IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
PP 64-69

Sentimental Analysis Based on the Features of Car Domain

Mrs. K. Chitra[1], Dr. A. Tamilarasi / HOD[2]
[1](Assistant Professor), Department of Computer Applications, Kongu Engineering College, Perundurai.
[2]Department of Computer Applications, Kongu Engineering College, Perundurai.

Abstract: *Opinion Mining is a type of natural language processing for tracking the mood of the people about any particular domain or product by review. The opinions are reviews from customers; comments are collected from web sites and user groups. The collected opinions are manipulated by various techniques, methods, algorithms and software tools to extract the polarity information. In this work, Support Vector Machine presents an approach for mining online user reviews to generate feature-based sentiment analysis on car domain that can guide a user in making an online purchase.*

Support Vector Machine is the supervised machine learning algorithm for opinion mining. It is used for classification purpose. In which features are extracted using various attributes of particular domain, then classify the selected attributes of car and predict the polarity.

Since the polarity of each feature is varied according to the domain. The outcome of the system is a set of reviews organized by their degree of positivity and negativity based on each feature. The polarity obtained for each feature from this approach is with good average accuracy.

Keywords: *Opinion Mining, Sentiment Analysis, Feature-Based Opinion Mining, Summarization,*

I. Introduction

Sentiment Analysis or Opinion Mining is an ongoing field of research in text mining field. The process of identifying and detecting subjective information using natural language processing, text classification. In short, the aim of sentiment analysis is to extract information on the attitude of the writer or speaker towards a specific product or service and predict the polarity or ratings about the particular product or service.

Sentiment analysis is extremely useful in social media monitoring as it allows the user to gain an overview of the wider public opinion behind certain topics. Collecting the customer reviews from social networks or any other web and apply the machine learning algorithms to classify the text for predict the polarity which is very useful for new customers.

In traditional model the opinions is considered as binary classification and uses the different machine learning methods in which SVM maintained to the best. Support Vector Machine (SVM) is a supervised learning method that is applicable for both classification and regression problems and gives the higher performance in terms of classification accuracy as compared to other data classification algorithms. The distinct property of the SVM is at the same time that minimizes the empirical classification mistakes and maximizes the geometrical margin.

One form of opinion mining in product reviews is to produce a feature-based summary. In this model, features of a particular domain are first identified, and positive and negative opinions on them are aggregated to produce a summary on the features. Features of a product are attributes, components and other aspects of the product, e.g., "engine quality", "battery life" and "mileage" of a particular car domain. In reviews (or any writings), people often use different words and phrases to describe the same product feature.

1 Evaluation of Opinion:

The Evaluation of Opinion is divided into two types.

1.1 Direct Opinion: It talks about only a single object. It gives positive or negative expressions about an object, product, topic or person directly [4]. For example, "This Bike has poor mileage" expresses a direct opinion about a single object i.e. Bike.

1.2 Comparison Opinion: It talks about multiple objects. This type of opinion expresses similarities or differences between more than one object. For example, "Car X is better than Y" expresses a comparison opinion between two objects i.e. Car.

II. Related Works

Grouping feature expressions, which are domain synonyms, is critical for effective opinion summary. Since there are typically hundreds of feature expressions that can be discovered from text for an opinion mining application, it's very time-consuming and tedious for human users to group them into feature categories. Some automated assistance is needed. Unsupervised learning or clustering is the natural technique for solving the

problem. The similarity measures used in clustering are usually based on some form of distributional similarity. Recent work also used topic modelling. However, we show that these methods do not perform well. Even the latest topic modelling method that consider pre-existing knowledge does not do well.

Obviously, thesaurus dictionaries can be helpful for finding synonyms but they are far from sufficient due to a few reasons. First, many words and phrases that are not synonyms in a dictionary may refer to the same feature in an application domain. For example, "appearance" and "design" are not synonymous, but they can indicate the same feature, design. Second, many synonyms are domain dependent. For example, "movie" and "picture" are synonyms in movie reviews, but they are not synonyms in camera reviews as "picture" is more likely to be synonymous to "photo" while "movie" to "video".

We exploit two pieces of natural language knowledge to achieve this:

Sharing words: Feature expressions sharing some common words are likely to belong to the same group, e.g., "battery life", "battery", and "battery power".

Lexical similarity: Feature expressions that are similar lexically based on WorldNet are likely to belong to the same group, e.g., "movie" and "picture". Note that synonyms are covered by lexical similarity as they will have very high lexical similarity.

Lexical resources such as SentiWordNet contain opinion bias scores based on individual terms, and when building a model based on this type of information there are certain challenges stemming from the nature of natural languages. Word sense disambiguation becomes relevant, since terms with potentially multiple meanings may carry different opinion bias depending on context and their use within a sentence.

