Global Journal of Computer Science and Technology: C
Software & Data Engineering
Volume 16 Issue 5 Version 1.0 Year 2016
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Review of Feature Selection and Optimization Strategies in Opinion Mining

By K. Venkata Rama Rao

Abstract- Opinion mining and sentiment analysis methods has become a prerogative models in terms of gaining insights from the huge volume of data that is being generated from vivid sources. There are vivid range of data that is being generated from varied sources. If such veracity and variety of data can be explored in terms of evaluating the opinion mining process, it could help the target groups in getting the public pulse which could support them in taking informed decisions. Though the process of opinion mining and sentiment analysis has been one of the hot topics focused upon by the researchers, the process has not been completely revolutionary. In this study the focus has been upon reviewing varied range of models and solutions that are proposed for sentiment analysis and opinion mining.

Keywords: *opinion mining, sentiment analysis, social web data, machine learning, social media.*

GJCST-C Classification: *C.2.1,C.2.4, H.2.8*

REVIEWOFFEATURESELECTIONANDOPTIMIZATIONSTRATEGIESINOPINIONMINING

Strictly as per the compliance and regulations of:

© 2016. K. Venkata Rama Rao. This is a research/review paper, distributed under the terms of the Creative Commons Attribution-Noncommercial 3.0 Unported License http://creativecommons.org/licenses/by-nc/3.0/), permitting all non-commercial use, distribution, and reproduction inany medium, provided the original work is properly cited.

Review of Feature Selection and Optimization Strategies in Opinion Mining

K. Venkata Rama Rao

Abstract- Opinion mining and sentiment analysis methods has become a prerogative models in terms of gaining insights from the huge volume of data that is being generated from vivid sources. There are vivid range of data that is being generated from varied sources. If such veracity and variety of data can be explored in terms of evaluating the opinion mining process, it could help the target groups in getting the public pulse which could support them in taking informed decisions. Though the process of opinion mining and sentiment analysis has been one of the hot topics focused upon by the researchers, the process has not been completely revolutionary. In this study the focus has been upon reviewing varied range of models and solutions that are proposed for sentiment analysis and opinion mining. From the vivid range of inputs that are gathered and the detailed study that is carried out, it is evident that the current models are still in complex terms of evaluation and result fetching, due to constraints like comprehensive knowledge and natural language limitation factors. As a futuristic model in the domain, the process of adapting scope of evolutionary computational methods and adapting hybridization of such methods for feature extraction as an idea is tossed in this paper.

Keywords: opinion mining, sentiment analysis, social web data, machine learning, social media.

I. Introduction

Expressions of sentiments are analyzed using some of the prevalent methods like the sentiment analysis or the opinion mining machines. Profoundly the feelings that are generated in human mind is circumstantial on the basis of numerous thoughts, events, occurrence and reactions, which are mostly subjective in nature like mood, emotion. Predominantly the expressions and the state of mind of a human can be envisaged with the combination of certain factors like the emotions, mood and also the bodily gestures like the facial expressions and postures, alongside some of the communication forms.

Opinions and the expressions have significant importance in the day to day living, and focusing on such expressions and opinions have become an integral part of communication and living. With the advent development of social media and digital communication trends, voicing of opinions over such communication medium and people comprehending such inputs has become a norm. Also, with the emergence of BI solutions, and the analytics in place, the emphasis on what is being expressed in the social networking and other online medium by the people.

Pertaining to their likes and dislikes, views and choices has led to conditions, where the opinion mining is gaining significant importance.For instance, the companies are profoundly relying upon such insights for gathering customer views on the brand and the products of the organization, which can be resourceful to them in their strategic planning [1].

In economic factors evaluation, for interpretations that are apart from the technical knowledge and fundamentals based analysis, focusing on diverse range of inputs like the reforms, impending announcements, response to market trends, emerging market conditions, surge or deflation in prices, and many other such factors becomes significant for effective decisions.

