



International Journal of Emerging Technologies in Computational and Applied Sciences (IJETCAS)

www.iasir.net

Summarizing and computing Orientation of user reviews using Bayesian Classification

Gaurav Dubey¹, Ajay Rana², Jayanthi Ranjan³
Amity University¹, Amity University², IMT Ghaziabad³
Sec125, Noida U.P.¹ Sec 125, Noida U.P.², IMT Ghaziabad³
INDIA

Abstract: Lots of Information is available on web today regarding the feature of a product. It has become a trend and necessity to explore web and study reviews about a product which gives us lots of information. Product reviews are in the form of text which contains number of opinion about a product, these opinions can be positive, negative or neutral, and studying reviews texts at the document level is useful in many cases. A positive review text on a particular object or product does not mean that the author has positive opinions on every aspect of the product. Likewise, a negative review does not mean that the author dislikes everything about the product. For example, in a product review, the reviewer usually writes both positive and negative aspects of the product, although the general sentiment on the product could be positive or negative. To obtain such detailed aspects, we need to go to the sentence level. This paper explains the model and techniques of extracting features from various user opinions.

Key Words: Sentiment Classification, Semantic Orientation, Opinion mining

I. Introduction

Web contains huge amount of data which is categorized as structured or unstructured in nature. Though structured data mining gives extraction of useful information, unstructured data which is mostly in form of text documents also give very useful information from the consumer and business point of view. Opinion mining[1] is now a days is very useful from various business perspective. It can be consider as the extraction of useful information from user feed back and reviews. Whenever a company launches any new product or service, a large number of user opinions and reviews can be found on web. These reviews or feedback contains positive, negative or neutral opinions about service or product. There are large number of sites which keep records of various reviews posted by various users, one such web is www.teambhp.com which gives thousands of reviews for various automobiles. All the reviews which are in the form of text documents are available in the form of text data. The task is not only technically challenging because of the need for natural language processing, but also very useful in practice. For example, businesses always want to find public or consumer opinions on their products and services. Potential customers also want to know the opinions of existing users before they use a service or purchase a product. Moreover, opinion mining can also provide valuable information for placing advertisements in Web pages. If in a page people express positive opinions or sentiments on a product, it may be a good idea to place an ad of the product. However, if people express negative opinions about the product, it is probably not wise to place an ad of the product. A better idea may be to place an ad of a competitor's product. The Web has dramatically changed the way that people express their opinions. They can now post reviews of products at merchant sites and express their views on almost anything in Internet forums, discussion groups, blogs, etc., which are commonly called the user generated content or user generated media.[1] A basic task in sentiment analysis is classifying the *polarity* of a given text at the document, sentence, or feature/aspect level-whether the expressed opinion in a document, a sentence or an entity feature/aspect is positive, negative, or neutral. Advanced, "beyond polarity" sentiment classification looks, for instance, at emotional states such as "angry," "sad," and "happy. The focus of our project is to mine the opinions on the Web(I-Phone in our case) and try to get a grip on the user sentiments expressed in that opinion. The task is not only technically challenging because of the need for natural language processing, but also very useful in practice.

For example, businesses always want to find public or consumer opinions on their products and services. Potential customers also want to know the opinions of existing users before they use a service or purchase a product. Moreover, opinion mining can also provide valuable information for placing advertisements in Web pages. If in a page people express positive opinions or sentiments on a product, it may be a good idea to place an ad of the product. However, if people express negative opinions about the product, it is probably not wise to place an ad of the product. A better idea may be to place an ad of a competitor's product.

The Web has dramatically changed the way that people express their opinions. They can now post reviews of products at merchant sites and express their views on almost anything in Internet forums, discussion groups, blogs, etc., which are commonly called the **user generated content** or **user generated media**. This online word-of-mouth behavior represents new and measurable sources of information with many practical applications. Techniques are now being developed to exploit these sources to help businesses and individuals gain such information effectively and easily.

The text documents which contains these reviews contains lots of text which contains positive, negative or neutral information in, Fig 1.1 showing a review of a phone which is showing that what you will like about the phone and what you don't.

