Feedback Classification with respect to Opinion Mining, for Online Forums using Hierarchical Clustering Techniques

[1]Siddharth P. Shanbhogue, [2]Vishnu Makam Gupta,[3]Dr Sumitradevi K.A.
[1,2]Students, 6th semester, Department of MCA,
[3]Director, Department of MCA,
RV College of Engineering,
Mysore Road, Bangalore - 59
[1]siddharthx64@linuxcandy.com, [2] kuroberos@gmail.com ,
[3] director-mca@rvce.edu.in

***Abstract* - Feedback is an important piece of information to the users of a system, who are looking to use or switch to a new product. All the products have the product info released by the company or institution, but the feedback gives a report of the user's experiences with the product, after they have used them for a considerable amount of time. Online forums are places where people from around the world share their experience of the products that they have been using. As online forums are independent of geographical location, they help in gathering experiences from a large group of users coming from various fields all of whom happen to use the same set of products for various purposes [1]. Due to the advent of the Internet, hundreds of thousands of online forums exist today. Millions of users regularly put forth their opinions, comments, thoughts and suggestions regarding the products/social topics and other related themes. The collective opinions of multiple individuals tend to be more accurate than the opinions of individual persons [2].**

Keywords: Sentiment Analysis, Hierarchical clustering, single linked agglomerative clustering, opinion mining

I. INTRODUCTION

The software project aims to demonstrate efficient usage of agglomerative hierarchical clustering algorithms in determining the general intent of a particular review. Based on this intent, the level of enthusiasm with which the product has been accepted by the market can be determined.

This project can work for both products and productized services. However, the general constraints for the project would be that the reviews would have to be free of slang/errors and would have to be only written in clean English language.

The major dependency would be availability of honest feedback and reviews. Without accurate feedback, the actual intent of the review may not be determinable and the whole point of the project would come to nothing. It is assumed that the reviews fed to the software are free from bias, and are as accurate as humanly possible.

The project involves distance calculations in two ways: Euclidean Distance to find distance between two points, and then finding the distance between two points. The same are explained below:

A. Euclidean Distance

To find the Euclidean distance between two points, the following formula is used

B. Distance between two points

This is calculated by using the two-dimensional approach of the Euclidean distance

In the above formulae, (p, q), (p1, q1), (p2, q2) are coordinates of points.

II. STATE OF ART DEVELOPMENTS

C. Knowledge Management

In today's competitive business structure a stock of knowledge is a considered as a prime strategic resource of the company. These stocks help the strategic managers to take positive and accurate decisions. The data are huge and get updated very frequently (in some cases few minutes or even seconds) which help in the dynamic capabilities of routine changes in the products displayed. This knowledge help taking speedy decisions making the user to have the most possible outcome. Also faster developments in technology helps better predictions and in some situations higher rate of patent when compared to others.

D. Knowledge Mapping

This is a process of gathering and analyzing information to discover the ownership and location. These knowledge maps are used to identify the scientific firm or the technological university. Also helps in keeping track of the current technological trends, with the capability of forecasting future technological developments.

III. EXISTING SYSTEM

E. Opinion Mining of Customer Feedback data on web

Opinion mining or web text mining is an extremely useful tool to extract relevant, useful information from a sea of vast data. The paper provides a detailed understanding of the various steps involved in opinion mining, the various techniques available to us, and finally, the tasks involved in developing a software system for the same. Through this paper, more

information can be gained regarding the procedure needed to apply in the project.

F. A Survey on Product Evaluation using Opinion Mining

In recent times, rating systems are being used to rate a given product positively or negatively in online forums by analysing the comments and reviews of users. Such methods are being used as a means of gaining more information regarding the product[3]. But such approaches are not enough for a customer to make a worry free decision about buying a product. In some recent publications, several approaches have been proposed, wherein some methods of rating the product and also the product features are provided. Such methods might involve having more customized ways of putting up the reviews. The approach also compares the products of different brands based on their ratings.

With the help of the surveys done in this paper, a lot can be learnt about the methods and procedures employed in the existing systems for rating and reviewing products through online forums,

G. Mining Opinion Features in Customer Reviews

Thanks to the Internet, there is no dearth of online forums for customers, potential buyers and reviewers of a particular product to meet up and share their opinions. This paper focuses on summarising these hundreds of opinions that a user usually finds on online forums, regarding customer reviews. The paper focuses mainly on extraction of the features of the product mentioned in the comments and reviews that the users have put up on the page, and not on the whole process of opinion mining per se [5]. The paper's outcome signifies that the opinion mining techniques available today are highly effective, and can be utilised for improved analysis of online feedback.

H. Unresolved Issues and Emerging Opportunities or new directions

The mining of forums and online discussions is a challenge on its own. The reason is the use of colloquial language, slang, abbreviations, various styles of writing (as the users are from all over the world) and the various knowledge level of the language. These issues make them different when they are compared with the text of an online newspaper [6].

