Get Set Shop

Aniket Phutane[1], Rakshanda Panzade[2], Sohan Karkera[3], Trishna Patil[4], Vanita Mane[5]
[1,2,3,4,5]Department of Computer Engineering, RamraoAdik Institute Of Technology, Nerul, Navi Mumbai, India
[1]aniketdphutane@gmail.com, [2]rakshandap24@gmail.com, [3]sohank1704@gmail.com,
[4]trishna.patil13@gmail.com, [5]vanitamane1@gmail.com

Abstract: A dynamic real time smart location based mobile shopping application for Android cell phones. The location information of the client will be retrieved using the Geo-position of their device. The stream of the application is that client gives his interests or inclination of items, and afterward Get Set Shop distinguishes the area and inquiries the item in the nearest neighbourhood dealers. The thought is to get the costs from every nearby store within stock data and smartly list the products. The mobile shopping application will provide precise and minimum error based searching and listing results with the help of proposed smart filtering algorithm.

1. INTRODUCTION

As there is increase in number of android smartphone users, people are more willing to use these smartphone devices over computers for searching, purchasing and many other purposes. People prefer to choose closer shopping malls like D-Mart, Big Bazar, etc. for buying the cheapest product for daily needs. Therefore, our attempt is to provide a platform for the local market to increase their business and profits using a real time geo location based dynamic android application. Sellers can sell their products with discounts using this application. Customers can find their products of interests anywhere and anytime on one click. The main reason behind developing this system is that nowadays, due to social media the discount rates provided by the superior malls and shops are already known to the people and so the lesser known local market shops are not able to publicize their product discount offers. In this project, we provide a platform for all the local mart shopkeepers to publicize their product discount rates and reap benefits from it.

2. RELATED WORK

To implement our real time dynamic application, the literature survey includes study of various research papers. Following research papers were studied based on the Global Positioning System (GPS), Clustering and Distance methods.

2.1 Global Positioning System (GPS) based methods

According to the authors, Amit Kushwaha and VineetKushwaha[1], each location based application needs to provide precise location information dynamically in a real time scenario with a personalized touch. In this era, the usage of tablets and smartphones is so high that the desktops and laptops will be replaced by them even for computational purposes. Such needs can only be catered with the help of Location Based Service which have a wide variety of applications which include security related jobs, general survey regarding traffic patterns, decision based on vehicular information for validity of registration and license numbers. Another application is surveillance where real time information is needed to identify whether a person is an actual threat or a mistaken target. A number of different software applications have been created which provide the user with information regarding a place he or she wants to visit, but are limited to desktops only and do have a need to be imported on mobile devices. All the information must be available in mobile device and in user customized format which will ensure that when visiting places, travel guides are not needed. According to Mahesh Kadibagil and Dr. H S Guruprasad [2], the Autonomous position detection and tracking system enhances the accuracy of locating the position of friends and family members by using GPS and standard web technology. As mentioned, this system includes a mobile client, a repository, a web client and a map service. The mobile client is used to find the location and send a Popup SMS to user once his/her friends or family members are in the vicinity of the user's area of direction. This location information can be stored on a server from where it can be viewed and managed by other users by using the web client.

2.2 Clustering based methods

In WISE-Cluster, Peng, Q., Meng, W., He, H. and Yu, C. [3] proposed a new approach to automatically clustering e-commerce search engines (ESEs) on the Web such that ESEs in the same cluster sell similar products. In this approach the ESE's were used as the base to build an ecommerce metasearch engine for each cluster and in accordance to theavailable features on the interface page of each ESE, clustering was performed.

The number of images and links, terms appearing in the search form and normalized price terms were the particular features on which they were operated and their experimental results, which were based on 300 ESE's, pointed towards achieving positive results. In a study by Raj bala, Sunil Sikka and Juhi Singh [4], comparison has been performed on the K-means, Hierarchical, EM and Density based clustering algorithms. Bank dataset was used for performing comparison. The relative results were represented in the form of tables and graphs. The process of gathering a set of data objects that are alike and based on similar attributes, within the same group is called clustering. There are many clustering algorithms. With the help of this paper we understood the fair analysis of four clustering algorithms namely Hierarchical algorithm, Density based algorithm K-means algorithm and Expectation & Maximization algorithm where WEKA tool was used to compare these algorithms by using efficiency and accuracy as the parameters. Both normalized and non-normalized data was

used for clustering and among all the algorithms K-means produced better results in terms of efficiency and accuracy.

3. PROPOSED METHOD

This system will be developed using Intelligent Filtering Algorithm which includes Greedy Search algorithm as a searching method, Agglomerative **Clustering** Algorithm along with Normalization. It also consists of Levenshtein Distance Algorithm and the algorithm for sorting of products. The route for the selected shop will be represented using Dijkstras Algorithm which will give the shortest possible path to the customer.

