Opinion Mining and Sentiment Analysis in Social Media: Challenges and Applications

Wenping Zhang[1], Mengna Xu[2(✉)], and Qiqi Jiang[3]

[1] School of Information, Renmin University of China, Beijing, China
wpzhang@ruc.edu.cn

[2] School of Economics and Management, Hangzhou Normal University, Hangzhou, China
xumena@163.com

[3] Copenhagen Business School, Frederiksberg, Denmark
qj.digi@cbs.dk

Abstract. It is a widely accepted truth there are great values embedded in the opinion and sentiment expressed by users on social media platforms. Nowadays, it is quite common for researchers or engineers to adopt opinion mining and sentiment analysis techniques to extract enriched emotional information from online text content. However, given the characteristics of social media, such as dynamic, short, informal and context dependent, applying general opinion mining and sentiment analysis techniques originally designed for static long text corpora would lead to serious bias. In many applications, even research that not specialized in opinion mining and sentiment analysis, this problem is ignored unintentionally or unintentionally. Such ignorance may contribute the failure of some designs or unexplainable results. In this paper, we summarized these challenges in social media sentiment analysis. Some potential solutions for these challenges are also discussed. Finally, we also introduced several state-of-the-art techniques in social media sentiment analysis.

Keywords: Opinion mining and sentiment analysis · Social media
Business applications

1 Introduction

There are 2.46 billion social media users worldwide up to 2017 (statistic 2018). There is no doubt that it is beneficial to analyze the content of social media. One of the most import analysis is to extract the opinion and sentiment expressed by users in social media posts. Many attempts haven been to make use of the public orientation expressed in social media to solve real social or business problems, such as election prediction (Agrawal and Hamling 2017), stock market analysis (Bollen et al. 2011) and personalized recommendation (Sun et al. 2015). However, due to characteristics of social media, like dynamic, short, informal and context dependent, it is risky to apply general opinion mining and sentiment analysis techniques directly to social media. There are several points that need special attentions and careful process when dealing with social media content. Ignorance of these points will bring serious bias in the analysis and finally lead to failure of some designs or unexplainable results.

F. F.-H. Nah and B. S. Xiao (Eds.): HCIBGO 2018, LNCS 10923, pp. 536–548, 2018.
https://doi.org/10.1007/978-3-319-91716-0_43

Given that most data in business practices are in text format, extracting emotions from text content plays a leading role in opinion mining and sentiment analysis. As a result, opinion mining and sentiment analysis generally can be considered as a branch of text mining. The situation is the same in social media. When referring to opinion mining and sentiment analysis in social media, we usually talking about extracting emotions from text content. Hence, it is necessary and important to make a deep understanding of text mining techniques before conducting opinion mining and sentiment analysis. However, there are some characteristics in social media content beyond the ability of common text mining. For instance, social media today is full of emoji, which is usually in non-text format but with rich emotions embedded. Emoji was treated as noise and removed directly in previous text mining (Kiritchenko et al. 2014). Apparently, it is not a suitable operation in up-to-date opinion mining and sentiment analysis in social media. Unfortunately, there is no good solution for emoji analysis due to its complexity and dynamics until recently. It is a challenge that we cannot bypass if we hope to make better understanding and usage of social media content.

In this paper, we summarize the points that need special attentions in opinion mining and sentiment analysis in social media. It is known to all that content in social media is short, meaningful and emotion enriched. Apparently, many statistics based opinion mining and sentiment analysis techniques that come from text mining don't work effectively anymore when applying directly to social media. If we treat each post in social media as one single document, and conduct analysis on it, there would be a large chance to get biased and inconsistent results (e.g. with large variance). Synonym is also one of the most well-known challenge in opinion mining and sentiment analysis. An opinion word may express different, even opposite, emotions in different places. Generally, the exact meaning can be deduced according to the context it affiliated. However, the context information is rather deficient when dealing with short social media posts, especially the method treating each post as an independent text fragment.

Finally, some typical applications of opinion mining and sentiment analysis in social media are provided. The use of opinion mining and sentiment analysis as a tool to assistant decisions in real world is not novel. However, the characteristics of social media could bring us unprecedented opportunities to do something important and interesting. For instance, the president election of U.S. in 2016 is one good example to show the power of public orientation analysis in social media. Compared with traditional methods, such as telephone questionnaire, it turns to be much faster and more accurate. We are in the era of social media. Catching opinions from social media could be a cheap, fast and effective way to collect feedbacks from users. In other words, there are great benefits contained in peoples' daily social media posts. The only problem is how to make good use of them.

