JOURNAL OF CRITICAL REVIEWS

ISSN- 2394-5125 VOL 7, ISSUE 15, 2020

Comparative Analysis of Emotion Mining Techniques

[1]Rishu Gupta [2]Dr.Williamjeet Singh

[1] Research Scholar, Department of Computer Science and Engineering, Punjabi University, Patiala

[2] Assistant Professor, Department of Computer Science and Engineering, Punjabi University, Patiala

ABSTRACT

Emotion is both prevalent in and essential to all aspects of our lives. It influences our decision-making, affects our social relationships, and shapes our daily behaviour. With the rapid growth of emotion-rich textual content, such as micro blog posts, blog posts, and forum discussions, rhymes etc, there is a great need to develop automatic tools for identifying and analyzing people's emotions expressed in text. In this study we have compared different emotion mining techniques, their working principles, advantages and disadvantages. Different application domains discussed indicates that emotion mining is being used in almost every aspect of day to day life, like it is playing a prominent role in Politics, tourism, media, sports etc. The study also focuses on the challenges related to emotion mining.

Keywords:

Application domain, Challenges, Emotion mining, Lexicon Approach, Machine Learning Approach, Techniques.

1. INTRODUCTION

Emotions are a binding part of human nature that can be considered as genetic. Also it has been found that expression of a particular emotion by different human being is identical. Some unrelenting emotions that last much longer result in mood. Mood can be a result of a combination of certain emotions of a person. On the whole emotions can be categorised into two: basic and complex.[66]

1.1 Type of Emotions

Type 1: Basic emotions are joy, sadness, anger, fear, disgust and surprise as defined by Ekman [22], [23], [24].

Type 2: The complex emotions are a combination of two or more basic emotions that are experienced by a person at an instance. Even though emotions have no clear boundaries but Plutchik [53], [54] proposes a theory with eight basic emotions which include Ekman's six along with trust and anticipation. Plutchik has organized the emotions in a wheel (Figure 1). The radius indicates intensity, closer to the centre, the higher the intensity. Plutchik [54] also proposed that the eight basic emotions form four contrasting pairs like joy & sadness, anger & fear, trust & disgust, and anticipation & surprise.

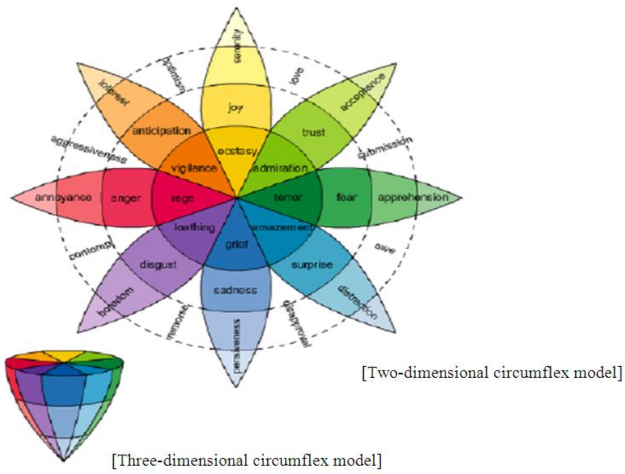


Figure 1: Plutchik Wheel of Emotions [54]

There is a scientific proof that atleast five different emotions (fear, disgust, anger, happiness, sadness) are noticeably different in logic of triggering different combination of brain parts. Ekman's evidence found in support of Emotions being a set of six general characteristics common to all basic emotions. From the human felt experience, it seems that there are two fundamental dimensions besides neurological levels. First, the demeanor of experienced emotion is the degree to which it is strongly positive or negative. Second is the level of stimulation felt that is the amount of energy alleged.

ISSN- 2394-5125 VOL 7, ISSUE 15, 2020

1.2 Motivation

This study is carried out in order to explore the existing emotion mining techniques and to know which are more relevant. This analysis is compulsory to make it possible to know which emotion mining techniques have been covered in past studies and helps to find the gaps.

This study aims at systematically reviewing the emotion mining techniques used in the existing studies. The result may help the researchers to get an overview of the status of emotion mining and highlights the research gaps.

1.3 Emotion categorisation by researchers

Basic Emotion Categories Identified by Researchers is depicted in Table 1.

Table 1: Basic Emotion Categories Identified by Researcher

Tomkins (1962)	**Izard (1977)**	**Plutchik (1980) [53],[54]**	**Ortonyet.al. (1988)**	**Ekman (1992) [22],[23],[24]**
joy anguish fear anger disgust surprise interest shame	enjoyment sadness fear anger disgust surprise interest shame shyness guilt	joy sorrow fear anger disgust surprise acceptance anticipation	joy sadness fear anger disgust surprise	happiness sadness fear anger disgust surprise

1.4 Challenges

1.4.1. Keyword Selection

Topic based classification usually uses a set of keywords to classify texts in different classes. In Emotion mining, text is classified into two classes (positive and negative) which are so different from each other. But coming up with a right set of keyword is not a easy task. This is because emotions can often be expressed in a subtle manner making it tricky to be identified when a sentence or document is considered in seclusion. [68]

1.4.2. Emotion is Domain Specific

Emotions are often domain specific and the meaning of emotion varies from sentence to sentence as change depends on the context they are used in.

