www.ijemr.net ISSN (ONLINE): 2250-0758, ISSN (PRINT): 2394-6962

Volume-7, Issue-2, March-April 2017

International Journal of Engineering and Management Research

Page Number: 257-263

An Approach towards Feature Specific Opinion Mining and Sentimental Analysis Across E-Commerce Websites

Lijisha T[1], Archana Balakrishnan[2], Sumayya K S[3], Arjun P K[4], Renjith Sunny[5]

[1,2,3,4]B.Tech Students, Department of Computer Science and Engineering, Ammini College of Engineering, Mankara, Palakkad, Keraka, INDIA

[5]Assistant Professor, Department of Computer Science and Engineering, Ammini College of Engineering, Mankara, Palakkad, Keraka, INDIA

ABSTRACT

The Sentimental Analysis And Opinion Mining Is One Among The Major Tasks Of Natural Language Processing (NLP). This Project Aims To Automate The Process Of Gathering Online, End Users Reviews For Any Given Product Or Services And Analyzing Those Reviews In Terms Of Sentiments Expressed About Specific Features . This Involves The Filtering Of Irrelavent And Unhelpful Reviews , Quantification Of The Sentiments Of Thousands Of Useful Reviews . Data And Knowledge Engineering (DKE) Stimulates The Exchanges Of Ideas And Interaction Between These Two Related Fields Of Interest. DKE Reaches a World Wide Audience Of Researches ,Designers, And Users .

Keywords— About four key words or phrases in alphabetical order, separated by commas

I. INTRODUCTION

The E-commerce websites has a great impact in our day to day life. Now a day they are preferred for shopping and it has made things so much easier that an individual can do his/her shopping from any corner of the world. The main highlights are that, anyone can do shopping without visiting any store directly and also it provides a variety of options for the consumers so he/she can select the best among the provided. The average consumer relies heavily on user reviews and opinions expressed (by peers) on the World Wide Web to learn about any product before making a decision to buy; be it a mobile phone or an automobile. A lot of studies where done to discover the influence of online recommendation on users' decision. But, from the point of view of a manufacturer, making sense out of such diverse data sets (reviews, blogs and user postings) spread across varied, unrelated domains is indeed a cumbersome task. At the same time, it is in the interest of all product manufacturers to ensure that the sentiments expressed online are positive. Moreover, the manufacturer needs to take into account all opinions expressed online. Hence, there is a need for an automated system that can gather, process, summarize and visualize those opinions to facilitate manufacturers and businesses in improving their products as per customer feedback.

II. LITERACTURE SURVEY

A detailed literature survey was done as the initial step of our project. And the following are the papers which we referred as a part of it.

- An approach towards comprehensive sentimental data analysis and opinion mining.

By Pooja Kherwa, Arjit Sachdeva, MSIT New Delhi, India.

- Consumers' reliance on product information and recommendations found in UGC.
- Sentiment analysis and opinion mining
- Opinion Mining and Sentiment Analysis in Social Networks: A Retweeting Structure-aware Approach, Lu Lin, Jianxin Li, Richong Zhang
- New Avenues in Opinion Mining and Sentiment Analysis, Yunqing Xia, Tsinghua University

An approach towards comprehensive sentimental data analysis and opinion mining

The world wide web can be viewed as a repository of opinions from users spread across various websites and networks, and today's netizens look up reviews and opinions to judge commodities, visit forums to debate about events and policies. With this explosion in the volume of and reliance on user reviews and opinions, manufacturers and retailers face the challenge of automating the analysis of such big amounts of data (user

reviews, opinions, sentiments). Armed with these results, sellers can enhance their product and tailor experience for the customer. Similarly, policy makers can analyze these posts to get instant and comprehensive feedback. Or use it for new ideas that democratize the policy making process. This paper is the outcome of the research in gathering opinion and review data from popular portals, e-commerce websites, forums or social networks; and processing the data using the rules of natural language and grammar to find out what exactly was being talked about in the user's review and the sentiments that people are expressing. This approach diligently scans every line of data, and generates a cogent summary of every review (categorized by aspects) along with various graphical visualizations. A novel application of this approach is helping out product manufacturers or the government in gauging response.

