Knowledge Acquisition System for Sentiment Analysis

Muhammad Sheharyar Liaqat[1], Ihtisham ul Haq[1], Muhammad Burhan[1], Shakir Mahmood Mayo[2]

[1]School of Systems & Technology, University of Management & Technology, Lahore, Pakistan

[2]Deapartment of City and Regional Planing, University of Engineering & Technology, Lahore, Pakistan

* Correspondence: Muhammad Sheharyar Liaqat – **Email ID: ch.shehri@gmail.com**

Sentiment Analysis is considered an advanced technology to predict and analyze people's opinions, attitudes, sentiments, and perceptions towards some topics, products, and services. Due to the fast evolution of Internet-based applications, opinion mining becomes a very substantial method nowadays. For this purpose, many systems have been developed so far; some use the statistical method whereas some use parsers to analyze the data. However, this article presents a complete study and uses a hybrid approach to analyze volumetric data in the form of sentences. The hybrid approach of sentiment analysis uses both deep learning (statistical methods) and knowledge-based methods. Furthermore, this article also uses a sentence structure approach that helps to overcome the linguistic effects in the knowledge base. In addition, this article also suggests some future directions using the sentence structure of natural language for an expert system.

Keywords: Expert System, Knowledge Acquisition, Sentimental Analysis, Opinion Mining, Common-Sense, Anaphore

INTRODUCTION

In the physical world, if somebody famine to make something or share the some of his ideas with the world then in response as a human being they hunger to know people's feedback to know what people reflect about their idea. In the world, every single person has a nature of sentiment or gives an opinion to others on any accomplishment, effort, or idea. As human beings, it's in our nature to demonstrate our sentiments about others' business. In the premature day, people contemplate it as a bad tradition if someone gives judgment about his effort but as the technology, science and business raise swiftly with the time this perception has been rehabilitated. Now according to the technology publicizing and business yield point of view they prerequisite people sentiments and thoughts about the product for the enhancement of their products and technology. That's why this tendency to accomplish the knowledge from people's judgment increasing so fast in recent years.

This knowledge Attainment has an excessive consequence on the advantages of technical perfection and financial point of view of the business. This progression of knowledge acquisition from people's responses is itself very enthusiastic and exposed blockades for solving the feebleness in technology and business [1], [2]. In now a day, organizations have identical relaxed ways to get sentiments from people by posting on the many social media podiums like Facebook, Twitter, Google+, YouTube, etc. but the foremost thing is data withdrawal of people's opinions and sentiments like feelings from the natural language is extreme level difficult, rough job to do it. Mining is a rough job because this progression needs good expertise in the natural language and its all characteristics like symmetry, irregulating, explicitly and semantic instructions, etc. There are many approaches depending upon the words like affects expression, the polarity of words, and how several times this definite word catches up.

The expert uses the training associated technology for the sentiment investigation like visionless confidence deprived of any hesitation but this mode of work entirely makes the grammar guidelines of language virtuously ignorable to the language dispensation. Before going into an additional conversation about the mawkishness examination let's just clear the two-term that have so reputation in this analysis conferring to the natural language. The first term is semantic which means associated the word with others by their meaning, sense, and relationship with any other sentimentality or we can say that perceptions do not follow the rules of natural language.

The other one is syntactic which means that effort renders to the rules of language, their grammar positions, rules, etc., and then attach with the sentimentality. It means working with a full of acquaintance base. There is a proportion of approaches working in now a day for the sentiment study but we conferred here the Sentic net method by [3] for such investigation and gather the knowledge of rules around the language. That means it apprehension the word like syntactic with the assistance of knowledge base and associates the people's actions, changes, essentials, etc. with their knowledge and give results. Sentic net does not work like theoretical technique by meanings of words or rate of recurrence of the word in a podium and accompanied with the common-sense knowledge of a system for withdrawal. This approach work on the multi-word general expression that associated the sentiments by given expressions as participation in natural language. Every characteristic has its restrictions on work.

