Lexicon A Linguistic Approach for Sentiment Classification

Ankita Sharma
USICT
Guru Gobind Singh Indraprastha University
New Delhi, India
Ankita.15116490019@ipu.ac.in

Udayan Ghose
USICT
Guru Gobind Singh Indraprastha University
New Delhi, India
udayan@ipu.ac.in

***Abstract*— It is known that social media has become a part and parcel in everyone's life. Human emotions are constantly being expressed in real time on various social networking sites. The availability of an enormous amount of opinion rich data from various social networking sites has fueled interest in opinion mining and sentiment analysis. There are mainly two approaches for performing sentiment analysis that is a lexicon-based approach and a machine learning based approach. In this paper, we chose to limit our study to just the lexicon-based approach of sentiment analysis. Lexicon based approach relies on the lexicons for classifying input data. Lexicon is a set of words, idioms, phrases, etc. having a semantic meaning. In this paper, prior research done in lexicon-based sentiment analysis has been studied; Also, a review of some state-of-the-art lexicon-based solutions have been presented for polarity classification of Sentiment Analysis. This paper is mainly oriented towards the various lexicons used for polarity classification.**

Keywords—sentiment analysis, lexicons, social media, natural language processing, text mining

I. INTRODUCTION

With advancement of technology and disclosure of Web 3.0, more and more people in leaps and bounds have started expressing their sentiments, emotions, opinions over internet. With huge amount of opinioned data available on the internet, opinion mining has gained popularity [1]. Sentiment analysis has become hotspot research area these days. Opinion mining has routed its way into the contemporary world by helping the people to analyze the people's purview. The study of attitudes and sentiments of the people is called sentiment analysis meaning thereby the aim is to draw out the polarity from the text. Sentiment analysis is one instance of text mining and in a text mining work-flow, the same kind of work-flow is followed. Some input data is read, next comes the enrichment part some additional information is added to the input document in form of tags. Those added tags work as a sticker for each word, a dictionary for positive words and a list or dictionary for negative words is used. The tagged words in input document are segregated into positive and negative lists according to the dictionary. The words which did not have any tags are deleted. After performing the above steps, finished rules are obtained. Now we can actually evaluate how good our rule-based approach is [2]. We have chosen to confine our study in this paper to only lexicon-based technique for sentiment analysis. As lexicon-based approach is top choice when working with messy data, big data; also when having limited time and resources [3].

The rest of the paper is organized as follows: In Section 2, a walkthrough is provided about natural language processing and sentiment analysis. In Section 3, a discussion on its prior work is done. In Section 4, lexicon-based approach is discussed in detail. In Section 5, brief evaluation of some popular lexicons is done. Finally, in Section 6, we have concluded our work.

II. A WALKTHROUGH ABOUT NLP AND SENTIMENT ANALYSIS

A. Natural Language Processing (NLP)

It is known that humans are most advanced species on earth and our success as humans is due to our ability to convey and share information, the subfield of artificial intelligence called NLP is focused on enabling computers to understand, interpret, manipulate and communicate etc. in human language. In other words, NLP is an AI approach of communicating and connecting with an intelligent system using human like language [4]. Just 21% of the data produced all over the web is structured in nature, rest of all data is unstructured. Most of this data is textual data, which is amorphous, unstructured and difficult to deal with algorithmically. So, in order to produce relevant useful insights from textual data, it is important to get acquainted with text analysis. Text analysis/text mining and natural language processing go hand in hand. Text mining refers to process of obtaining most relevant high-quality meaningful information from text. The process involves structuring input data, then deriving patterns from it and finally evaluating obtained output. NLP is applicable in various fields like text generation, speech recognition, topic modeling etc. Sentimental analysis is one such field where NLP is heavily used [5].

B. The Need of Era Sentiment Analysis (SA)

It is evident that Science can be visualized as an image of society, contemporary world imposes coinage of new research areas. One such fastest growing research area under NLP where text mining is used is SA; which means analyzing emotions out of given data [6]. Communication is essential for people, as it helps them to fulfil their emotional needs. Lately, focus of communication has shifted to social networking. People in leaps and bounds have started expressing their opinions on various social networking sites. Day and night people keep on posting their experiences and opinions related to products, restaurants, places, persons, things, politics, social issues and much more [7]. People express each and every emotion on various social networking sites. To get the right directions and to take the best decisions manufactures, politicians, company owners etc., across world are keen to know opinions of the folks, therefore mining such opinions from the text that is SA thus becomes important.