In economics and finance to understand beyond fundamental and technical knowledge analysis, sentiment analysis supporters suggest additionally it is essential to use information as diverse as, impending announcements, sudden surge in commodity prices, rumors and reports of a market collapse or break through, increase in the interest rates by central banks, fluctuations in dollar prices, etc. as these factors help in better estimating and forecasting situations of changes in market. In Fig 1 the process flow of opinion mining and sentiment analysis is shown. Such inputs could be gathered from vivid range of sources. Sentiment analysis and opinion mining kind of solutions can be very resourceful in planning such development.

Figure 1: Sentiment Analysis or Opinion Mining model architecture

In the communication process, one of the critical aspects is about expressing about an action, a description in the combination of emotion, mood and sentiment. Also, an affect is also expressed with text and speech that are combined with the descriptions and using some language, which could be very resourceful

Author: B.Sc, MCA, B.Tech(CSE), M.Tech(CSE) Assistant Professor.
e-mail: kvrrao20@gmail.com

© 2016 Global Journals Inc. (US)

in sentimental analysis. Sam Glucksberg [2], in his book "understanding figurative language", for instance, discuss a metaphor "my spouse's lawyer is a shark and my job is my jail" in a characteristic manner, which express an effect in a suggestive and distinct manner.

In terms of expressing an effect, the descriptions [3] are usually used in daily language, which could be popular even in the domain of fictional content and is extend to scientific languages too [4], which are more effective in terms of natural [5]. Also the language in terms of biological sciences [6] is also at times adapted in the expression similar to the usage of some technical aspects pertaining finance and economics too.

Key purpose of a sentiment analysis system is to identify and extract the opinion of an author/speaker and find how astute the author is, towards giving value to something like a value, an object or event state or a concept. For handling the notions towards quantitative semantics, spatial relationships are adapted using the cognitive capabilities of a computing system for emulating a human mind. [7]. The key objective of performing such sentimental analysis is about detecting the instinctive emotional response of a reader for a speaker note and tabulate such responses for effective analysis.

II. State of The Art in Feature Selection and Optimization Strategies of Opinion Mining

In classification the removal of a relevant feature reduces the classification performance. The presence of irrelevant features also affects the classification performance [8]. Redundancy in between features is found by measuring the correlation between features [9] and removal of redundant features improves classification performance. In feature generation features numbering in the thousands are produced and generally a huge number of these features do not give any information because of class specific irrelevancy or redundancy and have to be removed to enhance the classification performance [10][11][8]. A feature is outstanding if it is highly discerning and improves the predictive capabilities of the classifier in assigning a new sample to a class [12] in sentiment analysis.

The performance of the machine learning models depends not only on efficient methods of feature extraction and weights assignment, but also on effective methods of feature selection. Feature selection objective is for selecting most relevant and discriminating features through the removal of features that are noisy and irrelevant for classification. If higher feature vector length and noisy and irrelevant features [13] are present the performance of the machine learning techniques is decreased. To reduce the dimensionality of the feature vector feature selection techniques such as information gain (IG) [14], mutual information (MI) [14], etc., or feature transformation techniques, like singular value decomposition (SVD), LDA, etc. are used. The techniques of feature selection select important features based on a term goodness value and selects top features if they are above a threshold criteria, and by dropping remaining irrelevant features. The techniques of feature transformation are based on converting feature vector of high-dimensionality into a feature space of low-dimensionality where the reduced feature vector similar to the earlier feature vector consists of the contributions of every feature. Wang and Wan [15] performed dimensionality reduction using the technique latent semantic analysis (LSA) and with the reduced feature vector applied the classifier support vector machine (SVM) to enhance the sentiment analysis performance. The popular approach for reducing feature vector length is feature selection compared to feature transformation due to its simplicity and computational efficiency.

