, Media, Lexcal, Syntactic, Semantic, Sentiment.*

I. INTRODUCTION

According to a study, an average person spent approximately one-third of a day on social media. It takes about 8-9 hours. It includes some renowned applications such as Facebook commonly abbreviated as "fb", twitter, WhatsApp etc. It means we are addicted to them and in the upcoming years we couldn't think our lives without them. If we think of today, then also we acquire similar scenario.

Coming to the aspect of Sentiment analysis [1], it is defined as the chore of recognizing the favorable or non-favorable viewpoints, feelings, judgements in a given narrative. Viewpoints can be different for someone in one case and different for other in that same case or scenario. One can think that this work is fine, other may think that this work is worst, some other may think that it can be made better by putting extra efforts. When one has to communicate to a lot of people, the most efficient manner is the usage of social networking [2]. Whatever a person feels, he/she can post it on a social network. Words speaks for itself that they are positive or negative in nature. If a sentence written in a proper syntax then it is easy to understand that "what is written". That means the correct usage of formats that is discussed in the dimension of English Language i.e. Subject+ Verb + Object. If a sentence follows semantics that means its meaning is clear then its value becomes measureable. It means that "what is being said". And if all the three aspects come together it

becomes a perfect approach in the dimension of sentiment analysis.



Fig.1 The logo of Twitter and Facebook

II. RESEARCH BACKGROUND

A. LEXICAL ANALYSIS

The term "Lexical" is concerned with the dimension of concordance. Directly coming to the point that is in the field of social networking, let's take the case of Facebook, people use to post statuses on their wall so that their public can watch that. It could be happy, sad and angry or any kind of emotion or sentiment, we are free to post any feeling that we feel. This can be expressed by the means of words. Some example could be "I bought a new iPhone, it is amazing and I am happy", "Alas! My car broke down in an accident". In the first example, two words that are amazing and happy expresses happiness whereas in the second example, alas and accident expresses a feeling of sadness. People use this technique in the caption of their pictures other than statuses.

Coming to the aspect of twitter, "#" is used to indicate something. Let's take an example of a recent tournament that happened in Dubai. It was "Kabaddi Masters", if a person has to express a sentiment then he/she must use #kabaddikurfew in his/her tweet. An example could be, a tweet used by a legend player "Virendra Sehwaag", he tweeted "Another demolition. What a great performance by India to beat Pakistan once again. Ek aur jeet dila Diya humka Thakur, Ajay Thakur. 41-17 this time after 36-20 a few days ago. Well done, Team India #IndvPak #KabaddiKurfew". This was done when India had beaten Pakistan in a

match. It was a sentiment of happiness. In the field of compiler also, lexical analysis [3] is the first thing to do.



Fig.2 An example of a lexicon of sentiments

In the illustration mentioned above a sample of sentiment lexicons are mentioned. Great, Amazing and Wow expresses joy and happiness whereas Wrong and Mistake expresses bad feelings or sentiments.

B. SYNTACTIC ANALYSIS

This term is responsible for the syntactic connections among locutions. The Parts of Speech commonly abbreviated as (POS) [4] plays a deep impact role in this syntactic scenario. Let’s take the case of Verb that are basically of two types that are active verb and passive verb. If we consider a passive verb then an example of it could be “Person was good”. If we consider an active verb then the example of it could be “He protested”. They are syntactically correct but are semantically weak.

This syntactic approach [5] also has positive and negative viewpoints. An example of positive view point can be “Many Congratulations” and an example of negative viewpoint in this dimension “He Killed”. All these examples can be posted as facebook [6] statuses or as a caption of a picture. In the criteria of compiler phases, syntactic analysis comes in 2nd position after the lexical analysis.

Examining a sequence of string can be considered as syntactic analysis. Parsing is just a synonym of it.

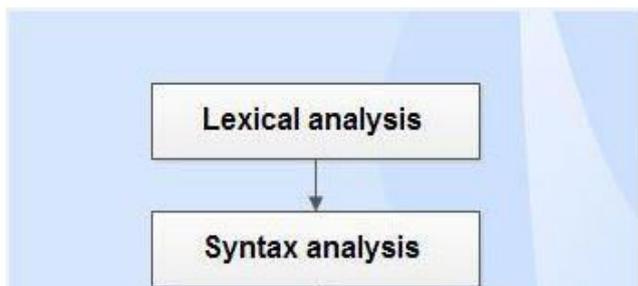


Fig.3 illustrating the appropriate use of syntax analysis that is after lexical analysis

Coming to the aspect of twitter [7], it is not an obvious thing that each and every time the tweets are syntactically correct that is Subject then Verb then Object. Many a times, tweets can be noisy and messy as well as not-structured that becomes a problem in maintaining a syntactic approach. This statement becomes true when we think of real world data (unprocessed raw facts and figures) is highly inconsistent.

In twitter also, like facebook can happen that syntactically a statement is correct but not semantically. And if a statement is not syntactically correct, there is very less probability that semantically it would be correct!

C. SEMANTIC ANALYSIS

The Third component that comes in this amalgamation process is semantics. Its meaning is concerned with rationality along with sense. It should be done after the syntax analysis have already been done.