This Sentic net has a limit of immediate oversimplifying the perception like play-Cricket or performs- Gym because it did not catalog except has the flawless match. The variety of Sentic net 3 pop the error of not instigate. This kind of error is not virtuous for any approach to solve this a new variety 4 familiarize that get the very straightforward concept of Noun and Verb for that kind of difficulties. In the above example, paly-Cricket or perform-Gym catalog as Play-Sports. This kind of example shows the knowledge of nouns and verbs. For example, Cricket and Gym are like a Noun commonly, and on the other hand play and performing is a kind of action so it's cataloged as Verb. There is a question ascend that why we use this verb and noun awareness for mining because generally, this Net has knowledge of perceptions and nevertheless that automatically pick that from clustering procedure. But some time above example conditions is not handled/controlled by these. So here's come the idea of Noun and Verb for such sentimentality that can disturb the acquaintance. The remaining paper's sections are divided into the literature reviews, Discussion, and results of discussed research. The last section is all about the conclusion and future work.

LITERATURE REVIEW

Let's deliberate some literature assessment about the knowledge Acquisitions about sentiments. These whole things are based on the profound study of other researchers about the sentimentality examination. As discoursing this examination these are alienated into two superlative approaches one is knowledge base and another count on the statistics [4]. Both methods have very respectable results in the sentiments investigation. The preliminary day's knowledge base was well-known for the analysis but in the current years, much research on sentimentality is done by the statistical model under substantial supervision.

Pang et al. [5] came up with approximately very diverse impressions to accumulate the appraisal of people. They use some machine learning mechanisms for assemblage. As a research of machine learning method applied on the catalog of sentiments about the film (Video documentary) appraisals given by the people. They used the positive and negative appraisals by the method to get 82% precision for that research.

Chauhan et al. [1] use opinion mining to predict election polling results from online social media platforms. This research highlights the challenges associated with the prediction of election results obtained from online content and suggests some open issues related to sentiment analysis.

Subsequently, so numerous years of this method researchers Socher et al. [6] used a diverse approach to the same assemblage of film appraisals. They implement the neural network for this and smear on the same dataset and accumulate the 85% precision result. In 2003 [7] used a method to discover sentence-level sentimentality mining. They use the semantic (word connotation relations to other) method to accumulate the result from the sentences. Hence it gets very reputation result from this procedure. In continuing conversation, the core focus is on the general knowledge Acquisitions of sentimentality but not go for the precise domain sentimentality information [8] offered a framework that produces very respectable results for the precise field/domain sentiment Acquisition. Nowadays, social media podiums are cumulative day by day and users also grow with the use of social media podiums. Social media podiums like Facebook, Twitter, YouTube, Blogs, Google+, etc. are undertaking a very significant part in the sentiment withdrawal. In these podiums, very sentiment tags, entrenched words, symbols, etc. are used to express the sentiments. To get assistance from these [9] use the neural network and make a system name

convolutional. This network works by the implanting tags, words manifestations regularity in the people's thoughts and also recently [10] worked on the previously entrenched words procedure for the sentiment Accusations Facet proportions.

Sideways with these researches [11]correspondingly work on neural network for the examination but diverse his consideration from embedding words to the petite sentence or small texts for the Knowledge Acquisitions. Continually since our conversation about the sentimentality withdrawal from the entrenched tags, words, URL, short texts, etc. but not talk what about the Audio and Video media. Sentiment mining in this category of media is categorically a tough job but [12] established a model for that audio and video media. They united both methods feature and decision level fusion it to abstract the sentiment Acquisitions. They experiment on YouTube videos as data set and get about 80% exceptional precise results and it accomplishes in more than 20% of states.

