

Bayesian Sentiment Analytics for Emerging Trends in Unstructured Data Streams

Najam us Sahar¹, Muhammad Sohail Irshad¹, Muhammad Adnan Khan^{2,*}

¹Department of Computer Science, Minhaj University Lahore.

²School of Computer Science, National College of Business Administration and Economics Lahore, Pakistan.

Abstract

Today the computational study of people's opinion expressed in free form written text is called the field of sentiment analysis and opinion mining. Various research areas such as Natural Language Processing, Data Mining, Text Mining lie in field of Sentiment Analysis and is also becoming major part of importance to organizations because of online commerce is included in their operational strategy. Due to excess of user's comments, feedback on web there is a need to analyze the user generated text. This research focuses on aspect level sentiment analysis in which identification of aspects and their related sentiments is being done. Opinion analysis helps to identify the polarity of the text and feature extraction. This study is being done to provide an effective and efficient framework to calculate the sentiments of written text by using Naïve Bayes approach. For sentiment analysis dataset of 1060 reviews of different restaurants from online website TripAdvisor.com is being used. The outcome achieved good accuracy 80.833 percent.

Keywords: Naïve Bayes, Bag of Words, Bag of Nouns, Term Frequency, Inverse Document Frequency.

Received on 06 May 2019, accepted on 28 May 2019, published on 04 July 2019

Copyright © 2019 Najam us Sahar *et al.*, licensed to EAI. This is an open access article distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/3.0/>), which permits unlimited use, distribution and reproduction in any medium so long as the original work is properly cited.

doi: 10.4108/eai.13-7-2018.159355

*Corresponding author. Email: madnankhan@ncbae.edu.pk

1. Introduction

Today the organizations have become very conscious about their customer's reviews toward their product or service. The opinions of the customer play an important role in an organization's image or in its value of demand for their products or services. It has different impacts for both non-profits, profit organizations such as for non-profits it generates the taxpayer support and political support. While in the later one it generates the revenues and increases market place. In short, customer's reviews are key to the success of any organization. In order to achieve customer's reviews, there are some methods adopted such as surveys, focus groups, observation etc. Usually, these are expensive and require more labor than available. As the internet has

become a very huge resource from which everyone can get survey tools readily available but accuracy and relevancy in data is still a difficult task to achieve. There are many online sites that are serving in providing customer's opinions about specific product or service of a company such as Yelp, Amazon etc. Online customer's reviews are very large in scale sometimes hundreds of thousands and continuously increases as the number of viewers increase. Opinions are very important and also influencers to whole human activities. It is a reality that beliefs and perceptions depend upon other's way of evaluation and seeing. For this purpose, the point of views of other people are found out which are helpful in making the decision not only for individuals but also for huge organizations. A feeling, an experience, an attitude or an opinion is called sentiment. Sentiment analytics includes several techniques such as Natural Language Processing (NLP), Data Mining (Data mining), Text Mining (TM), these techniques have become

more popular in organizations because of their ability to integrate online commerce. [1] Users or buyers want to know enough information about a product or service before to get it, user's reviews are more helpful in this case but it is difficult to read and analyze all reviews as it is a time-consuming and tiresome process. Similarly, an organization wants to know public comments about its product or a service that is provided by it and if they have to market or sell their product, prediction of new sale's trends and for managing its reputation they have to face several user's comments. In order to manage and analyzing such a huge available data and extract useful information/opinions. Sentiment analytics is a technique that will help in doing these tasks. This is the major reason that sentiment analysis is a broad topic today that has converted from computer science field to management and field of social sciences. It has many applications in different terms such as opinion oriented search engines, summarization of scrutiny and for correcting errors of wrong selection in user's ratings. There is a huge amount of documents that contain opinions regarding an issue. Information extraction systems, question answering systems, recommendation systems are another form of sentiment analytics. Today's study focused on sentiment analysis that deals with opinions which are in free text form and are processed so that the opinionated information is achieved. It has different sort of way to work than mining and information retrieval. Questionnaires are another form of achieving customer's overall gratification about any product or service. But in fact, it is expensive. Sometimes there is a restriction from law to public agencies for collection of these questionnaires. In its solution, there are online sites available which provide online services to its users like TripAdvisor. Due to all the above cited facts there is a need to develop a system that enhances scrutiny on products or services and provides clear view of them. [2]

1.1 Research objectives and applications

Using Bayesian approach create an effective model which is used to model the overall contributor's gratification/satisfaction. To examine various tourist websites that are rejoinder of contributor's can be secure by using sentiment analysis.

