Journal of Positive School Psychology http://journalppw.com
2022, Vol. 6, No. 4, 7254 – 7263

An assessment of data mining techniques reliability in predicting social media sentiments

[1]Dr. A. MALARVIZHI

[1]*Assistant Professor, P.G. and Research Department of Computer Science, H.H. The Rajah's College (A), Pudukkottai, malarvizhiselvam@gmail.com*

Abstract

SNSs (Social networking Sites) are the most popular medium for global communications. Internet users have been increasing in tandem with evolving technologies where their expression about organization, events, personalities and many more other discussions can be found in online review sites or social networks media or blogs. SAs (Sentiment Analyses) are a part of ongoing researches using DMTs (data Mining Techniques). They are computational treatments of opinions, sentiments and subjectivity of text. SNSs generate voluminous data that can play an essential role in decision making of individuals, organizations and even governments. It impossible to scrutinize texts or sentiments expressed on SNSs where SAs provide polarities to the text and classify text into positive or negative categories where DMTs can be used for categorizations though they may result in different accuracy percentages. The main aim of this study is to examine categorizing of sentiments expressed in SNSs by DMTs and evaluating them in terms of accuracies or speeds of executions.

Keywords: SNSs (Social networking Sites), SAs (Sentiment Analyses), DMTs (data Mining Techniques).

1. INTRODUCTION

SNSs (Social networking Sites) are the most popular medium for global communications. Internet users have been increasing in tandem with evolving technologies where their expression about organization, events, personalities and many more other discussions can be found in online review sites or social networks media or blogs. These new users SAs are a part of ongoing researches using DMTs. They are computational treatments of opinions, sentiments and subjectivity of text. SNSs generate voluminous data, and this voluminous data stores are used by individuals or business establishments or eve governments to harbour their decisions. Evaluating attitudes from words is a complex issue, and hence SAs assign polarities to words or messages for categorizing them into good or bad words. Thus SAs, computationally analyze attitudes, sentiments and opinions of the general public with regard to entities in their reviews. These expressed views [1] can be analyzed by using DMTs which extracts expressed sentiments from texts while SAs identify the hidden sentiments and categorize them based on their polarities. Figure 1 depicts the generic flow of SAs.

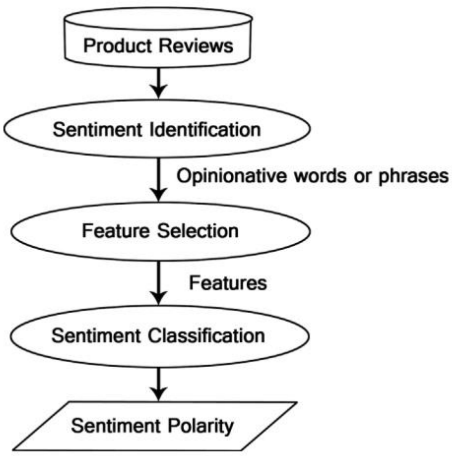

Fig.1 – *Flow of SAs*

Sentiments can be classified in three levels as documents, sentences and aspects. Documents can be classified as positive or negative based on expressed sentiments. These methods consider a complete document as one unit at the document level while sentence-levels consider a sentence as one unit in SAs. The basic objective of SAs is to classify sentiments expressed in sentences where the preliminary step is identifying sentences as subjective or objective. Subjective sentences can be positive or negative in expressed sentiments. Opinions can also be subjective in nature [2]. The demarcations between both sentence and document levels are marginal as sentences are short documents [3] making this demarcation unnecessary. The actual sentiments can be obtained only when aspects are studied. Aspect-level SAs classify sentiments based on certain entity aspects where the preliminary steps involve identifying entities followed by their aspects. Public opinions on the same entity might differ making SAs important for knowing opinions. The datasets utilised in SAs are especially important as they are the main resources and compiled from SNSs. These reviews are significant to many people like business owners who take business decisions based on customer opinions on their items found in the form of product reviews. Hence, SAs have gained popularity for analyses in various domains including financial [4,5], stories [6], and political arguments. [7]. Politicians can find out people's opinions about them or study other politicians with regard to public sentiments. Many studies and proposed methods are presented for SAs. Detailed study on SAs was done in [8] where challenges of SAs were detailed. Growing trends of SAs were discussed in [9] [10] [11]. Thus SAs have been researched and applied in many spheres for knowledge. SAs have undergone variety of enhancements in the techniques used. This extensive detailing of existing DMTs used in SAs can be useful for new comers in researches. Following this introductory section, the next section is an exhaustive review of literature followed by discussion in section three. This paper concludes with section four which is a summary of this study.