Laterally the multimode [13] did a profound study on the Facet proportion method in sentiment withdrawal. Previously that all studies straight get final consequences from mining but Facet Proportion is like a sub-task of sentiment Acquisitions because it directly marks the precise sentiment in the data set. They work with the help of a Neural Network and make a 7 Layer prototype that dispersed the Characteristic and Not Characteristic terms. They get very exceptional precise results from this prototype. Now days very well-known openly accessible tool for opinion withdrawal is Sentic net accessible by the [14]. Sentic net commonly mines Artificial Intelligence, logic, and common sense to interpret the positive and negative consequences connexion with the bag of concepts and represent them in the form of semantic knowledge. One of the distinct features of the tool is that dynamically alters the word expression into a new form that is helpful for the machine. That new form can be accessed by machine and the machine can progress this illustration. Subsequently that we can see numerous tools for the words and short text mining but for the big data mining Sentic net is a furthermost recent tool specifically focused on this era. As all distinguish that laterally the knowledge base technique there is alternative method statistical base works fine but [15] demonstrate that statistical approach feebleness. The statistical model works with the words, Keywords, and rate of recurrence words but only in given bulky text. This method does not work well with small text such as short text and sentences. One of the new perceptions "People Opinion Mining" was anticipated by Guo et al [16]. This impression essentially associates the people's opinions with dissimilar methods in every phase of mining and after comprehensive all phases, it can foresee the future of distinct domain with the help of people's opinions. This perception accomplishes some good consequences.

Online sources are the superlative informant for the structured and unstructured knowledge accessible all-time on the Internet. As human beings, we can read and comprehend the denotation of the massive volume of information from online resources but making the unstructured data meaningful to the machine is a categorically hard one. For a machine, this understanding of the data is rough because of the hole between the Natural Language data and concept level mechanism. To lessen the gap between these machines, prerequisite common-sense knowledge (Concept knowledge). However, some game theories and people sourcing they able to gather some valuable common-sense knowledge for Machine accessibility.

Word level knowledge acquisition is adopted by pattern directed or statical methods. Even some effective network like "ConceptNet" not deal effectively word level disambiguate [17]. On the other hand, MACQUIK is effective in terms of clear word level disamgiguities. The sentiment is playing an important role in different tasks of machine learning, in the current era, the importance of machine learning with sentiment analysis cannot be neglected, machine learning plays an important role in different fields of life [18]–[21].

HYBRID APPROACH METHODOLOGY

In each era of analysis, updated approaches are applied for sentimental knowledge. Some of the approaches get exceptional accurateness in the consequences. The authors suggest the hybrid method for the precise consequence of sentimental knowledge. The hybrid approach eliminates very mistakes in both procedures' knowledge and statistical approach.

The hybrid approach of sentimental analysis exploits both statistical methods and knowledge-based methods. It inherits high accuracy from deep learning (statistical methods) and stability from the lexicon-based approach. There are four main steps to processing sentimental analysis:

Step-1: Data Collection Step-2: Data Processing Step-3: Data Analysis Step-4: Data **Visualization**

The authors suggest the hybrid approach one supplementary thing about the sentence structures along with the verb and noun idea. In knowledge, base affected words are used for knowledge acquisition but don't recognize the sentence structure in the sense of linguistics if we use this approach sentence structure approach we can overcome the linguistic effects in the knowledge base, and beneath section, we show some results of hybrid approach results and graphs by improvement by Hybrid method. If the system has more sense about the grammar and sentence structure Hybrid can access accurately large text and short text for the sentimental knowledge acquisitions.
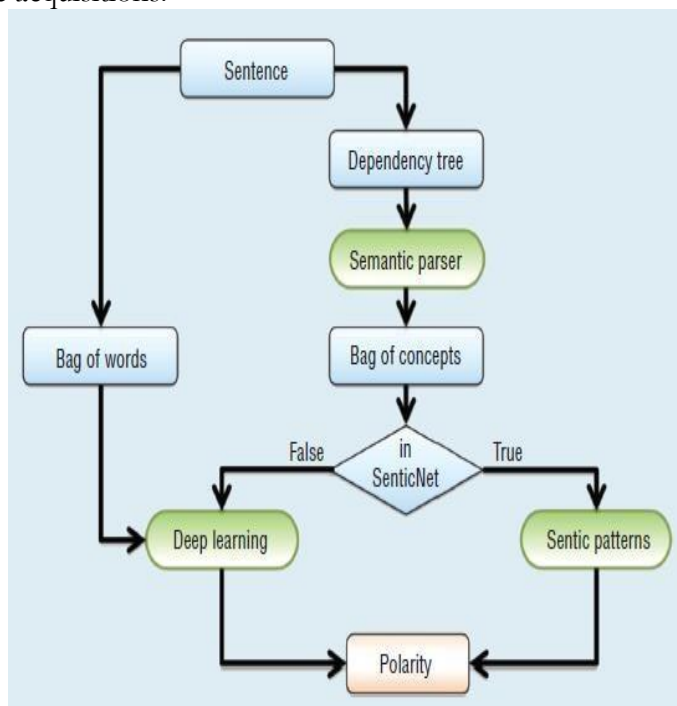