There is a need to make more clear system in sentiment analysis to enhance the scrutiny on products or services from different clients and to enhance the usefulness of aspects & opinion extraction.

Using ideology perform aspect level sentiment analysis We can know better results of products or services having various reviews that is it liked or not, how many

comments are in favor and also what are the good or bad points?

Using linguistic techniques identify relationships between extracted aspects and opinions.

This research helps in analyzing aspect level sentiments of customer's review about any product or service by implementing Bayesian approach to model overall satisfaction of a contributor/customer.

Use of sentiment classification to predict the customer's entire gratification and aspect identification. These methods are helpful for managers in assessing the importance of features that could be used in the derivation of customer's contentedness.

It will help to generate meaningful summary of opinions and individuals to take benefits from it to make their purchasing ability or decisions about services better.

It will also helpful for biggest organizations and companies in their selling decisions.

2. Paramount Stumbling Blocks

In This section contains the problem encountered in the sentiment analysis as well as in opinion mining and particularly in aspect level sentiment analysis.

2.1. Synonyms and Polysemy: Sometimes people use the same word which refers to different things and some words have several meanings. Synonymy and polysemy are two terms that are difficult to handle in sentiment analysis and opinion mining. There are many techniques adopted to resolve these issues.

2.2. Acerbity: It is difficult for humans as well as for machines to find and deal with acerbity. If the machine has the ability to identify acerbity in a reliable way then it can enhance the performance of tasks that are related to natural language processing, sentiment analysis, and opinion mining are also included. Acerbity can be defined like this, it is in which actual meaning is opposite to the predetermined.

2.3. Composite/Compound Sentences: In these sentences there are two sentences that are independent and combined by a conjunction. The conjunctions and, or, such as, but, for, semicolon also used in compound sentences. In sentiment analysis and opinion mining, it is difficult to handle compound sentences. For instance, "everyone shares posts on Facebook but I don't like to share". "Tomorrow day was very hot, but at night the weather was pleasant".

2.4. Raw and free/form data: The opinion holder is free to comment on any object or entity, which is

in free form text. Its transformation is done into semi-structured data. Semi-structured data is data which consists of tags and markers, that separate semantic factors. It's not in the form of raw material.

3. Related Work

This section contains the previous research work had done related to this topic and describes relevant techniques and methodologies.

In DL whole document is analyzed and overall polarity of the document's review is predicted and studied. In Aspect Level Sentiment Analysis (ALSA). First of all the aspects of the entity/object are identified, then the related sentiments must be discovered. There are many researchers who have been worked on the ALSA.[3]

The confrontation of the conversion of unstructured free form text to structured form a technique that is called unsupervised is used. In this, aspects are known prior and extraction of these aspects from unstructured text is done. The aspects of a specific product or a service that may influence the contributor's overall gratification must be known in advance. [4]

Zhu et al lodged Aspect Based Opinion Mining (ABOM) polling from an open textual customer scrutiny. In his research he used a model aspect based segmentation that separates multi aspect sentence and convert it into single aspect unit used for opinion polling. Behind this a major technique is involved multi aspect bootstrapping method and used for declaration of aspect having aspect related terms. Their research was on the Chinese restaurant scrutiny by utilizing the opinion polling algorithm. [5]

Chaovalit and Zhou had a research on movie review mining the method used was machine learning and semantic orientation. To classify the movie scrutiny in suggested ML some other techniques were used supervised classification and text classification. To illustrate the data in documents a corpus is composed. Classifiers are created by using that corpus. The suggested methodology is more coherent. [6]

Ghose and Ipeiritis had a research based on the product scrutiny using manufacturer ranking mechanism and were customer oriented. For ranking the expected helpfulness of the review is utilized and that ranking based on expected effect on sale. This suggested methodology depicts the scrutiny that has more impact. Affirm the information that the product contains and reviews with subjective point of view which is beneficial for experience goods. Econometric analysis with text mining and subjectivity analysis is used in this suggested methodology. Amazon.com is used as a source to assemble data such as product prices and sales ranking. [7]