C. Course of action involved in Sentiment Analysis

As it is mentioned that sentiment analysis is all about expressing opinions towards any subject, while it is known that opinion refers to the viewpoints, subject refers to the

2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence) | 978-1-6654-1451-7/20/$31.00 ©2021 IEEE | DOI: 10.1109/Confluence51648.2021.9377057

978-0-7381-3160-3/21/$31.00 ©2021 IEEE

target for an opinion and emotions implies the state of mind of a person [8].

- First step is to uncover the subject towards which feeling or emotion is directed.
- After identifying the subject, polarity of the emotion is calculated.
- Finally, the intensity of the sentiment words is calculated by assigning them the sentiment scores.

D. Brief overview of Lexicon Based approach (LBA) and Machine Learning approach (MLA)

It is known that SA is mainly performed using either MLA or LBA. Both the approaches have their pros and cons. MLA is all about train, test and then implement strategy [8]. In simple words, MLA needs a training data for training algorithm, in which inputs and their known outputs are provided. This is done before applying it to the actual data, so that they can work well with the new unknown data. For instance: "I am not excited", MLA will classify it as negative as the bigram "not excited" is negative as coded by humans. In spite of its numerous benefits it has got some drawbacks also, this method requires a large set of training data as well as testing data. Algorithm training is time consuming and is expensive method also, it requires massive resources to function. The performance of machine learning depends on how good is the match between the training and the testing data. Inadequate training or training with small set of data may produce biased or erroneous results [2][9].

From the literature review, it is found that MLA outperforms LBA in terms of accuracy and precision; in spite of that LBA methods are highly used [2]. They compete very well with MLA methods. LBA relies on the lexicons for classifying input data. As lexicons don't need training and testing, there are number of existing lexicons and researchers all over the world are creating new lexicons on the top of existing lexicons. This approach is widely used for the classification of the sentiments in input text using dictionaries that defines the opinion polarity of the words. This approach uses the language information like part of speech (POS), sentiment words and their polarity etc. to calculate the sentiment of the text. This approach is superior than the former mentioned approach in terms of pragmatic perspective [10]. E.g., "I am not excited", LBA will classify this as negative; as it has one positive word excited preceded by a negation word "not" that reverses its polarity. It is found that LBA is best to apply when working with big data [3]. These methods are very good at simulating the consequence of linguistic context. It is insensitive to amount and quality of data, meaning thereby quality and quantity doesn't affect its performance.

III. RELATED WORKS

The research done in LBA for Sentiment analysis have been studied and findings are presented below.

A lexicon of Sinhala Language using corpus-based approach is proposed. It was found that the made framework will be more precise with bigger text corpus and obtained accuracy is 69.23% [1].

A lexicon is proposed of Hindi language for movie and hotel domain. 5,200 reviews from movie and hotel domain were collected. It was found that there was an improvement in accuracy compared to existing Hindi SentiWordNet resource and obtained accuracy is 88% [5].

A Sentiment lexicon named CP-chunks is proposed. It aims at dealing with vagueness of lexical sentiments and is based on lexicon-based technique. LMRD and MRD dataset have been used. FCP-Lex obtained an accuracy of 82%. Also, it was found that the constructed lexicon is more effective in analyzing sentiments [6].

Author's tried to come up with a technique that can adequately perform big data sentiment analysis. A new LBA is introduced for drawing out sentiments using emoticons and hashtags. Proposed approach scales well to big data sets efficiently. Accuracy obtained is 73.5% [8].

The point in study was to develop an LBA for analysis of sentiments on faculty evaluation feedback given by students at end of course to deal with the open-ended questions based on uni-grams. 1,748 feedback were collected. Results show that score of sentiment is akin to Likert scale-based score. Accuracy obtained is 91.2% [15].

The point in study was to analyze sports related twitter communication using LBA. It is demonstrated that analyzing sentiments can proved to be good and effective and for sports content and it is deliberate to simulate discussion on Sentiment analysis in sports science. Accuracy obtained is 95% [16].

The point in study was to develop a WKWSCI Sentiment lexicon and compare its performance with the other five previously developed lexicon. This newly made lexicon has an accuracy of 69% [20].

Researchers have proposed an approach for finding word polarity using dictionary-based method and extraction of feature technique. Results implies that feature extraction perform better than dictionary-based methods. Accuracy obtained is 81% [25].