Intelligent Filtering Algorithm: 1) Remove unrelated products, 2) Calculate the representative price of the product, 3) Normalize the output of the step 2 (x), 4) Measure similarity score between search keyword and the product title (y), 5) Use modified agglomerative clustering algorithm, 6) Compute

$$a * x + b * \big((90 - y)90\big) + clustering\ score \quad \text{... (1)}$$

$$where, \quad a > b$$

3.1 Greedy Search Algorithm

Greedy Search Algorithm will be used for removing unrelated information from the data i.e. pre-processing of data (shown in the Figure 1) for which it will be modified according to our mobile shopping Android applications requirements. However the same algorithm and parameters in Greedy Search Algorithm will be used which will help in filtering the products.

3.2 Agglomerative Clustering Algorithm

Compared to Greedy Search Algorithm, the Agglomerative Clustering Algorithm includes some calculation and filtering. This increases its overall complexity. It will also be used for clustering the results which are made of product search. These results will enable the mobile user to see prices at the nearest electronic local stores in Get-Set-Shop mobile shopping Android application i.e. Intelligent Filtering of data (shown in Figure 1) for which the formula is shown below

$$P^I = \begin{cases} 100, \ if\ P \le 100 \\ \left\lfloor \frac{P}{m} - 0.5 \right\rfloor * m, \ m = 100^{\lfloor logP \rfloor} \ \ if\ P > 100 \end{cases} \quad \text{... (2)}$$

3.3 Normalization

The second operation is the normalization process which enhances the Clustering algorithm (shown in Figure 1). The output of the first operation which is the Agglomerative Clustering Algorithm will be the input of the normalization process. After the normalization process is performed, the first parameter (say x) will be generated. The price is scaled between 0 and 1 because it is used to present the products in a two dimension coordinate system and also the total calculation time is low. The formula for normalization is given as follows

$$Normalization(X_{price}) = \frac{X_{price}}{MaxPrice} \quad \text{... (3)}$$

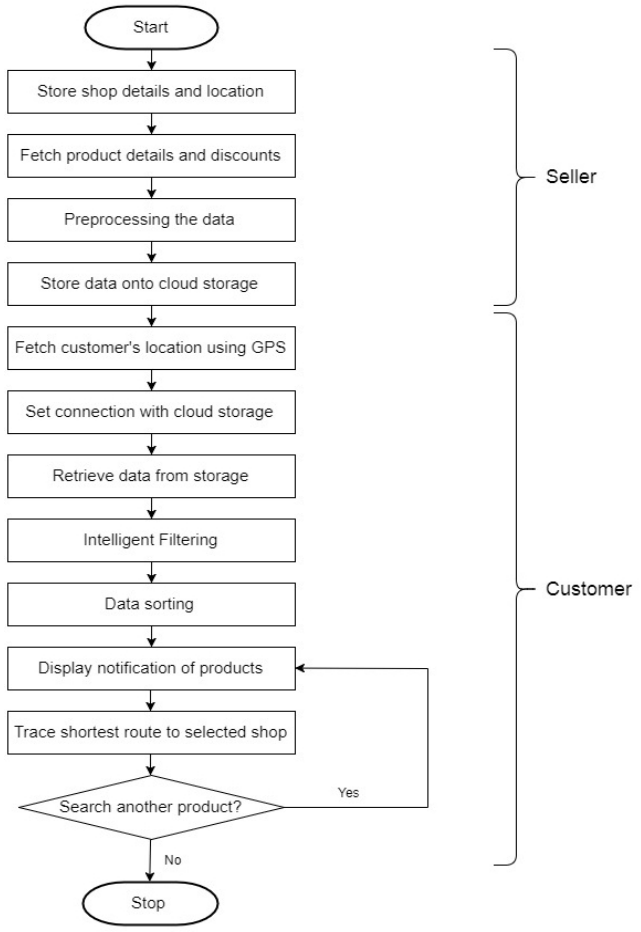

Fig. 1. Shows the Basic flow of the system

3.4Levenshtein distance algorithm

In this system, the Levenshtein distance algorithm, which is a string similarity function, will be used that It relies upon the number of changes required to transform single word into the other [5]. The point of this calculation is to gauge difference between two groupings or strings. For example, if aand b are two strings then the distance between them is given by,

$$lev_{a,b}(i,j) = \begin{cases} \max(i,j), \ if\ \min(i,j) = 0, \\ min \begin{cases} lev_{a,b}(i-1,j) + 1 \\ lev_{a,b}(i,j-1) + 1 \\ lev_{a,b}(i-1,j-1) + [a_i \neq b_i] \end{cases} \ otherwise. \end{cases} \quad \text{... (4)}$$

3.5 Scoring products

The products are sorted using merge sort **according** to the preference of customers with respect to the price in ascending or descending order (shown in Figure 1). Conceptually, a merge sort works as follows: 1) Divide the unsorted list into n sub lists, each containing 1 element (a list of 1 element is considered sorted). 2) Repeatedly merge sub lists to produce new sorted sub

lists until there is only 1 sub list remaining. This will be the sorted list.