Consumer's reliance on product information and recommendations found in UGC

The advent of the internet, the influence of online recommendations on consumer decision making has attracted great attention. You Tube and sites with blogging capabilities, such as Myspace and Facebook, are growing rapidly and frequently feature comments about brands and products. These comments, whether positive or negative, represent a form of user-generated content (UGC). Although recent research on peer recommendations considers electronic word of mouth, few studies focus on UGC. Using interviews with 17 participants, this study examines consumers' opinions of online recommendations embedded in UGC compared with those of producer-generated content. UGC remains a relatively new concept, and research regarding it is unlimited in nature. Because the goal of this research is to understand participants' point of view and attitudes toward UCG, indepth interviews provide a ready way to tap into appropriate information. Research should be made for products in cyberspace and seek UGC about products.

Sentiment analysis and opinion mining

The sentimental analysis and opinion mining paper mainly focuses on evaluating the performance of different methods used for opinion mining by calculating various metrics like precision, recall and F-measure. Precision in the fraction of retrieved instances that are relevant, while recall is the fraction of relevant instances that are retrieved. The two measures are sometimes used together in the F1 score (also F-score or F-measure) is a measure of a test's accuracy. Internet Movie Database (IMDb) is used for movie reviews and product reviews are downloaded from Amazon.com. Movie review mining is a more challenging application than many other types of review mining. The challenges of movie review mining lie in that factual information is always mixed with real-life review data and ironic words are used in writing movie reviews. Product review domain considerably differs from movie review domain because of two reasons.

Opinion Mining and Sentiment Analysis in Social Networks: A Retweeting Structure-aware Approach

Microblogs have become quick and easy online information sharing platforms with the explosive growth of online social media. Weibo, a Twitter-like microblog service in China, is characterized by timeliness and interactivity. A Weibo message carries the user's views and sentiments, particularly forms a fission-like spreading structure while being retweeted. Such structure accelerates information diffusion, and reflects different topics and opinions as well. However, current researches mainly focus on sentiment classification, which neither efficiently combine tree-like retweeting structure nor analyze opinion evolutions with a holistic view. In light of this, we build an opinion descriptive model, and propose an opinion mining method based on this model. With a microblog-oriented sentiment lexicon being constructed, a lexicon-based sentiment orientation analysis algorithm is designed to classify sentiments. Finally, we design and implement a prototype which can mine opinions with respect to retweeting tree structures and retweeting comments.

New Avenues in Opinion Mining and Sentiment Analysis

Opinions can be crucial when it's time to make a decision or choose among multiple options. When those choices involve valuable resources (for example, spending time and money to buy products or services) people often rely on their peers' past experiences Until recently, the main sources of information were friends and specialized magazine or websites. Now, the "social web" provides new tools to efficiently create and share ideas with everyone connected to the World Wide Web. Forums, blogs, social networks, and content-sharing services help people share useful information. This information is unstructured, however, and because it's produced for human consumption, it's not something that's "machine process able. "Capturing public opinion about social events, political movements, company strategies, marketing campaigns, and product preferences is garnering increasing interest from the scientific community (for the exciting open challenges), and from the business world (for the remarkable marketing fallouts and for possible financial market prediction). The resulting emerging fields are opinion mining and sentiment analysis. Although commonly used interchangeably to denote the same field of study, opinion mining and sentiment analysis actually focus on polarity detection and emotion recognition, respectively. Because the identification of sentiment is often exploited for detecting polarity, however, the two fields are usually combined under the same umbrella or even used as synonyms. Both fields use data mining and natural language processing (NLP) techniques to discover, retrieve, and distill information and opinions from the World Wide Web's vast textual information. Mining opinions and sentiments from natural language is challenging, because it requires a deep understanding of the explicit and implicit, regular and irregular, and

syntactical and semantic language rules. Sentiment analysis researchers struggle with NLP's unresolved problems: reference resolution, negation handling, anaphora resolution, named-entity recognition, and word-sense disambiguation

III. SYSTEM ANALYSIS

System analysis is the term used to describe the process of collecting and analyzing facts in respect of existing operation of the solution of the situation prevailing so that an effective computerized system may be designed and implemented of proved feasible. It also diagnosis the problems and using that information recommends improvement to the system. System analysis is the reduction of the entire system by studying the various operations performed and the relationship with the system and requirement of its successor. A system can be defined as an orderly grouping of independent component linked together according to a plan to achieve a specific objective. System analysis may be considered as an interface between the actual problem and computer. Before a computer can perform, it is necessary to investigations are called system analyst. System analysis also embraces system design which is an activity concerned with the design of a computerized application based on the facts disclosed during the analysis stage. The same person who knows as the system analyst carries out both activities. In feasibility study in most cases project is being driven by a problem in the business.