II. LITERATURE REVIEW

SAs need proper FSs (feature selections) for classification of sentiments where DMTs have been used. The first task in SAs is extracting textual features where features are individual words or their frequency of repetitions in a sentence called n-grams. Documents are seen as BOWs (Bag of Words) or as strings with word sequences by FSs. The individual words are given a binary weighting or term frequency weights are used for indicating their relative importance [12]. POSs (Parts of speeches) are the act of tracing adjectives which indicate opinions while words and phrases are generic phrases that convey positive/negative feelings like cost me an arm and a leg which does not indicate any opinion or sentiment directly. People while communicating also express negativity which change the orientations of expressed words like not bad is the same as good. Thus, SAs use a few parameters like Term Frequencies, N-gram Features, POSs and Positions of Term [13].

- Term Presence Vs Term Frequency: Term Frequencies find words counts in the corpus whereas Term Presence is a binary value implying a word's presence in sentence with 1 or 0. Term presences have greater significant than frequencies in SAs according to the study in [14].

- N-grams: These are features used by NLPs (Natural Language Processing) where n-grams indicate counts of specific words. The features can be single (unigram), dual (bigram), three (trigrams) and anything greater are n-grams. The study in [15] found that bi/tri grams showed better resulst.

- Tagging POSs: Sentiments expressed in the English language mainly contain verbs, adjectives, and adverbs where POSs find tagged words in a corpus. Adjectives, ad-verbs, and verbs are treated as features and irrelevant words are eliminated reducing the size of corpuses.

- Negations: Positive words turn negative, inverting their polarity when negations are used. For Example good is positive, nut not good changes its polarity from positive to negative.

- PMIs (Point-wise Mutual Information): Metrics for measuring mutual information are information theory based [16] and differentiate between characteristics and classes as in Equation (1)

$$M_i(w) = \log\left(\frac{F(w)\cdot p_i(w)}{F(w)\cdot P_i}\right) = \log\left(\frac{p_i(w)}{P_i}\right) \quad .....(1)$$

Where word is a w and i is a class. mutual independence and given by Pi F(w), and true co-occurrences are computed by F(w) pi(w) where ratio between these two values is mutual information. They might have positive correlations when Mi(w) > 0 and negative when < 0. PMIs have been used in many applications and enhancements like entropies have also be tried.

- Chi-squares: In n documents of a collection, if pi(w) is the probability of class i in documents with w and Pi is the global fraction of documents with class i while F(w) is the global fraction of documents with word w then the statistics between i and w (X2) can be defined as Equation (2) [17]

$$\chi_i^2 = \frac{n\cdot F(w)^2\cdot(p_i(w)-P_i)^2}{F(w)\cdot(1-F(w))\cdot P_i\cdot(1-P_i)} \quad ....(2)$$

Though PMIs and X2 are methods for assessing correlations in categories of words where the latter performs better due to normalization of values leading to examination of similarities.

- LSIs (Latent Semantic Indices): FSs attempt to reduce the dimensionality of data by way of selections from original attribute sets. Their transformations create smaller sets as a function of the original sets where LSIs are popular feature transformation methods [18]. LSIs transform texts into linear combinations while PCAs (Principal Component Analyses) do the same in original words [19]. PCAs decide which axes that have highest degree of information. Though LSI's classifications are unsupervised their fundamental shortcoming is the lack of knowledge on data and hence are not used much to segregate classes from documents.

Extracting required features from SNSs is the base for accuracy of classifiers. The evaluation results of FSs tested with classifiers is listed in Figure 2.

