Sentiment Analysis of Customer Reviews based on Hidden Markov Model

Swati Soni
M.Tech Scholar
Department of Electrical Engineering
National Institute of Technology, Raipur
Raipur, India
swatisoni087@gmail.com

Aakanksha Sharaff
Assistant Professor
Department of Computer Science & Engineering
National Institute of Technology, Raipur
Raipur, India
asharaff.cs@nitrr.ac.in

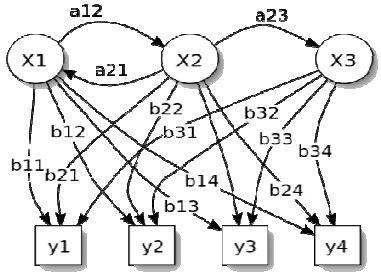
ABSTRACT
Presently there are various websites like Amazon.com, eBay, FlipKart, Snapdeal etc. which have large number of products available online. The seller or the manufacturer often ask their customers to share their opinions and hands-on experiences on the products they have purchased. Unfortunately, it is very difficult to go through all customer's reviews and to decide whether the overall performance of the product is satisfactory or not. This paper mainly focuses on the problem of sentiment analysis of customer's online reviews about the product. In this paper we aim to train our system to analyse whether the comment given by the customer is positive or negative. The work is divided in two phases: In first phase, we propose a training process of a Stochastic model namely Hidden Markov Model and secondly we test and reveal the individual comment for analyzing consumer opinions about the products. Our results indicates that the trained system is very promising in performing its tasks and we have achieved maximum possible Precision and Accuracy.

Keywords
Sentiment Analysis; Stochastic Model; Hidden Markov Model.

1. INTRODUCTION
With a rapid growth of online trading, there are number of products available on the internet for shopping. To provide proper customer satisfaction and to know about the product's performance, it is a very common practice to enable the customers to provide online review comments. So for a popular product, thousands of review comments are given. Now the problem arises when a new surfer or a new customers wants to know about the overall reviews for the product or the manufacturer/seller wants to know about performance of their product, then they have to look into each comment and take out conclusion that how well their product is doing. But this will not be a feasible method as it could be very tedious and time taking task to go through each and every comment and summarize overall feedback about the product. We believe that this problem will become increasingly important as more people are buying and expressing their opinions on the Web. We have proposed a model that can analyse the sentiment of online customer's review. Our experiment has been accomplished by training the Hidden Markov Model. The Hidden Markov Model (HMM) is a statistical Markov Model in which the system is supposed to be Markov process with (hidden) states. Hidden Markov Models are mainly known for their various pattern recognition techniques such as speech, handwriting, POS tagging, musical score following, gesture recognition, partial discharges and bioinformatics.

In HMM, the state is not directly visible to the observer, but the output, dependent on the state, is visible. Every state has a probability distribution over the possible output tokens. Thus the sequence of the tokens which is generated by HMM provides some information about the sequence of states. The adjective 'hidden' mainly refers to the state sequence through which the model is passed and not to the parameters of the model; the model can still be referred to as a 'hidden' Markov model even if all these parameters are known.

Fig 1. Probabilistic Parameters of Hidden Markov Model



X--states, y--possible observations, a--state transition probabilities, b- output probabilities.

ICARCSET '15, March 06 - 07, 2015, Unnao, India
Copyright 2015 ACM 978-1-4503-3441-9/15/03…$15.00
http://dx.doi.org/10.1145/2743065.2743077

This trained Hidden Markov Model can test for single (POS tagged) sentence, whether the comment is negative or positive oriented as well as its overall performance like (True Positive Rate)TPR, (True Negative Rate)TNR, (False Positive Rate)FPR, (False Negative Rate)FNR, Recall, Precision Accuracy and F-Measure can also be computed. It shows that how our model is working for sentiment analysis on the available dataset using some predefined MATLAB functions. This trained Hidden Markov Model will automatically extract the customer opinions present in the review comments on various product features. Our experimental results indicates that the proposed model is very promising in performing its tasks and we have tried to achieve maximum possible Precision and Accuracy rate.