All in all, this paper can be considered as a review that summarize work related to opinion mining and sentiment analysis. Besides these summarization, categorization and archive of existing work, we also discuss challenges that need special attentions when dealing with social media content. Some existing possible solutions are summarized to assist implementations in real business applications and research. We also provide some typical examples to support our introduction. We believe opinion mining and sentiment analysis techniques in social media could play a more important role in real business applications and research if handled properly.

2 Knowledge Foundation

2.1 Concept Foundation

Before the technical details of opinion mining and sentiment analysis, it is necessary to clarify some related basic concepts. Strictly speaking, there are certain differences between opinion mining and sentiment analysis. One evident difference is that opinion always has a holder and target, while sentiment doesn't have to. Intuitively, opinion mining is a process to identify someone's viewpoint (e.g. agree, disagree) to something. Sentiment analysis is to extract someone's attitude or feeling inside. Sometimes, it would be rather difficult to distinguish an opinion from a sentiment. One simple way is to image the possible responses. If one expression can be answered by "I agree/disagree", it must be an opinion. While if it can be responded by "I share your feeling/sentiment", it should be a sentiment. In most cases, it is not necessary to distinguish these two words explicitly. As a result, in most research and applications, these two terms are used alternatively. In this paper we also consider them as exchangeable if there is no special notifications.

Beside opinion and sentiment, three close terms are also worthy to be noticed: emotion, affection and mood. Compared with sentiment, emotion is more subjective. It can be described as mental activities reflected by degree of pleasure or displeasure. If someone express his/her feelings from the mental level, it would be reveal of emotions. Compared with above three widely used terms, affection and mood are not frequently to see in sentiment analysis yet. Affection can be defined as a state of mind that usually associated with a feeling (e.g. love). Mood usually stands for a conscious state of predominant emotion (e.g. in a good mood today). In most cases, it is unnecessary to distinguish these four terms. As a results, in many applications, these terms are considered as exchangeable during processing.

Nowadays, there is a trend that carrying study from group level to individual level. Along with this trend, emotion extract has become increasingly popular and important. In some specific research or application, it is necessary to extract peoples' emotions rather than general sentiments. For instance, in healthcare research, if we try to detect or predict depressions from users' social media, it is necessary and important to go to these delicate difference between emotion and opinion or sentiment. All in all, although there is no need to distinguish these four concepts in general applications, it is important to know the difference in case of any special situations.

2.2 Technology Foundation

Most business data are in text format (Breakthroughanalysis 2008). When referring to opinion mining, usually we mean extracting emotions from text content. Hence, here we have a brief introduction about text mining techniques. Similar to human comprehension processes, text mining techniques generally start with terms before analyzing the meaning of higher level structure (e.g. sentence, graph and document). Due to the characteristics of natural language, it is quite different for applying text mining techniques on different languages. Two typical ones are English and Chinese. Given that there are white space between sequential words in English, it is quite simple for the tokenization process for English text mining. The basic element of Chinese is word. However, the basic semantic element of Chinese is term, which is a combination of one to several words. Unfortunately, there is no white space for separation in Chinese. Hence, the first challenge in Chinese text mining is term segmentation. The performance of segmentation has a significant influence in the following analysis, e.g. opinion mining. However, it doesn't mean that English text mining is much easier that Chinese text mining. Since there are many derived word in English, it is always a problem to find their exact roots before further analysis. Such process is called stemmer. Without stemmer, same word in different formats would be considered as different ones. It will seriously decrease the performance of statistics based models.