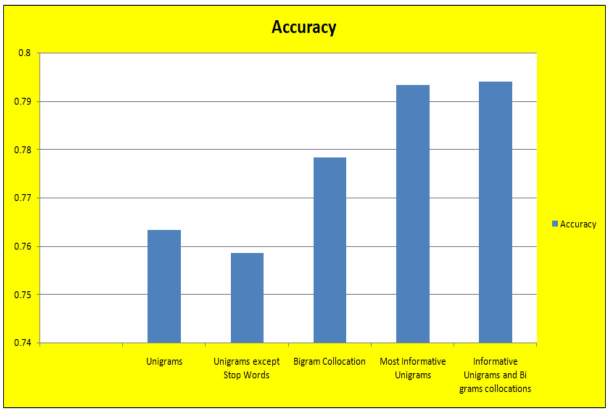

Fig. 2 - *Evaluation of FSs in terms of accuracy*

It is evident from Figure 2 that collations of informative unigram and bigram Unigrams words result in maximum accuracy (0.79407) of SA classification when compared to Unigrams (0.763266), Unigrams except Stop Words (0.758472), Bigrams Collocations (0.77826), Most Informative Unigrams (0.793356). Figure 3 and 4 depict substantiations of accuracy where it can be seen that precision values are higher and recall values are lower

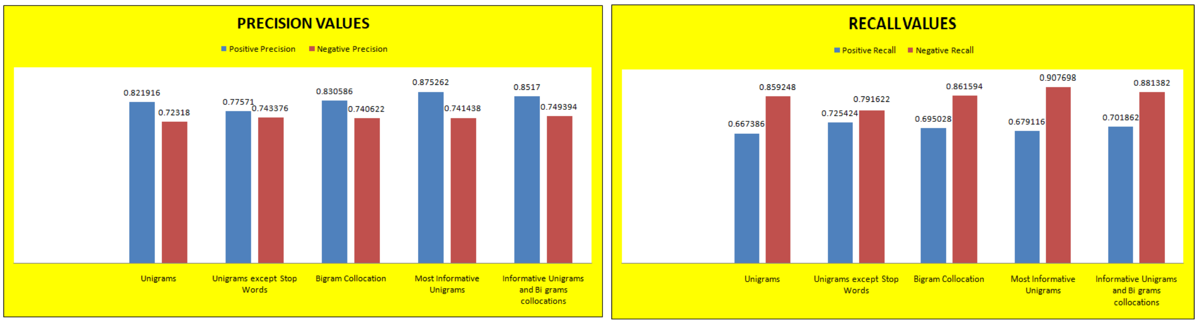

Fig. 3 - *Evaluation of FSs in terms of Precision and Recall Values*

Text classifications of sentiments in DMTs can be categorized into MLTs (Machine Learning Techniques) or LBTs (Lexicon based Techniques) or hybrid approaches [20] where MLTs can be supervised or unsupervised. Supervised methods need more labelled samples for their learning whereas unsupervised find it difficult to use labelled samples. LBTs depend on finding lexicons for their ext analyses. LBTs can be based on dictionaries where synonyms/antonyms whereas as begin with a seed list of words and search a larger corpus to in finding words in context to specific orientations in corpus-based approaches and where statistical or semantic approaches are used. Hybrid Approaches use a combination of methods and found more in classification of sentiment lexicons. Figure 4 depicts DMTs used in classification of sentiments from SNSs irrespective of the type of document structure used.