2. RELATED WORK

Sentiment analysis has been done by many researchers in recent years. Most of this work concentrated on finding the sentiment associated with a sentence (and in some cases, the entire review). There has also been some research on automatically extracting product features from review text. Though there has been some work in review summarization, and assigning summary scores to products based on customer reviews, there has been relatively little work on improving the precision of the system in evaluating the sentiment (positive or negative comment) about the review comment.

Our work is closely related to Minqing Hu and Bing Liu's work in [1]. To get the exact feedback, they have solved the problem of summarizing whole review comments for any given product on the Web using Feature Based Summarization (FBS) technique and used three steps

a) Feature extraction: This process involves mining of product features from the given set of sentence i.e. the comment given by the customer using Natural Language Processing (NLP). E.g. Sound quality of a music system.
b) Identification of comments: It is done to find whether the opinion is Positive or Negative. First of all adjective words are chosen out using NLP and then Opinion words are determined and then finally opinion orientation for each sentence is decided.
c) Summarization of the overall result.

The main disadvantage is that the feature tagging created issue in this research when user doesn't specified exact word for the feature. e.g. 'the mobile fits in pocket easily' doesn't mention the feature SIZE of the mobile. There are three main review formats on the Web. Different review formats may need different techniques to perform the above tasks.
Format (1) - Pros and Cons: The reviewer is asked to describe Pros and Cons separately. Clnet.com uses this format.
Format (2) - Pros, Cons and detailed review: The reviewer describes Pros and Cons separately and also writes a detailed review. Epinions.com uses this format.
Format (3) - free format: The reviewer can write freely, i.e., no separation of Pros and Cons. Amazon.com uses this format.
However, the techniques used by Minqing Hu and Bing Liu in Mining and summarization of customer reviews which are primarily based on unsupervised itemset mining, are only suitable for reviews of formats (3) and (1). Reviews of these formats usually consist of full sentences. The techniques are not suitable for Pros and Cons of format (2), which are very brief.

Bin Lu, Minquing Hu and J.Cheng in [2] proposed observation of review comments given by the customers for products available on web and implemented the use of Opinion Observer System. This Opinion Observer System helps in comparison and analysis of two or more product features and extracts out the product features from the given Pros and Cons in a review comment. This paper mainly focused on the Opinion Observer System. It compares customers opinion on various features of any product. E.g. if some customers are satisfied with the picture quality of one mobile and some are satisfied with the memory space of another mobile, then opinion observer helps to find out how many (strength) customers are giving positive/negative feedback for picture quality and memory space of both the product.

In Classification and Summarization of Pros and Cons for Customer Reviews [3] by X. Hu and Bin Wu, summarization of phrases are done rather than summarizing of sentence or words. It includes the weighing of a sentiments in all comments and then put positive in Pros and negative in Cons section and then take out generalised summary based on the two sets using key phrases extraction method. It assigns score to each word to detect the weightage of the sentiments.

Animesh Kar and Deba Prasad Mandal in [4] proposed the finding of strength of opinion polarity. It used Fuzzy Logic and special type of miner called Fuzzy Operation Miner for determining the intensity of the opinion about the product feature. Like Comment GOOD is inferior than EXCELLENT. Three subtasks involved in this process is as follows:-

a) Extracting and evaluating fuzzy weight to the product features.
b) Identifying the opinion about the product.
c) Determining the strength/intensity of the opinion of the product feature.

The work proposed in this paper is helpful in determining the ranking of the product by using Fuzzy Opinion Miner (FOM) with maximum precision.

In [5], Fangtao Li, Han, Huang and Zhu made use of movie and product review as dataset and also new machine learning framework by using CRF (Conditional Random Framework) which jointly extracts features and opinion for that feature in a unified way. CRF are used to make a particular structure model from the given sentence.