Neviarouskaya et al there is another case in which an opinion may have +ve and -ve sentiments. The situation of this kind in which an opinion has both +ve and -ve sentiment can be treated as a simple individual sentiment words but they need domain knowledge for the determination of their related orientations. Many researchers had been worked on this situation. Researchers introduced compositional rules such as sentiment reversal, aggregation, neutralization, propagation, intensification. They defined aggregation differently but it is same as the confliction. They proposed rule based approach to check the affect recognition from written text. [8]

Hu and Liu stated that for mapping aspects that are implicit. Hu and Liu described that there are two types of aspects one is implicit and second is explicit. But the explicit type of aspects are commonly dealt. Explicit aspects are known to be as the noun and noun phrases for instance, image quality in the sentence "the image quality of this LCD is very nice". While in implicit type all other expressions that must represent aspects lie. The implicit types are adjectives and adverbs. As the adjectives show the properties and attributes of the objects for instance cheap shows the "price" and ugly shows appearance. Verbs also lie in implicit type. For instance "this scenery will not be fit in a frame". While the words "fit in a frame" show the size of the entity. [9]

Hai et al there was another researcher who proposed a method for matching implicit aspects (sentiment words) with explicit features/aspects this method is known as two-phase co-occurrence association rule mining (TPCARM). Including each individual implicit, explicit aspect as a condition and consequence respectively this method generates association rules in 1st phase. Explicit aspects are clustered to activate more rules for implicit aspects in the 2nd phase. When a sentiment word is given without explicit aspect then testing is done and also for application. In this case it searches out more accurate rule cluster and its related word is determined as last aspect. [10]

Yu et al another researcher proposed a technique to organize the aspects of the specific product based on the knowledge of its domain and structure. The method includes different hierarchies, classifications and actual reviews of products to find out final aspect hierarchies. With the combination of a strategy set of distance measures is also utilized. A hierarchical organization is generated based on the derived hierarchy and consumer/customer reviews in which many aspects and their aggregation on aspects also generated. With the help of this hierarchical order people can get the

over review of customers on different aspects and also find out opinions on particular aspect. [11]

Lin et al had a research, they showed the application of sentiment analysis on twitter data and twitter's connection. They showed how sentiment analysis queries run. They performed different experiments on various queries. They implemented twitter sentiment analysis with different APIs of twitter. They showed that the neutral sentiments for tweets are higher. [12]

Ali et al had a research on different sentiment analysis techniques. They applied these approaches in the context of tourism. Their purpose of the study is to scrutinize and review different sentiment analysis techniques in tourism. For the identification of sentiment analysis in the context of tourism they used some combinations of keywords such as "sentiment analysis of tourism," "tourism sentiment data" on Google rather than using websites that relates to web science in order to searching and retrieving articles that are of interest and published on internet. [13]

Seghal and Agarwal performed sentiment analysis techniques of Big Data. They used Hadoop framework by using Twitter Data. As the Twitter is the very large social media which produces largest amount of data so it is difficult to handle and analyze such amount of data. So they used Hadoop for making the process of data analyzation easier. [14].

Sharma et al used sentiment analysis approach for live tweets. They presented web based application to visualize the current sentiments that associate with keywords such as #tag or any phrase that are used in Twitter's messages and plot them on a particular map. It helps in measuring the sentiment and mapping intensity of that sentiment with respect to the geography. Their purpose is to build such a system that provides services as an end to end system for analyzing sentiments of tweets.[15]

Shahid et al presented the overview of Opinion Mining and Sentiment Analysis in terms of its both technical and non-technical aspects. They called attention to both the technical and non-technical terms of Opinion Mining and Sentiment Analysis in the form of confrontations. Their purpose is to make the techniques of Opinion Mining and Sentiment Analysis more understandable in terms of human applications for text analysis. They used systematic literature review methodology. This technique differentiates traditional reviews with the reviews that are generated with this systematic technique. [16]