1.4.3. Multiple Emotions in a Sentence

Single sentence can contain multiple emotions along with subjective and accurate portions. It is helpful to segregate such clauses. It is also important to estimate the strength of emotions in these clauses so that we can find the overall mood of the sentence.

1.4.4. Negation Handling

Handling negation can be tricky in Emotion Mining. For example, I like this car and I don't like this car differ from each other by only one token but consequently are to be assigned to different and opposite classes. Negation words are called polarity reversers.

2. RESEARCH METHODS

2.1 Planning the Research

To conduct the review on the emotion mining techniques we firstly collected the several research papers based on emotion mining from existing authors. In each paper, we have evaluated different types of emotions and each paper has different kind of dataset with different techniques implemented on the data set to evaluate the results.

2.2. Sources of Information

For an extensive and broad coverage of the literature, a broad perspective is necessary. To increase the probability of highly relevant articles, a set of appropriate databases must be chosen. In this study following electronic databases are used to search relevant studies:

• IEEE eXplore (http://ieeexplore.ieee.org)
• ACM Digital Library (www.acm.org/dl)
• Science Direct (www.sciencedirect.com)
• Springer (www.springerlink.com)
•Wiley Interscience (www3.interscience.wiley.org)

Table2 shows the defined search strategy and number of results retrieved

Table 2: Search selection criteria

S.No	E-Resources	Studies Returned	Excluded			Keywords used
			Based on Title	Based on Abstract	Based on full text	
1	www.ieeexplore.ieee.org	59	18	18	5	Emotion mining, techniques
2	www.acm.org	48	18	15	5	Emotion mining, techniques
3	www.sciencedirect.com	8	1	2	2	Emotion mining, techniques
4	www.springerlink.com	41	23	11	4	Emotion mining, techniques
5	www.interscience.wiley.com	7	2	1	1	Emotion mining, techniques

2.3. Study Selection

The search engines in the above electronic databases hits number of studies, articles. In order to make the search goal oriented and removing the irrelevant papers, the search is limited to keywords, title and abstract. We tried to extract as much relevant literature as possible to ensure the completeness of study. The result shows the distribution of relevant papers over the studied years from 2009 to 2019 as shown in Table3.

Table 3: Distribution of papers over years

Resources \ Years	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
www.ieeexplore.ieee.org	2	0	2	4	2	3	2	5	5	8	25
www.acm.org	0	2	2	0	2	1	0	1	2	1	2
www.sciencedirect.com	2	2	5	2	2	2	3	5	2	0	0
www.springerlink.com	1	1	1	1	1	0	2	0	2	1	2
www.interscience.wiley.com	1	1	0	0	1	0	2	1	2	1	2

3. CLASSIFICATION

Emotion mining techniques can be classified as shown in figure 2.