In this, a tool to evaluate Arabic social content was developed based on LBA approach and POS tagging. Performance evaluation shows that tool gives more exact results, when applied to reviews which are Arabic and are domain based. 93.9% accuracy was obtained using this tool [32].

Emotion Analysis Platform (EAP) is proposed to be applied on Weibo to monitor the emotional pulse of China using fine grain emotions technique. 35,000 tweets of Sichuan earthquake were collected. EAP is efficiently applied to capture the emotional pulse from dimension of time and space. EAP obtained 80% of accuracy [33].

In this paper, Sentiment Analysis of online paper (SAOOP) is proposed to help researchers analyze online paper reviews using enhancement BOW model. Obtained results show that SAOOP is good for selecting efficient papers, it also evaluates topic domain parameters of scientific papers to evaluate total score of paper, also it obtained an accuracy of 82.5% [34].

IV. DELVING INTO LEXICON BASED METHOD FOR SENTIMENT ANALYSIS

In LBA, prior term polarity is extracted from lexical resources and summation of such polarities is done using NLP and linguistic rules are used to calculate sentiment from the given text [11][12].

A. There are two assumptions related to sentiment-based polarity classification:

- Prior polarity of terms is assigned independent of context.
- Prior polarity is expressed using numerical value.

With respect to these assumptions list of words can be generated along with their prior polarities, these lists of words are known as sentiment lexicons. SA using lexicon-based methods starts with creation of a lexicon or using some pre-existing lexicons or combination of both to find sentiment polarity in a piece of text. As we may appreciate that manually creating and validating a lexicon is not very easy, it is fairly a time-consuming task and that is the reason that there are just a handful of popular lexicons and most of the people use them or a combination of them [9][13]. In other words, lexicons can be said as linguistic approach that associate terms with its sentiment polarity like positive, negative or neutral polarity and their numerical scores indicate their sentiment strength. Majority of the lexicons are general purpose lexicons and are not tailored for any particular application; therefore they can be applied in diverse domains.

Given below is a framework showing general steps followed for lexicon-based sentiment analysis refer figure.1. First and the foremost task is to choose the lexicon, it may be a pre-existing lexicon or a newly made lexicon and then a dataset/document is inserted, preprocessing and enrichment is performed on it. Prior polarity is associated with each term in the document using sentiment lexicon that comes under subjectivity identification. Sentiment score calculation is done via prior polarities and it is tailored to give back polarities and aggregated to prognosticate polarity decision that is sentiment class (positive, negative or neutral).

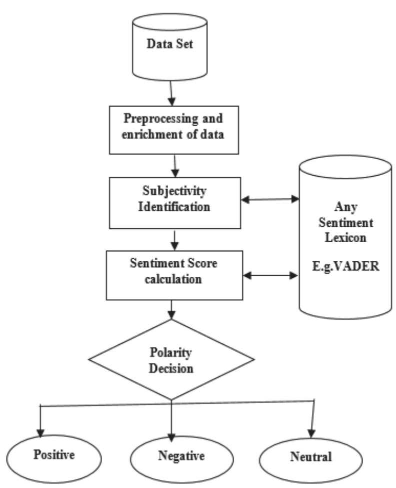

Fig. 1. Framework of Sentiment Analysis using lexicon-based approach.

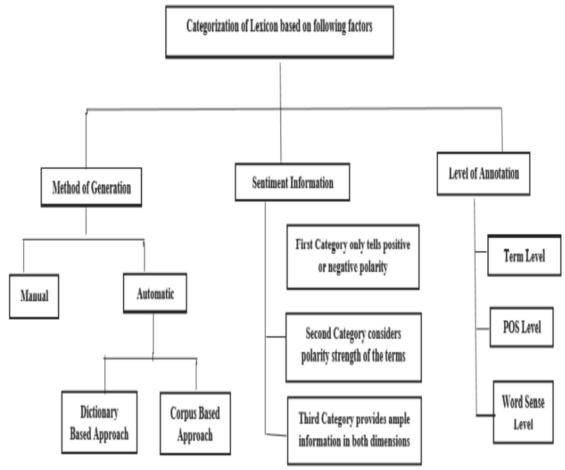

Fig. 2. Broad Categorization of Lexicon

B. Method of generation of Sentiment Lexicons

There are three methods for sentiment lexicon generation. One of them is manual and other two are automatic refer figure.2. For manually making lexicons, we require social scientists, psychologists, a group of experts etc. for manually collecting opinion words, they do so on basis of their language understanding and domain knowledge. Though manually created lexicons are more accurate but this process is very time consuming, also they have relatively lower term coverage. It is always a good idea to combine this approach with automated approach. Examples: opinion lexicon, GI etc. Automated approaches are Corpus based approach (CBA) and Dictionary based approach (DBA).