Min Peng et al proposed a model named Neural Sparse Topical Coding which focused on the replacement of the complex inference with back propagation. It makes the system able to exploring more extensions. In addition external semantic meaning of words added in the word embedding for improving the short text. They presented three extensions based on NSTC for showing the flexibility in Neural Network based systems and frameworks. They didn't reassumed inference algorithms. [17]

Qianqian Xie et al suggested a model named Bayesian Sparse Topical Coding with Poisson Distribution which differs from conventional STMs. While the STMs are the Sparse Topic Models used for understanding and learning semantic meaning of short texts. In this technique hierarchical sparse is imposed between different sparse coefficients. They also suggested sparsity-enhanced named Bayesian Sparse Topical Coding in addition with Normal Distribution by using mathematical approximation. They adopted superior hierarchical sparse for achieving the optimal solution. They used datasets of Newsgroups and Twitter and showed that both techniques are helpful in finding more clear semantic representations. [18]

Min Peng et al suggested Block Bayesian Sparse Topic Coding is used in finding latent semantic representation of the short text. This technique moderates the normalization constraints which lies in inferred indications with words embedding and blocks Sparse Bayesian Learning. It provides convenience to control the sparsity of words by exploiting the intra block correlations. They showed that it is helpful and improves accuracy of document classification and achieved better results on sparsity ratio of word codes.[19]

Jiahui Zhu et al presented Unsupervised Multiview Hierarchical Embedding (UMHE) a technique for revealing topical information and knowledge in social sites, events. They said that the Traditional Topical Methods didn't solve the problem of generating event-oriented topics with textual features and aspects. Their proposed method solves the problem of insufficient information and poor aspect depiction. Their framework firstly developed Multiview Bayesian tree for generating prior information for Latent Topics and relations among them. They proposed that with this prior information they designed Unsupervised Translation-Based Hierarchical Embedding technique for making more improved depiction of Latent Topics. They applied Self-adaptive Spectral Clustering on the Embedding Space. They extracted event-oriented topics in words for expressing social events. Their system was data driven,

unsupervised without any external information. They used TREC Tweets2011 and Sina Weibo as datasets for experiments for demonstrating that their framework can build hierarchical structure with better results. [20]

Jiangang Shu et al describes the problem of leakage of private and confidential information which lies in the context of crowdsourcing. They proposed Privacy-Preserving Task Scheme for crowdsourcing. This technique maintains task-worker matching by preserving and maintaining task privacy and worker privacy. They used in this technique polynomial function for the indication of multiple keywords of task requirements. They also constructed a method which is based on matrix decomposition for matching multiple keywords with multiple requesters. They indicated that with their system user accountability and user revocation gained coherently.[21]

Sudha et al raised the Domestic violence which is cause of concern because of threat it constitutes toward public health and Human rights. They highlighted the need of quick identification of the victims which is suffering from this condition so that Domestic violence crisis service can support in time. As the social media allowed Domestic violence victims to indicate stories and attain support from the high authority community which deals with the victims. They said it becomes difficult to handle manual browsing and large number of posts. They proposed automatic identification of Domestic violence victims. [22]

3. Research Methodology

In this section flow of the system is being described.

3.1. Data Collection: First of all data is being collected which is used for analysis and this is in the form of free text reviews about different restaurants. For the purpose of aspect based sentiment analysis online reviews of different people about different restaurants are used. Reviews are extracted from TripAdvisor official website and converted it into usable form which is used as a dataset for further processing.

3.2. Data Preprocessing: The data must be preprocessed prior to use for extracting some useful information. Data preprocessing consists of many operations such as finding missing values, noise removal, stop words removal, stemming, bag of words etc.