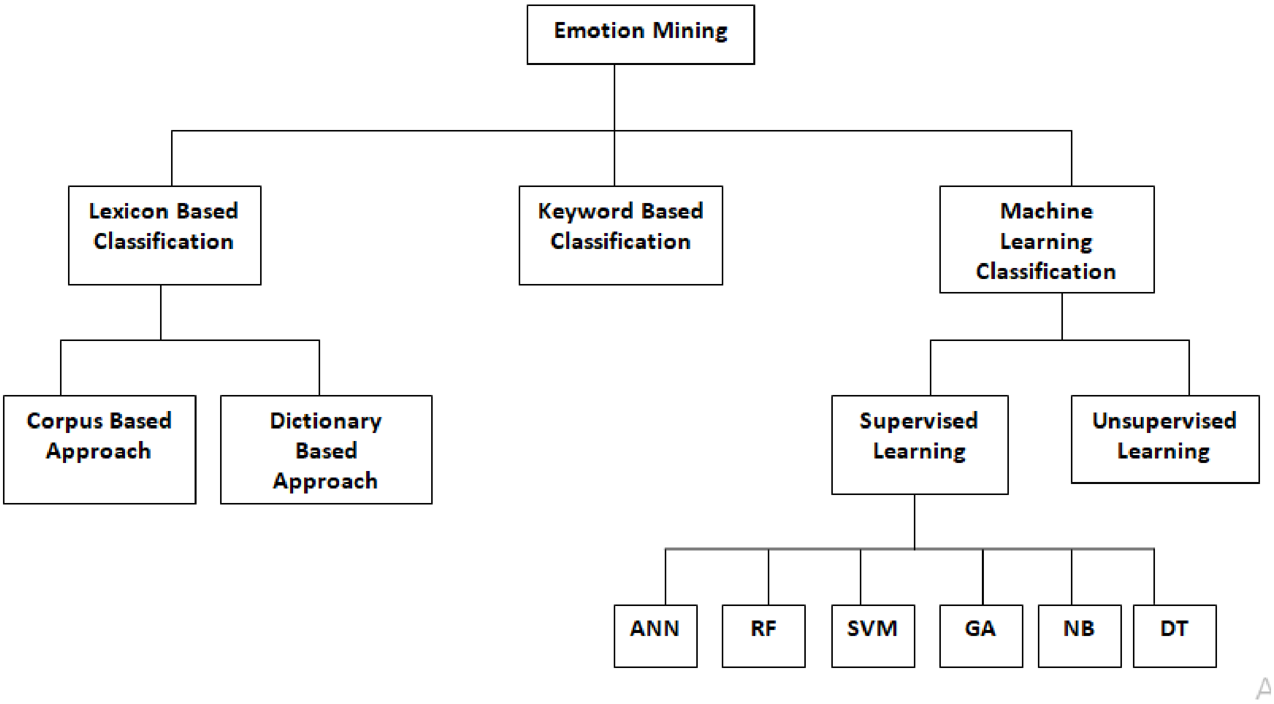
3.1 Keyword-based classification

This method classifies text based on the presence of positive or negative polarity words such as happy, joyful, delighted, miserable, sad, terrified, and uninterested [17]. The main drawback of keyword-based classification is the inability to steadfastly classify the negated words and polarity, as this approach depends on surface features [17]. Another drawback is that this approach is based on the obvious presence of positive or negative polarity. The good part is it is very simple and is user friendly.[68]

As was observed in [13] keyword-based emotion detection methods have three limitations described below.

3.1.1 Ambiguity in Keyword Definitions

Though using emotion keywords is a straightforward way to detect associated emotions, the meanings of keywords could be multiple and vague. Most words could change their meanings according to different usages and contexts. Moreover, even the minimum set of emotion labels (without all their synonyms) could have different emotions in some extreme cases such as ironic or cynical sentences.[46]



Where

ANN: Artificial Neural Network, RF: Random Forest, SVM: Support Vector Machine, GA: Genetic Algorithm, NB: Naïve Bayes, DT: Decision Tree

Figure 2: Classification of Emotion Mining Techniques

3.1.2 Incapability of Recognizing Sentences without Keywords

Keyword-based approach is totally based on the set of emotion keywords. Therefore, sentences without any keywords would imply they do not contain any emotions at all, which is obviously wrong. For example, "I passed my qualify exam today" and "Hooray! I passed my qualify exam today" should imply the same emotion (joy), but the former without "hooray" could remain undetected if "hooray" is the only keyword to detect this emotion.[46]

3.1.3 Lack of Linguistic Information

Syntax structures and semantics also have influences on expressed emotions. For example, "I laughed at him" and "He laughed at me" would suggest different emotions from the first person's perspective. As a result, ignoring linguistic information also poses a problem to keyword-based methods. In summary, keyword-based methods should also detect not only the existence of keywords, but also their linguistic information to detect emotions more accurately.[46]

3.2 Lexicon-based classification

Lexicon-based approaches construct lists of words manually labelled as having positive and negative polarity, and a polarity score for each word is created. This constructed lexicon is used to calculate the overall emotion score of a given text. The notable advantage of the lexicon based method is that these methods do not need training data (as the supervised machine-learning method does). The lexicon-based method is widely used in conservative text like reviews, forums, rhymes and blogs [45], [28]. However, they are less likely to be used for big data extracted from social media websites [28], [76]. The key reason is the unstructured format and nature of social media websites (the data contains textual peculiarities, informal and dynamic nature of language, new slang, abbreviations, and new expressions) [28],[76]. Even though this approach outperforms the keyword-based classification, it still has drawbacks.

3.2.1 Corpus Based Approach

Corpus based suggests data-driven approach [39] where there is access not only to emotional labels, but to a context which can be used to in an ML algorithm. It can certainly be a rule-based approach with NLP parsing too, or even a combination. It derives a set of abstract rules that govern a natural language from texts in that language, and explores how that language relates to other languages. Corpus also carries some domain specificity that can inform the algorithm of sentiment label variety for a word depending on its context / domain. In this the set of seed words is expanded by using a large corpus of documents from a single domain.

3.2.2 Dictionary based approach

ISSN- 2394-5125 VOL 7, ISSUE 15, 2020

Dictionary-based sentiment analysis is a computational approach to measuring the feeling that a text conveys to the reader. In the simplest case, sentiment has a binary classification: positive or negative, but it can be extended to multiple dimensions such as fear, sadness, anger, joy, etc. This method relies heavily on a pre-defined list (or dictionary) of Emotion-loaded words. In this the set of seed words is expanded by utilizing resources like WordNet. [73],[79]

Table 4: Comparison of various Lexicon Based Approach

Name of Lexicon Approach	Working Principle	Advantages	Disadvantages
Dictionary Based Approach [73],[79]	used to measuring the feeling that a text conveys to the reader	Simple, Can be extended to multiple dimensions	Impossible to think of all the relevant keywords and their variants, Lack of domain expertise
Corpus Based Approach [39]	use to suggests data-driven approach which have access not only to sentiment labels, but to a context	Provide degree of independence, helps learner encounter a range of real text examples	Noisy nature of large corpus, Idiomatic nature of the corpus

3.3 Machine learning-based approach

Machine learning research has become a significant task in many application areas. Machine learning reaches throughout recent years have magnificently created real world issues. Machine learning algorithms are grouped into supervised learning and unsupervised learning algorithms. [68]

Table 5: Comparison of various Emotion Mining Techniques

Approach	Working Principle	Advantages	Disadvantages
Keyword based Approach [13],[17],[46],[68]	used to classify text based on the presence of positive or negative polarity words	Simple to use	Ambiguity in Keyword Definitions, Incapability of Recognizing Sentences without Keywords, Lack of Linguistic Information
Lexicon Based Approach [28],[45],[68],[76]	Uses Predefined list or dictionary for classifying sentences or words individual	No training and modelling required	Lacks context or domain based classification sentences or words individual
Machine Based Approach [10],[68],[75]	Used to learn the polarity of neutral examples in the documents.	More Suitable for specific domains	Affected by linguistic variation problems

3.3.1 Supervised learning

Supervised learning is also known as machine learning or automatic learning. Supervised learning typically requires two types of datasets: (i) training set and (ii) testing set.