A number of feature selection techniques such as MI, IG, DF (document frequency), CHI (chi-square), etc. [16][17][18][19] have been experimented for performing sentiment analysis. A most easy method of feature selection is DF (document frequency). Document Frequency is a regularly used sentiment analysis technique [20] based using the most frequently occurring terms in the texts that are used to construct a feature vector. Tan and Zhang [19] tested the performance of 4 techniques of feature selection (MI, IG, CHI, and DF) in the sentiment analysis of Chinese language documents. They used 5 machine learning algorithms, such as, centroid classifier, KNN (K-nearest neighbor), NB (naive Bayes), winnow classifier, and SVM (support vector machine). The experimental results show, the performance of IG in feature selection is better compared to the other approaches, and SVM algorithm performance is the best. Abbasi et al. [21] experimented with IG and GA (genetic algorithm) using a movie review dataset and proposed a hybrid approach called EWGA (entropy weighted genetic algorithm) to enhance sentiment analysis accuracy. Nicholls and Song [22] developed a new approach for feature selection called document frequency difference and performed comparison of the proposed approach with other techniques of feature selection for sentiment analysis. In [23] to perform sentiment analysis a log-likelihood approach is utilized for selecting key features.

Agarwal and Mittal [24] devised feature selection approach CPPD (Categorical Probability Proportion Difference) selects features based on their relevance and the features that are class discriminative. The feature selection method CPPD or Categorical Probability Proportion Difference combines two different techniques, Categorical Proportional Different (CPD) and Probability Proportion Difference (PPD). Categorical Proportional Different (CPD) technique is based on

© 2016 Global Journals Inc. (US)

measuring the degree of contribution of a term in discriminating the class and in the classification only the topmost contributing terms are used [25]. The PPD technique is based on measuring the degree of belongingness/relatedness or probability that a term belongs to a specific class, and the difference is a measure of the ability to discriminate the class. In classification terms with higher degree of belongingness are chosen as candidate features. The advantage with CPD technique is it measures for a term its degree of class-distinguishing property that is a key attribute for a prominent feature. It is capable of removing terms that are unimportant for classification like terms that occur equally in both classes, and terms with high document frequency like stop words (i.e., a, the, an etc.). The advantage with PPD technique is that it is able to eliminate the terms with low document frequency, such as rare terms and are unimportant for sentiment analysis. Wang et al. [26] devised new Fisher's discriminant ratio-based feature selection technique for sentiment analysis of text data.

Duric and Song [27] introduced a novel feature selection technique based on a model of content and syntax that separates the reviewed entities and the opinion context (i.e., sentiment modifiers). The test performed with these features and using the maximum entropy classifier shows results comparable to that of state-of art methods. Agarwal and Mittal [17] approach for improving sentiment analysis performance is a hybrid approach of feature selection that combines the techniques, information gain and rough sets. Abbasi [28] devised feature selection approach intelligently explores and utilizes the syntactic and semantic information, and demonstrates how a heterogeneous feature set combined with a suitable technique of feature selection could enhance the sentiment analysis performance. O'keefe and Koprinska [29] proposed two techniques of feature selection, SentiWordNet Subjectivity Score (SWNSS) and SentiWordNet Proportional Difference (SWNPD). In sentiment analysis, SWNSS technique is capable of differentiating objective terms from subjective terms as only subjective terms may hold the sentiment, and SWNPD technique is capable of integrating feature selection with the ability to discriminate classes. Verma and Bhattacharyya [30] devised approach first prunes the data of terms that are semantically unimportant using a semantic score derived with SentiWordNet [31], and then information gain is used for selecting key features to enhance the accuracy of the sentiment classification.