Fig 1.1 Text containing review of iphone-5

Review1: *The iPhone 5 is a great piece of hardware accompanied by an operating system which has started to show its age. Apple makes it easy for consumers by launching just one new phone per year while the competition launches hundreds of devices every year. If you are using an iPhone 4S, you could give the iPhone 5 a pass but if you are using an iPhone 4 or any previous generation iPhone the iPhone 5 is a must have upgrade. Then again if you are one of those who love to get your hands on the best phones first, go ahead and get the iPhone 5. Good luck finding one. They are in great demand.*

Review 2: *I got my iphone 5 as a birthday gift. That was about 1 month ago. Until then i have never had an iphone. Been using nokia since 2000. I could choose between s4 s3 and htc one. I took a good friends advice and settled with the iphone. I honestly must say that I am amazed. The camera on the phone is brilliant, it records like a video camera. Also its really easy and intuitive to use. Everything seems to be where it should. Android based phones are more juiced up lets say but the feel and usage of the iphone are unbelievable. Furthermore it doesnt lag, i can open all my apps and still get the same response when wrtting an email or playing a video game. I have some cons also. The battery life is poor in my opinion, you can use the phone for just one day with wifi and 3g on. Ive been used with my e72 which could go on for a matter of several days. Besides that the iphone 5 is the best phone for me. Hope ive helped people who are in doubt on what to buy*

The text is called the evaluative text which contains the lots of English grammar which can be evaluated for meaningful results. Because of the importance of automatically extracting actionable knowledge from customer feedback data on web “opinion mining” has become a significant subject of research in the field of data mining the ultimate goal of opinion mining is to extract customer opinions on products and present the information in such a way that it is most effective and useful in terms of decision making. The method that are needed for feature extraction, sentiment classification, and opinion summarization are very new and upcoming, they consist lots of scope also for further research, these methods can be modified and can be applied for opinion mining.

II. Related Work

Now a days there is wide range of research has been done on customer reviews, the purpose of research is to identify opinions of user about the product the work has been done so far is on sentiment detection. Subjectivity in natural language refers to aspect of language used to express opinions and evaluations (Wiebe, 1994). The problem of subjectivity classification is to distinguish sentences used to prevent opinions and other form of subjectivity. Most prior work in learning to identify sentiment has focused on the binary distinction of positive vs. negative. Koppel and Schler (2005a, 2005b) show that it is crucial to use neutral examples in learning polarity for a variety of reason. As it become the common practice for online merchant to ask their customers to review the products that they have purchased. Since large number of peoples are using web to express their opinion, the number of reviews that a product receives grow rapidly. Most of the researcher about these reviews were focused on automatically classifying the products into “recommended” or “not recommended” (Pang, Lee, & Vaithyanathan, 2002; Ranjan Das & Chen, 2001; Terveen, Hill, Amento, McDonald, & Creter, 1997). A product has shortcomings in one aspect, may have merits in other place (Morinaga, Yamanishi, Tateishi, & Fukushima, 2002; Taboada, Gillies & McFetridge, 2006). Visual side by side comparison also helps to compare the various features of products. Liu, Hu, and Cheng (2005) proposed a novel

framework for analyzing and comparing consumer opinions of competing products. Another work which was related to the opinion mining is of summarization of reviews as there are large number of reviews available online for some popular product like Iphone. Opinion summarization (ku, Lee, Wu & Chen, 2005; Philip et al., 2004) summarizes opinions of articles by telling sentiment polarities, degree and the correlated events. With opinion summarization, a customer can easily see how the existing customers feel about a product, and product manufacturer can have different vies of customers. Hu and Liu (2004a, 2004b) conduct a work which is also similar to the previous work: given a set of customer reviews of a particular product, the task involve three subtasks, (i) identifying features of the product that customers have expressed their opinions; (ii) for each feature identify the polarity of sentence (iii) to finally produce the summary of products.

Ku, Liang and Chen (2006) investigated both news and web blog articles. Algorithms for opinion extraction at word, sentence and document level are proposed. Relevant sentence selection is discussed and then topical and opinionated information are summarized. Sometimes it is not just sufficient to find the polarity of opinions or aggregation and quantifying degree assessment of opinions available on web. Opinion reason mining tries to identify one of the critical element of online reviews to answer the question, “ what are the reasons that the author of this review likes or dislikes the product?” (Cardie, Wiebe, Wilson, & Litman 2003; Clarke & Terra, 2003; Li & Yamanishi, 2001; Stoyanov, cardie, Litman, & Wiebe, 2004) worked in the area of opinion reason mining.