3 Literature Review

3.1 General Techniques

Generally, sentiment analysis approaches can be classified into three categories: lexicon-based approach, linguistic-based approach and machine learning-based approach. In a lexicon-based approach, predefined sentiment lexicons that contain both positive indicators (term) and negative indicators are adopted to extract the number of corresponding sentiment indicators in a given text fragment (e.g. sentence, document) based on string matching (Pang and Lee 2008; Taboada et al. 2011). It has two remarkable advantages: ease of implementation and fast speed. Moreover, due to its intuitive outputs (e.g., a large number of positive terms indicates a strong positive orientation), the lexicon-based approach has been widely adopted in business practices and research (Liu et al. 2010). However, this method may suffer from low recall (the rate of detected targets on the total targets) because it strongly relies on the completeness of sentiment lexicons. Furthermore, it may cause confusions when dealing with synonymy and polysemy issues. One word may have different meanings (even opposite sentiment polarity) in different context. But in a lexicon, the meaning is fixed. In a machine learning-based approach, a set of labeled data (training data) is used to train classifiers to learn "rules" (Witten and Frank 2005). Then, these trained classifiers are used to predict the unlabeled data based on the "rules" they learned (Pak and Paroubek 2010; Pang et al. 2002). Such a process is generally carried on a text fragment level (e.g. sentence, document) instead of a word level. However, the overall learning rules or prediction processes are metaphorically described as a "black box" to users, which results in a dilemma when it comes to explaining or improving the algorithm. The principle of machine learning approach is that entrusting the machines, which is not easy to achieve in real business. In a linguistic approach, researchers have attempted to understand the semantic meaning of text and drew conclusions based on this meaning (Wilson et al. 2005). Such an approach is similar to the process of human cognition. However, due to the complexity and flexibility of human language (e.g., negation, idioms), this approach is not easily to implement in real-world applications.

Some researchers concentrated on improving detection accuracy of opinion mining from technical perspective. State of art machine learning techniques are adjusted and brought into the opinion mining tasks. For instance, Irsoy and Cardie (2014) applied recurrent neural networks (RNNs) for opinion mining tasks. Their experiments showed that the novel RNNs framework outperform previous conditional random field (CRF) based baselines. Similarly, Liu et al. (2015) developed a fine-grained token level opinion mining model that incorporating recurrent neural network. Cambria et al. (2015) adopted a vector space model—AffectiveSpace 2—to conduct concept-level sentiment analysis. Concept-level sentiment analysis concentrates on analyzing semantics of text utilizing web ontologies or semantic networks. Rather than gathering isolated opinion to an item, concept-level analysis allow for the inference of semantic and sentiment information associated with ontology features. In other words, it enables comparison of fine-grained features of items.

Although these advanced techniques could improve the accuracy of opinion mining significantly, these sophisticated models are still rare to see in the real business applications. One of the most important reason for such dilemma relates to model interpretability. In these machine learning models, the mining process is a black box for the users. It will lead to a feeling of "out-of-control", which prevent its applications in real business. As a result, the most widely used methods in real business is still lexicon-based approaches. As aforementioned, it is time and energy consuming to construct and maintain lexicons manually. As a result, low performance is one of the most serious weaknesses lexicon-based approaches must meet. To solve this problem, one possible solution is to construct lexicons automatically. One benefit along with these process is that context information can be incorporated in the sentiment lexicon construction or expansion process. Such context-aware sentiment lexicon, e.g. domain specific lexicon, could make contributions to increase opinion mining accuracy in certain applications. Losts of work has been done in this area. Du et al. (2010) developed a domain-oriented sentiment lexicon construction framework by adapting information bottleneck methods. Lu et al. (2011) proposed an optimization framework to combine different sources of information (e.g. context-dependent sentiment lexicons) by a unified and principled way. Huang et al. (2014) stated their constrained label propagation based domain-specific sentiment lexicon construction method.

Compared with automatic sentiment lexicon construction, one easier way is expanding and updating the existing widely used sentiment lexicon (e.g. SentiwordNet, HowNet, OpinionFinder). In practice, individuals tend to employ terms with similar or same meaning as alternatives in their expressions to avoid repetition. In other words, there will be a higher probability of co-occurrence for terms (words) with same or similar meaning in a short text fragment (e.g. a sentence, or a window). Based on this fact, sentiment lexicons can be expanded and updated incorporating context information. Generally, co-occurrence is not used directly in the expansion due to the problem of noise (e.g. frequently used but meaningless terms). Instead, a more advanced measure point-wise mutual information (PMI) are widely used (Read 2004; Su et al. 2006). PMI was defined as a measure of association between two objects (e.g. word) in information theory (see detail in Church and Hanks 1990). To calculate PMI, usually a slide window is used to catch the co-occurrence. There have been many applications that using PMI to expand sentiment lexicons and enhance analysis performance. For instance, Turney and Littman (2003) adopted PMI to extend positive and negative vocabulary. Khan et al. (2016) incorporated PMI to expand SentiWordNet to improve sentiment polarity detection. Manivannan and Kanimozhiselvi (2017) used PMI based integral classifier to conduct sentiment analysis cross domain.

