Natural Language Processing

Project Report

Jatan Patel (jjp140230)

Product Prediction Utilizing Activities of the
Customer on E-Commerce Website

Introduction

One of the biggest operational challenges faced by e-commerce industry today is to find the products which will interest the customer most and to predict what he/she is going to buy next? E-commerce organizations have been attracting the customers to buy more and more products online and have been offering many features to assist the shopping. Major functionalities provided are Product reviews, customer wish list, Product ratings, less expensive deals, Recommendations, High level product categorization, asking questions regarding the product features directly to the previous buyers and many more. A regular customer explores all the possible options before buying a certain product. There is a high probability that the organization may fail to consider some of the activities done by the customer before deciding to buy a particular product or not. Instead of suggesting irrelevant products, if enough concentration is given to the products which the customer is more likely to buy then considerable amount of increment in sales can be driven.

Reviews of customers who already have bought the product and rated the product are the major source of information as they have real experience of the product rather than what is written by the manufacturing company. A large percentage of new customers read the reviews carefully and based on the overall ratings the decide whether to go for the product or not.

Customers look for various features of the product. Some of the product features are more important to some customers than other features which are being branded. In my project I tried to help the customer who shops online by suggesting the products which have been graded better in the features he/she is looking for. My aim for this project is to maximize the product recommendations which are more relevant to a particular customer. Clearly, providing the products what the customer really wants will increase the sale.

Prior Research

I achieved my target by applying "Aspect Based Sentimental Analysis" on product reviews by customers on E-commerce website. ABSA is recommended by many research papers and has been proven effective to conduct the project which I proposed. Aspect Based Sentiment Analysis (ABSA) systems receive as input a set of texts (e.g., product reviews or messages from social media) discussing a particular entity (e.g., anew model of a mobile phone). The systems attempt to detect the main (e.g., the most frequently discussed) aspects (features) of the entity (e.g., 'battery', 'screen') and to estimate the average sentiment of the texts per aspect (e.g., how positive or negative the opinions are on average for each aspect). Although several ABSA systems have

been proposed, mostly research prototypes, there is no established task decomposition for ABSA, nor are there any established evaluation measures for the subtasks ABSA systems are required to perform. 3 main tasks are integral part of the logic behind ABSA:

1) Aspect Term Extraction
2) Aspect Term Aggregation
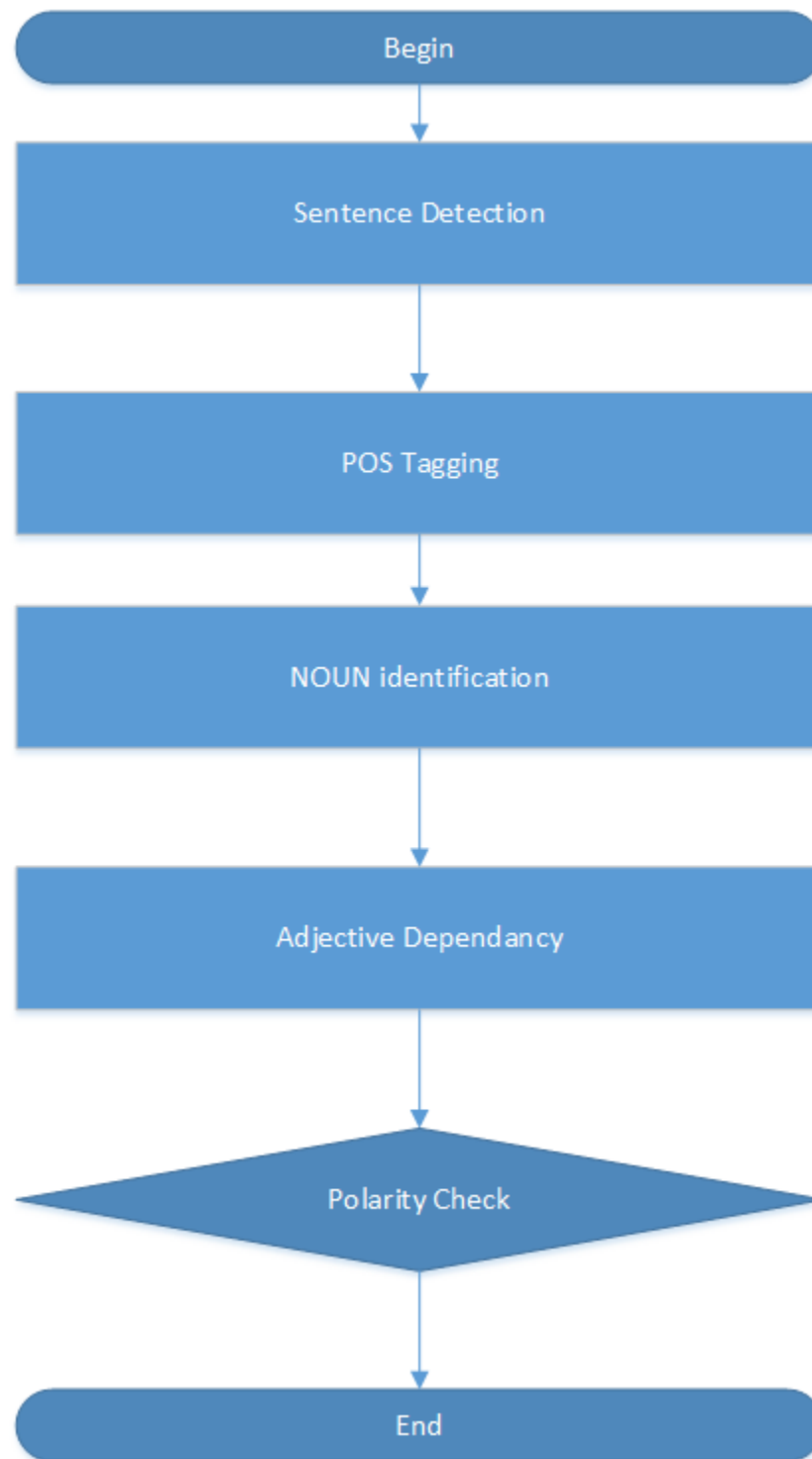3) Polarity Estimation of Aspect Term