Feasibility Study

A feasibility study is an evaluation of a proposal designed to determine the difficulty in carrying out a designated task. Generally, a feasibility study precedes technical development and project implementation. In other words, a feasibility study is an evaluation or analysis of the potential impact of a proposed project. Feasibility Study is performed to choose the system that meets the performance requirements at least cost. The most essential tasks performed by a Feasibility Study are the identification and description of candidate systems, the evaluation of the candidate systems and the selection of the best of the candidate systems. The best system means the system that meet performance requirements at the least cost. The most difficult part of a Feasibility Study is the identification of the candidate systems and the evaluation of their performances and costs. The new system has no additional expense to implement the system. The new system has advantages such as we can easily access files from any client in the Network, accurate output for accurate input and this application is more user friendly. We can use this application not only in this organization but also in other firms. So it is worth solving the problem.

Technical Feasibility

Technical Feasibility study is performed to check whether the proposed system is technically feasible or not. Technical feasibility centers on the existing computer system (hardware, software, etc.) and to what extent it can support the proposed addition. This involves financial consideration to accommodate technical enhancement. This system is technically feasible. All the data are stored in files. The input can be done through dialog boxes which are both interactive and user friendly.

Economical Feasibility

Economic Feasibility Study is the most frequently used method for evaluating the effectiveness of a candidate system. More commonly known as cost/benefit analysis, the procedure is to determine the benefits and savings that are expected from a candidate system and compare them with cost. This analysis phase determines how much cost is needed to produce the proposed system. As the organization has required machines and supporting programs for the application to execute itself.

Operational Feasibility

Operational feasibility is a measure of how well a proposed system solves the problems, and takes advantage of the opportunities identified during scope definition and how it satisfies the requirements identified in the requirements analysis phase of system development. The operational feasibility assessment focuses on the degree to which the proposed development projects fits in with the existing business environment and objectives with regard to development schedule, delivery date, corporate culture, and existing business processes. To ensure success, desired operational outcomes must be imparted during design and development. These include such design-dependent parameters such as reliability, maintainability, supportability, usability, reducibility, disposability, sustainability, affordability and others. These parameters are required to be considered at the early stages of design if desired operational behaviors are to be realized. A system design and development requires appropriate and timely application of engineering and management efforts to meet the previously mentioned parameters. A system may serve its intended purpose most effectively when its technical and operating characteristics are engineered into the design. Therefore, operational feasibility is a critical aspect of systems engineering that needs to be an integral part of the early design phases

II. SYSTEM SPECIFICATION

Hardware Requirements

- 1 GB Processor
- 3.5 inches or more screen size
- 1 GB RAM

Software Requirements

- Operating systems: Windows
- Front end: JAVA
- Back end: XML

IV. SYSTEM DESIGN

Use Case Diagram

A use case diagram is a graphic depiction of the interactions among the elements of a system. A use case corresponds to a sequence of transactions, in which each transaction is invoked from outside the website and engages internal objects to interact with one another and will the system's surroundings. Use case defines what happens in the system when the use case is performed. Use cases represent specific flows of events in the system. The use cases are initiated by actors and describe the flow of events that these actors set off. An actor is anything that interacts with a use case: it could be a human user, external hardware, or another system. An actor represents a category of user rather than physical user. A use case diagram is a graph of actors, a set of use cases enclosed by a system boundary, communication associations between the actors and the use cases, and generalization among the use cases.

A use case is a methodology used in system analysis to identify, clarify, and organize system requirements. In this context, the term "system" refers to something being developed or operated, such as a mail-order product sales and service website. Use case diagrams are employed in UML (Unified Modeling Language), a standard notation for the modeling of real-world objects and systems.

System objectives can include planning overall requirements, validating a hardware design, testing and debugging a software product under development, creating an online help reference, or performing a consumer-service-oriented task. For example, use cases in a product sales environment would include item ordering, catalog updating, payment processing, and customer relations. A use case diagram contains four components.