There is also an another form of review where opinion are presented in form of pros and cons, Kim and Hovy (2005) proposed a method for detecting opinion bearing expression .In their subsequent work (Kim & Hovy, 2006), collected large set of review data from epinions.com, which explicitly have pros and cons in the review. The automatic labeling system first collects phrases in pro and con fields and then searches the main review text in order to collect sentences corresponding to those phrases. Thomas, Pang and Lee (2006) try to determine from the transcript of US Congressional floor debates whether the speeches represent support of or opposition to proposed legislation. Mullen and Malouf (2006) describe a statistical sentiment analysis method on political discussion group posting to judge whether there is opposing political viewpoint to the original post.

III. Methodology

Approach: In this paper I am mainly discussing about **Bayesian classification** with the help of Bayes Probability theorem. The algorithm given in fig 1.4 uses Part- Of- Speech(POS) tagging. Some common POS categories in English grammar are: noun, verb, adjective, adverb, pronoun, preposition, conjunction and interjection. Number of other categories can be generated from these categories further. In this paper I am using Penn Treebank POS tags as shown table 1.1[10].

Table 1.1 Penn Treebank POS tags

Tag	Description	Tag	Description
CC	Coordinating conjunction	PRP\$	Possessive pronoun
CD	Cardial number	RB	Adverb
DT	Determiner	RBR	Adverb, comparative
EX	Existential <i>there</i>	RBS	Adverb, superlative
FW	Foreign word	RP	Particle
IN	Preposition or subordinating conjunction	SYM	Symbol
JJ	Adjective	TO	to
JJR	Adjective, Comparative	UH	Interjection
JJS	Adjective, superlative	VB	Verb, base form
LS	List item marker	VBD	Verb, past tense
MD	Modal	VBG	Verb, gerund or present participle
NN	Noun, singular or mass	VBN	Verb, past participle
NNS	Noun, plural	VBP	Verb, non-3 rd person singular present
NNP	Proper noun, singular	VBZ	Verb, 3 rd person singular present
NNPS	Proper noun, plural	WDT	Wh-determiner
PDT	Predeterminer	WP	Wh-pronoun
POS	Possessive ending	WP\$	Possessive wh-pronoun
PRP	Personal pronoun	WRB	Wh-adverb

POS tagging is the task of labeling (or tagging) each word in a sentence with its appropriate POS. The algorithm given extracts two consecutive words, where one member of the pair is an adjective/adverb and other is context word. Now we can generate a pattern which can extract two consecutive words based on their POS tags. Such patterns is given in table 1.2[10],

Table 1.2 Patterns of tags for extracting two-words phrases from

S.No	First Word	Second word	Third word(Not Extracted)
1	JJ	NN or NNS	anything
2	RB,RBR, or RBS	JJ	not NN nor NNS
3	JJ	JJ	not NN nor NNS
4	NN or NNS	JJ	not NN nor NNS
5	RB, RBR, or RBS	VB,VBD,VCN,VBG	anything

used for extracting two-words phrases from reviews. The tagging of the reviews which are given is done with the help of Stanford tagger. Fig 1.2 and Fig 1.3 shows the tagged output of the reviews given in Fig 1.1.

Fig 1.2 Text containing tagged review of iphone-5

Stunning_JJ hardware_NN design_NN Improved_VBN camera_NN New_NNP Ear_NN pods_NNS ear_NN phones_NNS Above_IN average_JJ battery_NN life_NN 2G_NN -LRB_-LRB- GPRS_NN ,_, EDGE_NN -RRB_-RRB- ,_, 3G_NN -LRB_-LRB- HSPA_NN ,_, HSPA_NN +_CC -RRB_-RRB- and_CC 4G_NN LTE_NN data_NNS supported_VBD Taller_JJR screen_NN might_MD be_VB a_DT bit_NN uncomfortable_JJ for_IN a_DT few_JJ iOS_NNP is_VBZ starting_VBG to_TO look_VB dated_JJ Maps_NNS and_CC Navigation_NNP is_VBZ of_IN very_RB poor_JJ quality_NN outside_IN the_DT US_NNP New_NNP nano_NNP SIM_NNP card_NN is_VBZ not_RB available_JJ widely_RB