Sentiment lexicon expansion using PMI could alleviate the domain dependence problem. However, it still suffer the problem of polysemy. The reason is that PMI is based on statistics and everything in statistic model is for certain (one word either belongs to one group or another). To solve this problem, probabilistic model could be one option. One typical method is topic modeling (e.g. latent Dirichlet allocation, LDA). In a topic model, terms with similar characteristics are probabilistically clustered into same topic trough sampling according to their context-aware co-occurrence (Blei et al. 2003). One word could belong to several topics with different probabilities. In this case, the exact meaning of a polysemous word could be inferred probabilistically according to the context it embedded. These unique characteristics of topic modeling are suitable for sentiment lexicon expansion. If the polarity of one topic is determined, then the words compose this topic would have a high probability to share the same polarity. The polarity of a topic can be indicated by a small sample of seeds (e.g. similarity to each kind of seeds). In real application, topic modeling users usually go one step further. Since a document can be seen as a distribution of topics, while a topic can be seen as a distribution of words. Once the polarities of topics are determined, the polarity of a document can be inferred according to the contributions of each kind of topics (positive, negative and neutral) to generate it. This method has become very popular in recent years. Li et al. (2014) proposed a supervised user-item based topic model for sentiment analysis. Nguyen and Shirai (2015) incorporated topic modeling based sentiment analysis on social media and used it for stock market prediction. Cao et al. (2016) designed a visual sentiment topic to extract sentiment from microblogs. Chen et al. (2017) utilized incremental hierarchical Dirichelt process (HDP) to conduct phrase level sentiment detection.

3.2 Challenges

There are several challenges to apply common opinion mining methods to social media. The first challenge refers to the length of the social media posts. Although the volume of social media is huge, each post is rather short. Usually, there are length limitations for each post on social media platforms, e.g. 140 characters for Tweets. Given the statistical information is rather limited in these short text, effective approaches suffer serious performance reduction when applying directly to social media. Generally, there are two ways to deal with these kind of short text. First, extracting the sentiment of each post, then calculate the overall sentiment according to the sentiment of each post. This method is very intuitive and explainable. But it would be very time consuming to process so many documents in real operation. Furthermore, the sentiment bias may be very large for each post, which will be reflected as a very large variance when calculating the overall sentiment. It will cast negative influence on the final analysis (e.g. weird or unexplainable results in the econometric analysis). Second, combining posts for a specific target in certain period as a "document" and extract the sentiment embedded in the "document". This is an easy but effective way that has been widely used the current social media analysis (Hong and Davison 2010).

The second challenge is social media content's strong context dependence. Social media posts usually are too short (even one or two words) to be self-explainable. It is quite common for people to look up the previous posts to catch the exact meaning of the current ones. In other words, the understanding of some terms seriously relies on the context they are in. Hence, it is important to handle context dependence problem in social media sentiment analysis. In machine learning-based approaches, this task is delivered to machines learning processes. The disadvantages is that we can never tell whether the context dependence problem is solved or not. In lexicon-based approaches, domain specific lexicon construction or expansion is effective to alleviate the context dependence problem. Thus widely used in social media sentiment analysis (e.g. Kouloumpis et al. 2011; Khan et al. 2015). Probabilistic topic models are also proved to be effective to this situation. However, this challenge is far from being solved due to the flexibility of social media. It is known to all that it is quite common for topic change in social media. A group of users may complain that some brand of notebook is too heavy. Here, "heavy" definitely expresses a negative sentiment. One moment later, they may be delighted to discuss the heavy investment that firm makes to improve their products. Here, "heavy" usually refers to a positive opinion. In this situation, even if we construct a domain specific lexicon on "computer" domain that can effectively recognize the polarity of "heavy" when refer to "notebook", it is helpless for the following "heavy" related to "investment". The key to handle this challenge is to catch the topic shift. One possible solution may be dynamic topic modeling (DTM) (Blei and Lafferty 2006). In DTM, topics evolves instead of static. Old topics may die while new topics are born with new data come (see detail in Blei and Lafferty 2006).