Several feature selection techniques have been applied in the context of sentiment analysis such as, MI, DF, CHI, IG, etc. These techniques of feature selection have demonstrated to be efficient for performing sentiment analysis. They remove noise and redundancy occurring in the feature vector directly enhancing sentiment analysis performance in terms of execution time and accuracy. The main effort of the current models of feature selection is for selecting features relevant for the sentiment classification, and not the redundancy between the features.

a) Heuristic Computation based Feature Selection Strategies

Balahur et.al [32] has focused on comparative study of varied methods and resources that are used for mining opinions. Despite the level of complexities involved, the crux of the study is about the evaluation metrics adapted in terms of using the annotated quotations from varied news that are offered by EMS news gathering engine. The study concludes that the generic opinion mining systems needs the large lexicons and also the specialized training or test data which could make impact on the accuracy levels of the models.

Bo et.al [33] has reviewed model of techniques and approaches which shall directly support opinion-oriented information-seeking systems. The study focused on methods which seek new challenges that are raised by the sentiment-aware applications, when compared to the ones that already have traditional models of fact-based analysis.

In the process of review, some of the discussions related to available resources, datasets that are of benchmark and also the evaluation campaigns were also chosen. The availability and popularity of varied opinion-rich resources like the online review sites and also the personal blogs, emerging opportunities and the challenges that are creeping up from the extensive adaptation of ICT trends are utilized for understanding the opinions. The level of surge in the opinion mining and sentiment analysis manages with computation treatment of opinion, text subjectivity and sentiments.

Xu et.al [34] proposed model towards improving aspect-level opinion mining towards conducting online customer reviews. The new generative topic model of JAS (Joint Aspect/Sentiment) model was proposed to extract aspects and the aspect-dependent sentiment lexicons that could be derived from the online customer reviews. Among the aspect dependent sentiment lexicon indicates to aspect-specific opinion words comprising aspect-aware sentiment polarities pertaining to specific aspects. Also, the extracted aspect-dependent sentiment lexicons which are applied to opinion mining tasks at the aspect-level comprising various factors like aspect identification, aspect-based extractive opinion summarization and the aspect-level sentiment classification factors. Experimental study of the JAS model depicts the efficiency and effectiveness of the model in an intrinsic manner.

Emerging developments like the Web 2.0 and the social media content has triggered the scope of generating exhaustive range of content like the opinions,

© 2016 Global Journals Inc. (US)

views, comments, reviews on varied socioeconomic factors, personal and lifestyle trends and market reviews. And the significant scope of focusing on such areas for sentiment analysis has been proposed by Zhang et.al [35].

If rightly adapted the online discussion forums, social networking postings and millions of tweets that are raised by the people can certainly support in gathering intrinsic insights in to varied emerging trends; Despite the fact that Opinion mining is a sub discipline of data mining models, still the level of computational and extraction techniques shall be very resourceful in gathering significant quantum of insightful inputs. Usage of sentiment analysis in the opinion mining can be very effective in terms of identifying the sentiment, affect and subjectivity, and also the emotional states using the online text. Zhai et.al [36] has proposed a model which focus on identifying product feature even before collecting the opinions of both positive and negative, for producing a summary of good or bad points. Aggregator sites of reviews and the ecommerce sites are some of the classic examples for business that is reliant on opinion mining for producing feature-based products quality summaries.

O'Keefe and Koprinska [29] have proposed the model of systematic evaluation for feature selectors and the feature weights using Naïve Bayes and the Support Vector Machine Classifiers. The mode inducts two new feature selection models and also proposes three feature weighting methods that could be adapted. Sentiment analysis focuses on whether the opinion in a document has to be positive or negative for a topic. The study also discuss that though there are many sentiment analysis applications that are proposed, still not many of them has the scope for handle large volume of features.

Abbasi et.al [21] has proposed the model of using the sentimental analysis methods for classifying the web for opinions in varied languages. For sentiment classification in both English and Arabic content, stylistic and syntactic feature extraction components are focused upon. Using the model of EWGA (Entropy Weighted Genetic Algorithm), a hybridized model of genetic algorithm for incorporating the information gain in terms of heuristic towards feature selection is also proposed in the study. It is imperative from the experimental study results of EWGA that it has outperformed in terms of features and techniques utility towards addressing document level sentiment classification.