Each task can be performed with different methods. The combination of all 3 sub tasks decides to narrow down the choices of products as per user preferences. Recommendations having more positive reviews regarding customer selections is the target here.

Natural Language processing concepts and techniques like Sentiment Analysis, Information Extraction as well as Information Retrieval were involved to solve the problem. I utilized over 100 Amazon customer reviews data of 'Digital' and 'SLR' cameras. I selected the product to be 'Camera' because of large number of features available. Professional photographers require different qualities than armatures. I distinguished 10 different aspects and implemented an ABSA using JAVA. JAVA was chosen due to the ability of comfortably prototyping an idea and support of the quality resources. Apache OpenNLP JAVA framework assisted me to implement the system.

System Design
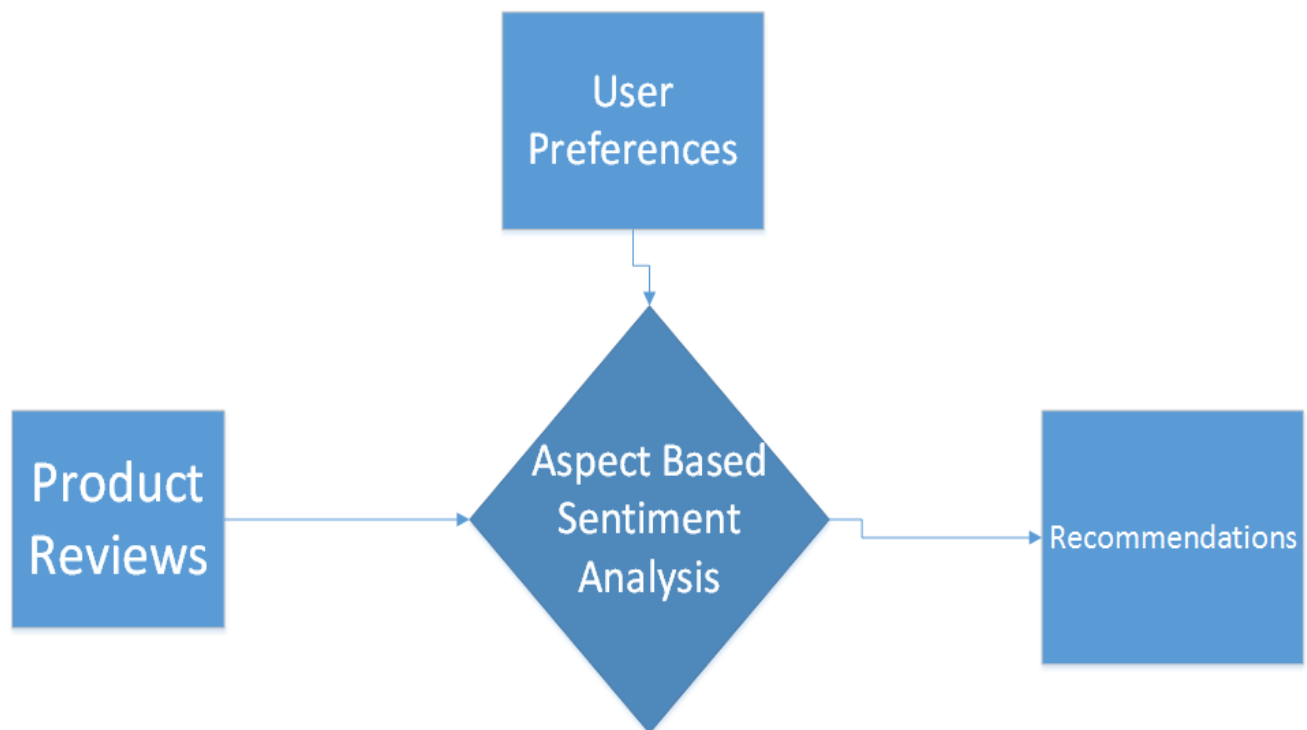
Above is the System design diagram upon which the logic was based. Following is the algorithm flow chart:

```
┌──────────────────────────────────────┐
│                Begin                 │
└──────────────────────────────────────┘
                    │
                    ▼
┌──────────────────────────────────────┐
│          Sentence Detection          │
└──────────────────────────────────────┘
                    │
                    ▼
┌──────────────────────────────────────┐
│             POS Tagging              │
└──────────────────────────────────────┘
                    │
                    ▼
┌──────────────────────────────────────┐
│          NOUN identification         │
└──────────────────────────────────────┘
                    │
                    ▼
┌──────────────────────────────────────┐
│         Adjective Dependancy         │
└──────────────────────────────────────┘
                    │
                    ▼
               ◇ Polarity Check ◇
                    │
                    ▼
┌──────────────────────────────────────┐
│                 End                  │
└──────────────────────────────────────┘
```