The third challenge is increasing usage of multimedia, such as emoji, image, and video. These non-text content has beyond the ability of traditional text mining based opinion mining techniques. In previous opinion mining, emoji usually was considered as noise and removed directly (Kiritchenko et al. 2014). With the popularity of emoji, this process is infeasible anymore. Nowadays, there is no surprise to see a conversation that only contains emoji but without one single word. Emoji is meaningful and cannot removed as noise anymore. Generally, emoji can be classified into three categories: coded emoji, static image emoji and dynamic image emoji. Coded emoji has their own character codes, usually released by large companies like Apple, Google, Samsung and Tecent. Since they have fixed codes as other words in the vocabulary, they can be simply processed as special words in text mining. Static image emoji is specially designed pictures, e.g. cartoon, and photo. Digital image processing techniques can be adopted to help processing this kind of emoji. Dynamic image emoji can be considered as a special kind of video (usually in gif format). While a video can be considered as a continuous display of sequential images. Hence, the key to solve such multimedia content is digital image processing. Additionally, audios are also very popular in social media. It is also a common phenomenon for a video to contain corresponding audio in it. The technology to handle audio refers to speech recognition, which has not been widely used in opinion mining.

Another challenge refers to the informal writing. It is quite common to see abbreviations and typos in social media expression. There is no problem for widely accepted common abbreviations. Due to the length limitation or just writing convenience, users may create their own abbreviations. Some of these abbreviations are difficult to recognize even for human beings no saying machines. Since there are no writing standards or following editing, typos are very common or inevitable. Some typos are even made deliberately to express some king of "cute". Currently, there has been no good solutions for this challenge. One possible option is to construct personalized knowledge graph for each user. This kind of personalized knowledge graph contains writing styles, expression habit, etc. of each user. When one unrecognized symbol (e.g. abbreviation, typo) comes, we can go to the personalized knowledge graph to search context information as supplement of their social media posts. Unfortunately, this kind of personalized knowledge graph is time and energy consuming to construct and maintain. As a result, it is not applicable in real business, at least in current stage.

4 Recent Developed Techniques

Nowadays, research and applications are no longer satisfied with the overall opinion. There is increasing demands to know details of the opinion. As aforementioned, each opinion should have a holder and a target. In real business, the opinion target is of great potential values. For instance, it is quite common for firms to collect firsthand customer feedbacks once they release a new product (Luo et al. 2013). An overall opinion only reflect general market response to this product. If they want to collect useful information for their further improvement, details, such as customer's feeling to one specific aspect of their product, are necessary and important. To meet this demand, aspect oriented sentiment analysis is proposed. Khan et al. (2015) combined lexicon-based and machine learning-based methods to achieve performance enhancement for entity-level Twitter sentiment analysis. This entity-level analysis could bring more detail to the related sentiment (e.g. product, brand). Lakkaraju et al. (2014) proposed a hierarchical deep learning framework to extract the sentiment and its associated aspect simultaneously. Poria et al. (2016) proposed a deep learning approach to extract aspect associated with each sentiment. More specifically, a 7-layer deep convolutional neural network is used to identify whether a word is an aspect or not. Lau et al. (2017) designed a parallel framework to accelerate the mining speed of product aspect and user's attitude to it. A parallel hierarchical Dirichlet process that incorporated Gamma-Gamma-Poisson process is proposed to keep the dependence of each processing. This design is especially suitable for large volume stream data (e.g. social media data).

There are increasingly more multimedia content used in social media. As a result, extracting sentiment from these multimedia content has become a challenging but promising task. Due to wide application prospects, many attempts has been made. You et al. (2015) adopted progressively trained and domain transferred deep networks to extract sentiment expressed by images. Wang et al. (2015) proposed an unsupervised sentiment analysis model for social media images to solve the challenge of lack of training data. As mentioned above, video can be considered as a continuous display of a sequence of related images. Based on this, Poria et al. (2016) proposed a temporal convolutional neural network (CNN) to deal with the time sequence problem of the video. Then a multiple kernel learning (MKL) method is used to select informative features and cluster them into groups. According to their experiments, MKL helped their model significantly outperform state-of-the-art other advanced multimodal emotion recognition and sentiment analysis on various dataset. Cao et al. (2016) designed a visual sentiment topic model to extract sentiment from images in microblog. Cai and Xia (2015) utilized convolutional neural network for multimedia sentiment analysis. Similarly, Luo et al. (2017) adopted deep neural network for social multimedia sentiment analysis.

5 Potential Applications

Sentiment expressed in social media reflects the public orientation. These public orientations are reveals of their true feelings about the event they involved. Profound prediction and analysis can be made utilizing these true feelings. Here I offer three examples to show the power of social media sentiment analysis.