Figure 1: Hybrid Approach Flow Chart

Figure 1 shows the semantic Diagram of anaphoric relationships of the hybrid approach used in our system. Anaphoric relations are the relations that exist between linguistic expressions. One's linguistic expression interpretation known as the anaphor relies on the other's linguistic expression interpretation known as the antecedent. Our approach takes the sentence as input then makes a bunch of words from the input and then applies parser and deep learning algorithms to obtain the required results.

Automatic Knowledge Acquisition by Assimilation

The seek for expressly or implicitly introduced identical ideas utilized in completely different elements of a text and their fusion into one linguistics representative throughout the consecutive transformation of this text into one integrated kilobyte is that the main task of the assimilation method.

In this technique, sentence meaning represented with semantic networks, The network nodes represent conceptual entities and the lines represent the relationship of the network.

The Resolution of References Induced by Preforms

The important types of the references are backward, forward and deictic and with the concept of proadverbs and pronouns. The text like "last/current year" often represent not explicitly introduced by the forgoing text.

Ontology Based References

Synonyms are often used for references text categorization, subordination and synonyms are features of ontology, and such references are so called "Ontological References". Subordinate are effective for form expression(article). The conference alternatives are described by "CORUDIS", and in this if background knowledge is missing then computation is too slow.

And The reference resolution technique is effective for non-pronoun references.

ROBUSTNESS of HYBRID APPROACH & DISCUSSION

Let's just recapitulate the complete conversation about the sentiment analysis from our point of view. Above all discussion can be characterized into three types Knowledge base, Statistical base, and Hybrid approach.

The knowledge base approach works by the affected words' knowledge. Affected words like "happy, sad, and angry" are used for the knowledge base approach. There are many conducts for characterizing the affected words with the help of Word Net effect, Sent Word Net, Sentic net, Lexicon, and many more. Apart from the strong side of knowledge base have some faintness. The knowledge base does not work well when linguistic influence comes up in the text. For example, when input as "I am happy today" it extracts the Happy affected word but in the linguistic "I am not happy today" it becomes fails to give a precise consequence. To overwhelm this, it involves the best knowledge about the rules of language. Thus, another problem of the knowledge base is boundaries because it does not footstep outside of his knowledge.

The other approach is statistical work with the help of cavernous machine learning algorithms. It also learns the affected words with the lexical effect and occurrence of affected words by profound learning. Apart from the strong side statistical approach also have some faintness.

The statistical model works with the words, Keywords, and rate of recurrence words but only in given bulky text. This method does not work well with small text such as short text and sentences. The best one come up with the use of amalgamation of both approaches as a