In DBA some popular dictionary is used for instance wordnet, it starts with some small set of sentiment words. Then it is enhanced by finding opposite meaning words, similar meaning words and glosses from dictionary. It is an iterative process; we keep on adding new words. The obtained dictionary will be more general purpose not so domain specific, meaning thereby DBA is not very good at identifying context specific sentiment. Examples: SentiWordNet, SenticNet, WordNet Affect etc. [2].

Second approach of automated method is CBA. To generate sentiment lexicons some sort of co-occurrence relations is used to find the sentiment polarities of the term. It is a kind of fill in the gaps that DBA cannot do well so they can be tailored to some domains, meaning thereby they remove drawback of DBA's inability to handle context specific classification of the opinion words. This approach uses some sort of labeled data. A seed set of opinion words with known polarity is used. The syntactic pattern is exploited to find new opinion words, the polarity of the opinion words in the corpus. DBA is more efficient than CBA, as CBA alone cannot identify all sentiment words because having and maintaining a corpus containing all sentiment words is a tedious job [14].

C. Categorization of the lexicons based on the amount of information given by them:

- The first category lexicons only tell positive or negative polarity information about terms, polarity strength of the terms is not taken into the consideration. E.g., good and best would be treated equal in sentiment as they both have positive valence. Example: MPQA subjectivity lexicon, GI, Opinion lexicon [15].
- Second category is that which considers polarity strength of the terms, example by using Likert scale from 1 to 5 in either positive or negative dimension. Prominent drawback is when term is unable to reach a maximum score then it often becomes ambiguous to decide whether the left-over score implies objectivity or opposite polarity. Example: SentiStrength.
- Third category of the lexicon is the category which provides ample information about sentiment in both positive and negative dimensions. Example SentiWordNet.

D. Level of annotation in Sentiment Lexicons

These are the linguistic properties that affects the assignment of the scores in lexicons. POS tags refers to grammatical type of word. For instance: verb, noun, adjective, adverb, article etc. POS indicates how a word functions in meaning as well as grammatically within the sentence. A word can have multiple POS based on context or the word sense in which it is used. There are three levels of viewing the sentiment lexicons refer figure2.:

- Term level
- POS level
- Word Sense level

Polarities are associated with terms in term level [16]. This level is not good as the other two levels as polarities can change on basis of POS or word sense [6]. While in POS level, lexicon annotation is determined at grammatical level [5]. At word sense level annotations are determined based on sense of the words.

E. Prominent drawbacks of Lexicon Based Approach

It is also important to highlight some of the drawbacks of lexicon-based approach: [17][18][35].

- The first drawback is that they are unable to process acronyms, initialism, emoticons, slangs etc. These are the things that are part and parcel of our everyday conversation. Lexicon based approach is unable to process this kind of colloquial language.
- Most of the lexicon does not consider the kind of factor of intensity of words. For instance: sentiment lexicons are unable to account for the sentiment intensity example good and exceptional have a very similar score.
- Lexicons are unable to deal with sarcasm. Processing sarcasm require a pretty high level of intelligence and LBA is not good at sarcasm processing.
- When using lexicon for SA, it is required that lexicon is maintained up to date, creating, updating and maintaining the lexicons requires a lot of time. If the lexicons are not maintained up to date then performance of the lexicons gradually drop.
- Question type of sentences do contain sentiment words, but they don't express any sentiment. For example: Who is mad? Does not implies any sentiment. Dealing with such sentences becomes tedious.
- Sentences not containing sentiment words can also imply sentiment. At times, words do express some sentiments but are not present in the lexicons. E.g., Stock market is like roller coaster these days.
- It is known that the domain greatly influences the score which is assigned to the term. A word may have opposite orientations in different domains, moreover not every domain specific term is present in the general lexicon leading to wrong calculation of the terms. Sentiment can be varied by domains; for instance: long can imply positive sentiment in product reviews like battery lives long. On the other hand, Queue for parcel is long, here long implies the negative sentiment.