Training set is used by classifier to learn a classification model; while testing set is used to validate the generated model.The main drawback of supervised machine learning algorithms is the responsibility to create a training example. The training example must be broad enough to make the algorithm effective and consistent enough to classify the illustration in test data.[68]

3.3.1.1 Artificial Neural Network (ANN)

ISSN- 2394-5125 VOL 7, ISSUE 15, 2020

The Artificial Neural Network (ANN) is a mathematical modelling approach that is stimulated by the operative processes of the human mind [75]. It is based on the artificial adaptive system which has some form of distributive architecture [10]. The system in ANN encompasses closely knit adaptive processing elements called artificial neurons or nodes which are proficient in carrying out enormous analogous and parallel computations for the purpose of information processing and Knowledge representation [19][68].

3.3.1.2 Random Forest

Random Forests is a classification and regression method based on the collection of a proliferation of decision trees [8].
Recently, attention has been given to ensemble learning, a method which can create several classifiers and produce aggregate results [41]. The two commonly known and used methods for the classification of trees are boosting as proposed by [58] and bagging as proposed by [40]. The latter proposed the general operating mechanism of the Random Forest (RF), which augments the additional layer of randomness to bagging. In RF, each tree is a standard Classification or Regression Tree (CART) Moreover; it picks the splitting predictor from a arbitrarily chosen subset of predictors. Each tree is made-up by using a bootstrap sample of the data and the prediction of all trees are finally accumulated by means of majority voting[8] [68].

3.3.1.3 Support Vector Machine

Support vector machines are considered to be universal learners. Generically they learn linear threshold functions. However, with the use of a suitable kernel function plug in, they can be used to learn in different applications in the form of radical basic function and sigmoid neural networks and be trained on polynomial classifiers [70]. SVMs were initially intended for binary classification, but research has extended it into a multiclass classification. There are two commonly used approaches for multiclass SVMs. The first deals with fabricating and conjoining the number of binary classifiers, and the second one directly involve keeping all data in one optimization construction [14] [68].

3.3.1.4 Genetic Algorithm

The idea of developing the Genetic Algorithm was initiated and developed by John Holland in the year 1975 in his book ``Adaptation in natural and artificial systems". Since then, the Genetic Algorithm (GA) has been increasingly documented as a popular evolutionary computational research technique [65]. It has gained recognition over the years as an optimization tool in a variety of research domains, including computer science, operational research, engineering, management, and social sciences[2][65]. A major reason behind the apprehension of this technique is its diversified applicability, efficiency of operations, and applicability on global scenario [16][43][68].

3.3.1.5 Naïve Bayes

The Naïve Bayes (NB) classifier has long been used in most applications of supervised machine learning. It is considered a tool for the retrieval of data [15]. It is based on a simple theorem of probability for making a probabilistic model of data. The mechanics of the NB algorithm are applied to numeric data [44]. It is simple, easy to understand, and quick for classification. It normally entails a minimal data set for training and then is used to foretell the parameters needed for classification purposes. [68]

3.3.1.6 Decision Tree

The decision tree (DT) classifier has been widely used for prediction and classification of tasks. The rules in creating the decision tree are easy to understand. The classifiers built through the decision tree are given in hierarchical representation.

The tree is composed of decision nodes, event nodes, edge, and path [37]. A variety of classifiers are used in a variety of applications. Some of the DT classifiers are ID3, C4.5, and C5.0, but a common problem that has been found in the DT classifier is the ability to incorporate all types of variations in data, including noise when trees get bigger and deeper. This problem is commonly known as over-fitting. In addition, the structure of the tree is deformed with the addition of data. To shun this problem, the random forest technique is recommended in which many trees are shaped and trained by dividing the training sets, and outcomes are generated through aggregation of all trees. [68]

Table 6: Comparison of various Lexicon Based Approach Machine learning supervised algorithms

Machine learning Supervised algorithms	Working Principle	Advantages	Disadvantages
Artificial Neural Networks[10][19][68][75]	Uses closely knitted adaptive processing elements which are proficient in carrying out enormous analogues and parallel computations	Better accuracy, enough data to train	Entail more processing power, mainly employed on graphics processing units
Random Forest [8][41][58][68][70]	Uses classification and regression method	Effective defence against over fitting, uses many features so no need of feature selection algorithm	Low model interpretability