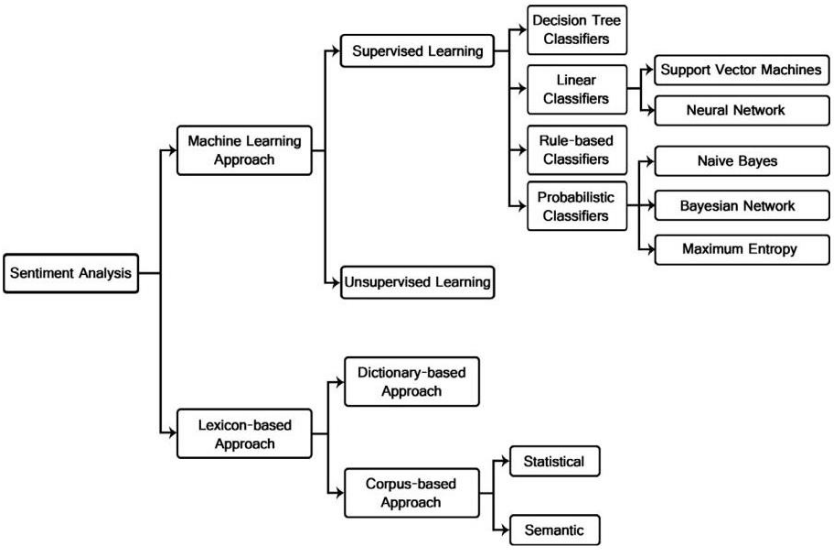

Fig. 4 – *DMTs and SAs*

MLTs rely on known algorithms for SAs which make use of syntactic and/or linguistic features for text classifications which can be defines as a set of training records T = {X1, X2, . . ., Xn} where X can be assigned to a class. Classifications relate features of records as belonging to one of the class labels and hence predicting of instances whose classes are unknown are assigned based on the labels. MLTs can be categorized into supervised or unsupervised MLTs where supervised learning methods depend on the existence of training document labels. They are used when labelled data available for training models. They use mainly two steps are namely training followed by predictions [21]. In training, labelled data is fed to the classification algorithms which give models as outputs which then predict based on test data> Certain supervised algorithms are detailed below:

• NBs (Naïve Bayes): These are probabilistic classification algorithms which consider each word as independent without considering locations in sentences. NBs are based on Bayes theorem to compute probabilities of term for their corresponding label and depicted mathematically as Equation (3).

$$P(label|features) = \frac{P(label) * P(features|label)}{P(features)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots..(3)$$

Where the prior probability of label in the dataset is p (label) and its prior probability of features getting associated with it is p(feature | label) while p(feature) is its prior probability of occurrence in features. The research in [22] used SentiWordNet Lexicon with NBs to improve twitter dataset classifications by using scores for positive and negative tweets.

• BNs (Bayesian Networks): The disadvantage of NBs is in treating each word as independent and thus missing out on semantic relationships between words which is overcome by BNs as they strongly consider dependencies between words. BNs depict these dependencies as acyclic directed graphs and where nodes represent words as variables while edges are dependencies between variables. The study in [23] used BNs for finding competitive outputs with other classifiers for satisfactory results.

• SVMs (Support Vector Machines): SVMs were introduced for binary classification issues and they focus on determining best hyper-plane which separate classes and act as separators of decisions between data points in terms of different classes. The hyperplane is a line which maintains maximum distances between two support vector classes as shown in figure 5. SVMs have the capability to manage linear, and non-linear classifications. The study [24] used SVMs for classifying various weighing schemes including term and binary occurrences. The study used chi-square in its FSs and reduced dimensionality while removing noises.

• ANNs (Artificial Neural Networks): ANNs mimic neuron structures of the human brain. ANNs typically comprise of input, hidden, and output layers. Figure 5 depicts a structure of ANNs. A vector ‘a (i)’ is given as input to neuron where in SAs vectors denote frequency of words in a document. Weight ‘A’, corresponding for each neuron is computed using a linear function x (i) =A. (a (i)). The resultant sign of x (i) is used to classify classes. ANNs follow forward/backward propagations where in forward propagations, inputs of the input layer’s neurons are multiplied by random numbers and functions normalize outputs in the interval [0,1]. Outputs are then compared with target values and on finding errors between these two values backward propagations are performed. In back propagations inputs are multiplied by error values for adjustment of weights and learning is based on errors. The study in [25] proposed a technique where LSAs (latent semantic analyses) converted words to vectors and classified using ANNs to obtain 87% accuracy of classifications in SAs.

• DTs (Decision trees):. DTs use tree like structures where non- terminal nodes represent features and terminal nodes represent labels and routing paths are based on conditions. This is iterated until it reachs terminal nodes which label inputs. One significant challenge DTs is to choose the attribute as a root node which is handled by using information gains or Gini indices. DTs work well in SAs as they provide good results on voluminous data. Common examples of DTs include CART, CHAID, and C5.0. DTs divide their data hierarchically based on conditions which are attribute values. These divisions are recursive and stopped only when terminal nodes have minimal features which are used in classifications. Figure 5 depicts DT’s structure. DTs were used in [26] to block false content on internet sites. Words were weighed to assess its importance while a binomial classifier classified a word to specific categories.