V. EVALUATION OF SOME POPULAR SENTIMENT LEXICONS

This section provides an overview of some widely used sentiment lexicons.

A. Opinion Lexicon (Hu and Liu,2004)

- Hu and Liu developed a general-purpose sentiment lexicon, which classifies the English words into two categories that is positive words and negative words [36].
- It contains nearly 6,800 words out of which about 2,006 words classified as positive and 4,783 words as negative respectively [18].
- This polarity-based lexicon has evolved over several years, based on customer reviews from different domains [19].
- There is no part of speech tagging in this lexicon.
- Though it is applicable to social media text and product review. But lack of initialisms, emoticons, acronyms decreases its performance for sentiment analysis.

B. LIWC (Linguistic Inquiry and Word Count)

- LIWC pronounced as "Luke", classifies words into psychologically meaningful categories.
- It is a high quality, comprehensive lexicon and a text analysis program mainly used in psychology for inspecting the psychological perspective [18].
- The Sociologists, linguists and psychologists have been working with LIWC.
- It is an extensively validated lexicon, therefore people from different domains find it appealing.
- Its simple word lists and easy to understand dictionary make it an attractive option to use to extract emotional polarity from text; also it is easy to understand, inspect and extend if needed.
- This lexicon has eighty language categories and it comes under semantic orientation (polarity-based) lexicons.
- It is known that language plays a chief role in understanding a person's state of mind. It was found that LIWC is very good at identifying emotion in the language use.
- In spite of the benefits, this lexicon has some limitations also. Initialisms, emoticons, slang words,

acronyms which are very important for opinion mining are ignored.

- Also this lexicon is unable to distinguish between sentiment intensity of the words, for instance: "This book is good" and "This book is amazing" both sentences will be assigned equal score by LIWC, as both of them contains one positive word each.

C. General Inquirer (GI)

- GI is a word-affect association, having a flavor of categories.
- By word-affect it is meant that a word is associated with affects or some words do have multiple sense.
- GI is used as a gold standard and there are about 11,896 entries with about 190 attributes related to sentiments and affects [18].
- This lexicon has two valence categories: About 1,915 words are of positive outlook and 2,291 words are of negative outlook [20].
- GI is one of the oldest manually created lexicon and is used by various scientists, researchers for objectively finding particular characteristics in messages like sentiment properties of text [21].
- Its two prime sources for tag categories is Harvard IV-4 dictionary and the Lasswell value dictionary [22][23]. Recently newly constructed categories are also added like "marker" categories for disambiguation.
- GI has no limit on the number of categories it can deal with, also each category is assigned a unique name. Name of the category is case sensitive; but has no restrictions on the length of name.
- To distinguish the Harvard dictionary category from similar Lasswell dictionary category. '@' sign is appended to represent the former category and "Lw" is appended for the latter category. Other than this marker categories are represented in upper case [24].
- Mainly GI finds its application in social science content analysis, but can also be applied to other domain like sentiment analysis also.
- The absence of sentiment relevant lexical features, also inability to differentiate among the intensity of opinion bearing words taint its performance.

D. VADER (Valence Aware Dictionary for sEntiment Reasoning)

- VADER is a gold standard lexicon for English Sentiment classification and has been validated by human beings [18].
- It is based on rules with the mélange of qualitative and quantitative methods specifically attuned for sentiment analysis in tweet like data [25].
- This lexicon is large in size, quick, simple, easily understood, parsimonious and applied without extensive training, also it is easy to modify and extend if needed without compromising accuracy.
- It is applied extensively in social media domain, but it can be generalized and performs well in other domains as well.
- The accessibility of the rules and lexicon used by VADER makes linguists, psychologists, sociologists and researchers in various domains use VADER. Its horizons are not limited to only computer science community.
- It not only gives the polarity score but also describes how positive or negative express opinion in input data.
- It also considers the boosting words, if boosting words like very, extremely etc. are present in the input text then its valence is increased. Also, valence is boosted by the presence of capital words, phrases, idioms etc. [26].
- Range of normalized valence is between -1 and 1. The valence less than 0 and -1 is for negative, valence 0 is for neutral and valence greater than 0 and up to 1 is for positive polarity.