Fig 1.3 Text containing tagged review of iphone-5

I_PRP got_VBD my_PRP\$ iphone_NN 5_CD as_IN a_DT birthday_NN gift_NN ._. That_DT was_VBD about_RB 1_CD month_NN ago_RB ._. Until_IN then_RB i_FW have_VBP never_RB had_VBN an_DT iphone_NN ._. Been_VBN using_VBG nokia_NN since_IN 2000_CD ._. I_PRP could_MD choose_VB between_IN s4_NN s3_NN and_CC htc_NN one_CD ._. I_PRP took_VBD a_DT good_JJ friends_NNS advice_NN and_CC settled_VBN with_IN the_DT iphone_NN ._. I_PRP honestly_RB must_MD say_VB that_IN I_PRP am_VBP amazed_VBN ._. The_DT camera_NN on_IN the_DT phone_NN is_VBZ brilliant_JJ ,_, it_PRP records_VBZ like_IN a_DT video_NN camera_NN ._. Also_RB its_PRP\$ really_RB easy_JJ and_CC intuitive_JJ to_TO use_VB ._. Everything_NN seems_VBZ to_TO be_VB where_WRB it_PRP should_MD ._. Android_NN based_VBN phones_NNS are_VBP more_RBR juiced_VBN up_RP lets_VBZ say_VBP but_CC the_DT feel_NN and_CC usage_NN of_IN the_DT iphone_NN are_VBP unbelievable_JJ ._. Furthermore_RB it_PRP doesnt_JJ lag_NN ,_, i_FW can_MD open_VB all_DT my_PRP\$ apps_NNS and_CC still_RB get_VB the_DT same_JJ response_NN ehen_NN wrtting_VBG an_DT email_NN or_CC playing_VBG a_DT video_NN game_NN ._. I_PRP have_VBP some_DT cons_NNS also_RB ._. The_DT battery_JJ life_NN is_VBZ poor_JJ in_IN my_PRP\$ opinion_NN ,_, you_PRP can_MD use_VB the_DT phone_NN for_IN just_RB one_CD day_NN with_IN wifi_NNS and_CC 3g_NN on_IN ._. Ive_RB been_VBN used_VBN with_IN my_PRP\$ e72_NN which_WDT could_MD go_VB on_IN for_IN a_DT matter_NN of_IN several_JJ days_NNS ._. Besides_IN that_IN the_DT iphone_NN 5_CD is_VBZ the_DT best_JJS phone_NN for_IN me_PRP ._. Hope_NNP ive_NN helped_VBD people_NNS who_WP are_VBP in_IN doubt_NN on_IN what_WP to_TO buy_VB

The tagged input as mentioned is computed by the algorithm given in fig 1.4

Fig 1.4 Algorithm To Extract Phrases Based On Bayes Classification

ALGORITHM TO EXTRACT PHRASES & ORIENTATION BASED ON BAYESIAN CLASSIFICATION

Step 1.Input the tagged review from user and store it in a string array //String tagged<- Tagged review

Step 2. Find length of tagged review and if review is single word then display error message to user show("Review cant be a Single Word, Try Again!")

Step 3. Separate the tags from review

```
ArrayList tag1 <- identify and separate tags from the tagged review
    if (tag1.isEmpty()) {
        Error( "No tags found, Please entered a tagged review using POS tagger"); }
    else {
        result = extract(pos_arr[i], pos_arr[i + 1], tagged,s_tag); //extracts phrases based on Rules
    }
```

Step 4. Display the extracted phrases to the user extracts and stores the phrases based on Rules given in table 1.2

```
total_adj <- total number of adjectives in the review
e_adj <- total number of extracted adjectives from the review
```

Step5. For each extracted adjective Classify the extracted adjectives as positive or neagative and store their count.