5.1 Case 1: Election Prediction

In 2016, president election of United Sates, according to the polls conducted by traditional media, such as ABC News and Washington Post, the favorability of Clinton is always much higher than Trump (ABC News 2016). There are still a large gap just before the election. On the contrary, according to Twitter trend, Trump won more supports. Finally, the election result proved the power of the public orientation on social media. Inspired by this fact, much research has been done to mine the value behind this phenomenon (e.g. Agrawal and Hamling 2017; Bessi and Ferrara 2016; Schumacher et al. 2016).

5.2 Case 2: Stock Market Analysis

According to Keynes's famous castle-in-the-air theory, firms' stock prices reflect investors' confidence and expectations instead of their business performance in a long term (Lawlor 1998). It is exactly in the very center of the ability of social media sentiment analysis. Many attempt has been made to predict the stock market utilizing public opinions from the social media. One well-known and controversial study was carried by Bollen et al. in 2011 (Bollen et al. 2011). They extracted a set of Google-Profile of Mood States (GPOMS) that measured mood in terms of 6 dimensions (Calm, Alert, Sure, Vital, Kind, and Happy) to predict the changes in Dow Jones Industrial Average (DJIA) closing values. According to their experiments, they achieved 86.7% accuracy in predicting daily up and down changes. Although their work attracted many disputes from the supporters of Efficient Market Hypothesis (EMH), it did proved the power of social media sentiment analysis. Following their steps, more attempts have been made to incorporate the power of social media sentiment analysis to stock market prediction (e.g. Azar and Lo 2016; Chen et al. 2014; Nguyen and Shirai 2015).

5.3 Case 3: Personalized Recommendation

People are always free to express their true feelings on social media. Expressions like "The camera of my current cellphone is terrible. I really need a new one." are frequently to see in social media posts. These kinds of expressions reflect the real needs of potential customers. More accurate product recommendation can be made once this kind of information is extracted. Some attempts has been made to make use of such value information. For instance, Sun et al. (2015) developed a novel sentiment-aware social media recommendation framework that utilize collaborative filtering method to improve recommendation performance. Ashok et al. (2016) applied machine learning techniques to extract sentiment from social media content to enhance performance of personalized recommender system. From the application perspective, many ecommerce platforms have built partnership with social media platforms to make use of this kind of personal information. Although it is of great benefits to use these data, critics pointed out that it would be a serious threat for people's personal privacy.

6 Conclusion

In this paper, we made a brief review on sentiment analysis in social media. Applying common opinion mining techniques to social media would suffer serious performance reduction. Some simple operations could improve the performance obviously. For instance, when dealing with short social media posts, we can combine related ones as a document instead of processing them one by one to avoid serious bias. Common coded emoji can be processed as normal words, while more advanced multimedia processing techniques, such as digital image processing, speech recognition, are needed to deal with sophisticated emoji. It is also a popular and promising trend to leverage multimedia processing techniques to social media sentiment analysis.

With the development of advanced techniques, new business needs emerge. People are no longer satisfied with an overall sentiment. Aspect oriented sentiment analysis could help to specify the opinion target. This technology is beneficial for firms to collect first hand customer feedbacks online. Compared with traditional methods (e.g. survey), it is much faster and cheaper. Moreover, utilizing this kind of detail information could also contribute to more accurate personalized recommendation.

All in all, there are great values embedded in people's daily social media posts. However, it is a very challenging task to extract these useful information from large volume of dataset. Applying opinion mining techniques that perform well in other domain directly to social media is not a very wise idea. Adjustment and improvements are necessary to make these techniques effective. Emotions expressed in social media is a reflection of their true feelings. Once extracted, there will be unprecedented opportunities to carrying on significant social and business analysis utilizing this enriched information.

Acknowledgement. The work was fully supported by the following grants: Zhejiang Provincial Natural Science Foundation of China (No. LQ14G020012), Beijing brain research project of Beijing Municipal Science & Technology Commission (No. Z171100000117009), the Fundamental Research Funds for the Central Universities and the Research Funds of Renmin University of China (No. 17XNLF05), the National Nature Science Foundation of China (NSFC 71702133).