E. SentiStrength

- SentiStrength has an extended sentiment dictionary covering a variety of data coming from social websites such as Myspace, BBC, Digg and Runners World etc.
- It has been trained on more than 4,200 tweet text and around 3,400 texts from YouTube. The number of sentiment terms was increased from 890 to 2489 in current version [27].
- It has been being tested on six different social media data sets; hence it is quite robust and an accurate in sentiment classification. It can detect positive and negative sentiment strength in microblogs easily and has a near human accuracy.
- The output score provided by the SentiStrength is in the range of 1 to 5, whereas 1 and -1 represents weakest positive and the weakest negative emotion respectively. Whereas 5 and -5 corresponds to the strongest emotion both positive and negative respectively [28].
- SentiStrength works well with short text like microblog like content, but is not good at dealing sarcasm.

F. SenticNet

- SenticNet is a popular publicly available semantic lexicon, mainly used for aspect level SA [4].
- This lexicon is based on semantic web techniques and artificial intelligence.
- SenticNet 5 is its latest version and successor of SenticNet 4.
- Using this SA is performed on both syntactic and semantic level, also it uses dimensionality reduction concept to find common sense concepts polarity [18].
- It is much more precise than SentiWordNet.
- It has around fourteen thousand two hundred twenty-two common sense concepts along with the information related to polarity, sentimental and semantic associations.
- SenticNet 5 has three levels of semantic networks that is primitive level, concept level and the entity level. In first level, actions and basic statics are defined with help of primitives. While level two deals with interconnection of common-sense concepts via semantic relationships and level three is comprised of named entities that is connected to common sense concepts [29].

- Also, SenticNet 5 does not blindly follow word co-occurrence and keywords, instead it depends on implicit meaning of common-sense concepts.

G. SENTIWORDNET 3.0

- SentiWordNet is an affective lexicon resource made on basis of WordNet in English [18].
- It is one of the widely used lexical resource and is made up of synsets i.e., sets of synonyms. Every synsets has a score in range 0 to 1 and total is 1.0 for every synsets [30].
- Positive score is for positive valence and negative score is for negative valence. Scores are assigned using classifiers trained on three WordNet subsets i.e., positive, negative and neutral synsets [25].
- This lexicon is not a gold standard resource, it is very noisy in nature and even many of synsets have no positive, negative polarity. It also lacks sentiment carrying lexical features.
- But this lexicon is interesting as it gives positive and negative polarities of terms for different senses and at a deeper level word sense [31].
- SentiWordNet 3.0 is successor of SentiWordNet 1.0 and is more accurate than first version.
- SentiWordNet 3.0 annotates WordNet 3.0 and uses random walk process for correctly classifying scores[6].

TABLE I. A CONCISE SUMMATION OF SURVEY OF WORK DONE ON LEXICONS

Author and Year	**Technique Used**	**Domain Oriented**	**Lexicon Type Used and Data Set**
Chetan and Atul (2014)	Lexicon Based Technique	Yes	6,74,412 tweets 560 Chinese Review
Mohammed et al. (2014)	Lexicon Based Approach based on POS tagging	Yes	Domain Oriented Lexicon Dataset of Chat
Duyu et al. (2014)	Fine grain emotions	Yes	Chinese lexicon 35,000 tweets of Sichuan Earthquake
Doaa et al. (2015)	Enhancement BOW model	Yes	New Lexicon Three Datasets of around 11,500 tweets
Quratulain et al. (2016)	Uni grams-based lexicon-based technique	Yes	MPQA 1,748 Feedback reviews by students
Deepali et al. (2016)	Context Oriented Lexicon	Yes	Hindi SentiWordNet 5,200 reviews from movie and hotel domain
Khoo et al. (2018)	Lexicon Based Technique	No	WKWSCI Lexicon Amazon Product Review
Deepa et al. (2019)	Machine Learning and Lexicon based technique	Yes	SentiWordNet, VADER Twitter US Airline dataset
Chathuranga et al. (2019)	Machine Learning and Lexicon based technique	No	SentiWordNet 3.0 Sinhala News Corpus
Yin et al. (2020)	Lexicon Based Technique	Yes	Lexicon FCP – Lex LMRD and MRD dataset
Wunderlich and Daniel (2020)	Lexicon Based Approach	Yes	LIWC, QDAP dictionary, SenticNet4 Lexicon 10,000 tweets

VI. CONCLUSION

The availability of enormous amount of opinion rich data from various social media platforms have fueled interest in opinion mining and sentiment analysis. It is known that sentiment analysis is performed mainly using machine learning approach and lexicon-based approach. The point of study in this paper is around lexicon-based method for analyzing sentiments. It is a linguistic approach in which prior term polarity is extracted from lexical resources and summation of such polarities is done using NLP and linguistic rules are used to calculate sentiment from the given text. Lexicons can be categorized using number of factors. In this paper lexicons are categorized based on three factors i.e. method of generation of sentiment lexicons, sentiment information (polarity, strength or both) and level of annotation in sentiment lexicons. A survey on prior research done in lexicon-based sentiment analysis have been studied in this paper. Also, this paper has attempted to bring out an overview of some lexicon-based solutions for polarity classification for analyzing sentiments. It would be promising to further explore our wings in this field by applying these lexicons on datasets and also the expansion of the systematic literature review circle larger with futuristic research unendingly.