Step 6.Calculate semantic orientation of the extracted adjectives using Bayes' Formulae

$P(A|B)=P(B|A)*P(A)/P(B)$ where,
 $P(A|B)$ is conditional probability that training data B belongs to class A
 $P(B|A)$ is the conditional probability of B given AP(A) is the probability of class A
 $P(B)$ is the probability of training data B

Step 7. Display the semantic orientation to the user as positive , negative or mixed ; end;

Fig 1.5 Example showing how Bayesian Classification Compute orientation

Given Data Set:			
Total Adjective	Extracted Adjective	Positive	Negative
30	22	15	7
28	8	9	9
19	15	8	7

we have to find $P(\text{positive/extracted adjective})$, $P(\text{negative/extracted adjective})$ {using Naive Baye's}

Three Classifiers: Positive, Negative & Mixed

Formulae:

$$P(+/E) = (P(E/+) * P(+)) / P(E) \quad // \text{ for positive orientation}$$

$$P(-/E) = (P(E/-) * P(-)) / P(E) \quad // \text{ for negative orientation}$$

example 1

$$P(22/+)= (15/22 * 15/30)P(E) = 0.340/P(E)$$

$$P(22/-)= (7/22 * 7/30)P(E) = 0.0742/P(E)$$

Since $P(22/+) > P(22/-)$
therefore, dataset 1 belongs to class positive.

IV Results

The above proposed method has been applied on various set of reviews of i-phone and other mobile phones, the input review is first processed by Stanford tagger into tagged input. I have collected around hundred reviews of i-phone and other phones from web , all these reviews first stored into separate txt files. Using tagger all these input files are converted into tagged input and stored separately. All the tagged input file are then taken as input to the program developed by me which computes the summary of reviews and their orientation. Table 1.3 below shows some of the results computed using above proposed method.

Table 1.3 Shows the output of the various customers reviews.

Tagged Input	Summary of Review	Orientation
Nice_JJ looking_VBG design_NN , looks_VBZ to_TO have_VB decent_JJ battery_NN life_NN , light_JJ weight_NN and_CC decent_JJ speed_NN on_IN wifi_NNS and_CC LTE_NN A_DT good_JJ improvement_NN over_IN the_DT iPhone_NNP 4_CD but_CC not_RB the_DT 4S_NN	light weight good improvement	The Review has positive orientation
The_DT iPhone_NNP 5_CD is_VBZ no_DT massive_JJ innovation_NN , it_PRP is_VBZ super_JJ sleek_JJ to_TO hold_VB , attracts_VBZ eyes_NNS , and_CC is_VBZ extremely_RB lightweight_JJ for_IN its_PRP\$ capability_NN . Its_PRP\$ well_RB made_VBN and_CC looks_VBZ good_JJ , performs_VBZ doubly_RB better_JJR than_IN the_DT iPhone_NNP 4S_NN . However_RB its_PRP\$ not_RB a_DT lot_NN different_JJ from_IN its_PRP\$ predecessor_NN for_IN a_DT normal_JJ user_NN , who_WP will_MD not_RB notice_VB the_DT difference_NN b/w_VB them_PRP , apart_RB from_IN the_DT obvious_JJ styling_NN and_CC looks_VBZ . The_DT camera_NN is_VBZ much_RB better_JJR and_CC the_DT video_JJ quality_NN is_VBZ much_RB more_RBR improved_VBN .	massive innovation super sleek extremely lightweight well made lot different normal user not notice obvious styling video quality more improved	The Review has mixed orientation.
Stunning_JJ hardware_NN design_NN Improved_VBN camera_NN New_NNP Ear_NN pods_NNS ear_NN phones_NNS Above_IN average_JJ battery_NN life_NN 2G_NN -LRB_-LRB- GPRS_NN , EDGE_NN -RRB_-RRB- , 3G_NN -LRB_-LRB- HSPA_NN , HSPA_NN +_CC -RRB_-RRB- and_CC 4G_NN LTE_NN data_NNS supported_VBD Taller_JJR screen_NN might_MD be_VB a_DT bit_NN uncomfortable_JJ for_IN a_DT few_JJ iOS_NNP is_VBZ starting_VBG to_TO look_VB dated_JJ Maps_NNS and_CC Navigation_NNP is_VBZ of_IN very_RB poor_JJ quality_NN outside_IN the_DT US_NNP New_NNP nano_NNP SIM_NNP card_NN is_VBZ not_RB available_JJ widely_RB	Stunning hardware average battery bit uncomfortable dated Maps poor quality	The Review has a negative orientation
The_DT iPhone_NNP 5_CD is_VBZ a_DT great_JJ piece_NN of_IN	great piece	The Review has positive