Swaminathan et.al [37] proposed the model of using unstructured text being used as primary means for publishing biomedical research results. To extract and integrate data, text mining process can be applied in a routine manner. Some of the key aspects in the process is to extracting relationships between bio-entities like the foods and diseases. Also, the studies depict on stop short of how extracted relationships like the polarity and certainty levels could have impact.

Jeong et.al [38] has proposed the model of FEROM which focus on the process of feature extraction and refinement, and correct features towards reviewing data using exploitation of grammatical properties and also the semantic characteristic for feature words. The experimental studies of FEROM model depict that the process has been producing results in more accurate way and relative to functional opinion mining process.

Ozkis and Babalik [39] sounded the model of A-ABC (Accelerated Artificial Bee Colony) in which the two modifications are predominantly used on ABC algorithm for ensuring local search availability and also the levels of convergence speed. Modifications in the stream are depicted as Step Size (SS) Modification Rate (MR). Performances of A-ABC and the standards ABC are compared to varied benchmarking functions have resulted with better efficiency and outcome.

In extension to the models many more models similar to the A-ABC model has evolved. For instance, the models like MABC (Modified Artificial Bee Colony) [40], and another novel feature selection method proposed in [41] [42] also focus on optimal feature subset configurations and has achieved effective levels of classification accuracy when compared to the benchmarking models.

b) Feature Weighting and Representation Schemes

Feature weighting methods are of high importance for the approaches based on machine learning. These methods assign weights to features in terms of the importance levels of sentiment for enhanced classification of the sentiments data. There are different methods for assigning weights like, term frequency (TF), term-frequency-inverse document frequency (TF-IDF), binary weighting scheme etc. [18]. The binary weighting scheme assigns a feature value equal to 1 in the presence of a term else a value of 0 is given. The TFIDF method assigns weights to every term in terms of how rarely the terms occur in the remaining documents [14].

A comparison of these methods shows that for a maximum number of topic-based text classifications, TF-IDF weighing method performs efficiently, however for sentiment analysis, binary weighting method outperforms the other frequency-based methods [20]. An explanation for this may be because people write reviews usually using different sentiment words of expression. E.g. a person reviews a camera as, "This Nikon camera is great. Picture quality is clear and it looks nice." In this review 3 different sentiment words are used that are "great," "clear," and "nice." There are very less chances that same review will be written by him using only a single sentiment word "great", "this Nikon camera is great. Picture quality is great and its looks are also great." So the presence/absence of a

© 2016 Global Journals Inc. (US)

term has a higher importance more than its frequency of occurrence in sentiment analysis.

Deng et al. [43] devised a method to weight a term based on its importance in the document and the term importance in expressing sentiment for sentiment analysis based on supervised approach. Martineau and Finin [44] devised a new method of finding feature weights called Delta TF-IDF that measures feature importance of a term based on its class distinguishing property and the terms that are unevenly distributed between positive and negative classes are assigned more weights with evenly distributed feature assigned a zero value. If a feature is more unevenly distributed among classes, the feature must be having more importance. In the sentiment analysis Delta TF-IDF outperformed the methods of term frequency and TF-IDF weighting.

Dai et al. [45] approach highlights the sentiment feature through an increase of their weights. Additionally with the bagging technique multiple classifiers are built on several feature spaces and a combination of all of these creates one aggregate classifier that is used to enhance the sentiment analysis performance. Paltoglou and Thelwallm [46] produced a complete study of several weighting techniques used in sentiment analysis. In the experiments performed by them, the classic $tf - idf$ variant methods showed to improve the sentiment analysis performance. Authors in [47] developed an improved mutual information based feature weighting method that assigns terms weights using sentiment scores. In the tests performed by them MI based weighting method is seen to be having higher effectiveness over the other methods. A few researchers have worked on including the words positional information in the text. One such research work by Raychev and Nakov [48] introduced a novel language-independent method of calculating term weights that uses the word position and its possibility of being subjective. Further multinomial naive Bayes is applied using the position information of the word in a feature set. This approach is tested with a movie dataset and the objective is to show that a particular word based on its position of occurrence in a document may have varied subjective power. O'Keefe and Koprinska [29] explored several feature weighting methods such as, feature frequency, feature presence (FP), and TFIDF, and based on the words grouped by their SentiWordNet (SWN) values, devised three new feature weighing methods, SWN Word Score Group (SWN-SG), SWN Word Polarity Groups (SWN-PG), and SWN Word Polarity Sums (SWN-PS).