IV. Proposed System

The application makes use of Single Linkage Agglomerative Hierarchical Clustering to obtain a general opinion regarding the given product or service. Single Linkage Agglomerative Hierarchical Clustering allows for linkage between clusters based on the shortest distance between the nearest two points of the Clusters (one point from each cluster).

The link between two clusters is made by a single element pair, namely those two elements (one in each cluster) that are closest to each other. The shortest of these links that remains at any step causes the fusion of the two clusters whose elements are involved.

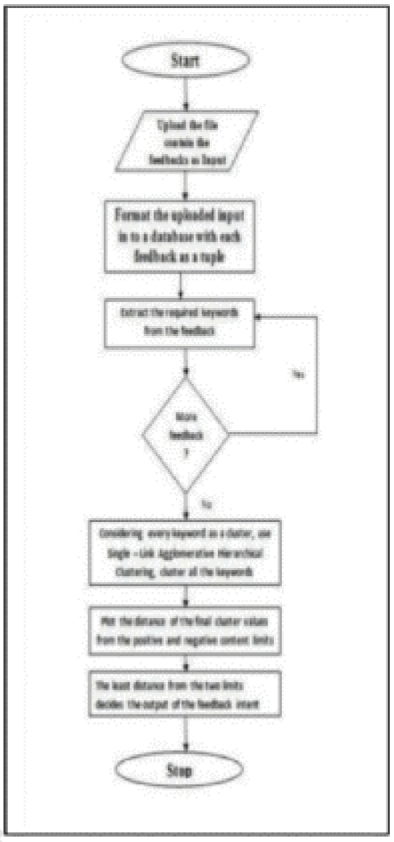

Fig 1: Flowchart depicting the process of the application

I. Principles Used

For the purpose of identifying the general intent of any customer feedback or review, the keywords used in the text must be determined. Along with the keywords, the qualifiers used around it must also be detected. For instance, a keyword "good" gives a completely new and opposite meaning when the qualifier "not" is found right before it. These kind of details need to be detected and analyzed.

Clustering Techniques have been immensely useful in determining internal relationships among various seemingly unconnected elements. When it comes to Opinion Mining or sentiment analysis, this is all the more effective, since there is a need to correlate several elements based upon their proximity with other elements. Hierarchical clustering works with either top-down or bottom-up approach. In the bottom-up approach, each keyword is treated as a cluster. Each pair of clusters is then successfully merged until all clusters are merged into a single cluster. The Agglomerative Hierarchical Clustering Algorithm was chosen for this application. The figure 2 shows a generic use-case diagram for the application.

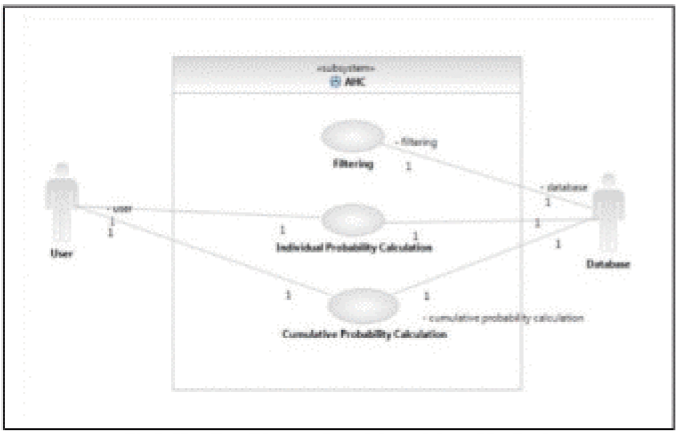

Fig 2: Use-Case Diagram for opinion mining of feedback

J. *Algorithms Used: Hierarchical Clustering Algorithm*

Connectivity based clustering, also known as hierarchical clustering, is based on the core idea of objects being more related to nearby objects than to objects farther away. As such, these algorithms connect "objects" to form "clusters" based on their distance. A cluster can be described largely by the maximum distance needed to connect parts of the cluster. At different distances, different clusters will form, which can be represented using a dendrogram, which explains where the common name "hierarchical clustering" comes from: these algorithms do not provide a single partitioning of the data set, but instead provide an extensive hierarchy of clusters that merge with each other at certain distances. In a dendrogram, the y-axis marks the distance at which the clusters merge, while the objects are placed along the x-axis such that the clusters don't mix.

The generic algorithm that is followed in Single Linkage Agglomerative Hierarchical Clustering is given below:

A. Start with a disjoint clustering having level $L(0) = 0$ and sequence number m = 0.

B. Find the most similar pair of clusters in the current clustering, say pair (r), (s), according to $d[(r),(s)] = \min d[(i),(j)]$ where the minimum is over all pairs of clusters in the current clustering.