REFERENCES

[1] Chathuranga, P. D. T., S. A. S. Lorensuhewa, and M. A. L. Kalyani, "Sinhala Sentiment Analysis using Corpus based Sentiment Lexicon", in International Conference on Advances in ICT for Emerging Regions (ICTer), Sri Lanka, 2019, Vol. 250, pp. 1-7.

[2] Hailong, Zhang, Gan Wenyan, and Jiang Bo, "Machine learning and lexicon-based methods for sentiment classification: A survey", in 11th Web Information System and Application Conference, China, 2014, pp. 262-265.

[3] Akter, Sanjida, and Muhammad Tareq Aziz, "Sentiment analysis on facebook group using lexicon-based approach", in 3rd International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), Bangladesh, 2016, pp. 1-4.

[4] Hardeniya, Tanvi, and Dilipkumar A Borikar, "Dictionary based approach to sentiment analysis-a review", International Journal of Advanced Engineering, Management and Science (IJAEMS), 2016.

[5] Mishra, Deepali, Manju Venugopalan, and Deepa Gupta, "Context specific Lexicon for Hindi reviews", in Procedia Computer Science 93, India, 2016, pp.554-563.

[6] Yin, Fulian, Yanyan Wang, Jianbo Liu, and Lisha Lin, "The Construction of Sentiment Lexicon Based on Context-Dependent Part-of-Speech Chunks for Semantic Disambiguation", 2020, pp.63359-63367.

[7] Rezapour, Rezvaneh, Lufan Wang, Omid Abdar, and Jana Diesner, "Identifying the overlap between election result and candidates' ranking based on hashtag-enhanced, lexicon-based sentiment analysis", in IEEE 11th International Conference on Semantic Computing (ICSC), USA, 2017, pp. 93-96.

[8] Kaushik, Chetan, and Atul Mishra, "A scalable, lexicon based technique for sentiment analysis", in International Journal in Foundations of Computer Science & Technology (IJFCST), 2014, arXiv preprint arXiv:1410.2265.

[9] Sahu, Tirath Prasad, and Sarang Khandekar, "A Machine Learning-Based Lexicon Approach for Sentiment Analysis", in International Journal of Technology and Human Interaction (IJTHI), 2020, pp.8-22.

[10] Mudinas, Andrius, Dell Zhang, and Mark Levene, "Combining lexicon and learning based approaches for concept-level sentiment analysis", in Proceedings of the first international workshop on issues of sentiment discovery and opinion mining, 2012, pp. 1-8.

[11] Mashuri, Mochamad, "Sentiment Analysis In Twitter Using Lexicon Based and Polarity Multiplication", in International Conference of Artificial Intelligence and Information Technology (ICAIIT), 2019, pp. 365-368.

[12] Kamal, Rashid, Munam Ali Shah, Carsten Maple, Mohsin Masood, Abdul Wahid, and Amjad Mehmood, "Emotion classification and crowd source sensing; a lexicon based approach", in IEEE Access 7, 2019, pp.27124-27134.

[13] Drus, Zulfadzli, and Haliyana Khalid, "Sentiment Analysis in Social Media and Its Application: Systematic Literature Review", in Procedia Computer Science 161, 2019, pp.707-714.

[14] Verma, Binita, and Ramjeevan Singh Thakur, "Sentiment analysis using lexicon and machine learning-based approaches: A survey", in Proceedings of international conference on recent advancement on computer and communication, Singapore, 2018, pp. 441-447.

[15] Rajput, Quratulain, Sajjad Haider, and Sayeed Ghani, "Lexicon-based sentiment analysis of teachers' evaluation", in the Applied Computational Intelligence and Soft Computing, 2016.