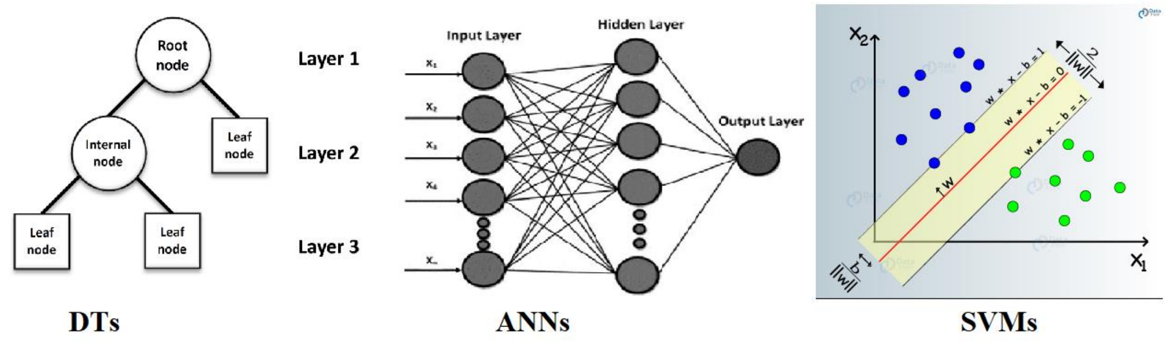

Fig. 5. *Structures of DTs, ANns and SVMs*

• Rule-Based Classifier. These classifiers generate models from a set of framed rules followed by new rule predictions from the framed sets. Thus using antecedent (left-hand side) and consequent (right-hand side) rules , classes are predicted. Equation (4) [27] depicts a rule form.

{w1 ∧ w2∧w3} →{+|-}…………………(4)

Words as a part of rules and expressing sentiments are depicted as Equation (5)

{Good} →{+} {Bad} → {-}……………..(5)

Features present in terms are a part of IF while THEN parts turn into classification labels. Rule formations are based on their Confidence and Support where support defines instance counts associated with a rule while training while confidence implies likelihood of feature's label in a class. The work in [28] suggested Association Rule based Text classifications of two itemsets where words that did not overlap with other classes and overlapped ones were demarcated for construction of rules from frequent itemsets resulting in 95% accuracy.

When data is filled with labels that cannot be trusted, unsupervised learning procedures are used as they generate labels and subsequently, category's key words lists assist in classifications. Unsupervised methods analyze and classify domain dependent data very easily. Unsupervised spectral clustering approach for SAs was used by the study in [29] where tweets were grouped into positive and negative groups. Positive opinions are considered while negative opinions are to be discarded. Lexicons are formed from predefined words with polarity scores and the most straightforward ways for categorizing sentiments and used by classifiers for word matches which subsequently categorize words where vocabulary sizes determine classification effectiveness. Lexicons can be in two types and are detailed below:

• Dictionaries: Approaches using dictionaries, select seed words which subsequently identify synonyms and expand word set sizes using online dictionaries. Seed words are unique corpus words and significant to opinions. Seed words and expanded words then become the base for SAs. Examples of dictionaries include WordNets, SentiWordNets, SentiFuls, and SenticNets. The work in [30] constructed lexicon thesaurus using three online dictionaries and only terms that co-occurred in lexicons were saved from dependability.

• Corpuses: In corpuses labels and contexts of words are identified. Approaches based on corpuses, initially generate list of seed words, which subsequently produce new subjective words using listed word's syntactic pattern where syntactic patters are group of words that occur in the same order or at the same time. Corpuses works both statistically and semantically. These approaches were examined by the author in [31] where the use of SVMs with corpuses resulted in high accuracies.