hardware_NN accompanied_VBN by_IN an_DT operating_NN system_NN which_WDT has_VBZ started_VBN to_TO show_VB its_PRPS age_NN ._. Apple_NNP makes_VBZ it_PRP easy_JJ for_IN consumers_NNS by_IN launching_VBG just_RB one_CD new_JJ phone_NN per_IN year_NN while_IN the_DT competition_NN launches_VBZ hundreds_NNS of_IN devices_NNS every_DT year_NN ._. If_IN you_PRP are_VBP using_VBG an_DT iPhone_NNP 4S_NN ._. you_PRP could_MD give_VB the_DT iPhone_NNP 5_CD a_DT pass_NN but_CC if_IN you_PRP are_VBP using_VBG an_DT iPhone_NNP 4_CD or_CC any_DT previous_JJ generation_NN iPhone_NNP the_DT iPhone_NNP 5_CD is_VBZ a_DT must_MD have_VB upgrade_NN ._. Then_RB again_RB if_IN you_PRP are_VBP one_CD of_IN those_DT who_WP love_VBP to_TO get_VB your_PRPS hands_NNS on_IN the_DT best_JJS phones_NNS first_RB ._. go_VB ahead_RB and_CC get_VB the_DT iPhone_NNP 5_CD ._. Good_JJ luck_NN finding_VBG one_CD ._. They_PRP are_VBP in_IN great_JJ demand_NN	new phone previous generation null go Good luck great demand	orientation
The_DT phone_NN is_VBZ great_JJ especially_RB for_IN web_NN browsing_NN and_CC its_PRPS so_RB fast_JJ and_CC quite_RB easy_JJ to_TO work_VB ._. The_DT only_JJ major_JJ bad_JJ thing_NN about_IN this_DT phone_NN is_VBZ the_DT battery_NN life_NN ._. With_IN all_PDT the_DT functions_NNS the_DT phone_NN has_VBZ you_PRP want_VBP to_TO use_VB it_PRP all_PDT the_DT time_NN like_IN a_DT computer_NN but_CC battery_NN life_NN is_VBZ so_RB poor_JJ i_FW have_VBP to_TO have_VB a_DT battery_NN charger_NN on_IN me_PRP all_PDT the_DT time_NN ._. It_PRP wont_VBD even_RB last_JJ half_PDT a_DT day_NN if_IN your_PRPS using_VBG it_PRP alot_NN ._. They_PRP should_MD have_VB made_VBN it_PRP thicker_JJR to_TO add_VB a_DT bigger_JJR battery_NN life_NN ._. It_PRP needs_VBZ to_TO be_VB doubled_VBN in_IN order_NN to_TO be_VB up_RB to_TO the_DT standard_NN of_IN the_DT phones_NNS capabilities_NNS ._.	so fast quite easy only major bad thing so poor even last	The Review has a negative orientation
In_IN terms_NNS of_IN usability_NN and_CC reliability_NN ._. just_RB go_VB for_IN iphone_NN 5_CD ._. Iphone_NNP 5_CD hardware_NN maybe_RB low_JJ compared_VBN to_TO most_RBS high_JJ end_NN android_NN phones_NNS ;. iOS_NNP makes_VBZ better_JJR out_IN of_IN it_PRP ._. iphone_NN 5_CD may_MD only_RB have_VB 1Gigabyte_NN of_IN ram_NN but_CC most_RBS 3d_JJ games_NNS runs_VBZ smoothly_RB in_IN iOS_NNP than_IN a_DT 2Gigabyte_JJ ram_NN on_IN android_NN ._. Apple_NNP products_NNS have_VBP only_RB one_CD disadvantage_NN ._. proprietary_NN ._. If_IN you_PRP dont_VBP mind_VB the_DT bucks_NNS and_CC only_RB care_NN for_IN quality_NN in_IN both_CC hardware_NN and_CC software_NN ;. go_VB for_IN iphone_NN 5_CD ._.	just go maybe low high end only have 3d games 2Gigabyte ram	The Review has mixed orientation
The_DT weight_NN of_IN the_DT handset_NN is_VBZ surprising_JJ ._. for_IN its_PRPS size_NN I_PRP expected_VBD it_PRP to_TO weigh_VB more_JJR than_IN it_PRP does_VBZ ._. Its_PRPS a_DT beautiful_JJ phone_NN with_IN cracking_VBG features_NNS ._. like_IN auto_NN brightness_NN and_CC customisable_JJ notification_NN leds_NNS ._.	beautiful phone customisable notification	The Review has positive orientation