- The boundary, which defines the system of interest in relation to the world around it.
- The actors, usually individuals involved with the system defined according to their roles.
- The use cases, which the specific roles are played by the actors within and around the system.
- The relationships between and among the actors and the use cases diagram.

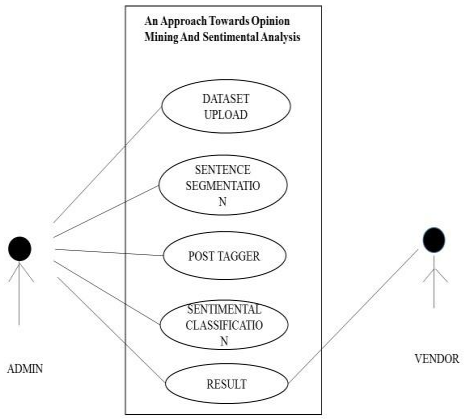


fig.usecase diagram

Data Flow Diagram

A data flow diagram (DFD) is a graphical representation of the "flow" of data through an information system, modelling its process aspects. A DFD is often used as a preliminary step to create an overview of the system, which can later be elaborated. DFDs can also be used for the visualization of data processing (structured design). Data flow diagrams are as well-known and widely used notation for specifying the functions of an information system and how data flow from functions. They describe systems as collections of functions that manipulate data. DFD describes the flow of data and the processes that change or transform data throughout a system. It is a structural analysis and design tool that can be used for flowcharting in place of, or in associations with, information oriented and process oriented system flowcharts. The DFD reviews the current physical system, prepares input and output specifications, specifies the implementation plan etc.

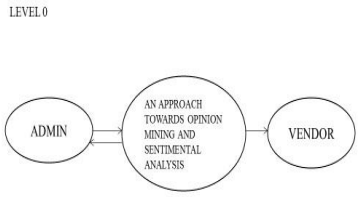
A DFD shows what kind of information will be input to and output from the system, where the data will come from and go to, and where the data will be stored. It does not show information about the timing of process or information about whether processes will operate in sequence or in parallel (which is shown on a flowchart). Data flow diagrams are also known as bubble charts.DFD is a designing tool used in the top down approach to Systems Design. This context-level DFD is next "exploded", to produce a Level 1 DFD that shows some of the detail of the system being modeled. The Level 1 DFD shows how the system is divided into sub-systems (processes), each of which deals with one or more of the data flows to or from an external agent, and which together provide all of the functionality of the system as a whole. It also identifies internal data stores that must be present in order for the system to do its job, and shows the flow of data between the various parts of the system.

Data flow diagrams are one of the three essential perspectives of the structured systems analysis and design

ISSN (ONLINE): 2250-0758, ISSN (PRINT): 2394-6962

method SSADM. The sponsor of a project and the end users will need to be briefed and consulted throughout all stages of a system's evolution. With a data flow diagram, users are able to visualize how the system will operate, what the system will accomplish, and how the system will be implemented. Data flow diagrams can be used to provide the end user with a physical idea of where the data they input ultimately has an effect upon the structure of the whole system from order to dispatch to report. How any system is developed can be determined through a data flow diagram model.

The old system's dataflow diagrams can be drawn up and compared with the new system's data flow diagrams to draw comparisons to implement a more efficient system. In the course of developing a set of levelled data flow diagrams the analyst/designer is forced to address how the system may be decomposed into component sub-systems, and to identify the transaction data in the data model.



Fi .2 Data Flow

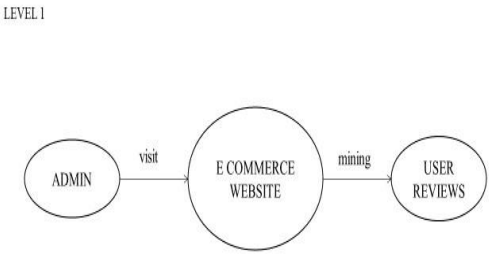
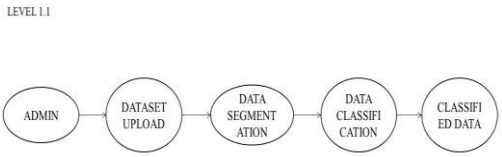