The objective of the research for devising different weighting methods is of utilizing the information of feature polarity for assigning feature weights. The literature has numerous weighting methods produced such as, TF, binary weighting scheme, TF-IDF, etc. and an evaluation of these various weighing methods for sentiment analysis has shown binary weighting method to be the best.

III. Conclusion

In this paper, emphasis is on methodical approach of reviewing the varied models of sentimental analysis and opinion mining solutions that were earlier proposed in the studies. With the emerging trends of digital communication trends, adaptation of ICT trends and huge volume of opinion related data that is generated, the focus on sentiment analysis and opinion has been high in terms of research that is carried out. Review of the models of feature selection strategies and optimization techniques that are adapted in the machine learning based sentiment analysis depict that more models has been focused in such dimension. It is imperative from the review that due to some of the constraints like comprehensiveness and the knowledge of the problem, performing the tasks of sentiment analysis are much complex. Also, the issues like natural language processing limitations are also the other factors that impact the outcome of sentiment analysis. Despite the fact that considerable developments have taken place in the domain, still there is potential scope for improvement. One of the potential solutions that could be considered is about using the scope of evolutionary computational methods and adapting hybridization of such methods for feature extraction towards developing effective sentiment analysis model.

References Références Referencias

1. Candillier, Laurent, Frank Meyer, and M. Boullé. 2007. Comparing state-of-the-art collaborative filtering systems. International conference on Machine Learning and Data Mining MLDM 2007, Leipzig/Germany. Lecture Notes in Computer Science, 2007, Volume 4571/2007:548– 562. Berlin: Springer.
2. Glucksberg, Sam. 2001. Understanding figurative language: From descriptions to idioms. Oxford: Oxford University Press.
3. Goatly, Andrew. 1997. The language of descriptions. London: Routledge.
4. Miller, A.L. 1984. Imagery in scientific thought. Boston, Basel, Stuttgart: Birkhaeuser.
5. Pullman, B. 1988. The atom in the history of human thought. Oxford: Oxford University Press.
6. Verschuuren, G.M. 1986. Investigating the life sciences: An introduction to the philosophy of science. Oxford: Pergamon Press.
7. Hobbs, Jerry R.1990. Literature and cognition. Lecture notes, number 21, Center for the Study of Language and Information, Stanford, California.
8. Yu L, Liu H (2004) Efficient feature selection via analysis of relevance and redundancy. J Mach Learn Res 5(1): 1205–1224.
9. Peng H, Long F, Ding C (2005) Feature selection based on mutual information: criteria of max-depen-

© 2016 Global Journals Inc. (US)

