Sentiment Analysis On Facebook Group Using Lexicon Based Approach

Sanjida Akter
Lecturer in CSE dept.
Northern University Bangladesh
Dhaka, Bangladesh
Sanjida.akter@gmail.com

Muhammad Tareq Aziz
Software engineer
goBD
Dhaka, Bangladesh
tareq@gob.co

Abstract—Internet is one of the primary sources of Big Data. Rise of the social networking platforms are creating enormous amount of data in every second where human emotions are constantly expressed in real-time. The sentiment behind each post, comments, likes can be found using opinion mining. It is possible to determine business values from these objects and events if sentiment analysis is done on the huge amount of data. Here, we have chosen FOODBANK which is a very popular Facebook group in Bangladesh; to analyze sentiment of the data to find out their market values.

Keywords— social media, sentiment analysis, market bucket analysis, supervised learning, probabilistic classifier, lexicon based approach

I. INTRODUCTION

Sentiment analysis is the concept: "how a person reacts on something or some events". It is always an important issue "what people think". People's thoughts are available in social medias to be used as data source of market basket analysis. Social media plays an important role here because we can have data from people even we don't know. Such analysis can help to determine recent market trend or market value of a particular product or event. based on people's interest. Here we have utilized FACEBOOK which is a very popular social media platform as our data source and. "FOODBANK" which a well-known FACEBOOK group.

For opinion mining or sentiment analysis some methods are applied like – Naive Bayes Machine Learning Classifier, Senti Word Net, Support Vector Machine. Here we applied both machine learning approach and Lexicon Based Dictionary. Most of the work on opinion mining is done with Naïve bayes. As our data set contains Bangla language; we built a Dictionary on top of the data set. We count the occurrence of sentimental words and featured words which are tagged as the target of calculated sentiment value. This approach can find 73% cases correctly whether post is positive or negative. After comparing two results, we found that lexicon based algorithm works better here.

At the end of analysis; we can determine recent trends and characteristics of people's food habit. It also can be identified that "what will be the food habit in future" and "where investors should put their attention".

II. SENTIMENT ANALYSIS

In recent time, sentiment analysis is a popular research topic because many real-life problems are part of this subject matter. It is also highly challenging as NLP research topic that covers many novel sub problems. Additionally, there was almost no significant research done before the 2000 in either NLP or linguistics. It is because of lack of availability of opinion or text in digital forms. Since the year 2000 the field has grown rapidly to become one of the most active research areas in NLP. It is also widely researched in data mining, Web mining, and information retrieval technologies. In fact, it has spread from computer science to management and business intelligence.

There are mainly three methods or level of research on Sentiment Analysis.

Document level [1]: analyze the overall sentiment expressed in the text and determine if the overall sentiment is positive or negative.

Sentence level [2] - examine the sentiment expressed in sentences and determines whether each sentence expressed be a positive, negative, or neutral opinion.

Aspect level [3] - aspect level performs better analysis when document level and the sentence level analysis do not find what people liked and did not like. Aspect level also called feature level.

Aspect level does not work on document, paragraph sentences or clauses, it directly finds out the sentiment and the target. The sentiment can be positive, negative or neutral towards that specific target or entity. Realizing the importance of opinion targets also helps us understand the sentiment analysis problem better. For example, "I love to eat chicken".

Here we can see a positive sentiment and the target of this sentiment or the entity is "chicken". There is another challenge which is that we can classify opinion in two types: Regular opinion and Comparative opinion [4].

III. LEXICON BASED

Undoubtedly, the most important indicator of sentiments are sentiment words. These are words that are commonly used

to express positive or negative sentiments. For example, good, wonderful, and amazing are positive sentiment words, and bad, poor, and terrible are negative sentiment words. Sentiment words are instrumental to sentiment analysis for obvious reason. A list of such words are called a sentiment lexicon (or opinion lexicon). Over the years, researchers have designed numerous algorithms to compile such lexicons.

Although sentiment words are important for sentiment analysis, only using them is not sufficient. The problem is much more complex.

Some issues related to lexicon based approach are given bellow:

- A positive or negative sentiment word may have opposite orientations in different application domains.

    For example, "dangerous" is a negative word but it can also be highly positive when someone posts "dangerous tasty".

- A sentence containing sentiment words may not express any sentiment. This happens frequently in question types of sentences.

    For example, "Suggest a good coffee shop around gulshan n banani..."