In [6], Zhongwu Zhai & Peifa Jia have used clustering of features having similar meaning. Most of the papers worked on product feature extraction only but in this paper , they have used clustering of features. For e.g. Picture quality and Visual quality are similar feature of a mobile. For this problem semi supervised leaning method Using EM Algorithm (based on Naive Bayse Classification) has been used in this paper. Two procedures have been used here:

a) Sharing of words having similar groups like battery, battery life, battery power.
b) Lexically Similar words grouping based on WordNet, e.g. movie and picture belongs to same group.

In [7], a feature based product ranking technique is used to find the product feature within the product category and also its subjective and comparative sentence in the review. After the sentiment extraction of those review comments, they have modelled a relationship among various products by plotting a weighted graph and then relative quality of a product as compared to another product has been calculated. Particularly, the main contributions are:

a) Natural Language Processing & some dynamic programming techniques are used to identify subjective/comparative sentences in review comments and determine their opinion.
b) Sentence classification techniques are also used to build a graph which shows the inherent quality of the products in terms of their quality/feature.
c) A ranking algorithm that uses that graph to develop a ranking list of the chosen products based on each considered product feature i.e., the end result of the algorithm is a ranking list that is useful for the

potential customer to know the best products among all other products.

Feature-based Opinion Mining and Ranking [8], done by Eirinaki, Shamita Pisal and J. Singh presented an opinion search engine system that incorporates two novel opinion mining algorithms. The proposed framework not only classifies a review, but also extracts the most representative features of each reviewed item, and assigns opinion scores to them.

A paper introduced OPINE [9], an unsupervised information extraction system that gives solution to Opinion Mining Task. OPINE is built on top of the Know-It-All Web information-extraction system. This paper focused on three review mining subtasks and their contributions are as follows:

a) They have introduced OPINE, a review mining system whose novel component helps to find the semantic orientation, in context of given product feature.

b) Compared OPINE with other previously used review mining system.

c) Used OPINE for opinion phrase extraction and opinion phrase polarity determination.

In [10], J.Yu, Zheng-Jun Zha dedicate to the topic of aspect ranking, which aims to automatically identify important product aspects from online consumer reviews. They have developed an Aspect Ranking Algorithm and experimented on 11 popular product's reviews.

3. METHODOLOGY

The experiment is implemented on MATLAB software package and the steps involved in our proposed work is as follows:

1. The dataset we have used is taken from Amazon.com. This dataset consists of review comments on various popular products and is in Part of Speech (POS) tagged format. This tagged dataset is useful to train the Hidden Markov Model by reading and analysis of the data. Some ratio of dataset is taken for training purpose and rest is used for testing.

2. For every entry in the dataset, our system reads and identifies all the tags and classes using "strfind" function in MATLAB. Here tags refers to Noun, Pronoun, Adjective, Verb, Determinant etc. and the sentiment i.e. positive or negative belongs to classes.

3. All the positive or negative tags are then stored in Data_vect and Data_class variable.

4. The “hmmestimate” function is used to evaluate the Sequences and States, which will give the Transition and Emission probabilities.

5. For analysis of sentiment of a review comment, a loop will run which will match the desired output of the comment with all the previously obtained Emission Probabilities.

6. After analysis of the sentiment in any comment, we can also examine the overall performance of our trained HMM. All the performance measures like Accuracy, Precision, Recall, F-Measures can be calculated.

TABLE 1. Dataset & Variables used in the MATLAB coding.

Tag Set	Corresponding Entities
<data_vector>	Tags, symbols
<data_class>	Class (+ve or -ve opinion)
<data_vect_len>	States
<data_class_n>	Known data class
<data_class_p>	Predicted data class
<trans>	Transition probability
<emis>	Emission Probability

The Hidden Markov Model is a generative, probabilistical model which can handle large variations in the input values. Basically there are two functions in MATLAB which we have used for the training purpose. First is the 'hmmestimate' which is required to know the sequence of states. The following function takes the emission and state sequences and returns estimates of the transition and emission matrices:

[TRANS_EST, EMIS_EST] = hmmestimate (seq, states).