Hybrid approach. This approach gives the best way to use both approaches combined and get more accuracy in work. The knowledge base approach is like a bag full of knowledge and a statistical like full of concepts. When machine learning and knowledge work together it can cover the flaws of another approach so effortlessly and give the more precise consequences. Below is the figure show hybrid architecture and how this approach works with the help of mutual approaches.
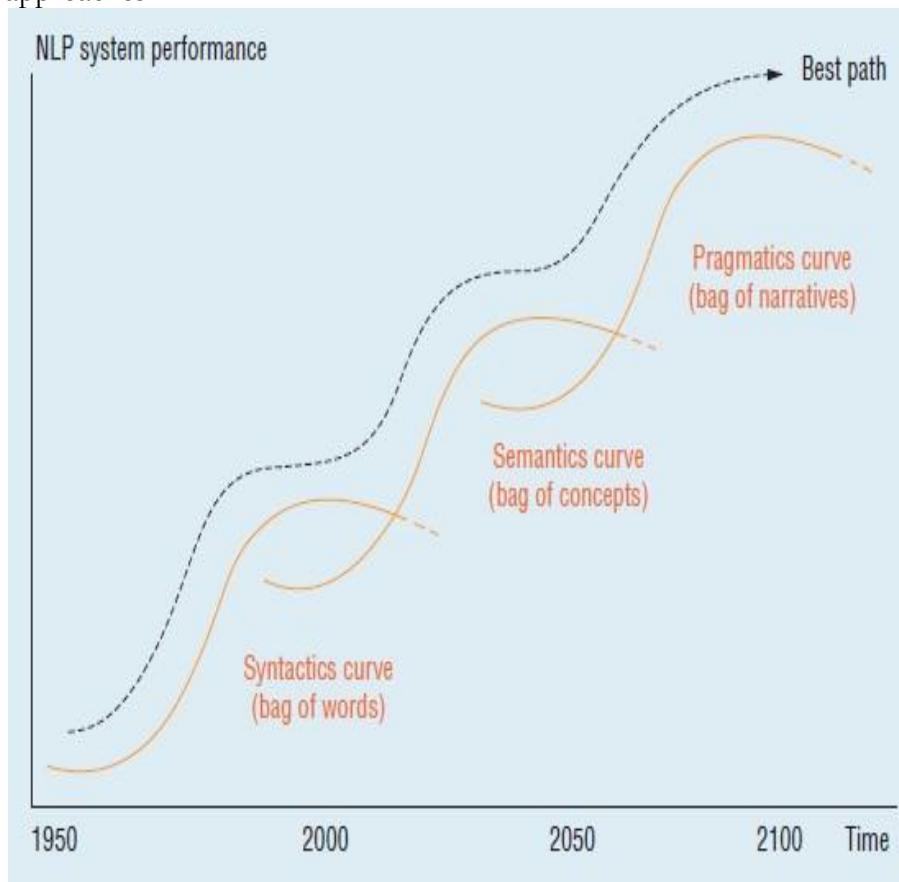

Figure 2: Result of Hybrid approach in Graphical Form

CONCLUSION

This article presented an overview of different approaches to analyzing sentiments of people's opinions. Furthermore, sentimental analyses are becoming popular now a days to think like human beings. A hybrid approach is used in this paper to handle and eradicate sentence structure problems. This stated approach is a combination of statistical and knowledge-based methods that apply deep learning algorithms to a given sentence. Besides, results are shown in the form of a graph and methodology in the form of a flow chart. This approach helps to reduce the linguistic problem.

In the future, if the system could acquire the sentence structure of natural language, then it might be more effective in the etymological complications and learn more common-sense knowledge to understand more about the feeling and emotions of human nature. It can accomplish through deep learning to become an expert system.

REFERENCES

[1]   P. Chauhan, N. Sharma, and G. Sikka, "The emergence of social media data and sentiment analysis in election prediction," *J. Ambient Intell. Humaniz. Comput.*, vol. 12, no. 2, pp. 2601–2627, 2021, doi: 10.1007/s12652-020-02423-y.

[2]   Y. Lin, X. Wang, and A. Zhou, "Opinion spam detection," *Opin. Anal. Online Rev.*, no. May, pp. 79–94, 2016, doi: 10.1142/9789813100459_0007.

[3]   Erik Cambria, D. Olsher, D. Rajagopal, and E. Cambria, "Proceedings of AAAI," *SenticNet 3 A common common-sense Knowl. base Cogn. Sentim. Anal.*, pp. 1515–1521, 2014.

[4]   E. Cambria, "Affective Computing and Sentiment Analysis," *IEEE Intell. Syst.*, vol. 31, no. 2, pp. 102–107, Mar. 2016, doi: 10.1109/MIS.2016.31.

[5]   G. B. Fischer, "Pneumocystis carinii, Aspergillus fumigatus) •," *Empir. Methods Nat. Lang. Process.*, no. October, pp. 1631–1642, 2004.

[6]   H. Yu and V. Hatzivassiloglou, "Towards answering opinion questions," pp. 129–136, 2003, doi: 10.3115/1119355.1119372.

[7]   B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment Classification using Machine Learning Techniques," pp. 79–86, 2002, doi: 10.3115/1118693.1118704.