The basic algorithm which was designed is as follows:

1) Apply Sentence Splitter using OpenNLP Framework
2) Apply Part-Of-Speech Tagger on the detected sentence
3) Extract tokens which are identified as 'Noun'. These tokens have potential of being the targeted entity or aspect in the sentence.
4) Find the sentiment of the adjectives
5) Apply Dependency parser on (Noun, Adjective).
6) Find the polarity of the sentence.

Here the entity is what the user is looking for in the product. Reviews are filtered based on this algorithm and the resultant products are suggested to the customer.

System Input

System is taking 'User Preference' as an input as shown below:

Example:

What kind of camera you are looking for?

1. Digital 2.SLR

User Response: 1

Which of the following aspects you are looking for?

1. Zoom 2.Picture Quality 3. Color 4. Weight 5. Noise Reduction 6. Point & Shoot 7. Memory 8. Battery 9. Control 10. Resolution

User Response: 1

System Output

The output of the system is 'Suggested Products' based on the input as shown below:

Example:

Fetching Results...

Following products are best suitable for you:

Product No : 1

Product ID : b00004w3yx

Title : kodak dc4800 3.1mp digital camera w/ 3x optical zoom

Summary : fantastic

Rating : 5.0

Review :   this camera is amazing! it had everything we were looking for, 3.1 mp, 3x zoom, and rechargable batteries. moreover, you can override the auto control and manually select the f-stop, iso, shutter speed, color saturation, exposure, sharpness and exposure metering (center-weighted, etc...). it even allows for you to take black and whites (with 3 filter options, none, red or yellow) and sepia photos. it is certainly the most powerfull of the kodak line, it is a shame they seem to be moving away from these features and going after the market share where people don't want all these options. that isn't to say that you can't just point and click... you can, it has auto-focus and auto exposure/speed controls... but i like the ability to override those, especially when i have backlighting or want to play with the focal depth. oh, and another rare feature is that it has a video output jack that you can run to a tv to view your pictures on. this has come in handy when visiting friends or relatives and i wanted to show them pictures we had in our camera but didn't want to download them to a computer. the menu system is amoung the more intuitive that i have seen. all manual settings can be changed with just a couple button presses and i have yet to delete a picture by accident. some options, like f-stop and exposure control can be controled from dials on the camera without even going into any menu. the battery is good, recently i took the camera out of town for a long weekend and forgot the battery charger. we took pictures at many events over a few days and just barely got by before the battery ran out. of course, we have learned over time some tips on how to extend the battery life such as minimizing lcd viewing of pictures. the 3x optical zoom is nice but i would prefer a little more. i don't like the quality digital zoom provides, but fortunately there is a setup option to disable digital zoom so that you don't accidently use it.as close to perfect as it is, there are some minor gripes. the manual exposure control dial seems a little flimsy like it might come off one day, fortunately it is rarely used. second there is a little door that you open that covers the usb and battery ports that also seems a little flimbsy. also, the button you use to turn on/off the camera is a bit too close to the shutter button. if someone else is taking our picture, they invariably press the wrong button unless we make it clear which button to press. the flash is somewhat weak (although comparible to small film cameras). another gotcha is that you have to press the shutter button half way down to lock in the auto focus, speed and f-stop settings before pressing the button the rest of the way down to take the picture. although this is standard with nearly all but the most inexpensive cameras, it is something that you have to carefully explain to anyone else who you want to take a picture or the picture will likely be out of focus. but then again, it is digital so you can look at the picture right away and determine if it needs to be re-shot. finally, i think that the 16 meg card is much too small for any type of vacation use. at the highest