3.3. Feature Selection: First of all in ABOM the aspects and features of a particular product/service are identified about which the users showed their opinion. The clustering over selected sentences is being done for the identification of aspects. For this purpose proposed technique is to use bag of nouns rather than bag of words. Using the technique Bag of words by a researcher did not produce satisfactory result. Further K means algorithm is used for the extraction of similar sentences from the group of multiple sentences that are posted as a review about a particular entity. The usage of Bag of noun enhances the clustering algorithm for aspect identification. It is necessary to inline the sentences in a proper way for implementation of the clustering prior. There are different methods: Term frequency with weighting terms and inverse document frequency. Aspect and Feature recognition: This portion contains the steps for aspect/feature extraction and before applying clustering over sentences some preprocessing procedure are being described one by one step. The term "Review" is used which shows the user's opinion. Clustering is helpful in order to identify the aspects/features of a specific entity about which users generate their opinion, thought in form of text. It is probable that identical/similar sentences in a review are about identical/similar features or aspects, if there is a possibility of finding identical/similar sentences. By indicating individual sentence in appropriate manner with a vector (V) after that implementation of the clustering algorithm to these vectors will help out to describe that sentences in clusters are about the identical aspects.

3.4. Sentiment Recognition: When the aspects are identified then the next step is to find out the sentiments that relates to those aspects. Here machine learning (ML) techniques are used to find out the sentiments. ML based techniques takes sentiments recognition stumbling block as a classification stumbling block. Feature/aspect uprooting, selecting the type of the classifier are 2 major tasks in the design of a classifier. To represent the stumbling block under control properly one must has a set of features in feature recognition step. Then selection of appropriate classifier must be done from existing classifiers.

3.5. Bag of words depiction: First of all bag of words approach is used as the Feature Set (FS) in which entire document is considered in the corpus, Vocabulary List (VL) is built and each individual document is depicted with a vector which represents the presence of a term in the document. Each word of document is converted into a vector representing a numerical value.

3.6. Bag of nouns depiction: Bag of nouns are basically represent the lower case tokens and in this article used later as the new feature set. In this entire document is

considered as a lower case corpus. Term frequency and inverse document frequency are different methods for weighing each phrase. The presence of the word in the document is depicted as 1 and absence is represented as 0. There are term frequency-Inverse document frequency formulas as follows:

$$tf(t,d) = 1 * \frac{f(t,d)}{\max_{t \in d} f(t,d)} \quad (1)$$

$$idf(t,D) = \log \frac{|D|}{|\{d \in D: t \in d\}|} \quad (2)$$

$$tfidf(t,d,D) = tf(t,d) \times idf(t,d,D) \quad (3)$$

In equation (1) f is the frequency of the term t and d is the document in which this term t is formalized by max frequency of t in d. Where |D| in (2) depicts the total number of documents.

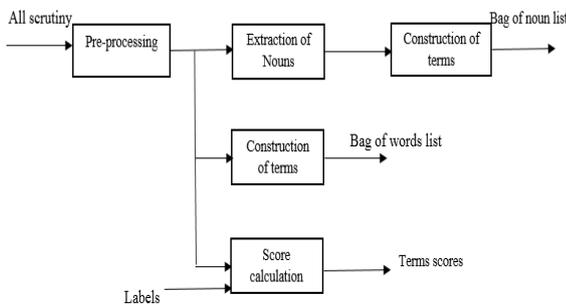


Figure 1(a). Proposed System diagram

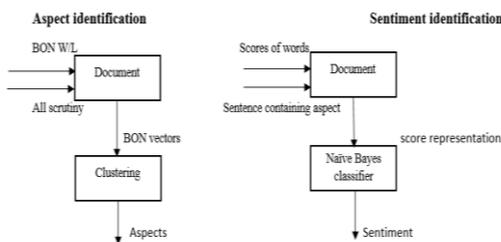


Figure 1(b). Proposed System Diagram

4. Speculative Outcomes

This portion contains speculative outcomes that are achieved by comparing bag of noun clustering with bag of word clustering. Dataset of 1060 reviews is being used and Weka tool used for manipulating the data.

Table 1. Proposed Clustering with BOW

Bag of words cluster instances		
0	598	100%
1	2	0 %

Table 2. Proposed clustering with BON

Bag of noun cluster instances		
0	300	50%
1	300	50%

5. Differentiation between BOW and BON with Naïve Bayes Classifier

It is stated earlier that bag of nouns contains the entire words with nouns and lower terms while in bag of words all upper terms are considered in construction of the word list. BON deals with lower case vectors. BON yields more simple benefits of computation and also helpful to clustering algorithm in finding further clusters. In the below table correctly and incorrectly predictions are shown. There are total 279 positive instances, 288 negative and 33 neutral. This below table shows the correctly and incorrectly predictions that are obtained by applying Naïve Bayes classifier with Bag of words technique.