V Conclusion

In this paper I have discussed a approach based to find the outcome from the user reviews which is based on the Bayesian classification. The proposed method not only summarizes the reviews but it also compute thr orientation of the review, hence saving lots of time to find a conclusion of a review. The main advantage of document level Bayesian classification is that it provides good opinion about an product , but it has certain shortcomings also which are like it does not give details on what people liked or disliked. Compared to other method which has been used so far to compute the orientation of the customer reviews, the Bayesian is easy to implement. In a typical evaluative text such as a review or feedback, the customer usually writes specific aspects of an object that he/ she dislikes. Now as it has been discussed earlier in this paper that are two more mining task are there which also give the mining summarization based on product review. In later cases the product is being considered as object which contains the explicit and implicit features of the product. The opinion mining is the upcoming field in data mining and still lots of scope is there for further research.

VI References

- [1] Bing Liu, Web data mining, Exploring hyperlinks , Contents and Usage Data, Springer.
- [2] P. Turney. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In *Proc. of the Meeting of the Association for Computational Linguistics (ACL '02)*, pp. 417–424, 2002
- [3] B. Santorini. *Part-of-Speech Tagging Guidelines for the Penn Treebank Project*. Technical Report MS-CIS-90-47, Department of Computer and Information Science, University of Pennsylvania, 1990.
- [4] V. Stoyanov and C. Cardie. Toward Opinion Summarization: Linking the Sources. In *Proc. of the Workshop on Sentiment and Subjectivity in Text*, pp. 9–14, 2006.
- [5] J-T. Sun, X. Wang, D. Shen, H-J. Zeng, and Z. Chen. CWS: A Comparative Web Search System. In *Proc. of the 15th Intl. Conf. on World Wide Web (WWW'06)*, 2006.
- [6] K.-C. Tai. The Tree-to-Tree Correction Problem. *Journal of the ACM*, 26(3), pp. 422–433, 1979.
- [7] P. -N. Tan and V. Kumar. Discovery of Web Robot Sessions Based on their Navigational Patterns. *Data Mining and Knowledge Discovery*, 6(1), pp. 9–35, 2002.
- [8] P. -N. Tan, V. Kumar and J. Srivastava. Selecting the Right Interestingness Measure for Association Patterns. In *Proc. of the ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining (KDD '02)*, pp. 32–41, 2002.
- [9] P.-N. Tan, M. Steinbach, and V. Kumar. *Introduction to Data Mining*. Addison-Wesley 2006
- [10] The Penn Treebank site <http://www.cis.upenn.edu/~treebank/home.html>
Pedro Domingos and Michael J. Pazzani. 1997. On the optimality of the simple Bayesian classifier under zero-one loss. *Machine Learning*, 29(2-3):103– 130.
- [11] Aidan Finn, Nicholas Kushmerick, and Barry Smyth. 2002. Genre classification and domain transfer for information filtering. In *Proc. of the European Colloquium on Information Retrieval Research*, pages 353–362,
- [12] Glasgow.Vasileios Hatzivassiloglou and Kathleen McKeown. 1997. Predicting the semantic orientation of adjectives. In *Proc. of the 35th ACL/8th EACL*, pages 174–181.
- [13] Vasileios Hatzivassiloglou and Janyce Wiebe. 2000. Effects of adjective orientation and gradability on sentence subjectivity. In *Proc. of COLING*.
- [14] Marti Hearst. 1992. Direction-based text interpretation as an information access refinement. In Paul Jacobs, editor, *Text-Based Intelligent Systems*.
- [15] Lawrence Erlbaum Associates. Alison Huettner and Pero Subasic. 2000. Fuzzy typing for document management. In *ACL 2000 Companion Volume: Tutorial Abstracts and Demonstration Notes*, pages 26–27.
- [16] Thorsten Joachims. 1998. Text categorization with support vector machines: Learning with many relevant features. In *Proc. of the European Conference on Machine Learning (ECML)*, pages 137– 142.
- [17] Thorsten Joachims. 1999. Making large-scale SVM learning practical. In Bernhard Schölkopf and Alexander Smola, editors, *Advances in Kernel Methods - Support Vector Learning*, pages 44–56. MIT Press.
- [18] Jussi Karlgren and Douglass Cutting. 1994. Recognizing text genres with simple metrics using discriminant analysis. In *Proc. of COLING*.