ISSN- 2394-5125 VOL 7, ISSUE 15, 2020

Support Vector Machine [14][68][70]	Used to learn different application which are in form of radical basic function	provides considerably fast classifiers, has potential generalisation ability	selection of kernel is difficult
Genetic Algorithm [2][16][43][65][68]	Genetic Algorithms are search and optimization techniques in a multifaceted search space.	Flexible, robust and efficient in nature	Global maxima is difficult to find, requires much more time
Naïve Bayes [15][44][68]	Used for making the probabilistic model of data and is used for numeric data	simple, easy to understand, quick technique	Susceptibility to Bayesian poisoning, decrease effectiveness of the system
Decision Tree [8][33]	Used to recognise different type of patterns	Easy to understand	Inability to corporate all types of variations in data including noise when trees are bigger and deeper

3.3.2 Unsupervised Learning

Another type of machine learning is unsupervised learning algorithms. The working principle of this algorithm is to identify the hidden associations in unlabeled data. The unsupervised learning methods are based on calculating resemblance differences between data. For example, it calculates k-means in which similarity between data is computed based on proximity measures, such as Euclidean distance. [68]

Table 7: Application Domains

Category	dataset	Area
Health Care [20][32][56]	Facebook, Daily Strength and Twitter	Medical Science, Drugs and Health Care
Politics [1][11][58]	Weblogs, Twitter	Policy making
Financial Sector [33][34][74]	Yahoo message box, Tweets	Stock Market
Tourism [21][34][64]	Trip advisor, Twitter	Tourism, Hospitality
Sports [61][77]	Twitter	Sports
Reviews [12][25][35][56]	Amazon, Trip Advisor, Youtube and Yahoo blog	Product Review, Hotel Review and Travelling Review
Marketing and Sales [7][18][72]	appannie.com, epinion.com, Facebook Page	sales
Assessment and Evaluation [5][38]	Thai Stories Dataset	Assessment
Media [12][25][35][44][56]	Local Newspaper	Media

3.4 Emotion Mining in Big Data

JOURNAL OF CRITICAL REVIEWS

ISSN- 2394-5125 VOL 7, ISSUE 15, 2020

Most of existing emotion datasets are relatively small, of the order of thousands of entries, which fail to provide a complete coverage of emotion-triggering events and situations [28][76].

Various big data mining techniques are as follows:

Table 8: Various Big Data Mining Techniques

Big Data Mining Techniques	Description
Clustering [9],[10][26][36][44][51][57] [67]	It is the process of making a group of objects into classes of similar objects
Classification [11] [36][51][67][69]	It is derivation of model which determines the class of an object based on its attributes.
Regression [9] [10][36][44][51][67][69]	It is used to predict a range of numeric values given in a particular dataset
Pattern recognition [9][30][36][60]	Pattern recognition is the automated recognition of patterns and regularities in data.
Probabilistic Algorithm [30][44][60]	A probabilistic algorithm is an algorithm where the result and/or the way the result is obtained depend on chance.

Table 9: Summary of common methods used in Emotion Mining Techniques

Machine Learning Approach	**ANN**	[10],[19],[68],[75]
	RF	[8],[40],[41],[58],[68]
	SVM	[4],[14],[26],[27],[29],[42],[51],[57],[60],[68],[70]
	GA	[2],[16],[43],[65],[68]
	NB	[6],[42],[44],[68],[80]
	DT	[37],[42],[68],[80]
Lexicon Based Approach	**DBA**	[28].[30],[31],[36],[45],[62],[71],[73],[79]
	CBA	[28],[30],[31],[36],[39],[45],[62],[71],[76]

Where

ANN: Artificial Neural Network, RF: Random Forest, SVM: Support Vector Machine, GA: Genetic Algorithm, NB: Naïve Bayes, DT: Decision Tree, DBA: Dictionary Based Approach, CBA: Corpus Based Approach

4. DISCUSSION

The Study depicts that large no. of papers used SVM supervised machine learning emotion mining technique followed by dictionary based approach. The study indicates that Lexicon-based approach outperforms supervised machine learning approach not only in terns of accuracy and precision but also in terms of economy of time and efforts used as Lexicon approach avoids the laborious step that are needed for labelling the data to train the classifiers whereas SML is capable to adapt and then create models for which the classifiers are being trained. Many studies indicates that the SML approach is having high precision whereas others indicate that Lexicon approach is equally competitive due to less hard work required for classifiers training. Therefore, the supremacy and the limitations of the technique used may depend upon the situation and objectives for which it is being used and the result of the same may vary from techniques to technique used. Several problems such as handling of implicit product features and dealing with negation, ambiguity and lack of linguistic information etc. are the major challenges discussed in this paper.

5. CONCLUSION

Emotion Mining is the active area of research and has promising future. Many techniques of Emotion Mining are being discussed in this paper. The result of this study may help new users in understanding the wide range of available techniques.

The study showed that there is some consistency between the different proposed techniques and these techniques continue to increase their consistency. It is also observed that different techniques can be combined to overcome each other's limitation and provide a better classification all around. The study focused on the emotion mining hot applications. In addition to being able to create various applications, the ability to detect emotions automatically can increase human-computer interaction. Therefore, it is concluded that Lexicon-based approach is performing better than SML approach as it can efficiently handle data from different domains. SML approach can handle data from different domains efficiently only if enough data is collected in each domain which is quite time consuming process.

The future work may be to use the existing techniques and to develop a new technique to mine emotions.

6. REFERENCES

[1] A Ceron, L. Curini, S. M. Iacus, and G. Porro, "Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France," New Media Soc., vol. 16, no. 2, pp. 340_358, 2014.