resolution but least compressed jpg, each photo is about 700-900 kb. we purchased an additional 64mb card which are quite cheap now and have yet to run out of memory (and if we did we could go out and buy another from any electronics store).all in all, a great camera!!

Product No : 2

Product ID : b00004w3yx

Title : kodak dc4800 3.1mp digital camera w/ 3x optical zoom

Summary : versatile camera

Rating : 5.0

Review : has many features that you can manually adjust or go completely automatic. you can burst shoot, shoot in two modes of color, b&w or sepia tones. nice 3x optical zoom lense as well as less desirable 2x digital zoom. it is light weight, easy to handle and even has a closeup feature in the lcd of the picture you just took. it even has six choices of quality (compression settings).all in all i highly recommend this camera to all.

Product No : 3

Product ID : b00004w3yx

Title : kodak dc4800 3.1mp digital camera w/ 3x optical zoom

Summary : great camera - still comparable to todays model

Rating : 4.0

Review : excellent photos. simple and easy to use, yet still technically advanced and customizable for the hardcore group. ability to change apeture, light intake , red eye reduction, b&w or sephia photos also.can change an average joe into a photography nut. the included rechargeable battery is great but i still wish you could use regular aa batteries in emergency situations.auto focus is wonderful, manual settings need time to master. sweet zoom lense. good software. able to upgrade the software and the internal firmware via the kodak website. included 16 mb compact flash will give you a max of 47 fair photos (sometimes more) adjustable to 32 great photos or 8 excellent photos.(the great photos are more than acceptable). upgrade to the 128 mb compact card and the 290 great photos will have you smiling forever.

Product No : 4

Product ID : b00004w3yx

Title : kodak dc4800 3.1mp digital camera w/ 3x optical zoom

Summary : everything i thought i wanted and more.

Rating : 5.0

Review : i wouldn't write a review except i read some of the negative reviews below. all i can say is these people must have gotten a lemon from the first day of the assembly line or something. i bought the 4800 and it is awesome. the autofocus on mine was the first thing that impressed me when i looked at the output. it has not occured to me it has ever made a mistake. the depth of field is impressive. the thing is laid out perfectly and fits my hand like a glove. the low light capability is like a spy camera out of a movie. the pictures, the most important thing, which come out of this camera are very very nice.the default color choice out of the box i guess is suppose to emulate kodachrome. sometimes it is a little too much, but the camera has a easy override which takes neutral coloration pictures when you want them. i think people and their clothes look better with the saturated color on but when you need accurate color it can be yours immediately. the zoom has a stop before it gets to the digital, it isn't that hard to detect it for me. the flash works with the auxiliary lenses which block the nikon 990.i will say that while the flash is not physically blocked with any of the auxiliary lenses, the professional quality wide angle lense is very large and does cast a shadow on wide angle shots of art work in the lower right corner from a distance of three or four feet. how many times you use wide angle lenses in a dark room from four feet with the flash will determine how big a problems this is. the camera has such good low light capability you can, i feel, get around this.the batteries are suppose to be expensive. why? because they allow you to shoot, with the lcd on, 40 pictures or so without a problem. the cheap aa batteries belong on a cheap camera. this is suppose to be a nice camera and have a power system that has some oomph. it does. this camera should be about 1100 dollars with the other top prosumer cameras. that it costs only what it does is the achievement to fixate on.

Conclusion

Opinion Mining and Sentiment analysis can be strongly applied for various reasons on social media and e-commerce. Advance and accurate development of tools utilizing Customer Feedback System can be a great input for predictions and Natural Language

Processing can play main role in this. Different statistical methods can be formed to achieve considerable results.

References

1) "ASPECT BASED SENTIMENT ANALYSIS" by Ioannis (John) Pavlopoulos, DEPARTMENT OF INFORMATICS ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS
2) Apache OpenNLP (https://opennlp.apache.org/)
3) The Stanford Parser: A statistical parser (http://nlp.stanford.edu/software/lex-parser.shtml)
4) "Mining and Summarizing Customer Reviews" by Minqing Hu and Bing Liu, Department of Computer Science, University of Illinois at Chicago