III. DISCUSSIONS

People assess what is happening on around them, and with the increased use of social media, governments, organisations, businesses, and even manufacturers have begun to examine people's views toward goods, locations, and events. These feelings are conveyed as sentiments on social media sites, while SAs are automated approaches for extracting topic sentiments. These data are being used to gain real-time insights into people's feelings. Blogs, online forums, newspaper comment sections, and social networking sites like Facebook and Twitter are all examples of social media. These social media platforms have the ability to capture the opinions or word of mouth of millions of individuals. In computational linguistics and social network analysis, communication and the availability of these real-time perspectives from individuals all over the world has ushered in a revolution. For businesses, social media is becoming an increasingly significant source of information. People, on the other hand, are more eager and glad than ever before to share details about their lives, knowledge, experiences, and ideas with the rest of the world via social media. They actively participate in events that occur in society by expressing their ideas and making statements. This technique of sharing their knowledge and feelings with society through social media leads businesses to gather more information about their enterprises, goods, and how well-known they are among the public, allowing them to make better informed decisions about how to run their businesses efficiently. As a result, it's evident that SAs play an important role in a variety of decisions, including creative Customer Experience Management and Customer Relationship Marketing. Additionally, firms trying to sell their products, uncover new prospects, and maintain their reputation can use this tool. As companies seek to automate the process of filtering out noise, comprehending discussions, recognising valuable material, and taking action, Many people are increasingly interested in sentiment analysis. Having access to enormous amounts of information is no longer an issue in the period we live in, commonly referred to as the information age or the knowledge society, as seen by the mountains of fresh material published every day on the internet. Information has become the primary trade object for many businesses in this period. If we can design and use systems to search for and retrieve important data and information, as well as mine it to convert it to knowledge with accuracy and timeliness, we will be able to make the most use of the vast amount of data at our disposal. The majority of today's solutions depend on basic Boolean words to express sentience in tweets, Facebook wall posts etc. But this is not enough to address the above mentioned problems in the area of sentiment analysis and it will not generate precise and timely knowledge for aggregate sentiments. In order to get accurate knowledge after analyzing a sentiment, it should thoroughly consider solving the issues mentioned above. Most other systems that try to give solutions for these issues are still on research level, some systems also try to analyze sentiments from multiple languages and few systems which address some of the above mentioned drawbacks are available commercially also. Table 1 lists an aggregation of datasets, techniques used for SAs and their accuracies in classifying sentiments.

Table 1 – *Aggregation of SAs used with their accuracies*

Method	Dataset	Technique	Accuracy Percentage
Lexicon based	Movie review data	SentiWord-Net	65
MLTs	Movie review data	NBs	82
MLTs	Movie review data	SVMs	77
MLTs	Twitter	SVMs	82.52
Maximum Entropy classifier	Blogs	Statistical	82.8
Lexicon based	Tweets on JAIHIN D	Statistical	73.53
Lexicon based	Tweets World cup 2015	Statistical	81.97
MLTs	Twitter	ANNs	86

MLTs	Twitter	NBs	63.9
MLTs	Stanford Twitter Sentiment dataset	CNNs	87.6
Lexicon based	Movie review data	SentiWord-Net	85.4
Hybrid	Movie review data	SentiWord-Net + NBs	89
Hybrid	Movie review data	SentiWord-Net + SVMs	76
MLTs	Bangladeshi Facebook pages	ANNs	83.79

MLTs are used in solving problems for their simplicity and ability to learn from training data which make them adapt to multiple domains. Lexicon based algorithms have been used often in SAs due to their scalability and computational efficiency. Binary classifications are nice first steps, as they involve distinguishing between two polarities. SAs are mostly done on Product Reviews with usage on other kinds of data in recent studies. Many studies have also shown that domain-dependent data produces more accurate findings than domain-independent data [32]. For the sake of simplicity, the researchers normally operate in a domain-independent manner, as illustrated in Fig. 8. As a result, the domain-dependent issue, also known as a context-based SA, has been a continuing subject of search. Researchers have increasingly been drawn to SAs that use non-English languages, posing a new set of hurdles for those attempting to construct lexica, corpora, and dictionaries for other languages. Figure 6 shows the accuracies of several approaches used in SAs.