[2] A Kumar, R. M. Pathak, and Y. P. Gupta, "Genetic algorithm based approach for file allocation on distributed systems," Comput. Oper. Res., vol. 22, no. 1, pp. 41_54, 1995.

[3] A M. Azmi and S. M. Alzanin, Aara_"A system for mining the polarity of Saudi public opinion through e-newspaper comments," J. Inf. Sci., vol. 40, no. 3, pp. 398_410, 2014.

[4] A Misra, A. Sharma, P. Gulia, and A. Bana, "Big data: Challenges and opportunities," Int. J. Innov. Technol. Exploring Eng., vol. 4, no. 2, pp. 41_42, 2014.

[5] A Muhammad, N. Wiratunga, and R. Lothian, "Contextual sentiment analysis for social media genres," Knowl.-Based Syst., vol. 108, pp. 92_101, Sep. 2016.

[6] A Qazi, A. Tamjidyamcholo, R. G. Raj, G. Hardaker, and C. Standing, "Assessing consumers' satisfaction and expectations through online opinions: Expectation and discon_rmation approach," Comput. Hum. Behav., vol. 75, pp. 450_460, Oct. 2017.

[7] A Stavrianou and C. Brun, "Expert recommendations based on opinion mining of user-generated product reviews," Comput. Intell., vol. 31, no. 1, pp. 165_183, 2015.

[8] A.-L. Boulesteix, S. Janitza, J. Kruppa, and I.König, "Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics," Wiley Interdiscipl. Rev., Data Mining Knowl. Discovery, vol. 2, no. 6, pp. 493_507, 2012.

[9] Accenture. (2014). Big Data Analytics in Supply Chain: Hype or Here to Stay? [Online]. Available:https://acnprod.accenture.com/_acnmedia/Accenture/ConversionAssets/DotCom/Documents/Global/PDF/Dualpub_2/Accenture-Global OperationsMegatrends-Study-Big-Data-Analytics.pdf

[10] C.A. L. Bailer-Jones, R. Gupta, and H. P. Singh, "An introduction to artificial neural networks," in Automated Data Analysis in Astronomy, vol. 52. New Delhi, India: Narosa Publishing House, Sep. 2001, p. 18.

[11] C. Alfaro, J. Cano-Montero, J. Gómez, J. M. Moguerza, and F. Ortega,"A multi-stage method for content classi_cation and opinion miningon weblog comments," Ann. Oper. Res., vol. 236, no. 1, pp. 197_213,2016.

[12] C. Chiu, N.-H. Chiu, R.-J. Sung, and P.-Y. Hsieh, "Opinion mining of hotel customer-generated contents in Chinese weblogs," Current Issues Tourism, vol. 18, no. 5, pp. 477_495, 2015.

[13] C.-H. Wu, Z.-J. Chuang, and Y.-C. Lin, (2006), "Emotion Recognition from Text Using Semantic Labels and Separable Mixture Models", ACM Transactions on Asian Language Information Processing (TALIP), Vol. 5(2),Jun.2006.

[14] C.-W. Hsu and C.-J. Lin, ``A comparison of methods for multiclass support vector machines," IEEE Trans. Neural Netw., vol. 13, no. 2, pp. 415_425, Mar. 2002.

[15] D. D. Lewis, ``Naive (Bayes) at forty: The independence assumption in information retrieval," in Proc. Eur. Conf. Mach. Learn., 1998, pp. 4_15.

[16] D. E. Goldberg, Genetic Algorithms in Search, Optimization, and Machine Learning. Reading, MA, USA: Addison-Wesley, 1989.

[17] E. Cambria, B. Schuller, Y. Xia, and C. Havasi, ``New avenues in opinion mining and sentiment analysis," IEEE Intell. Syst., vol. 28, no. 2,pp. 15_21, Mar. 2013.

[18] E. D. Avanzo and G. Pilato, ``Mining social network users opinions `to aid buyers' shopping decisions," Comput. Hum. Behav., vol. 51, pp. 1284_1294, Oct. 2015.

[19] E. Grossi and M. Buscema, ``Introduction to artificial neural networks,"Eur. J. Gastroenterol. Hepatol., vol. 19, pp. 1046_1054, Dec. 2007.

[20] E. H.-J. Kim, Y. K. Jeong, Y. Kim, K. Y. Kang, and M. Song, ``Topic-based content and sentiment analysis of Ebola virus on Twitter and in the news,"J. Inf. Sci., vol. 42, no. 6, pp. 763_781, 2016.

[21] E. Marrese-Taylor, J. D. Velásquez, and F. Bravo-Marquez, ``A novel deterministic approach for aspect-based opinion mining in tourism products reviews," Expert Syst. Appl., vol. 41, no. 17, pp. 7764_7775, 2014.

[22] Ekman, P. "An argument for basic emotions", Cognition and Emotion, 6(3), pp. 169-200, 1992.

[23] Ekman, P. "Emotion in the Human Face", Oxford University Press, 2005.

[24] Ekman, P. and Friesen, W. V. "Unmasking the Face: A Guide to Recognizing Emotions From Facial Expressions", Malor Books, 2003.

[25] F. Lazhar and T.-G. Yamina, "Mining explicit and implicit opinions from reviews," Int. J. Data Mining, Modelling Manage., vol. 8, no. 1, pp. 75_92, 2016.