dency, max-relevance, and min-redundancy. IEEE Trans Pattern Anal Mach Intell 27(8): 1226–1238.
10. Agarwal B, Mittal N (2013) Optimal feature selection for sentiment analysis. In: Proceedings of the 14th international conference on intelligent text processing and computational linguistics (CICLing 2013), vol 7817, no 1, Samos, pp 13–24.
11. Forman G (2003) An extensive empirical study of feature selection metrics for text classification. J Mach Learn Res 3(1): 1289–1305.
12. Hoque N, Bhattacharyya DK, Kalita JK (2014) MIFS-ND: a mutual information-based feature selection method. Expert Syst Appl 41(14): 6371–6385.
13. Aphinyanaphongs Y, Fu LD, Li Z, Peskin ER, Efstathiadis E, Aliferis CF, Statnikov A (2014) A comprehensive empirical comparison of modern supervised classification and feature selection methods for text categorization. J Assoc Inf Sci Technol 65(10): 1964–1987.
14. Manning CD, Raghvan P, Schutze H (2008) Introduction to information retrieval. Cambridge University Press, Cambridge.
15. Wang L, Wan Y (2011) Sentiment classification of documents based on latent semantic analysis. In: Proceedings of the international conference on advanced research on computer education, simulation and modeling (CESM), Wuhan, pp 356–361
16. Agarwal B, Mittal N (2012) Text classification using machine learning methods-a survey. In: Proceedings of the 2nd international conference on soft computing for problem solving (SocPros-2012), vol 236, no 1, Jaipur, pp 701–710.
17. Agarwal B, Mittal N (2013) Sentiment classification using rough set based hybrid feature selection. In: Proceedings of the 4th workshop on computational approaches to subjectivity, sentiment and social media analysis (WASSA'13), NAACL-HLT, Atlanta, pp 115–119.
18. Pang B, Lee L (2008) Opinion mining and sentiment analysis. Foundations and trends in information retrieval, vol 2, no 1–2. Now Publishers, Hanover, pp 1–135.
19. Tan S, Zhang J (2008) An empirical study of sentiment analysis for chinese documents. Expert Syst Appl 34(4): 2622–2629.
20. Pang B, Lee L, Vaithyanathan S (2002) Thumbs up? Sentiment classification using machine learning techniques. In: Proceedings of the conference on empirical methods in natural language processing (EMNLP), Prague, pp 79–86.
21. Abbasi A, Chen H, Salem A (2008) Sentiment analysis in multiple languages: feature selection for opinion classification in Web forums. ACM Trans Inf Syst 26(3): 1–34.
22. Nicholls C, Song F (2010) Comparison of feature selection methods for sentiment analysis. In: Proceedings of the 23rd Canadian conference on advances in artificial intelligence. LNCS, vol 6085, no 1, Ottawa, pp 286–289.
23. Gamon M (2004) Sentiment classification on customer feedback data: noisy data, large feature vectors, and the role of linguistic analysis. In: Proceedings of the 20th international conference on computational linguistics, Geneva, pp 841–848.
24. Agarwal B, Mittal N (2012) Categorical probability proportion difference (CPPD): a feature selection method for sentiment classification. In: Proceedings of the 2nd workshop on sentiment analysis where AI meets psychology, COLING 2012, Mumbai, pp 17–26.
25. Simeon M, Hilderman R (2008) Categorical proportional difference: a feature selection method for text categorization. In: Proceedings of the 7th Australasian data mining conference, Glenelg, pp 201–208.
26. Wang S, Li D, Song S, Wei Y, Li H (2009) A feature selection method based on Fisher's discriminant ratio for text sentiment classification. In: Proceedings of the international conference on web information systems and mining (WISM), Shanghai, pp 88–97.
27. Duric A, Song F (2011) Feature selection for sentiment analysis based on content and syntax models. In: Proceedings of the 2nd workshop on computational approaches to subjectivity and sentiment analysis, ACL-HLT, Portland, pp 96–103.
28. Abbasi A (2010) Intelligent feature selection for opinion classification. IEEE Intell Syst 25(4): 75–79.
29. O'keefe T, Koprinska I (2009) Feature selection and weighting methods in sentiment analysis. In: Proceedings of the 14th Australasian document computing symposium, Sydney, pp 67–74.
30. Verma S, Bhattacharyya P (2009) Incorporating semantic knowledge for sentiment analysis. In: Proceedings of the international conference on natural language processing (ICON), Hyderabad, pp 1–6.
31. Esuli A, Sebastiani F (2006) SentiWordNet: a publicly available lexical resource for opinion mining. In: Proceedings of 5th international conference on language resources and evaluation (LREC), Genoa, pp 417–422.
32. Balahur, A., Steinberger, R., Goot, E. V. D., Pouliquen, B., and Kabadjov, M. "Opinion mining on newspaper quotations", In Web Intelligence and Intelligent Agent Technologies, Vol. 3, pp. 523-526, 2009.
33. Bo, Pang., and Lillian, Lee. "Opinion Mining on Newspaper Quotations", Foundations and Trends in Information Retrieval, Vol.2, No.1-2, pp. 1–135, 2008.
34. Xu, X., Cheng, X., Tan, S., Liu, Y., and Shen, H. "Aspect-level opinion mining of online customer