[8]   T. Durmus, J. Stöckel, T. Slowinski, A. Thomas, and T. Fischer, "The hyperechoic zone around breast lesions - An indirect parameter of malignancy," *Ultraschall der Medizin*, vol. 35, no. 6, pp. 547–553, 2014, doi: 10.1055/s-0034-1385342.

[9]   D. Tang, F. Wei, B. Qin, T. Liu, and M. Zhou, "Coooolll: A Deep Learning System for Twitter Sentiment Classification," *8th Int. Work. Semant. Eval. SemEval 2014 - co-located with 25th Int. Conf. Comput. Linguist. COLING 2014, Proc.*, no. SemEval, pp. 208–212, 2014, doi: 10.3115/v1/s14-2033.

[10] D. Tang, F. Wei, N. Yang, M. Zhou, T. Liu, and B. Qin, "Learning Sentiment-Specific Word Embedding," *Acl*, pp. 1555–1565, 2014.

[11] S. Poria, E. Cambria, N. Howard, G. Bin Huang, and A. Hussain, "Fusing audio, visual and textual clues for sentiment analysis from multimodal content," *Neurocomputing*, vol. 174, pp. 50–59, Jan. 2016, doi: 10.1016/J.NEUCOM.2015.01.095.

[12] C. N. Dos Santos and M. Gatti, "Deep convolutional neural networks for sentiment analysis of short texts," *COLING 2014 - 25th Int. Conf. Comput. Linguist. Proc. COLING 2014 Tech. Pap.*, pp. 69–78, 2014.

[13] S. Poria, E. Cambria, and A. Gelbukh, "Aspect extraction for opinion mining with a deep convolutional neural network," *Knowledge-Based Syst.*, vol. 108, pp. 42–49, Sep. 2016, doi: 10.1016/J.KNOSYS.2016.06.009.

[14] S. Poria, E. Cambria, A. Gelbukh, F. Bisio, and A. Hussain, "Sentiment Data Flow Analysis by Means of Dynamic Linguistic Patterns," *IEEE Comput. Intell. Mag.*, vol. 10, no. 4, pp. 26–36, Nov. 2015, doi: 10.1109/MCI.2015.2471215.

[15] A. Casey *et al.*, "A systematic review of natural language processing applied to radiology reports," *BMC Med. Informatics Decis. Mak. 2021 211*, vol. 21, no. 1, pp. 1–18, Jun. 2021, doi: 10.1186/S12911-021-01533-7.

[16] K. Guo, L. Shi, W. Ye, and X. Li, "A survey of Internet public opinion mining," *PIC 2014 - Proc. 2014 IEEE Int. Conf. Prog. Informatics Comput.*, pp. 173–179, Dec. 2014, doi: 10.1109/PIC.2014.6972319.

[17] H. Liu and P. Singh, "Commonsense reasoning in and over natural language," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol.

3215, pp. 293–306, 2004, doi: 10.1007/978-3-540-30134-9_40.

[18]   M. A. Arshed, S. Mumtaz, M. S. Liaqat, and I. Haq, "LSTM Based Sentiment Analysis Model to Monitor COVID-19 Emotion LSTM Based Sentiment Analysis Model to Monitor COVID-19 Emotion," no. May, 2022.

[19]   M. A. Arshed, S. Mumtaz, O. Riaz, W. Sharif, and S. Abdullah, "A Deep Learning Framework for Multi-Drug Side Effects Prediction with Drug Chemical Substructure," *Int. J. Innov. Sci. Technol.*, vol. 4, no. 1, pp. 19–31, 2022.

[20]   M. T. Ubaid, A. Kiran, M. T. Raja, U. A. Asim, A. Darboe, and M. A. Arshed, "Automatic Helmet Detection using EfficientDet," *4th Int. Conf. Innov. Comput. ICIC 2021*, 2021, doi: 10.1109/ICIC53490.2021.9693093.

[21]   B. Liu *et al.*, "Comparison of Machine Learning Classifiers for Breast Cancer Diagnosis Based on Feature Selection," *Proc. - 2018 IEEE Int. Conf. Syst. Man, Cybern. SMC 2018*, pp. 4399–4404, Jan. 2019, doi: 10.1109/SMC.2018.00743.