[26] F. Provost and T. Fawcett, "Data science and its relationship to big data and data-driven decision making," Big Data, vol. 1, no. 1, pp. 51_59, 2013.
[27] F. R. Lucini et al., "Text mining approach to predict hospital admissions using early medical records from the emergency department," Int. J. Med. Inf., vol. 100, pp. 1_8, Apr. 2017.
[28] Giachanou and F. Crestani, "Like it or not: A survey of Twitter sentiment analysis methods," ACM Comput. Surv., vol. 49, no. 2, p. 28, 2016.
[29] H. Chen, R. H. L. Chiang, and V. C. Storey, "Business intelligence and analytics: From big data to big impact," MIS Quart., vol. 36, no. 4, pp. 1165_1188, Dec. 2012.
[30] H. Sundmaeker, P. Guillemin, P. Friess, and S. Woelfé, "Vision and challenges for realising the Internet of Things," Cluster Eur. Res. Projects Internet Things, Eur. Commision, vol. 3, no. 3, pp. 34_36, Mar. 2010.
[31] H. Yu and V. Hatzivassiloglou, "Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences," in Proc. Conf. Empirical Methods Natural Lang. Process., 2003, pp. 129_136.
[32] I Korkontzelos, A. Nikfarjam, M. Shardlow, A. Sarker, S. Ananiadou, and G. H. Gonzalez, "Analysis of the effect of sentiment analysis on extracting adverse drug reactions from tweets and forum posts," J. Biomed. Inform., vol. 62, pp. 148_158, Aug. 2016.
[33] I Li and P. Meesad, "Combining sentiment analysis with socialization bias in social networks for stock market trend prediction," Int. J. Comput. Intell. Appl., vol. 15, no. 1, p. 1650003, 2016.
[34] I Philander and Y. Y. Zhong, "Twitter sentiment analysis: Capturing sentiment from integrated resort tweets," Int. J. Hospitality Manage., vol. 55, pp. 16_24, May 2016.
[35] J. Chen, D. P. Huang, S. Hu, Y. Liu, Y. Cai, and H. Min, "An opinion mining framework for Cantonese reviews," J. Ambient Intell. Hum. Comput., vol. 6, no. 5, pp. 541_547, 2014.
[36] J. Manyika et al., "Big data: The next frontier for innovation, competition, and productivity," McKinsey Global Institute, Seoul, South Korea, Tech. Rep., 2011.
[37] J. R. Quinlan, "Induction of decision trees," Mach. Learn., vol. 1, no. 1, pp. 81_106, 1986.
[38] K. Pasupa, P. Netisopakul, and R. Lertsuksakda, "Sentiment analysis of Thai children stories," Artif. Life Robot., vol. 21, no. 3, pp. 357_364, 2016.
[39] Keith Stuart, Ana Botella, and Imma Ferri, "A Corpus-Driven Approach to Sentiment Analysis of Patient Narratives", EPiC Series in Language and Linguistics Volume 1, 2016.
[40] L. Breiman, "Bagging predictors," Mach. Learn., vol. 24, no. 2, pp. 123_140, 1996.
[41] Liaw and M.Wiener, "Classification and regression by random forest," R News, vol. 2, no. 3, pp. 18_22, 2002.
[42] M. Batty et al., "Smart cities of the future," Eur. Phys. J. Special Topics, vol. 214, no. 1, pp. 481_518, Nov. 2012.
[43] M. Mitchell, "An introduction to genetic algorithms," Sadhana, vol. 24, nos. 4_5, pp. 293_315, Aug. 1999.
[44] M. Sahami, S. Dumais, D. Heckerman, and E. Horvitz, "A Bayesian approach to filtering junk e-mail," in Proc. Learn. Text Categorization, Papers Workshop, vol. 62, 1998, pp. 98_105.
[45] M. Taboada, J. Brooke, M. To_loski, K. Voll, and M. Stede, "Lexiconbased methods for sentiment analysis," Comput. Linguistics, vol. 37, no. 2, pp. 267_307, 2011.
[46] Mr. Nilesh M. Shelke. "Approaches of Emotion Detection from Text" International Journal of Computer Science and Information Technology Research, April 2014.
[47] Munawar Hasan, "Genetic Algorithm and its applications to Big Data Analysis", July 1991.
[48] Mustafa Hajeer, dipankar Dasgupta, "Distributed Genetic Algorithm to big Data clustering", Dec 2016.
[49] Navdeep Singh Gill,"Xenonstack blog", 2018.
[50] Neelam Muktar, Mohammad Abid Khan, "Lexicon based approach outperforms supervised machine learning approach for urdu sentimental analysis in multiple domains", Aug 2018.
[51] P. Zikopoulos and C. Eaton, "Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data". New York, NY, USA: McGraw-Hill, 2011.
[52] P.Amudha, S.Sivalumar, "Big data Analytics using SVM" jan 2018
[53] Plutchik R. "A general psych evolutionary theory of emotion". Emotion: Theory, research, and experience, 1(3), pp. 3-33, 1980.
[54] Plutchik R. "The psychology and biology of emotion", New York: Harper Collins, 1994.
[55] Poonam Bonde, Sachin Barahate, "Classification of Big Data using SVM", julu 2017
[56] Q. Zhou, R. Xia, and C. Zhang, "Online shopping behavior study based on multi-granularity opinion mining: China versus America," Cogn. Comput., vol. 8, no. 4, pp. 587_602, 2016.
[57] R. E. Bryant, R. H. Katz, and E. D. Lazowska, "Big-data computing: Creating revolutionary breakthroughs in commerce, science and society," Comput. Community Consoritum,Washington, DC, USA, Tech. Rep. Version 8, Dec. 2008.
[58] R. E. Schapire, Y. Freund, P. Bartlett, andW. S. Lee, "Boosting the margin: A new explanation for the effectiveness of voting methods," Ann. Stat., vol. 26, no. 5, pp. 1651_1686, 1998.

