Markov Process for Behaviour Analysis in Social Media

Ashwin Perti
Associate Professor,
ABES Engineering College, Ghaziabad

Munesh Chandra Trivedi, PhD
Professor (CSE),
ABES Engineering College, Ghaziabad

ABSTRACT
Social Network Analysis (SNA) provides tools to examine relationships between people. Here Markov Process is used to examine the behavior of the people involved in Social Networks. Markov process is the stochastic process in which the behavior evolves over time in a probabilistic manner. In this paper, Markov process is used to describe people's interests and to find out the people with similar interests. It is found that the expected positive relationship between sharing interests and communicating among people. Here focus is on analyzing the behavior of people who are using Social Networking sites.

General Terms
Natural Language Processing, AI

Keywords
Markov Process, Stochastic Process.

1. INTRODUCTION
Online social networks have been one of the most exciting events in this decade. Many popular online social networks such as Twitter, LinkedIn and Facebook have become increasingly popular. In addition, a number of multimedia networks such as Flickr have also seen an increasing level of popularity in recent years. Many such social networks are extremely rich in content, and they typically contain a tremendous amount of content and linkage data which can be leveraged for analysis. The linkage data is essentially the graph structure of the social network and the communications between entities; whereas the content data contains the text, images and other multimedia data in the network. The richness of this network provides unprecedented opportunities for data analytics in the context of social networks.

The field of Opinion Mining and Sentiment Analysis is well suited to various types of intelligent applications. Business Intelligence is one of the main factors behind corporate interests in this field.

In classification, it is predicted what class an instance of data should fall into and another task in machine learning can be regression. Regression is the prediction of a numeric value. Most people have probably seen an example of regression with a best-fit line drawn through some data points to generalize the data points. Classification and regression are examples of supervised learning. This set of problems is known as supervised because here it is told to the algorithm what to predict.

In Data Mining the various Algorithms are divided into the following three categories:

1. Supervised Methods

2. Unsupervised Methods

3. Semi-supervised Methods

The opposite of supervised learning is a set of tasks known as unsupervised learning. In unsupervised learning, there's no label or target value given for the data. A task where grouping of similar items is done together is known as clustering. In unsupervised learning, researchers may also want to find statistical values that describe the data. This is known as density estimation. Another task of unsupervised learning may be reducing the data from many features to a small number so that it can be properly visualized it in two or three dimensions.

Here the problem is defined based on Markov decision processes– which is categorized into Unsupervised learning. This formulation will laid the basis for the solution of analyzing the behaviour using the Social Network. In a Markov decision process (MDP) the person can perceive a different set of S states. The states in a Markov process the person can be in one of these states – where four cases are considered.

2. REVIEW OF THE RELATED WORK
- As in [1] the author has talked about the sentiment Analysis and has helped many researchers for Sentiment Analysis. Sentiment Analyzer: Extracting Sentiments about a given topic using Natural language processing techniques. This has made significant thrust for the Natural Language Processing (NLP) in the field of Sentiment Analysis.

- In [2] the Sentiment Analysis is being done by the author using supervised and unsupervised learning method. The approach followed in this paper is also supervised approach where the output is known (i.e. the requirement of the companies) and input is the Sentiment Analysis of the candidate are supposed to take from the social sites (like Facebook / twitter).

Input   **Output**

Facebook data   Sentiment Analyses

3. COMPETENCY MAPPING
Sentiment Analysis is the process where detection of the information is done by the extraction from unstructured text. Generally the text which is taken from the social sites like facebook and twitter is in unstructured form which is composed of text, audio, images, video and many more formats which makes it unstructured. This paper focuses on

unstructured text form. The sentiments are mainly based on the behavior of the candidate for which competency is checked.

Sentiment Analysis procedure is as follows:

1. Agent Features Extracted
2. Tokenization of the features Extracted
3. Feature Selection based on the various mood that a person can attain
4. Comparing those features which gives positive or negative
5. Sentiment Polarity based on various mood
6. Classification of the Sentiments

4. PROPOSED METHODOLOGY

In this paper the first step is Natural Language Processing tasks like – tokenization, stop-words removal, stemming to filter the ambiguous text. For the said purpose nltk package can be used in python or tm package provided in R software.