3.6 Dijkstra's Algorithm

Dijkstra's algorithm is an algorithm for finding the shortest paths between source (customers location) and target (shops location) in the Geo Map which represents road networks. The initial node is the node at the start. It can be used in Google maps to find the shortest possible path from customers current location to target shops location (shown in Figure 1). We can consider any geographical map as a graph, locations in the map as vertices in algorithm and road between these locations as edges. Then the weights of edges here will be the distance between those two locations.

4. EVALUATION PARAMETERS

The efficiency parameters that will be used to tally the results are as follows: True Positive (TP) - These are the correctly predicted positive values which means that the Product name correctly identified as Product name. True Negative (TN)

- When the Product is correctly identified as Product. False Positive (FP) - When the Product is incorrectly identified as Product name. False Negative (FN) - When the Product name is incorrectly identified as Product. Once these four parameters have been defined, we can calculate Accuracy, Precision and Recall.

Accuracy

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad \ldots (5)$$

Precision

$$Precision = \frac{TP}{TP+FP} \quad \ldots (6)$$

Recall

$$Recall = \frac{TP}{TP+FN} \quad \ldots (7)$$

5. CONCLUSION

This paper gives description of the android application for mobile users for smart shopping that uses location information to search and list the product as per the user's desire.The main idea is to use location information to get the prices from each local store around the user's local vicinity and smartly list the products. The efficiency of the Get-Set-Shop is optimized by limiting the product data gathering time. Collecting the data and extracting the information from internet resources takes a long time for which the Java HTML Parser called Jsoup is used for collecting data from internet resources in the Get-Set-Shop mobile shopping Android application. The required data is collected by using this parser as it uses HTTP protocol to connect to the internet resources. Every connection to the internet resources has a default 3 second timeout period. As in future scope, Get-Set-Shop can be expanded easily to add other sectors such as local shops for gifts, books, hardware, clothing, etc.

ACKNOWLEDGMENT

We would like to thank our project supervisor Mrs. Vanita Mane, coordinator Mrs. Aditi Chhabria, PrincipalDr. Ramesh Vasappanavara and the H.O.D. of Computer department, Dr. Leena Ragha and all those who have helped us in visualizing this project.

7. REFERENCES

[1] Amit Kushwaha and VineetKushwaha, International Journal of Advances in Engineering Technology (2011), Location Based Services using Android Mobile Operating System, Vol. 1,Issue 1,pp.14-20

[2] Mahesh Kadibagil and Dr. H S Guruprasad, International Journal of Computer Science and Information Technology Security (2014), Vol. 4, No. 3

[3] Peng, Q., Meng, W., He, H. and Yu, C. (2004) WISE-Cluster: Clustering ECommerce Search Engines Automatically. WIDM Workshop. [4]Raj bala, Sunil Sikka and Juhi Singh, International Journal of Computer Applications (2014), A Comparative Analysis of Clustering Algorithms, Volume 100 No.15

[5] Levenshtein Distance. http://en.wikipedia.org/wiki/Levenshteindistance

[6] PayPal (2013) Mobilising Sales Making Money in the Mobile Commerce Revolution.

[7] Marker, F. and Chan, Y.H. (2009) A Survey on Android vs. Linux.

[8] Bell, M. (2008) Introduction to Service-Oriented Modeling. In: ServiceOrientedModeling: Service Analysis, Design and Architecture, Wiley Sons, Hoboken.

[9] Jonathan Hedley (2009-2013) Jsoup Java HTML Parser, with Best of DOM, CSS, and Jquery. http://jsoup.org/

[10] Hastie, T., Tibshirani, R. and Friedman, J. (2009) The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, Berlin. http://dx.doi.org/10.1007/978-0-387-84858-7

[11] Huang, A. (April 2008) Similarity Measures for Text Document Clustering. In: New Zealand Computer Science Research Student Conference (NZCSRSC), Christchurch.

[12] Hamming Distance. http://en.wikipedia.org/wiki/Hammingdistance

[13] H Pew Research Center (2011) How People Learn About Their Local Community.