© 2016 Global Journals Inc. (US)

reviews", Communications, China, Vol.10, No.3, pp.25-41, 2013.
35. Zhang, Y., Yu, X., Dang, Y., and Chen, H. "An Integrated Framework for Avatar Data Collection from the Virtual World: A Case Study in Second Life", IEEE Transaction on Intelligent Systems, Vol.25, No.6, pp.17-23, 2010.
36. Zhai, Z., Liu, B., Wang, J., Xu, H., and Jia, P. "Product Feature Grouping for Opinion Mining", IEEE Transaction on Intelligent Systems, Vol.27, No.4, pp.37-44, 2012.
37. Swaminathan, R., Sharma, A., and Yang, H. "Opinion mining for biomedical text data: Feature space design and feature selection", In Proceedings of the 9th International Workshop on Data Mining in Bioinformatics, 2010.
38. Jeong, H., Shin, D., and Choi, J. "Ferom: Feature extraction and refinement for opinion mining", ETRI Journal, Vol.33, No.5,pp.720- 730, 2011.
39. Ozkis, A., and Babalik, A. "Accelerated ABC (A-ABC) Algorithm for Continuous Optimization Problems", Lecture Notes on Software Engineering, Vol.1, No.3, pp.262- 266,2013.
40. Anandhakumar, R., Subramanian, S., and Ganesan, S. "Modified ABC Algorithm for Generator Maintenance Scheduling", International Journal of Computer Science and Electrical Engineering, Vol.3, No.6, pp.812- 819, 2011.
41. Palanisamy, S., and Kanmani, S. "Artificial Bee Colony Approach for Optimizing Feature Selection", International Journal of Computer Science, Vol.9, No.3, pp. 432-438, 2012.
42. Karaboga, D., and Ozturk, C. "Fuzzy clustering with artificial bee colony algorithm", Scientific research and Essays, Vol.5, No.14, pp.1899- 1902, 2010.
43. Deng ZH, Luo KH, Yu HL (2014) A study of supervised term weighting scheme for sentiment analysis. Expert Syst Appl 41(7):3506–3513.
44. Martineau J, Finin T (2009) Delta TFIDF: an improved feature space for sentiment analysis. In: Proceedings of the third AAAI international conference on weblogs and social media, pp 258–261.
45. Dai L, Chen H, Li X (2011) Improving sentiment classification using feature highlighting and feature bagging. In: Proceedings of the 11th IEEE international conference on data mining workshops (ICDMW), Vancouver, pp 61–66.
46. Paltoglou G, Thelwallm M (2010) A study of information retrieval weighting schemes for sentiment analysis. In: Proceedings of the 48th annual meeting of the Association for Computational Linguistics, Uppsala, pp 1386–1395.
47. Lin Y, Zhang J, Wang X, Zhou A (2012) An information theoretic approach to sentiment polarity classification. In: Proceedings of the 2nd joint WICOW/AIRWeb workshop on web quality, Lyon, pp 35–40.
48. Raychev V, Nakov P (2009) Language-independent sentiment analysis using subjectivity and positional information. In: Proceedings of the international conference recent advances on natural language processing (RANLP), Borovets, pp 360–364.

© 2016 Global Journals Inc. (US)

This page is intentionally left blank

© 2016 Global Journals Inc. (US)