Table 3. Proposed Confusion matrix of BOW

Classified as	a	b	c
a= positive	210	62	7
b= negative	44	241	3
c=neutral	4	5	24

In the Table 4. Proposed methodology for sentiment identification. This technique correctly classified instances 79%.

Table 4. Proposed Naïve Bayes Classification with BOW

Classes	Precision Recall f1score			TP Rate	FP Rate
	Precision	Recall	f1score		
Positive	0.814	0.753	0.782	0.753	0.150
Negative	0.782	0.837	0.809	0.837	0.215
Neutral	0.706	0.727	0.716	0.727	0.018
Weighted average	0.793	0.792	0.791	0.792	0.174

Table 5 shows the correctly and incorrectly predictions made by proposed Naïve Bayes classifier with Bag of noun approach. There are also same instances used 279 positive, 288 negative, 33 neutral.

Table 5. Proposed Confusion matrix of BON

Classified as	a	b	C
a= positive	221	54	4
b= negative	44	242	2
c=neutral	5	6	22

The below table 6 shows the results achieved by applying Naïve Bayes classification with Bag of noun approach, term frequency and inverse document frequency. With this instances are 80% correctly classified as compare to bag of words

Table 6. Proposed Naïve Bayes Classification with BON

Classes	Precision Recall f1score			TP Rate	FP Rate
	Precision	Recall	f1score		
Positive	0.819	0.792	0.805	0.792	0.153
Negative	0.801	0.840	0.820	0.840	0.192
Neutral	0.786	0.667	0.721	0.667	0.011
Weighted Average	0.808	0.808	0.808	0.808	0.164

6. Outcomes

The proposed framework predicts sentiment with a decent precision. With the Bag of nouns and Naïve Bayes classification instances correctly classified 80.333% and with Bag of words the outcome is 79.1667%. With the extensive dataset, the best precision is as high as 0.80 acquired by Naïve Bayes. This shows that proposed technique is powerful and that descriptive words can be considered as imperative element for opinion investigation. It is trusted that the distinctions in the outcomes are so high that essential test set won't be essential.

7. Discussion

It is not known how to figure out and locate an arrangement of target text and how the classifier will characterize the data consummately. Due to this, the outcomes achieved by assessing measurements to be improbable high. At the end the outcomes are not decisive and further experiments are not performed. By observing the techniques it can be seen that using BON with Naïve Bayes it yields better results and outcomes. Even in general it can yields the better results. In this article small set of data was being used containing 1066 reviews. In the beginning it is not known how proposed approach will scale with real world datasets containing millions of reviews. But if the results are seen from the datasets used, proposed approach could be considered feasible. However, at this size classification takes a long time to complete. Outcomes are so high that essential test set won't be essential.

8. Conclusion

This section contains the study of aspect identification and sentiment identification using three-class bag of noun sentiment classification. For aspect identification it is suggested that use clustering over sentences with bag of nouns and bag of words and compared their results. Outcomes depicts that using bag of noun technique helps in achieving more useful aspects than bag of words. And for the sentiment identification it is suggested to select bag of nouns with weighting schemes term frequency and inverse document frequency with Naïve Bayes classifier instead of BOW. Their comparison also showed above.

9. Future Work

Within the dataset experienced a lot of trouble as well as the objective text. The dataset did not operate as a noise. It would be powerful to prove that proposed system could be efficiently perform. It can be said that this is able to be natural extension of the corpus. Today the data generated by the user have been increasing on the social media sites but its reliability is not assured. Customer gratification is partially known by the organizations and they are continuously working hard to meet customer's requirements. About a particular service/product having a negative review or an experience may leads to a bad impact on an organization. Sometimes people, or an organization may generate fake reviews about another organization in order to attract more customers for competition.