3.1 Estimation of Posterior State Probabilities

The posterior state probabilities of an emission sequence 'seq' are the conditional probabilities that the model is in a particular state when it generates a symbol in sequence (seq), given that the sequence is emitted. We compute the posterior state probabilities with hmmdecode function as:

PSTATES = hmmdecode (seq, TRANS, EMIS).

The output PSTATES is an M x L matrix, where M is the no. of states and L is the length of sequence "seq". PSTATES(i,j) is the conditional probability that the model is in state i when it generates the jth symbol of "seq", given that "seq" is emitted. "hmmdecode" begins with the model in state 1 and step 0, prior to the first emission. PSTATES(i,1). To return the log of the probability of the sequence "seq", we have used the second output argument of hmmdecode as :

[PSTATES,logpseq]=hmmdecode(seq,TRANS,EMIS).

The probability of a sequence tends to 0 as the length of the sequence increases, and the probability of a sufficiently long sequence becomes less than the smallest positive number your computer can represent. " hmmdecode" returns the logarithm of the probability to avoid this problem.

3.2 Sentiment Classification

Sentiment Classification of reviews is useful to shoppers, but also crucial to product manufacturers. For experiment purpose, we have taken the dataset from Amazon.com for variety of products. The review comments available are already POS tagged, it means that the product features are usually nouns or noun phrases in review sentences. The process also identifies simple noun and verb groups (syntactic chunking). Also the review already has a class (e.g., some quantitative or binary ratings). Thus the part-of-speech tagging is crucial.

The following shows an online customer's review comment with POS tagging and defined classes. In the below stated customer's review comment (taken from Amazon.com) the starting of the comment is defined with <SENTENCE>, the tagged sentence is identified by <POS>. After the tagging of the

sentence it is then arranged and combined in proper sequence and each TAG is closed in curly brackets { } which we will use to detect each TAG one by one. Once the TAGS are arranged properly qualities of the product and its features are extracted. Also the last sentence contains a binary number or any number to let the system easily identify the Sentiments about the product. Here in this dataset [1] is used for positive opinion and [2] is used for negative comment. Our system has used these tagging and sequencing for analysis of sentiment given in the customers review for any particular product.

<SENTENCE> I mean the camera is amazing, i actually think that it takes better or equal pics with my 6mp Sony.
<POS> I/PRP mean/VBP the/DT camera/NN is/VBZ amazing/JJ ,/, i/FW actually/RB think/VBP that/IN it/PRP takes/VBZ better/JJR or/CC equal/JJ pics/NNS with/IN my/PRP$ 6mp/CD Sony/NN ./.
<SEQUENCE> {I, PRP}{mean, VBP}{the, DT}{camera, NN}{is, VBZ}{amazing, JJ}{i, FW}{actually, RB}{think, VBP}{that, IN}{it, PRP}{takes, VBZ}{better, JR}{or, CC}{equal, JJ}{pics, NNS}{with, IN}{my, PRP$}{6mp, CD}{Sony, NN}
<INST_NUM = 1> 1_it, 3_pics, 4_better, [1]

4. EXPERIMENTAL RESULT

In this section we evaluate the performance of trained HMM based Sentiment analyzer system. We conducted our experiments on customer reviews on different product categories (Digital Camera, Mobile, Laptop, Music System etc) extracted from Amazon.com. Performance measures like accuracy, precision, recall and F-measure have been computed for analyzing the classification on customer review based on Hidden Markov Model. Below are the Performance Measures:

- Accuracy is related to systematic error and is defined as degree of closeness as compared to a standard quantity. Its formula is given by (TP + TN)/(P + N).
  (True Positive + True Negative)/(Positive + Negative) i.e. (TP + TN)/(P + N).

