Boğaziçi University
Software Engineering

SWE 578 DIRECTED STUDIES I

Opinion Mining and Dynamics in Online Digital Communities

Instructor:
Asst. Prof. Suzan Üsküdarlı

Student: Kaan CAN 2021719030

1. INTRODUCTION

For most of the people, one of the most prominent development over the humankind is web technology. With the innovation of this technology, the habits, knowledge level and attributes of people have risen up exponentially. At the beginning of web technology, there were static web pages and users were not allowed to change the content, make some contributions or comments here. These were generally for people to gather information at specific domains. With the development in the programming technology, these web pages have obtained the attribute of being dynamic. In other words, the users were able to give feedback, make some comments and as a result make come connections with the owner of these sites and other users.

Obtainment of feedback from the users has brought great advantages to the owners of the products since the shortcomings or open-to-develop features were gathered and adapted with the following products. However, evaluation of these opinions manually become impossible with exponential and progressive utilization of internet usage and as result huge amount of data. In order to handle with these data automatically and systematically, a new method has been introduced and this is opinion mining.

Opinion mining can be defined as subfield of NLP which is Natural Language Processing technique[1]. Opinion mining help to discover or generate human thoughts and perceptions in unstructured context like user reviews and comments. In the literature, there is well-accepted definition of opinion mining by Saleh. He has expressed this as ““The automatic processing of documents to detect opinion expressed therein, as a unitary body of research”[2].

2. TASK AND APPLICATIONS OF OPINION MINING

There are lots of tasks performed by opinion mining and some of them can be listed as sentiment analysis, opinion extraction, sentiment mining, subjection analysis, affect or emotion analysis and review mining. In sentiment analysis, the main purpose is to recognize the sentiment and extract the public opinion. For opinion extraction used in the web applications, its purpose is to explore the thoughts and opinions of the users. In sentiment mining, there are two main objectives. First one is to determine whether the context is objective or subjective and the second one is the classification of the opinions and thoughts in the categories like negative, neutral and positive. In the emotion analysis, the purpose is the specification of emotions in the text by utilizing natural language process. For the review mining which is the sub-topic of text sentiment analysis, the purpose is the extraction of aspects in the users’ sentiments as a final summary of the sentiments.

Regarding the procedures of opinion mining, some sub-steps are generally performed and these items can be listed data collection, opinion identification, aspect extraction, opinion classification, judgment of whether they are positive or negative, production summary and lastly evaluation respectively. The scheme of this process is shown in the Figure 1 below,

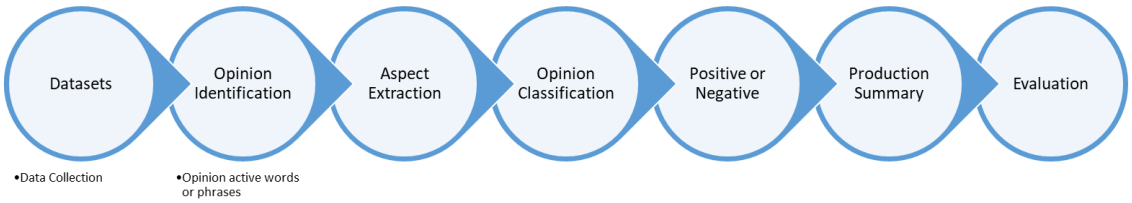

Figure 1 The Procedure of Opinion Mining [1]

When it comes to the application areas of opinion mining, main ones can be thought as commercial products, politic areas and stock market. For the commercial products, it can be helpful for the customer to have a summary of comments or reviews with opinion mining instead of reviewing all the comments in the massive amount. In addition, producers can obtain good field experience from the first end users by analyzing the reviews and comments and they can define their next step in a confident attitude easily. In politics area, the people make comment or share their thoughts for the different political and social issues freely and the summary of these opinions with opinion mining can help the governments to develop the public weal with the appropriate implementations. Regarding the utilization of opinion mining in the stock market, these are utilized to forecast the value of stocks in the market. There is a related study by Bollen examining the relationship between the twitter comments and Dow and Jones Industrial Average closing[3].

3. OPINION MINING LEVELS

Based on the literature, the levels of opinion mining can be listed as document, sentence, aspect and concept levels.

In the document level, a general sentiment analysis performed in order to determine whether the whole document is positive or negative. Since this kind of opinion mining is not detailed it can be resulted fast however this level is not appropriate for price evaluation or comparison. The main purpose of opinion mining in document level is to categorize the document in the positive or negative as a summary.