Fig2.Data Flow



Fi .2 DataDiag

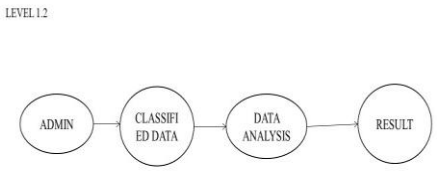
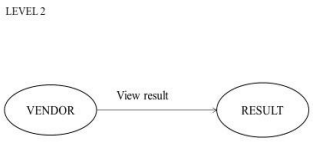


Fig2. Data Flow



Fg .2.Data Flow

Activity Diagram

Activity diagrams are graphical representations of workflows of stepwise Fig 8.2.5 Data Flow Diagram activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams are intended to model both computational and organizational processes (i.e. workflows). Activity diagrams show the overall flow of control. Activity diagrams are constructed from a limited number of shapes, connected with arrows. The most important shape types:

- rounded rectangles represent actions;

ISSN (ONLINE): 2250-0758, ISSN (PRINT): 2394-6962

- diamonds represent decisions;
- bars represent the start (split) or end (join) of concurrent activities;
- a black circle represents the start (initial state) of the workflow;

Arrows run from the start towards the end and represent the order in which activities happen. Activity diagrams may be regarded as a form of flowchart. Typical flowchart techniques lack constructs for expressing concurrency. However, the join and split symbols in activity diagrams only resolve this for simple cases; the meaning of the model is not clear when they are arbitrarily combined with decisions or loops. While in UML 1.x, activity diagrams were a specialized form of state diagrams, in UML 2.x, the activity diagrams were renormalized to be based on Petri net-like semantics, increasing the scope of situations that can be modeled using activity diagrams. These changes cause many UML 1.x activity diagrams to be interpreted differently in UML 2.x.

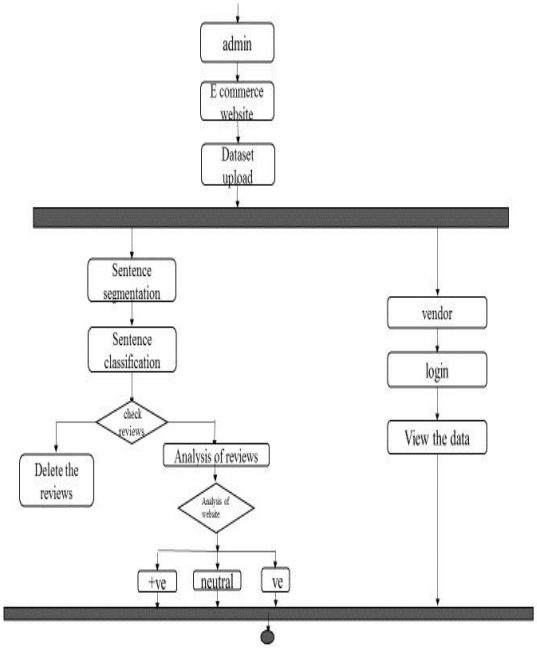


Fig Activity Diagram

A Sequence diagram is an interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. A sequence diagram shows object interactions arranged in time sequence. It depicts the objects and classes involved in the scenario and the sequence of messages exchanged between the objects needed to carry out the functionality of the scenario. Sequence diagrams are typically associated with use case realizations in the Logical View of the system under development. Sequence diagrams are sometimes called event diagrams or event scenarios. A sequence diagram shows, as parallel vertical lines (lifelines), different processes or objects that live simultaneously, and, as horizontal arrows, the messages exchanged between them, in the order in which they occur. This allows the specification of simple runtime scenarios in a graphical manner

Testing

Testing is the name given to the checking and analysis process that ensure that the software conforms to its specification. In this project few testing techniques where applied at different stages of the project so as to test the performance of the project. The testing activity can be carried at the implementation to verify that the project behaves as intended by the designers. During these activities the software reliability was checked to a large extent. The successful testing process must confirm that the system is free of defects and is ready to use.

Module Testing

It is a process of testing the system module, what are all the inputs given and what are all outputs produced and whether they are required. Here after completing each module a testing was performed.