[16] Wunderlich, Fabian, and Daniel Memmert, "Innovative Approaches in Sports Science—Lexicon-Based Sentiment Analysis as a Tool to Analyze Sports-Related Twitter Communication" in Applied Sciences, 10(2), 2020, p.431.

[17] John, Annet, Anice John, and Reshma Sheik, "Context deployed sentiment analysis using hybrid lexicon", in 1st International Conference on Innovations in Information and Communication Technology (ICIICT), India, 2019, pp. 1-5.

[18] Hutto, Clayton J., and Eric Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text", in Eighth international AAAI conference on weblogs and social media (ICWSM-14), 2014, vol. 81, p. 82.

[19] Ding, Xiaowen, Bing Liu, and Philip S. Yu," A holistic lexicon-based approach to opinion mining", in Proceedings of the international conference on web search and data mining, 2008, pp. 231-240.

[20] Khoo, Christopher SG, and Sathik Basha Johnkhan, "Lexicon-based sentiment analysis: Comparative evaluation of six sentiment lexicons", in Journal of Information Science 44.4, 2018, pp.491-511.

[21] Lorge, Irving, and Edward L. Thorndike, "The teacher's word book of 30 000 words", in Columbia University Press, 1963.

[22] Dunphy, Dexter Colboyd, Cedric G. Bullard, and Elinor EM Crossing, "Validation of the General Inquirer Harvard IV Dictionary", in Harvard University Library, 1974.

[23] Lasswell HD and Namenwirth JZ, "The Lasswell value dictionary", vols 1–3, Yale University, 1968.

[24] Karadeniz, Talha, and Erdogan Dogdu, "Improvement of General Inquirer Features with Quantity Analysis", in IEEE International Conference on Big Data (Big Data), 2018, pp. 2228-2231.

[25] Deepa, D., and A. Tamilarasi, "Sentiment Analysis using Feature Extraction and Dictionary-Based Approaches", in Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC), India, 2019, pp. 786-790.

[26] Amin, Al, Imran Hossain, Aysha Akther, and Kazi Masudul Alam, "Bengali vader: A sentiment analysis approach using modified vader", in International Conference on Electrical, Computer and Communication Engineering (ECCE), Bangladesh, 2019, pp. 1-6.

[27] Thelwall, Mike, "The Heart and soul of the web? Sentiment strength detection in the social web with SentiStrength", in Cyberemotions, 2017, pp. 119-134.

[28] Islam, Md Rakibul, and Minhaz F. Zibran, "SentiStrength-SE: Exploiting domain specificity for improved sentiment analysis in software engineering text", in Journal of Systems and Software 145, 2018, pp.125-146.

[29] Cambria, Erik, Soujanya Poria, Devamanyu Hazarika, and Kenneth Kwok, "SenticNet 5: Discovering conceptual primitives for sentiment analysis by means of context embeddings", in Thirty-Second AAAI Conference on Artificial Intelligence, 2018.

[30] Baccianella, Stefano, Andrea Esuli, and Fabrizio Sebastiani, "Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining", in Lrec, vol. 10, 2010, pp. 2200-2204.

[31] Gatti, Lorenzo, Marco Guerini, and Marco Turchi, "SentiWords: Deriving a high precision and high coverage lexicon for sentiment analysis", in IEEE Transactions on Affective Computing 7.4, 2015, pp.409-421.

[32] Al-Kabi, Mohammed N., Amal H. Gigieh, Izzat M. Alsmadi, Heider A. Wahsheh, and Mohamad M. Haidar, "Opinion mining and analysis for Arabic language", in International Journal of Advanced Computer Science and Applications 5 (IJACSA), 2014, pp.181-195.

[33] Tang, Duyu, Bing Qin, Ting Liu, and Qiuhui Shi, "Emotion analysis platform on chinese microblog", 2014, arXiv preprint arXiv:1403.7335.

[34] El-Din, Doaa Mohey, Hoda MO Mokhtar, and Osama Ismael, "Online paper review analysis", in International Journal of Advanced Computer Science and Applications (IJACSA), 2015, pp.220-229.

[35] Aung, Khin Zezawar, and Nyein Nyein Myo, "Lexicon Based Sentiment Analysis of Open-Ended Students' Feedback", in International Journal of Engineering and Advanced Technology (IJEAT), 2018.

[36] Ankita Sharma, and Udayan Ghose, "Sentimental analysis of twitter data with respect to general elections in India", in Procedia Computer Science 173, India, 2020, pp. 325-334.