References

- Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1), 1-167
- King, R. (2011). Sentiment analysis gives companies- insight into consumer opinion. *Business Week*, 1
- Pang, B., Lee, L., & Vaithyanathan, S. (2002, +July). Thumbs up?: sentiment classification using machine learning techniques. *Paper presented at the Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*
- Abbasi Moghaddam, S. (2013). Aspect-based opinion mining in online reviews. *Applied Sciences: School of Computing Science*.
- Zhu, J., Wang, H., Zhu, M., Tsou, B. K., & Ma, M. (2011). Aspect-based opinion polling from customer reviews. *IEEE Transactions on Affective Computing*, 2(1), 37-49.
- Chaovalit, P., & Zhou, L. (2005). Movie review mining: A comparison between supervised and unsupervised classification approaches. *Paper presented at the System Sciences, 2005. HICSS'05. Proceedings of the 38th Annual Hawaii International Conference on*
- Ghose, A., & Ipeirotis, P. G. (2007). Designing novel review ranking systems: predicting the usefulness and impact of reviews. *Paper presented at the Proceedings of the ninth international conference on Electronic commerce*
- Neviarouskaya, A., Prendinger, H., & Ishizuka, M. (2011). Affect analysis model: novel rule-based approach to affect sensing from text. *Natural Language Engineering*, 17(1), 95-135.
- Hu, M., & Liu, B. (2004, August). Mining and summarizing customer reviews. *Paper presented at the Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*.
- Hai, Z., Chang, K., & Kim, J.-j. (2011). Implicit feature identification via co-occurrence association rule mining. *Paper presented at the International Conference on Intelligent Text Processing and Computational Linguistics*.
- Yu, J., Zha, Z.-J., Wang, M., Wang, K., & Chua, T.-S. (2011). Domain-assisted product aspect hierarchy generation: towards hierarchical organization of unstructured consumer reviews. *Paper presented at the Proceedings of the Conference on Empirical Methods in Natural Language Processing*.
- Lin, Z., Feng, M., Santos, C. N. D., Yu, M., Xiang, B., Zhou, B., & Bengio, Y. (2017). A structured self-attentive sentence embedding. *arXiv preprint arXiv:1703.03130*.
- Alaei, A. R., Becken, S., & Stantic, B. (2019). Sentiment analysis in tourism: capitalizing on big data. *Journal of Travel Research*, 58(2), 175-191.
- Sehgal, D., & Agarwal, A. K. (2018). Real-time sentiment analysis of big data applications using Twitter data with Hadoop framework. *In Soft computing: theories and applications*. 765-772. *Springer, Singapore*.
- Sharma, N., Pabreja, R., Yaqub, U., Atluri, V., Chun, S., & Vaidya, J. (2018, May). Web-based application for sentiment analysis of live tweets. *In Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age*. 120. *ACM*.
- Shayaa, S., Jaafar, N. I., Bahri, S., Sulaiman, A., Wai, P. S., Chung, Y. W., ... & Al-Garadi, M. A. (2018). Sentiment analysis of big data: Methods, applications, and open challenges. *IEEE Access*, 6, 37807-37827.
- Peng, M., Xie, Q., Zhang, Y., Wang, H., Zhang, X., Huang, J., & Tian, G. (2018, July). Neural sparse topical coding. *In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 2332-2340).
- Peng, M., Xie, Q., Wang, H., Zhang, Y., & Tian, G. (2018). Bayesian sparse topical coding. *IEEE Transactions on Knowledge and Data Engineering*.
- Peng, M., Shi, H., Xie, Q., Zhang, Y., Wang, H., Li, Z., & Yong, J. (2018, May). Block Bayesian Sparse Topical Coding. *In 2018 IEEE 22nd International Conference on Computer Supported Cooperative Work in Design ((CSCWD))* (pp. 271-276). *IEEE*.
- Peng, M., Zhu, J., Wang, H., Li, X., Zhang, Y., Zhang, X., & Tian, G. (2018). Mining event-oriented topics in microblog stream with

- unsupervised multi-view hierarchical embedding. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 12(3), 38.
21. Shu, J., Jia, X., Yang, K., & Wang, H. (2018). Privacy-preserving task recommendation services for crowdsourcing. *IEEE Transactions on Services Computing*.
 22. Subramani, S., Wang, H., Vu, H. Q., & Li, G. (2018). Domestic Violence Crisis Identification From Facebook Posts Based on Deep Learning. *IEEE access*, 6, 54075-54085.