[59] R. G. Rodrigues, R. M. das Dores, C. G. Camilo-Junior, and T. C. Rosa, "SentiHealth-Cancer: A sentiment analysis tool to help detecting mood of patients in online social networks," Int. J. Med. Inform., vol. 85, no. 1, pp. 80_95, 2016.
[60] R. L. Villars, C. W. Olofson, and M. Eastwood, ``Big data: What it is and why you should care," IDC, Framingham, MA, USA, White Paper 228827, Jun. 2011.
[61] R. P. Schumaker, A. T. Jarmoszko, and C. S. Labedz, Jr., "Predicting wins and spread in the Premier League using a sentiment analysis of Twitter," Decision Support Syst., vol. 88, pp. 76_84, Aug. 2016.
[62] R. Piryani, D. Madhavi, and V. K. Singh, "Analytical mapping of opinion mining and sentiment analysis research during 2000_2015," Inf. Process. Manage., vol. 53, no. 1, pp. 122_150, 2017.
[63] R. R. B. López, S. Sánchez-Alonso, and M. A. Sicilia-Urban, "Evaluating hotels rating prediction based on sentiment analysis services," Aslib J. Inf. Manage., vol. 67, no. 4, pp. 392_407, 2015.
[64] R. R. B. López, S. Sánchez-Alonso, and M. A. Sicilia-Urban, "Evaluating hotels rating prediction based on sentiment analysis services," Aslib J. Inf. Manage., vol. 67, no. 4, pp. 392_407, 2015.
[65] S. N. Sivanandam and S. N. Deepa, "Genetic algorithms," in Introduction to Genetic Algorithms. New York, NY, USA: Springer-Verlag, 2007.
[66] Sanjeev Dhawan, Kulvinder Singh, Deepika Sehrawat. "Emotion Mining Techniques in Social Networking Sites",International Journal of Information & Computation Technology. ISSN 0974-2239 Volume 4, Number 12 (2014)
[67] Sarpreet Kaur, Dr. Williamjeet Singh, "Systematic Mapping study of Big Data Mining tools and techniques", 8th ICCCNT, July 2017.
[68] Shahid Shayaa, Noor Ismawati Jaafar , Shamshul Bahri, Ainin Sulaiman, Phoong Seuk "Sentiment Analysis of Big Data: Methods, Applications, and Open Challenges", International Journal of Computer Science and Information Technology Research, June 2018
[69] T. Hai, K. Shirai, and J. Velcin, "Sentiment analysis on social media for stock movement prediction," Expert Syst. Appl., vol. 42, no. 24, pp. 9603_9611, 2015.
[70] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features," in Proc. Eur. Conf. Mach. Learn., vol. 98, 1998, pp. 137_142.
[71] T. T. Thet, J.-C. Na, and C. S. G. Khoo, "Aspect-based sentiment analysis of movie reviews on discussion boards," J. Inf. Sci., vol. 36, no. 6, pp. 823_848, 2010.
[72] T.-P. Liang, X. Li, C.-T. Yang, and M. Wang, "What in consumer reviews affects the sales of mobile apps: A multifacet sentiment analysis approach," Int. J. Electron. Commerce, vol. 20, no. 2, pp. 236_260, 2015.
[73] Tanvi Hardeniya, Dilip kumar A. Borikar, "Dictionary based approach to sentiment analysis", Jan 2016.
[74] W. Chen, Y. Cai, K. Lai, and H. Xie, "A topic-based sentiment analysis model to predict stock market price movement using Weibo mood," Web Intell., vol. 14, no. 4, pp. 287_300, 2016.
[75] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," Bull. Math. Biophys., vol. 5, no. 4, pp. 115_133, 1943.
[76] Wenbo Wang, Lu Chen, Krishnaprasad Thirunarayan, Amit P. Sheth, " Harnessing Twitter 'Big Data' for Automatic Emotion Identification" september 2012.
[77] Y. Yu and X. Wang, "World cup 2014 in the Twitter world: A big data analysis of sentiments in US sports fans' tweets," Comput. Hum. Behav., vol. 48, pp. 392_400, Jul. 2015.
[78] Yan Hou, "Decision Tree Algorithm for Big Data Analysis", 2018
[79] Youness Madani, Mohammed Erritali,"Classification of Social Network Data using Dictionary based approach" Jan 2019.
[80] Z. Khan and T. Vorley, "Big data text analytics: An enabler of knowledge management," J. Knowl. Manage., vol. 21, no. 1, pp. 18_34, 2017.