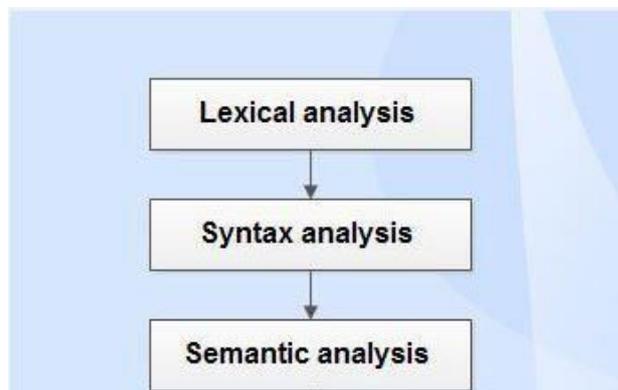


Fig.4 illustrating the correct stage of semantic analysis that is after syntactic analysis

In the dimension of social networking basically talking about facebook and twitter, the semantics plays a major role.

One term that meant here is “Contextual Semantics” [8]. It particularly means that “Terms that happen in homogeneous situations basically meant similar”. In the year 1953, it was told by “Wittgenstein”. An Illustration of it can be depicted with a figure given below-

In the positive manner, if the word “Trojan Horse”[9] is used, in social network then as a status or post it man refer to Greek Culture, A Tale, History, Class,, Wooden, Troy etc.

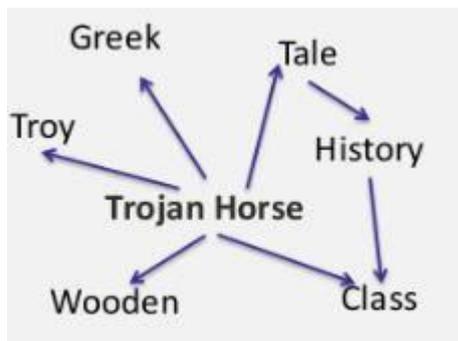


Fig.5 Positive meaning of words (Trojan horse)

If taken in a negative manner, it may be used for some code, malware, threat, hack, program etc.

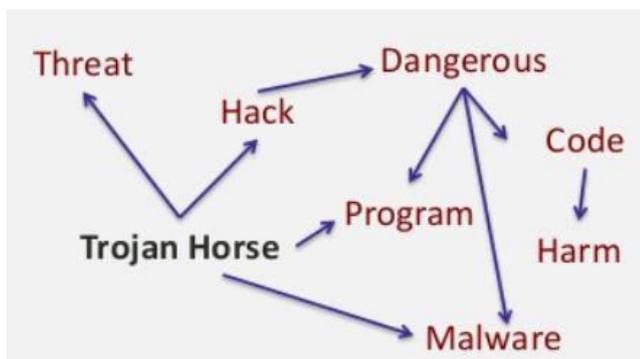


Fig.6 Negative Meaning of words (Trojan horse)

The point is that, on social media positive things may become viral or not but negative things become viral at a very alarming rate!!

This word can be used as a post, status or a tweet on social media and can be used in both ways that is in favorable or non-favorable aspect.

Some structure that belongs to Contextual Meaning Emotions/Sentiments [10] also plays a vital role in this scenario. It can be depicted as few terms that are mentioned in variable tweets with similar meaning, logics and emotions tend to form clusters. An illustration of this is depicted in Fig.7.

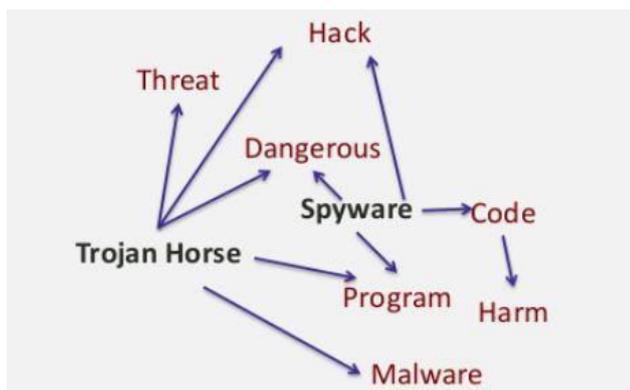


Fig.7 Trojan horse and Spyware are combined to form a new scenario.

The Fig.7 follows a negative contextual pattern and if presented on social media, it will expand exponentially!

III. SENTIMENT ANALYSIS

One way of expressing sentiments is by emoticons, Fig. 8 depicts the emoticons [11], their meanings along with their strength.

Emoticon	Meaning	Strength
:D	Big grin	1
BD	Big grin with glasses	1
XD	Laughing	1
\m/	Hi 5	1
:).=).:-)	Happy, smile	0.5
:*	kiss	0.5
:	Straight face	0
:\	undecided	0
:(sad	-0.5
</3	Broken heart	-0.5
B(Sad with glasses	-0.5
:'	crying	-1
X-(angry	-1

Fig.8 Emoticons

Some Verbs and adverbs also play an important role in determining the value of semantic scores [12]. This can be depicted in Fig.9.