The methodology that we have followed in this paper is discussed below:
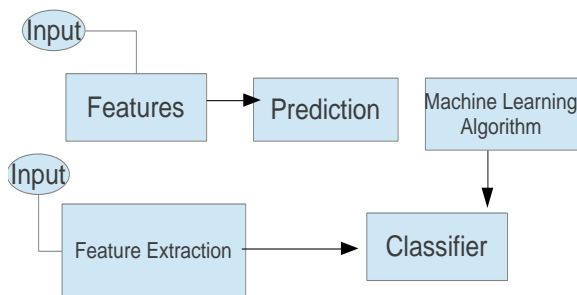
4.1 Training

In this step, the input is provided as a text in form of tokens. During the training period the parameters are collected as per the requirement that the comment given by the customers were in which mood.

4.2 Prediction

After the text collection, we go for prediction for the sentiments that reseachers had collected.In this phase the data collected from the social media like twitter, Facebook, LinkedIn and other social sites are used as input and features are extracted from them. The matching process is done at this level where the parameters are matched and the Sentiments so desired is taken out.

In more detailed form is explained below:

1. Create an app on facebook / twitter / or any other social site
2. Mining the data from the Social Site
3. Matching the quality with the parameter decided below
4. On the basis of the matching the parameters with the Sentiments of the individual

5. The Recommendation process starts here after this step
6. Finally the Behaviour of the person can be judged by this procedure

4.3 Behaviour Classification techniques

As Sentiment classification techniques are divided into machine learning approach [5]. The parameters that have to be mapped with the posts, likes or tweets are summarized into the table. These parameters matched with the mood of the agents (people) who are using the Social Media Platform for their comments.

S.No.	Parameters
1	Very Happy (in jolly Mood)
2	Good mood
3	Sad Mood
4	Confused State
5	Neutral

Let us assume that the probability of these moods can be in the order as given below:

1. Very Happy – 0.2
2. Good Mood – 0.1
3. Sad Mood – 0.1
4. Confused State – 0.1
5. Neutral – 0.5

So considering the various states with their Transition probabilities, following are the findings:

States	Very Happy	Good Mood	Sad Mood	Confused	Neutral
Very Happy	0.2	0.1	0.1	0.1	0.5
Good Mood	0.1	0.1	0.2	0.1	0.5
Sad Mood	0.1	0.2	0.1	0.1	0.5
Confused	0	0.1	0.2	0.1	0.5
Neutral	0.1	0.2	0.1	0.1	0.5

So, with the use of Markov Chain Process the behaviour of the People may be predicted who are using the Social Media platform and can analyze what type of work they are going to do on the Social Media – that would also include their statement.

5. CONCLUSION

So it may be concluded that the tweet (or post) given by the agent (people) says about the mood of the person who has tweeted or given his comments. Features can be extracted from various inputs that are coming from various social websites like Facebook, twitter, LinkedIn and many more

other websites. The same procedure can also be applied for the Sentiment analysis also. The behaviour of different people taken in the above said paper is the views of the author itself.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] B. Pang, L. Lee, and S. Vaithyanathan. 2002. Thumbs up? sentiment classification using machine learning techniques. In Proceedings of EMNLP, pages 79–86.

[2] Rupali P. Jondhale and Manisha P. Mali, (Feb. 2015), "Study on Distinct Approaches for Sentiment Analysis", in IJCA, Volume 14, Pg. 21-24.

[3] Walaa Medhat, Ahmed Hassan , Hoda Korashy. Sentiment analysis algorithms and applications: A survey. Published on 27 May, 2014, hosting by Elsevier B.V. on behalf of Ain Shams University.

[4] F. Li, M. Huang, and X. Zhu. 2010. Sentiment analysis with global topics and local dependency. In Proceedings of AAAI, pages 1371–1376.

[5] C. Lin and Y. He. 2009. Joint sentiment/topic model for sentiment analysis. In Proceeding of the 18th ACM. Conference on Information and Knowledge Management, pages 375–384.