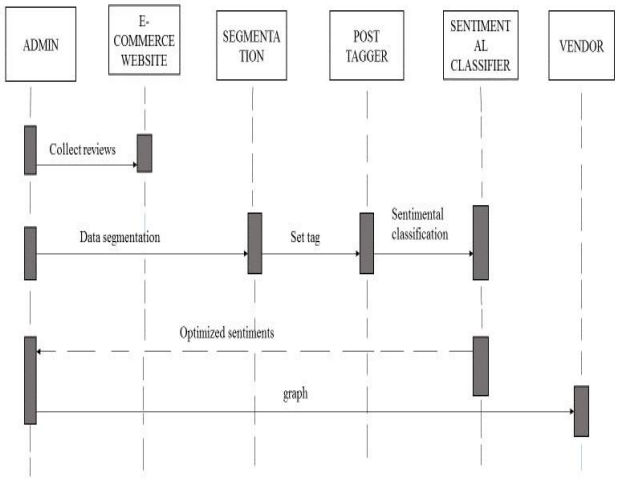


Fig8.4.1 Sequence Diagram

System Testing

It is a series of different tests whose primary purpose is to exercise the computer based system. It also tests to find discrepancies between system and its original objective, current specification. The requirements of our project where analyzed and a test activity was performed whether the system as the necessary software to run this project. If not, then the software is then installed in to the system and project will be implemented on it.

Integrated testing

This is a very impotent testing process in any project. This testing mainly focuses on the combination of

several parts of the project and making it to work. Here this testing activity was done at the end of the project and based on the result of this testing the output will be determined.

White box testing

It focuses on the program control structure .Test cases are derived to ensure that all statements in the program have been executed at least once during testing and that all logical conditions have been exercised. Basic path testing, a white box technique, makes use of program graphics to derive a set of linearly independent test that will ensure coverage. Condition and data flow testing further exercise program logic and loop testing complements other white box techniques by providing a procedure for exercising loops of varying degrees of complexity.

Black box testing

It was designed to validate functional requirements without regard to the interval working of the program. This technique focuses on the information domain of software deriving test case coverage. Equivalence partitioning divides input domain in to classes to data that are likely to exercise specific software function.

A classical system testing problem is finger pointing. This occurs when a defect is uncovered and one system element developer blames the other for the problem.

Test results

The listed tests were conducted at various development stages. Through put was conducted. The errors were debugged. The integration testing will be performed once the system is integrated. The results were and analyzed and alternations were made. The test result proved positive and henceforth the application is feasible and the test is approved.

V. CONCLUSION

Sentimental analysis and opinion mining is a field of study that analyzes people's sentiment, attitude's or emotions towards certain entities. This project on sentimental analysis and opinion mining tackles a fundamental problem of sentiment analysis, sentiment polarity categorization. Online product reviews from E-commerce website flip cart are selected as the input to this project. This will enable the system to analyze and then generate results (for reviews) which have been sorted according to various geographic regions. The system shall present a temporal analysis of reviews as well, to enable manufacturers to look at the sentiments as a function of time to judge the improvements or deteriorations of the same over time.

REFERENCES

[1] W. Bruce Croft, Donald Metzler, Trevor Strohman, Search Engines: Information Retrieval, 2009.

[2] Laura A Granka, Matthew Feusner, Lori Lorigo, Eye Monitoring in online Search, 2008.

[3] I.Qunis, G. Amati, V. Plachouras, B. He, C. Macdonald and C. Lioma, A High Performance and Scalable Information Retrieval Plateform, In SIGR Workshop on Open Source Information Retrieval, 2006

[4] T. Fagni, R. Perego, F. Silvestri, and S. Orlando., Boosting the performance of Web Searching Engines : Caching and pre-fetching Query Results by Exploring historical usage Data, ACM Trans. Inf. Syst., 24(1): 51-78, 2006

[5] X. Long and T. Suel, Three Level Caching for Efficient Query Processing in Large Web Search Engines, 2005

[6] R.A. Baeza Yates and F. Saint-Jean, A Three Level Search Engine Index Based in Query Log Distribution , In SPIRE 2003.

[7] R. Lempel and S. Mohan, Predictive Caching and Pre-fetching of Query Results in Search Engines, 2003.

[8] S. J. Caughey, D. B. Ingham and M. C. Little, flexible open caching for the web, dec. 1997.