Verb	Strength	Adverb	Strength
Love	1	complete	+1
adore	0.9	most	0.9
like	0.8	totally	0.8
enjoy	0.7	extremely	0.7
smile	0.6	too	0.6
impress	0.5	very	0.4
attract	0.4	pretty	0.3
excite	0.3	more	0.2
relax	0.2	much	0.1
reject	-0.2	any	-0.2
disgust	-0.3	quite	-0.3
suffer	-0.4	little	-0.4
dislike	-0.7	less	-0.6
detest	-0.8	not	-0.8
suck	-0.9	never	-0.9
hate	-1	hardly	-1

Fig.9 Strengths of Verbs and Adverbs (Twitter)

The main implementation can be depicted by a well-renowned classifier. It is known as Naive Bayes Classifier [13]. In this all the nodes are independent. X has a sole parent C which is a root and having no parents. It is better than other classifiers.

Implementation can be illustrated with the following figure-

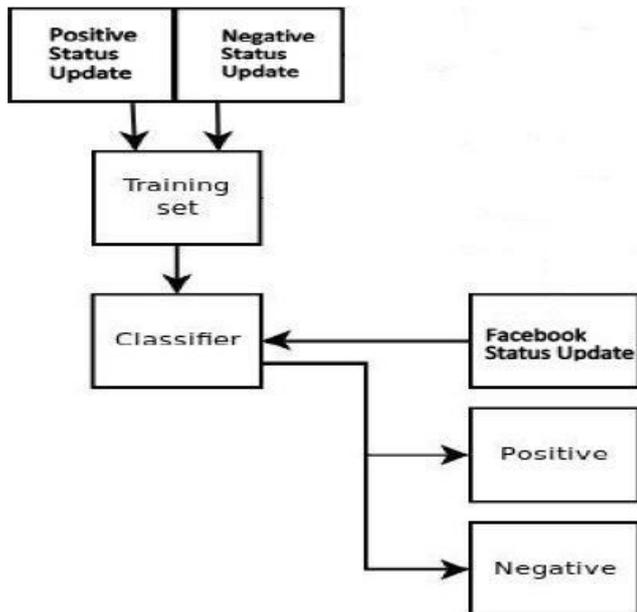


Fig.10 Main Implementation of Naïve Bayes Classifier

IV. CONCLUSION

From my point of view, this amalgamation approach is quite a good approach in maintaining the sentiment analysis. The sentiment analysis can be performed with the well-known classifier known as Naïve Bayes Classifier. The individual functionalities of lexical analysis, syntactic analysis and semantic analysis is mentioned in this paper. Emoticons as well as Strengths of verbs and adverbs are also mentioned in this paper. In the phase of compiler also, some processed data is mentioned in this paper. The positive as well as negative approaches in all aspects in also mentioned in this paper. All these information will certainly contribute to the research and development industry.

REFERENCES

1. Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1), 1-167.
2. Lenhart, A., & Madden, M. (2007). Social networking websites and teens: An overview.
3. Church, K., Gale, W., Hanks, P., & Hindle, D.

(1991). Using statistics in lexical analysis. *Lexical acquisition: exploiting on-line resources to build a lexicon*, 115, 164.

4. Croft, W. (2005). Word classes, parts of speech, and syntactic argumentation. *Linguistic Typology*, 9(3), 431-441.

5. Brayer, J. A. R. A. J., DeMori, R., SL, H. Y. F. K. F., Moayer, H. B., Swain, T. P. W. S. P., & Varnos, T. (1982). Syntactic pattern recognition, applications.

6. Ortigosa, A., Martín, J. M., & Carro, R. M. (2014). Sentiment analysis in Facebook and its application to e-learning. *Computers in human behavior*, 31, 527-541.

7. Kumar, A., & Sebastian, T. M. (2012). Sentiment analysis on twitter. *International Journal of Computer Science Issues (IJCSI)*, 9(4), 372.

8. Saif, H., He, Y., Fernandez, M., & Alani, H. (2016). Contextual semantics for sentiment analysis of Twitter. *Information Processing & Management*, 52(1), 5-19.

9. Butler, R. W. (1990). Alternative tourism: pious hope or Trojan horse?. *Journal of Travel Research*, 28(3), 40-45.

10. Kelly, J. R., & Barsade, S. G. (2001). Mood and emotions in small groups and work teams. *Organizational behavior and human decision processes*, 86(1), 99-130.

11. Walther, J. B., & D'Addario, K. P. (2001). The impacts of emoticons on message interpretation in computer-mediated communication. *Social science computer review*, 19(3), 324-347.

12. Fischer, P., Gatterer, G., Marterer, A., & Danielczyk, W. (1988). Nonspecificity of semantic impairment in dementia of Alzheimer's type. *Archives of neurology*, 45(12), 1341-1343.

13. Troussas, C., Virvou, M., Espinosa, K. J., Llaguno, K., & Caro, J. (2013, July). Sentiment analysis of Facebook statuses using Naive Bayes classifier for language learning. In *Information, Intelligence, Systems and Applications (IISA), 2013 Fourth International Conference on* (pp. 1-6). IEEE.