Domain-specific terms are also an issue, since they may indicate a different bias than that of their more commonly seen uses. Even every domain has its own set of features, and polarity of each feature is varies according to the domain. The above issues naturally impose limitations to the effectiveness of sentiment classification using SentiWordNet.

In various research works, SentiWordNet is used to do sentiment classification reviews, but there is need to the feature based sentiment classification. If these user reviews classified appropriately and summarized based on features can play an instrumental role in influencing the buyer's decision.

III. Proposed Work

In the proposed work section, clearly explain the system design of the feature based summarizer and polarity classification implemented by us. This proposed work focus on car reviews from various car websites which post car reviews. As shown in Figure 1, this approach has following phases which explained next. These phases are

(1) Feature extraction phase by domain ontology, (2) Review collection and pre-processing phase,
(3) Support vector machine algorithm

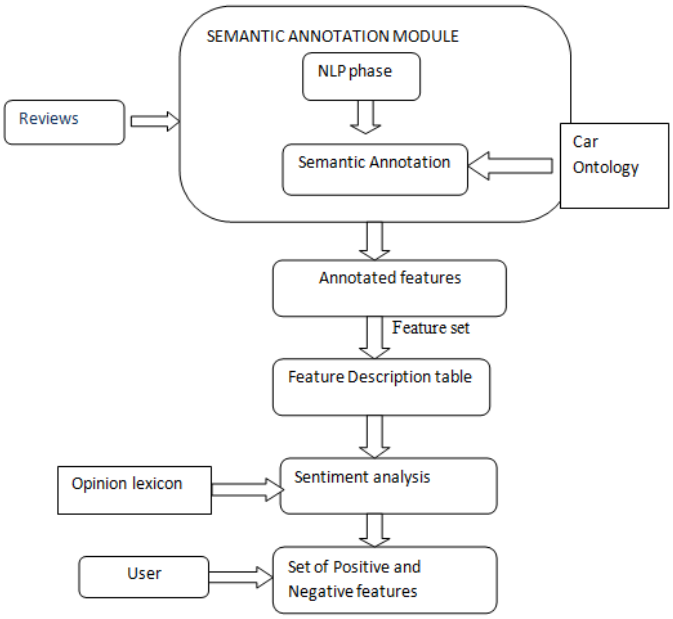

Fig 1 System Architecture

Our feature based opinion mining system needs three basic components: a lexical resource L of opinion expressions, a lexical ontology where each concept and each property is associated to a set of labels that correspond to their linguistic realizations and a review R.

3.1 Feature extraction using domain ontology

The main aim of ontology is to provide knowledge about specific domains that are understandable by both the computers and developers. It also helps to interpret a text review at a finger granularity with shared meanings and provides a sound semantic ground of machine understandable description of digital content. Ontology improves the process of information retrieval and reasoning thus results in making data interoperable between different applications.

For each expression, the system tries to find out whether the expression under question is an individual of any of the classes of the car ontology. Then, system retrieves all the annotation respective to expression. Each class in the ontology is defined by means of a set of relations and data type properties. Then, when an annotated term is mapped onto an ontological individual, its data type and relationships constitute the potential information which is possible to obtain for that individual. It helps to define meaning, concepts, relationships and entities that describe a domain with unlimited number of terms. These sets of terms are very helpful for extracting explicit and implicit features. For example, in the following restaurant review: cold and not tasty the negative opinion not tasty is ambiguous since it is not associated to any lexicalized feature.

However, if the term cold is stored in the ontology as a lexical realization of the concept quality of the cuisine, the opinion not tasty can be easily associated to the feature cuisine of the restaurant.

It provide structure for these features through their concept hierarchy but also their ability to define many relations linking these concepts. This is also a valuable resource for structuring the knowledge obtained during feature extraction task. This way it helps to extract domain features.

Table 1 Feature Description Table

Feature	Positive polarity	Negative polarity
comfort	Complacent, enjoyable, satisfying, useful, cozy,	Dissatisfied, unpleasant, unsuited, unfriendly, discontented, hopeless, pitiable, disagreeable, troubled,
price	Cheap, reasonable, Valuable, inexpensive, worthy, affordable, good, low	Expensive, pricey, extravagant, invaluable, worthless, exorbitant, costly, high-priced, high.
speed	Fast, quick, swift, speedy, nimble, brisk, high-speed, high, good, rapid	Slow, bad,
quality	Awesome, cool, good, satisfactory, superfine, bang-up, fabulous, fantastic,	Terrific,bad, inferior, low-grade, substandard, unsatisfactory, atrocious, awful, execrable, pathetic, poor,
performance	High, good	Poor,terrible,pathetic,bad,worst
engines	Good, awesome, cool, satisfactory, superfine, bang-up, fabulous, quiet	Bad,worst,terrific,loe-grade,unsatisfactory,pathetic,poor,terrible