- Many sentences without sentiment words can also imply opinion.

    For example, "They took a lot of time to serve us". This sentence has a negative sentiment. In recent time many researchers are working on lexicon based approach [5].

IV. SENTIMENT ANALYSIS ON SOCIAL MEDIA GROUP

It is possible to understand the real world movement by analyzing data of social media groups. [5] Sentiment analysis tools have been applied to examine the relationship between release of products, the 'discussion' online, and actual sales of products with the outcome being that such data can be used to predict sales volumes [6].

As we have mentioned that sentiment analyses are about opinion or expression or imply positive or negative sentiment or emotion towards an object. So the elements of Sentiment analysis are:

"Opinion" which are a judgment, viewpoint, or statement about matters commonly considered to be subjective.

"Entity" which are a product, service, topic, issue, person, organization, or event. It is the target of an opinion.

"Subjectivity and emotions" which are the state of mind of a person and instinctive responses.

Let us use a Facebook post on "FoodBank" group to describe opinion properly. The post is given bellow with the time when it was posted.

Tanvir RahmanMay 8, 2015 at 04.30 pm

"Today we visit KFC. The decoration was awesome. We ordered Chicken Zinger burger. It was so tasty. But my sister think it contains too much calories."

Observation:

1. This post contains a number of opinion.
2. It has more than one target object or Entity.
3. It has both positive and negative opinion.
4. It has two opinion holders.
5. This post is holding a date.

V. SENTIMENT ANALYSIS USINGMACHINE LEARNING APPROACH

almost 90% researchers work with different machine learning algorithms to detect sentiment or for opinion mining. Some common algorithms that are used for sentiment analysis are – Naïve Bayes, Maximum Entropy, support vector machine etc. [8]. Naive Bayes classifiers are highly scalable, requiring a number of parameters linear in the number of variables (features/predictors) in a learning problem. Maximum-likelihood training can be done by evaluating a closed-form expression, which takes linear time, rather than by expensive iterative approximation as used for many other types of classifiers.

VI. SENTIMENT ANALYSIS USING LEXICO BASED APPROACH

The lexical approach is a method of teaching foreign languages described by M. Lewis in the 1990s. At recent days researchers also working with lexicon based approach for sentiment analysis.[9][10] The basic concept on which this approach rests is the idea that an important part of learning a language consists of being able to understand and produce lexical phrases as chunks. Students are thought to be able to perceive patterns of language (grammar) as well as have meaningful set uses of words at their disposal when they are taught in this way.

Machine translation can use a method based on dictionary entries, which means that the words will be translated as a dictionary does i.e. word by word, usually without much correlation of meaning between them. Dictionary lookups may be done with or without morphological analysis or lemmatization. While this approach to machine translation is probably the least sophisticated, dictionary-based machine translation is ideally suitable for the translation of long lists of phrases on the sub sentential (i.e., not a full sentence) level, e.g. inventories or simple catalogs of products and services.

VII. OUR APPROACH AND EXPERIMENT RESULT

Apart from many processes, we have used two main approaches to analyze sentiments of Social Network Data. These are:

- Naïve Bayes Machine Learning classifier.
- Dictionary Based Approach.

A. Collecting Data

As per our interest we decided to do our sentiment analysis experiment on Social network data. But social network is full of chaos. So we needed a group where people would talk on specific topic.

We found there are groups like BDCyclists, FoodBank where people discuss on same kind of topic. we considered "FoodBank" for this work.

To collect Data, there is an API for developers in Facebook called Graph API. This API lets programmer to do different kinds of programmatic activities on top of Facebook data. So we wrote a C# console application to download all the posts that is available on FoodBank since the beginning.

B. Formatting Data

Before using the data in any of the sentiment analyzing process, we needed to do some formatting on the data. The data we downloaded were in JSON format.

1) Formatting Data for Naïve Bayes Classification: We write another c# console application that would extract and make a list of Messages of all posts for naïve Bayes classification. We write another Visual GUI application in C# that would let our team mates do the manual classification since all machine learning classifiers are heavily depend on trained data set.

2) Formatting Data for Naïve Bayes Classification: To do this, we tokenize each and every words. We made a list of unique words and there were 9000 words. Then we manually classified all these words based on sentiment rating like neutral, positive, negative, super positive, super negative, food name or location name etc.