C. Increment the sequence number: $m = m + 1$. Merge clusters (r) and (s) into a single cluster to form the next clustering m. Set the level of this clustering to $L(m) = d[(r),(s)]$

D. Update the proximity matrix, D, by deleting the rows and columns corresponding to clusters (r) and (s) and adding a row and column corresponding to the newly formed cluster. The proximity between the new cluster, denoted (r,s) and old cluster (k) is defined as $d[(k), (r,s)] = \min d[(k),(r)], d[(k),(s)]$.

and the results are maintained in a better way

RESULTS

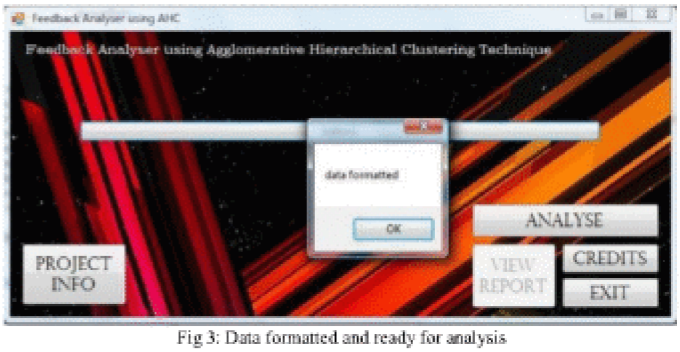

Fig 3: Data formatted and ready for analysis

Figure 3 shows the raw input uploaded and formatted into a required format by the application, so the data can be analyzed

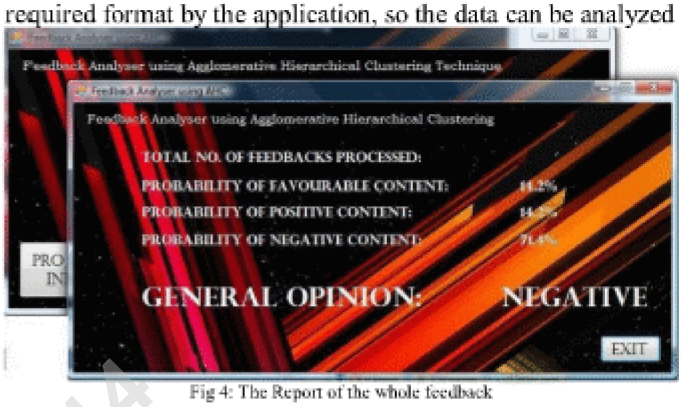

Fig 4: The Report of the whole feedback

Figure 4 displays the report of the whole feedback, providing a clear understanding of the overall general feedback that was uploaded and analyzed by the application. In this case, the general opinion about the product/service appears to be negative in nature.

V. CONCLUSION

A. The project successfully manages to determine the general intent of provided feedback, and allows the user to gain an accurate understanding of where the product stands in the global market.

B. By implementing this project, a more effective and more accurate alternative can be offered, when compared to the older, imprecise Star Rating system that is being heavily followed today.

C. This project can be successfully utilized in a variety of scenarios, including (but not limited to) product reviewing and market analysis, experience reviewing (such as holiday destinations, hotels, etc.), performance reviewed (work performance review of employees), tool efficiency reviews based on multiple separate user experience accounts, and so on.

As part of future modifications and enhancements, a module could be integrated that would allow the application to make note of a new word that describes emotion, both negative and positive in nature, that has not been used before

REFERENCES

[1] Inbeom Hwang, Minsuk Kahng, Sang-goo Lee, Exploiting User Feedback to Improve Quality of Search Results Clustering, 2011, Proceedings of the 5th International Conference on Ubiquitous Information Management and Communication (ICUIMC 2011)

[2] Dr.N.Rajalingam, Dr. K.Ranjini, "Hierarchical Clustering Algorithm - A Comparative Study", International Journal of Computer Applications (0975 – 8887) Volume 19– No.3, April 2011.

[3] Bing Liu, Lei Zhang. "A Survey of Opinion Mining and Sentiment Analysis" book chapter in Mining Text Data, Kluwer Academic Publishers, 2012

[4] Ahmad Barirani, Bruno Agard, "Competence Maps using agglomerative hierarchical clustering", Journal of Intelligent ManufacturingVolume 24, Issue 2 , pp 373-384, 2013-04-01

[5] Minqing Hu and Bing Liu, "Mining Opinion Features in Customer Reviews", American Association for Artificial Intelligence, Proceedings of Nineteenth National Conference on Artificial Intelligence (AAAI).

[6] Anna Stavrianou — Jean-Hugues Chauchat,"Opinion Mining Issues and Agreement Identification in Forum Texts", Atelier FODOP 2008, pages 51 to 58