Fig. 6- *Accuracies of Different Methodologies on SAs*

IMDB and Amazon.com are two well-known review data providers. Movie reviews may be found on IMDB, while product reviews can be found on Amazon.com. In SAs, these data sources are employed. Twitter is often used since it is a well-known social networking site where tweets convey people's ideas and have a limited length of 140 characters. Eastern languages are also utilised more frequently, according to the study. Regional Indian languages still have a scarcity of resources. As a result, it is now a very good research trend. NLPs can help SAs run more smoothly and produce more accurate results. This establishes a new research trend of employing NLP as a pre-processing stage prior to sentiment analysis. Working with a domain-specific corpus produces better outcomes than working with a domain-independent corpus. In the topic of domain-specific SAs, also known as context-based SAs, there is currently a dearth of study. This is due to the fact that creating a domain-specific corpus is more difficult than utilising a domain-independent one.

IV. CONCLUSION

This paper compared the various stages of SAs and the approaches utilised for feature selection and sentiment classification. This article brings together papers from SAs and allied topics that apply SA approaches to a variety of real-world problems. It is apparent that improvements to FS algorithms are still a work in progress. The most often utilised MLTs in SAs are NBs, SVMs, and ANNs. They are used as benchmarks against which many proposed algorithms are evaluated. Interest in languages other than English is developing in this sector, despite a scarcity of resources and study on these languages. WordNet, which is available in languages other than English, is the most often used lexicon source. Many natural languages still require the creation of resources for usage in SAs. Microblogs, blogs, and forums, as well as news sources, are frequently utilised in SAs. This type of media material is extremely useful in communicating people's sentiments or ideas about a certain issue or product. The use of social networking sites and microblogging sites as data sources still need further investigation. In many applications, it's critical to take into account the text's context as well as the user's choices. As a result, additional study on context-based SA is required. This survey's contribution is valuable for a variety of reasons. First, according to the techniques used, this survey provides sophisticated categorization of a large number of recent articles. This perspective might aid researchers who are familiar with certain strategies in implementing them in SAs and selecting the best methodology for a given application. Second, the numerous SA approaches are classified, with brief descriptions of the algorithms and their sources. This can provide newcomers to SAs with a bird's-eye view of the entire field. Finally, the various benchmark data sets are explored and classified according to their application.

Reference

[1] Tsytsarau Mikalai, Palpanas Themis. Survey on mining sub-jective data on the web. Data Min Knowl Discov 2012;24:478–514.

[2] Wilson T, Wiebe J, Hoffman P. Recognizing contextual polarity in phrase-level sentiment analysis. In: Proceedings of HLT/ EMNLP; 2005.

[3] Liu B. Sentiment analysis and opinion mining. Synth Lect Human Lang Technol 2012.

[4] Yu Liang-Chih, Wu Jheng-Long, Chang Pei-Chann, Chu Hsu-an-Shou. Using a contextual entropy model to expand emotion words and their intensity for the sentiment classification of stock market news. Knowl-Based Syst 2013;41:89–97.

[5] Michael Hagenau, Michael Liebmann, Dirk Neumann. Auto-mated news reading: stock price prediction based on financial news using context-capturing features. Decis Supp Syst; 2013.

[6] Tao Xu, Peng Qinke, Cheng Yinzhao. Identifying the semantic orientation of terms using S-HAL for sentiment analysis. Knowl-Based Syst 2012;35:279–89.

[7] Maks Isa, Vossen Piek. A lexicon model for deep sentiment analysis and opinion mining applications. Decis Support Syst 2012;53:680–8.

[8] Pang B, Lee L. Opinion mining and sentiment analysis. Found Trends Inform Retriev 2008;2:1–135.

[9] Cambria E, Schuller B, Xia Y, Havasi C. New avenues in opinion mining and sentiment analysis. IEEE Intell Syst 2013;28:15–21.

[10] Feldman R. Techniques and applications for sentiment analysis. Commun ACM 2013;56:82–9.

[11] Montoyo Andre´s, Martı´nez-Barco Patricio, Balahur Alexandra. Subjectivity and sentiment analysis: an overview of the current state of the area and envisaged developments. Decis Support Syst 2012;53:675–9. W. Medhat et al.