C. Formatting Data

In machine learning, naive Bayes classifiers are a family of simple probabilistic classifiers based on applying Bayes' theorem with strong (naive) independence assumptions between the features. Since it's a probabilistic classifier, it heavily depends on the probability mathematics. We used Python programming language for this classification technique. We took about 3600 random statuses from the list and classified them manually based on human judgment.

Problems with Naïve Bayes Classifier:

In different scientific papers it has been found that Naïve Bayes machine learning classifier is extremely successful when applied on well-formed text corpuses. Like movie reviews or novels.

But in Facebook, Naïve Bayes has not performed well, at least according to our experimentation. When we gave a very close look to the poor performance of Naïve Bayes Classifier, we found that the Facebook Group Posts are extremely noisy. Since people post in random lengths with lots of spelling mistakes. For example:

Status	Problems
The delicious pizza.. Bt must add..the mushroom was very nice...☺☺	No mention of what pizza and where
Ajke kheye ashlam Comic Cafe er set menu no.81 OMG OMG !!! that was sooooo tasty !! Quantity etttoo beshi j eka kheye shesh kora possible na. Amar shesh korte onkkk koshto hoise. 320/= ekdom usul hoy trust me!! Place- Comic Cafe Price - 10/10 (320 tk) Service- 9/10 (10mins wait korte hoise seat pawar jonno)	Bangla written in English form. Frequent in Social Media
Best place for dinner in Gulshan or may be Banani. The food must be real good! Urgent reviews needed! Tia	Nearly 50% of query is comprised of queries like this. As a result many of the sentences do not express exactly any sentimental value

Fig. 1. *variation in post on social network*

Since variation and combination of words in local Facebook groups are way too high; we needed a huge amount of training sample in order to reach to a satisfactory level of accuracy. But at this moment we have found simple counting positive and negative words. Specialized key words are giving better performance than Naïve Bayes classifier. So, for the rest of the sentiment analysis technique we emphasized on Lexicon Based approach.

D. Approach II (our proposed solution: Lexicon based Sentiment Analysis)

In a nutshell, it is nothing more than counting key positive and negative and neutral key words We found this approach is extremely efficient since ordinary people posts on this FoodBank and they are not comprising of complex sentence structures that would let us to satirical or ironic posts. Point to be noted here that people on FoodBank uses a useful pattern to review something. For example:

my review on #fazitas sub-sandwich nd nachos... #nachos:nachos were preety good... actually great *grin emoticon* quite similar like Mexican one.Enjoyed it *smile emoticon* taste:10/10 price:10/10 #sandwich:nacho meat nd sandwich meat was same *colorthree emoticon* they gave just a single slice of capsicum *frown emoticon* not enough ketchup *cry emoticon* (I think after getting such reviews on sub-sandwich of fazitas my expectations were high) my sis didn't like it at all *colorthree emoticon* taste:5.5/10 price:9/10
মাওয়া ঘাঁটের তাজা ইলিশ না খাইলে মিস্ ☺ ফেরী ঘাঁটে যখন যাবেন খাবার নেওয়ার আগে দাম টা জিজ্ঞেস করে নিবেন। কারন ওখানে যার থেকে যেমন নিতে পারে তেমন নিয়ে যায়। সাধারণত বাছাই করে কাঁচা ইলিশ নিলে ওরা এক পিচ ইলিশ ৮০ টাকা করে নিয়ে থাকে। আবার কেউর কাছে ১০০, ১২০ তার বেশিও নিয়ে থাকে। আমি নিজের পছন্দ মত নিয়ে ছিলাম কাঁচা ইলিশের পিচ তারপর তাজা তাজা বেজে দিলো। কিন্তু আমার কাছ থেকে ৬০ টাকা করে নিয়েছে। *smile emoticon* Price- 10/10 Taste-9/10 Service-8/10

Fig. 2. *Ratting pattern on FOODBANK*

We can observe that sometime people rate foods or location on a scale of 10. This is an interesting keyword to determine what people think about certain feature words. But since people write such kind of rating in different manners; sometimes it is hard to find out actually what is rated and if there are multiple feature words. So, we consider such sentiment rating as a basis of judging the whole messages. Our

approach calculates the sentiment of the status in subjective manner rather than objective.

E. The proposed work

1. Download all the data from FoodBank.

2. Make a list of unique words both that are used as a whole in all of the statuses.

3. Noise out all the words that doesn't represent any sentimental value or feature value like location and food.

4. Rate the rest of the words that do represent the sentimental or feature value, make a wordlist out of it.

5. Check for the occurrence of words of our word list.

6. Tag each status with those words to each status, the sentimental values represent the sentiment of whole status. The food or location indicates to whom the sentiment is expressed.