In order to perform a deeper analysis, the sentence level is proposed since the document level is too coarse. The main purpose in the sentence level is to obtain a classification for each sentence. In the process of sentence, there are two main steps. First one is the determination of if the sentence is objective or subjective and the second is determination of polarity of the sentence. Even though the document is reviewed in the piece level as a sentence to have better accuracy this serve much more challenges as a result due to the complexity.

The shortcomings of document and sentence level addressed with the new level which is aspect level. For the document and sentence level, being positive or negative about the sentiments accept that the author is positive or negative for all entity. However, this cannot be the reality. Accept level in opinion mining gives more importance on the opinion itself. Its purpose is to define and extract all the aspects and then make polarity of them. For the documents containing more than one aspect, this level can give much accurate results compared the previous two[4].

Concept level open mining can be regarded as deep understanding of the natural language texts by machine. At this level, analysis of emotions are performed according to conceptual information about emotions and sentiments[5].

4. OPINION MINING CLASSIFICATION TECHNIQUES

As a general, the methods are divided in two categories which are machine learning and lexicon based. In the machine learning methods, there are 3 categories which are supervised learning, semi-supervised learning and unsupervised learning. The supervised learning also divided into 2 categories such as probabilistic classification and non-probabilistic classification. In the probabilistic classification, some well-known machine learning methods are utilized and these are Naïve Bayes, Bayesian Network and Maximum Entropy. For the non-probabilistic classification, Support Vector Machine, Neural Network, Nearest Neighbor, Decision Tree and Rule-based methods are generally utilized[1].

When it comes to the lexicon based approaches, there are two mainly preferred

approaches which are Corpus-based and dictionary based approaches[1]. As a summary of these techniques, it is shown in Figure 2 below.

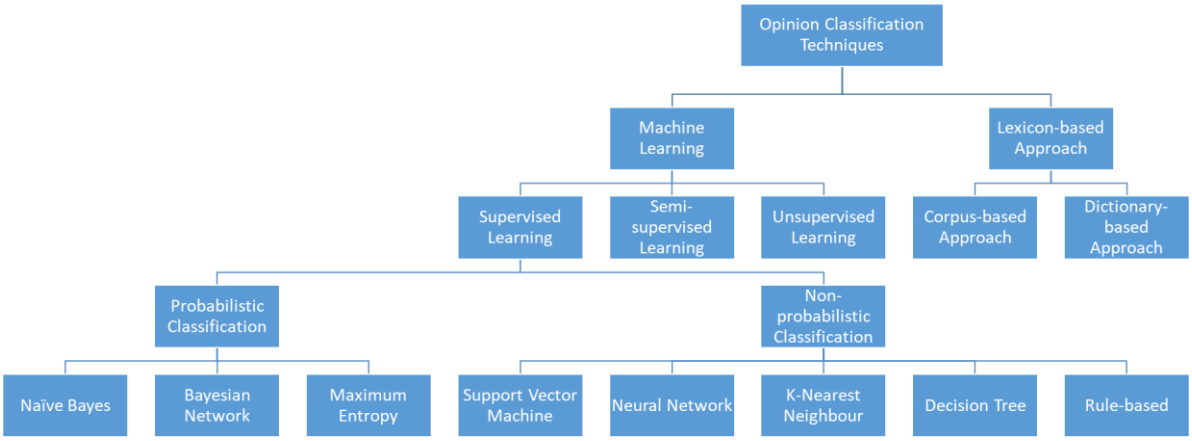

Figure 2 General View of Opinion Mining Methods[1]

5. FUTURE WORK

Regarding the future work, the applications, advantages and disadvantages of these opinion mining techniques will be investigated and presented based on the literature studies.

6. REFERENCES

1. Hemmatian, F., Sohrabi, M.K.: (2017) ”A survey on classification techniques for opinion mining and sentiment analysis.” In: Artificial Intelligence Review
2. Saleh MR,Martín-Valdivia MT,Montejo-Ráez A, Ureña-López LA (2011) “Experiments with SVM to classify opinions in different domains.” Expert Syst Appl 38(12):14799–14804
3. Bollen J, Mao H, Zeng X (2011) “Twitter mood predicts the stock market.” J Comput Sci 2(1):1–8
4. Liu B (2012) “Sentiment analysis and opinion mining.” Synthesis lectures on human language technologies. Morgan & Calypool Publishers, pp 1–167.
5. Cambria E. (2013) “An introduction to concept-level sentiment analysis. In: MICAI 2013: Advances in soft computing and its applications.” Mexican international conference on artificial intelligence, pp 478–483