[12] Yelena Mejova, Padmini Srinivasan. Exploring feature definition and selection for sentiment classifiers. In: Proceedings of the fifth international AAAI conference on weblogs and social media; 2011.

[13] Mejova, Y., & Srinivasan, P. “Exploring feature definition and selection for sentiment classifiers”. In Proceedings of the fifth international AAAI conference on weblogs and so-cial media 2011.

[14] Pang, B., & Lee, L. “Opinion Mining and Sentiment Analysis”. In Foundations and Trends Information Retrieval, vol. 2, pp. 1–135, 2008.

[15] Dave, K., Lawrence, S., and Pennock, D. “Mining the Peanut Gallery: Opinion

Extraction and Semantic Classification of Product Reviews". 2003.
[16] Cover TM, Thomas JA. Elements of information theory. New York: John Wiley and Sons; 1991.
[17] Yi Hu, Li Wenjie. Document sentiment classification by explor-ing description model of topical terms. Comput Speech Lang 2011;25:386–403.
[18] Jolliffee IT. Principal component analysis. Springer; 2002.
[19] Griffiths Thomas L, Steyvers Mark, Blei David M, Tenenbaum Joshua B. Integrating topics and syntax. Adv Neural Inform Process Syst 2005:537–44.
[20] Diana Maynard, Adam Funk. Automatic detection of political opinions in tweets. In: Proceedings of the 8th international conference on the semantic web, ESWC'11; 2011. p. 88–99.
[21] J Kaur & M Sabharwal. "Spam Detection in Online Social Networks Using Feed Forward Neural Network". In RSRI Conference on Recent Trends in Science and Engineering, vol.2, pp. 69-78, 2018.
[22] Goel, A., Gautam, J., & Kumar, S. "Real time sentiment analysis of tweets using Naive Bayes". In 2nd International Conference on Next Generation Computing Technologies (NGCT), pp. 257-216, 2016.IEEE.
[23] Al-Smadi, M., Al-Ayyoub, M., Jararweh, Y., Qawasmeh, O.: Enhancing aspect-based sen-timent analysis of Arabic hotels' reviews using morphological, syntactic and semantic fea-tures. Inf. Process. Manag. (2018).
[24] Zainuddin, N., & Selamat, A. "Sentiment analysis using Support Vector Machine". In In-ternational Conference on Computer, Communications, and Control Technology (I4CT), pp. 333-337, 2014, IEEE.
[25] Patil, S., Gune, A., & Nene, M. "Convolutional neural networks for text categorization with latent semantic analysis". In International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), pp. 499-503, 2017, IEEE.
[26] Kotenko, I., Chechulin, A., & Komashinsky, D. "Evaluation of text classification tech-niques for inappropriate web content blocking". In 8th International Conference on Intelli-gent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), pp. 412-417, 2015, IEEE.
[27] Xia R, Xu F, Yu J, Qi Y, and Cambria E. "Polarity shift detection, elimination and ensem-ble: a three-stage model for document-level sentiment analysis ". In Information Pro-cessing and Management, vol. 52, pp. 36–45, 2016.
[28] Buddeewong, S., & Kreesuradej, W. "A new association rule-based text classifier algo-rithm". In 17th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'05), 2005.
[29] Unnisa, M., Ameen A., & Raziuddin, S. "Opinion Mining on Twitter Data using Unsuper-vised Learning Technique". In International Journal of Computer Applications, pp.0975 – 8887, Vol. 148, 2016.
[30] Park, S., & Kim, Y. "Building thesaurus lexicon using dictionary-based approach for sen-timent classification". In IEEE 14th International Conference on Software Engineering Re-search, Management and Applications (SERA), 2016.
[31] Abdulla, N. A., Ahmed, N. A., Shehab, M. A., & Al-Ayyoub, M. "Arabic sentiment anal-ysis: Lexicon-based and corpus-based". In IEEE Jordan Conference on Applied ElectricalEngineering and Computing Technologies (AEECT), 2013.
[32] Cruz Fermı´n L, Troyano Jose´ A, Enrı´quez Fernando, Javier Ortega F, Vallejo Carlos G. Long autonomy or long delay?' The importance of domain in opinion mining. Expert Syst Appl 2013;40:3174–84.