References

ABC News (2016). http://abcnews.go.com/Politics/clinton-trump-leaves-10-unhappy-contest-tightens-conventions/story?id=40615476

Agrawal, A., Hamling, T.: Sentiment analysis of tweets to gain insights into the 2016 US election. Columbia Undergrad. Sci. J. **11** (2017)

Ashok, M., Rajanna, S., Joshi, P.V., Kamath, S.: A personalized recommender system using machine learning based sentiment analysis over social data. In: 2016 IEEE Students' Conference on Electrical, Electronics and Computer Science (SCEECS), pp. 1–6. IEEE, March 2016

Azar, P.D., Lo, A.W.: Practical applications of the wisdom of Twitter crowds: predicting stock market reactions to FOMC meetings via Twitter feeds. Pract. Appl. **4**(2), 1–4 (2016)

Bessi, A., Ferrara, E.: Social bots distort the 2016 US presidential election online discussion (2016)

Blei, D.M., Lafferty, J.D.: Dynamic topic models. In: Proceedings of the 23rd International Conference on Machine Learning, pp. 113–120. ACM, June 2006

Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. J. Mach. Learn. Res. **3**(Jan), 993–1022 (2003)

Bollen, J., Mao, H., Zeng, X.: Twitter mood predicts the stock market. J. Comput. Sci. **2**(1), 1–8 (2011)

Breakthroughanalysis (2008). https://breakthroughanalysis.com/2008/08/01/unstructured-data-and-the-80-percent-rule/

Cai, G., Xia, B.: Convolutional neural networks for multimedia sentiment analysis. In: Li, J., Ji, H., Zhao, D., Feng, Y. (eds.) NLPCC 2015. LNCS (LNAI), vol. 9362, pp. 159–167. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-25207-0_14

Cambria, E., Fu, J., Bisio, F., Poria, S.: AffectiveSpace 2: enabling affective intuition for concept-level sentiment analysis. In: AAAI, pp. 508–514, January 2015

Cao, D., Ji, R., Lin, D., Li, S.: Visual sentiment topic model based microblog image sentiment analysis. Multimed. Tools Appl. **75**(15), 8955–8968 (2016)

Chen, H., De, P., Hu, Y.J., Hwang, B.H.: Wisdom of crowds: the value of stock opinions transmitted through social media. Rev. Financ. Stud. **27**(5), 1367–1403 (2014)

Chen, Y., Lin, Y., Zuo, W.: Phrase-based topic and sentiment detection and tracking model using incremental HDP. KSII Trans. Int. Inf. Syst. **11**(12) (2017)

Church, K.W., Hanks, P.: Word association norms, mutual information, and lexicography. Comput. Linguist. **16**(1), 22–29 (1990)

Du, W., Tan, S., Cheng, X., Yun, X.: Adapting information bottleneck method for automatic construction of domain-oriented sentiment lexicon. In: Proceedings of the Third ACM International Conference on Web Search and Data Mining, pp. 111–120. ACM, February 2010

Hong, L., Davison, B.D.: Empirical study of topic modeling in Twitter. In: Proceedings of the First Workshop on Social Media Analytics, pp. 80–88 (2010)

Huang, S., Niu, Z., Shi, C.: Automatic construction of domain-specific sentiment lexicon based on constrained label propagation. Knowl.-Based Syst. **56**, 191–200 (2014)

Irsoy, O., Cardie, C.: Opinion mining with deep recurrent neural networks. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 720–728 (2014)

Khan, A.Z., Atique, M., Thakare, V.M.: Combining lexicon-based and learning-based methods for Twitter sentiment analysis. Int. J. Electron. Commun. Soft Comput. Sci. Eng. (IJECSCSE) 89 (2015)

Khan, F.H., Qamar, U., Bashir, S.: SentiMI: introducing point-wise mutual information with SentiWordNet to improve sentiment polarity detection. Appl. Soft Comput. **39**, 140–153 (2016)

Kouloumpis, E., Wilson, T., Moore, J.D.: Twitter sentiment analysis: the good the bad and the omg! Icwsm **11**(538–541), 164 (2011)

Lakkaraju, H., Socher, R., Manning, C.: Aspect specific sentiment analysis using hierarchical deep learning. In: NIPS Workshop on Deep Learning and Representation Learning (2014)

Lau, R.Y.K., Zhang, W., Xu, W.: Parallel aspect-oriented sentiment analysis for sales forecasting with big data. Prod. Oper. Manag. (2017)