Table 2 Overall accuracy

Features	Using FD table		Using SENTIWORDNET		Average accuracy of polarity(FD)	Average accuracy of polarity(SENTIWORDNET)
	Total	Total negative	Total positive	Total negative		
cost	17	08	09	16	0.6	0.3
speed	20	9	15	7	0.8	0.6
safety	16	09	14	12	0.6	0.5
comfort	18	07	12	18	0.7	0.4
mileage	15	12	11	13	0.6	0.4
Lighting system	21	6	16	10	0.8	0.6
Internet service	12	11	11	10	0.6	0.4
seat	19	8	18	9	0.7	0.7
Total					0.6	0.4

3.2 Reviews collections and pre-processing:
It collects the various car related reviews which contains positive and negative opinions and stores it in database. Document Term Matrix is used to pre-process the data. The data is converted into document term matrix form.

3.2(a) Document Term Matrix
A document term matrix or term document matrix is a mathematical matrix that describes the frequency of terms that occur in a collection of documents. In a document-term matrix, rows correspond to documents in the collection and columns correspond to terms.

3.3 SVM Algorithm
Support Vector Machine (SVM) is a powerful, state-of-the-art algorithm based on linear and nonlinear regression. SVM has strong regularization properties. Regularization refers to the generalization of the model to new data. Support Vector Machine (SVM) is a supervised machine learning algorithm which can be used for both classification and regression challenges. However, it is mostly used in classification problems. In this algorithm, plot each data item as a point in n-dimensional space (where n is number of features) with the value of each feature being the value of a particular coordinate. Then, perform classification by finding the hyper-plane that differentiates the two classes very well.

3.3(a) SVM in Car Dataset
In Support Vector Machine, always look for two things:
1. Setting a training dataset and build the SVM model for it.
2. Test the data with testing and training dataset.

3.3(b) Document Term Matrix
Load the training dataset in the particular environment R and create a document term matrix with the help of header variable in the dataset. The function create_matrix is used to create the document term matrix. It contains the information about the number of row and columns are involved in the preprocessed data. The container will be created for configuring the training dataset by using create_container package.

3.3(c) SVM Model
Applying SVM Linear Classifier, in the created container dataset with the low cost value. SVM model will be created by giving the parameter kernel as linear. Classifier is achieved with the support of train_model package.

3.3(d) Test Data
Predict the list of test data and load into the particular container which is used to test the training dataset. Again create the document term matrix for the test data which is defined by user. Prediction container will be created with the size of prediction matrix (document term matrix of test data).

3.3(e) Classify the Model
SVM model is classified by using the container
Test data and SVM classifier of Training data. Test data will be classified according to the classification of training dataset. It classifies the various number of classes which is predicted in the training data.
A research work on car dataset for analysing the polarity of car in various attributes. The results are shown with the help of graphs on running for various iterations. After selecting the particular attributes of car data set, which is applied on SVM algorithm to classify the data then the classified results are shown.

Classification Performance
To judge the correctness of proposed work, the classifier performance evaluator phases have facilities the calculated of various classification performance. These measures are as follows:

Reference					
Prediction	1	2	3	4	5
1	8	0	0	0	0
2	1	11	1	0	0
3	0	2	16	0	0
4	0	0	0	9	1
5	0	0	0	0	7

Table 3 Confusion Matrix

where 1,2,3,4,5 are the classes which predicts the ratings of car attributes such as mileage and acceleration. Confusion matrix provided in table 1 is used to estimate these terms. The column in the matrix represents the instances of a predication class while the row represents the instances of an actual class.

IV. Graph Representation Based On The Review Evaluation

For this work, applied Pre-processing steps like sentence boundary detection, spell-error correction on the review dataset as explained earlier in the pre-processing phase. We obtained the feature set as explained in the feature extraction phase, generated the feature description table and determined the polarity according to the features of domain. Since polarities of features are varied according to domain our approach helps to get results with high accuracy.

SVM classifier then calculate accuracy, kappa value of the classifier.

Overall Statistics
Accuracy : 0.9107
95% CI : (0.8038, 0.9704)
No Information Rate : 0.3036
P-Value [Acc > NIR] : < 2.2e-16
Kappa : 0.8854
Mcnemar's Test P-Value : 0.6025370

Table 4 Accuracy Results

Since our aim is to do the feature based sentiment analysis we have done the polarity classification based on features as follows:

The results we obtained show that polarity classification using feature description table has high accuracy then polarity.