7. When all the messages sentiment is detected, a macro level overview like food/ restaurant popularity, trends etc. is possible.

We have followed all these steps to analyze sentiment in this lexicon based technique.

VIII. APPLICATIONS OF SENTIMENT ANALYSIS

There are many applications of Sentiment analysis. A short list of application is given bellow.

1. Customer Opinion mining

2. Manipulation

3. Political Observation

4. Social Study

5. Crowd Sourcing

6. Artificial Intelligent Research

7. Business Insight

IX. ADVANTAGES OF SENTIMENT ANALYSIS USING LEXICON BASED APPROACH

Advantages of sentiment analysis on a social media are given bellow.

1. Making decision based on knowing others interest.

2. Low cost and time saving method of getting consumer insight.

3. A faster way of getting customer data.

4. Ability to act on customer suggestion.

5. Helps to Identify Organization's Strength, Weakness, Opportunities and Threats. (SWOT) analysis.

6. More accurate and insightful customer perception and feedback

X. CONCLUSION

In this paper, we have shown how to predict the sentiment behind a status post of Facebook which in nature of unstructured dataset, cross language domain and noisy. So, it can be determined that traditional opinion mining is not efficient enough to find out sentiments from social media's like FACEBOOK. But lexicon based dictionary approach works efficiently in such kind of works. As sentiment analysis has many practical advantages so a structured process need to be developed.

References

[1] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan "Thumbs up? Sentiment classification using machine learning techniques" Proceedings of EMNLP, pp. 79--86, 2002.

[2] Steven Bethard, Hong Yu, Ashley Thornton, Vasileios Hatzivassiloglou and Dan Jurafsky Wilson, Wiebe and Hwa, "Automatic Extraction of Opinion Propositions and their Holders" Computing Attitude and Affect in Text: Theory and Applications Volume 20 of the series The Information Retrieval Series pp. 125-141.

[3] Wilson, Wiebe, Hwa "Just how mad are you? finding strong and weak opinion clauses" presented at the 19th national conference on Artificial intelligence, pp. 761-767.

[4] Minqing Hu and Bing Liu "Feature-based opinion mining and summarization" KDD '04 Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining pp.168-177

[5] Nitin Jindal and Bing Liu "Mining Comparative Sentences and Relations" (2006), AAAI'06 proceedings of the 21st national conference on Artificial intelligence - Volume 2, pp.1331-1336.

[6] Moreo A, Romero M, Castro JL, Zurita JM. "Lexicon-based comments-oriented news sentiment analyzer system", Expert Syst Apple 2012; pp. 66–80.

[7] Walaa Medhat, Ahmed Hassan, Hoda Korashy "Sentiment analysis algorithms and applications: A survey", Ain Shams Engineering Journal, Volume 5, Issue 4, December 2014, pp. 1093–1113.

[8] Bo Pang, Lillian Lee, Shivakumar Vaithyanathan "Sentiment classification using machine learning techniques" Proceedings of EMNLP, 2002, pp. 79—86.

[9] Prabu Palanisamy, Vineet Yadav, Harsha Elchuri "Serendio: Simple and Practical lexicon based approach to Sentiment Analysis" presented at Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Seventh International Workshop on Semantic Evaluation (SemEval 2013), Atlanta, Georgia, June 14-15, 2013. pp. 543–548.

[10] Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, Manfred Sted "Lexicon-Based Methods for Sentiment Analysis", Computational Linguistics Volume 37, Number 2, 28 September 2010, pp. 268-307.

[11] Lincoln C. Wood, Torsten Reiners, Haris S. Srivistava "Expanding Sales And Operations Planning Using Sentiment Analysis: Demand And Sales Clarity Form Social Media" ANZAM 2013, December 4-6, 2013.

[12] Mohammad Ali Abbasi, Sun_Ki Chai, Huan Liu, Kiran Sagoo "Real-World behavior analysis through a social media lens" SBP'12 Proceedings of the 5th international conference on Social Computing, Behavioral-Cultural Modeling and Prediction, pp. 18-26.

[13] Bo Pang, Lillian Lee "Opinion mining and sentiment analysis", Foundations and Trends in Information Retrieval Vol. 2, No 1-2 (2008) pp.1–135.