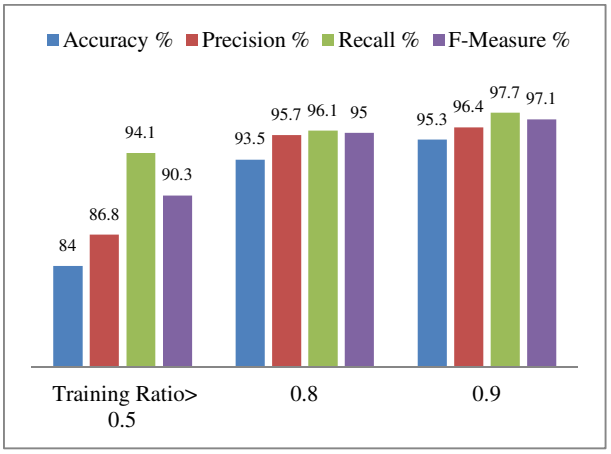
- Precision is related to random errors and refers to exactness of the measured quantity. Formula for Precision in terms of True and false Positive rate is given by (TP)/(TP + FP).
  (True Positive)/(True Positive + False Positive) i.e. (TP)/(TP + FP).

- F-measure is the Harmonic mean of Precision and recall and its formula is given by 2*Precision*Recall/(Precision + Recall).

- Recall refers to completeness of the measured quantity as compared to standard value and is given by following formula (True Positive)/(True Positive + False Negative) i.e. (TP)/(TP + FN) i.e. TP/(TP+FN).

In above definitions,
TP = Number of Positives correctly detected by the system.
TN = Number of Negatives correctly identified by the system.
FP = Number of Positives identified wrongly as Negative by the system.
FN = Number of Negatives wrongly identified as Positive by the system.
P = Number of Positives comments present in the dataset.
N =Number of Negatives comments present in the dataset.

Performance Calculation.



5. CONCLUSION & FUTURE WORK

In this paper, we proposed a technique for developing Hidden Markov Model based sentiment analyzer which will help in analysing online customer reviews. The objective is to provide a Sentiment-based result for a large number of customer reviews of a products sold online. Our experimental results indicate that the proposed technique is very promising in performing its tasks. Previously any model that has been used, were little less efficient than our model.

In our future work, we plan to further improve and refine our techniques, and deal with the outstanding problems of feature extraction. Also we will look into summarization of customer reviews. We believe that summarization will be particularly useful to product manufacturers and to the customers because they want to know about overall performance (positive or negative) comments of their available products.

6. ACKNOWLEDGMENT

The author thanks the National Institute of Technology, Raipur for providing the infrastructure and its support to carry out the Research work.

7. REFERENCES

[1] Minquing and Liu,. *Mining and summarizing customer reviews.* In Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-*2004), 2004.*

[2] Liu, Hu and Cheng. *Opinion Observer: Analyzing and comparing opinions on the web".* WWW 2005, Chiba, Japan, May 10-14, 2005.

[3] Hu and Wu. *Classification and Summarization of Pros and Cons for Customers Review*. IEEE 2009.

[4] Kar, Mandal.*"Finding Opinion Strength Using Fuzzy Logic on Web Reviews".* International Journal of Engineering and Industries, vol 2, March, 2011.

[5] Li, Han, Huang, Zhu. *Structure Aware Review Mining and Summarization.* Proceedings of the 23rd International Conference on Computational Linguistics, Beijing, August 2010.

[6] Zhai, Liu, Xu, and Jia. *Clustering Product Features for Opinion Mining.* WSDM'11, Hong Kong, China, February 9–12, 2011.

[7] Zhang, Narayanan, and Choudhary. *Mining Online Customer Reviews for Ranking Products.* Technical Report, EECS department, Northwestern University, 2009.

[8] Eirinaki, Pisal, Singh. *Feature Based Opinion Mining and Ranking.* Journal of Computer and System Sciences, 10.1016/j.jcss.2011.10.007.

[9] Popescu and Etzioni. *Extracting Product Features and Opinions from Reviews*.

[10] Yu, Zha, Wang, Chua. *Aspect Ranking: Identifying Important Product Aspects from Online Consumer Reviews.* Proceedings of the 49th Annual Meeting of the Association for Computational LinguisticsPortland, Oregon, June 19-24, 2011.