Domain: CAR

Graph represents the polarity classification using our system:

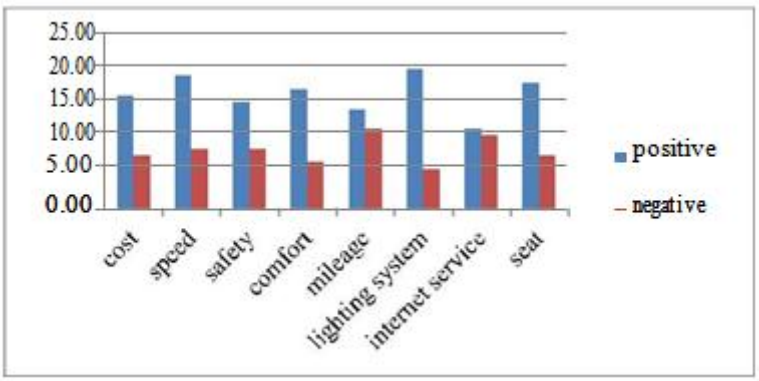

Fig 2 Sentiment classification of features

V. Conclusion And Future Work

Opinion mining has become a fascinating research area due to the availability of a huge volume of user-generated content, e.g., reviewing websites, forums, and blogs. Aspect-based opinion mining, which aims to extract item aspects and their corresponding ratings from online reviews, is a relatively new sub-area that attracted a great deal of attention recently. The extracted aspects and estimated ratings not only ease the process of decision making for customers but also can be utilized in other opinion mining systems.

Opinion mining is an important field of data mining that has been emerging with a boom. The proposed methodology is supported by natural language processing methods to annotate car reviews, then the features are classified into positive and negative polarity with the help of predicted ratings. Various features of the car are manipulated by SVM algorithm and predict the accurate results. Also the various factors are analysed in which the factors like confusion matrix, accuracy with global cost are applied on the data. The outcome of the system is a set of reviews organized by their degree of positivity and negativity based on each feature and achieve the higher accuracy when compared to existing system in the same domain. This system helps to reduce the manual effort of evaluating reviews according to features in which user is interested.

The opinions matter a lot while mining the sentiments from social media, any forums or websites and so on. The proposed system helps to give the uniform accuracy. In future, extend feature based opinion mining focus on various domains. Also like to extend the work to find out the strength of various features which help to increase the accuracy of sentiment analysis.

References

[1]. A. Nisha Jebaseeli, Dr. E. Kirubakaran "M-Learning Sentiment Analysis with Data Mining Techniques", International Journal of Computer Science And Telecommunications, Volume 3, Issue. 8, Aug. 2012.

[2]. Arti Buche Dr.M.B. Chandak, Akshay Zadgaonkar, "Opinion Mining And Analysis : A Survey", International Journal on Natural Language Computing (IJNLC), Volume 2, No. 3, June 2013.

[3]. "Data Mining Concepts and Techniques" Jiawei Han, Micheline Hamber Morgan Kaufman Publishers, 2003.

[4]. G. Angulakshmi, Dr. R. Manickachezian, "An Analysis on Opinion Mining : Techniques and Tools", International Journal of Advanced Research in Computer and Communication Engineering, Volume 3, Issue. 7, July 2014.

[5]. Nidhi Mishra, Dr. C.K. Jha, "Classification of Opinion Mining Techniques", International Journal of Computer Applications, Volume 56, No.13, October 2012.

[6]. Sumathi.T, Karthik.S, Marikannan.M "Performance Analysis of Classification Methods for Opinion", International Journal of Innovations in Engineering and Technology (IJIET), Volume 2, Issue 4, August 2013.

[7]. G. Vinodhini, RM.Chandrasekaran, "Sentiment Analysis and Opinion Mining: A Survey", International Journal of Advanced Research in Computer and Communication Engineering, Volume 2, Issue 6, June 2012.

[8]. K.G. Nandakumar, Dr.T.Christopher "Opinion Mining: A Survey", International Journal of Computer Applications, Volume 113, No.2, March 2015.

[9]. Zhu Zhang, 2008 Weighing Stars, "Aggregating Online Product Reviews For Intelligent E-Commerce Applications", IEEE Intelligent Systems, 42-49.

[10]. Ziqiong Zhang, Qiang Ye, Zili Zhang, Yijun Li, "Sentiment Classification of Internet Restaurant Reviews written in Cantonese", Expert Systems with Applications, 2011.

[11]. B.Liu. 2010 "Sentiment Analysis and Subjectivity", Second Edition, The Handbook of Natural Lanugage Processing.

[12]. N.Mishra and C.K.Jha 2012, "An Insight into task of opinion mining", Second International Joint Conference on Advances in Signal Processing and Information