Lawlor, M.S.: Keynes's uncertain revolution. Hist. Polit. Econ. **30**(4), 683–686 (1998)

Li, F., Wang, S., Liu, S., Zhang, M.: SUIT: a supervised user-item based topic model for sentiment analysis. In: AAAI, vol. 14, pp. 1636–1642, July 2014

Liu, P., Joty, S., Meng, H.: Fine-grained opinion mining with recurrent neural networks and word embeddings. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pp. 1433–1443 (2015)

Liu, Y., Chen, Y., Lusch, R., Chen, H., Zimbra, D., Zeng, S.: User-generated content on social media: predicting market success with online word-of-mouth. IEEE Intell. Syst. **25**(1), 75–78 (2010)

Lu, Y., Castellanos, M., Dayal, U., Zhai, C.: Automatic construction of a context-aware sentiment lexicon: an optimization approach. In: Proceedings of the 20th International Conference on World Wide Web, pp. 347–356. ACM, March 2011

Luo, J., Borth, D., You, Q.: Social multimedia sentiment analysis. In: Proceedings of the 2017 ACM on Multimedia Conference, pp. 1953–1954. ACM, October 2017

Luo, X., Zhang, J., Duan, W.: Social media and firm equity value. Inf. Syst. Res. **24**(1), 146–163 (2013)

Manivannan, P., Kanimozhiselvi, C.S.: Pointwise mutual information based integral classifier for sentiment analysis in cross domain opinion mining. J. Comput. Theor. Nanosci. **14**(11), 5435–5443 (2017)

Nguyen, T.H., Shirai, K.: Topic modeling based sentiment analysis on social media for stock market prediction. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, vol. 1, no. 1, pp. 1354–1364 (2015)

Pang, B., Lee, L.: Opinion mining and sentiment analysis. Found. Trends Inf. Retrieval **2**(1–2), 1–135 (2008)

Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up?: sentiment classification using machine learning techniques. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 79–86 (2002)

Pak, A., Paroubek, P.: Twitter as a corpus for sentiment analysis and opinion mining. In: Proceedings of the Seventh Conference on International Language Resources and Evaluation, pp. 1320–1326 (2010)

Poria, S., Chaturvedi, I., Cambria, E., Hussain, A.: Convolutional MKL based multimodal emotion recognition and sentiment analysis. In: 2016 IEEE 16th International Conference on Data Mining (ICDM), pp. 439–448. IEEE, December 2016

Read, J.: Recognising affect in text using pointwise-mutual information. Unpublished M.Sc. Dissertation, University of Sussex, UK (2004)

Schumacher, R.P., Jarmoszko, A.T., Labedz Jr., C.S.: Predicting wins and spread in the premier league using a sentiment analysis of Twitter. Decis. Support Syst. **88**, 76–84 (2016)

Statistic 2018 https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/

Su, Q., Xiang, K., Wang, H., Sun, B., Yu, S.: Using pointwise mutual information to identify implicit features in customer reviews. In: Matsumoto, Y., Sproat, R.W., Wong, K.-F., Zhang, M. (eds.) ICCPOL 2006. LNCS (LNAI), vol. 4285, pp. 22–30. Springer, Heidelberg (2006). https://doi.org/10.1007/11940098_3

Sun, J., Wang, G., Cheng, X., Fu, Y.: Mining affective text to improve social media item recommendation. Inf. Process. Manag. **51**(4), 444–457 (2015)

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., Stede, M.: Lexicon-based methods for sentiment analysis. Comput. Linguist. **37**(2), 267–307 (2011)

Turney, P.D., Littman, M.L.: Measuring praise and criticism: inference of semantic orientation from association. ACM Trans. Inf. Syst. (TOIS) **21**(4), 315–346 (2003)

Wang, Y., Wang, S., Tang, J., Liu, H., Li, B.: Unsupervised sentiment analysis for social media images. In: IJCAI, pp. 2378–2379, July 2015

Wilson, T., Wiebe, J., Hoffmann, P.: Recognizing contextual polarity in phrase-level sentiment analysis. In: Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing, pp. 347–354 (2005)

Witten, I.H., Frank, E.: Data Mining: Practical Machine Learning Tools and Techniques. Elsevier Inc., Oxford (2005)

You, Q., Luo, J., Jin, H., Yang, J.: Robust image sentiment analysis using progressively trained and domain transferred deep networks. In: AAAI, pp. 381–388, January 2015