International Journal of Advanced Technology in Engineering and Science www.ijates.com
Volume No.02, Special Issue No. 01, September 2014 ISSN (online): 2348 – 7550

AN ENHANCED PRODUCT RECOMMENDATION SYSTEM BASED ON USER RATINGS AND OPINIONS

Madhu K P[1], D Manjula[2]

[1]*Research Scholar, Department of Computer Science and Engineering,*
Anna University, Chennai, (India)
[2]*Professor, Department of Computer Science and Engineering,*
Anna University, Chennai, (India)

ABSTRACT

In the latest trend of Internet, optimized suggestions for the search are anticipated by every individual. Crowd sourcing, a large human resource provider network, is useful in search suggestions that helps to find out common phrases that other people have searched for. We identify an interesting real time problem finding the best products through suggestions with individual user rating to a particular brand and the generalized crowd sourcing opinion. The system comprise of product to product co-relation under a brand, user to user co-relation under common attributes, crowd sourcing opinion by means of key factors obtained for the product in order make the suggestions more optimal. The system is implemented for recommendation of cars. Adaptive weighting method helps in identifying the dominating attributes for recommendation.

Keywords: Adaptive Weighting, K-Nearest Neighbor, Opinion Mining, Rating-based Recommendation, User Co-relation

1. INTRODUCTION

With the widespread use of social networks and the high availability of huge amount of media contents, there is a radical growth in the domain of online sales and marketing. This unprecedented progress and innovation provides consumers with a wide variety of choices. Consumer items such as cars, books, cameras, and movies come in various subjects, features, and genres. Online shopping provides access to these items to anyone with an internet connection. Consequently, sellers anywhere can reach consumers anywhere, and consumers have access to an increasing number of products. In making adoption decisions, users rely on one another to organize the complex information on the web. This is evident from the abundant amount of user-generated content, such as tags, ratings, and reviews, all of which collectively aim to allow items to be more easily discovered by other users. Social networks have also become a conduit for discovering relevant information. The success of virtual social communities on the web and the increasing resistance and avoidance of customers towards traditional forms of advertising have led marketers to turn to new forms of marketing such as social media marketing and viral marketing. Users choices are increasingly driven not only by personal preferences, but also by the preferences of others in their social networks. This gives rise to the phenomenon of social correlation; where by users who are socially related tend to make similar choices. At the initial stage the user can choose the multiple items from the list or menu, some users can choose their reliable familiar products and the user conveys the information to everyone for advertising the product. The system is meant to provide recommendation of cars to the potential customers.

2. RELATED WORK

In the current scenario, the concepts of mining is predominant to match the preference and tastes of one user with another in order to make suggestions for the search. Generalised rating of the product has been gathered in the existing system to suggest any user according to the mining strategies. There is no specific consideration of rating in order to suggest the user's search that could serve as an optimal suggestion for the search. Also, search histories of similar users who are in no way acquainted with the user who is currently searching is not factored in, thus denying the opportunity for an individual from being influenced by an unknown, yet very similar user.

Freddy Chong et al [1], proposed generative models for item adoptions using social correlation. Users today are faced with an ever increasing array of choices of products on the web. The direct effect is that consumers have a harder time making purchase decisions while sellers don't know what to sell and who to sell it to. To assist this information overload, retailers attempt to assist consumers by putting aids such as bestseller lists. These lists make the recommendations skewed towards popular items. Amazon and Netflix have put in place a personalized recommendation system based on users past interests. However, this does not help users with little search history. This gives rise to the phe-nomenon of social correlation where socially related users make similarchoices. This work analyses how social correlation affects item adoption.

Based on the proceedings from the conference [2] by Kempe, Kleinberg and Tardos, it has been noted that people tend to have at-tributes similar to those of their friends. There are two underlying rea-sons for this. First, the process of social influence leads people to adopt behaviors exhibited by those they interact with; this effect is at work in many settings where new ideas diffuse by word-of-mouth or imitation through a network of people. A second, distinct reason is that people tend to form relationships with others who are already similar to them. This phenomenon, is often termed as selection. Online communities, however, provide an excellent opportunity to study large-scale social phenomena.

Badrul Sarwar et al [3], states that the Item-based techniques first analyze the user-item matrix to identify relationships between different items and then use these relationships to indirectly compute recommen-dations for user. They analyze the different item-based recommendation generation algorithms. They have looked into different techniques for computing item-item similarities and different techniques for obtaining recommendations from them. Finally, they have experimentally evaluated their results and compared them to the basic K-Nearest Neighbor approach. This work also suggests that the item-based algorithms provide dramatically better performance than user-based algorithms, while at the same time providing better quality than the best available user based algorithms.

3. RECOMMENDATION SYSTEM ARCHITECTURE

System Analysis is a combined process dissecting the system responsibilities that are based on the problem domain characteristics and user requirements. An extensive survey is performed among the potential customers with the information pertinent to user's personal data as well as the domain related information. Based on the information gathered, users as well as products can be grouped into different communities. This information can be moved to the database for future use. Collaborative Filtering is the basic technique used in the recommender system. But it is enhanced with an adaptive weighting method to give more preference to the most preferred attributes of the items which are mostly included in the ratings.

3.1 System Components

The various component modules involved in the recommendation system are discussed below. The main components are the clustering modules, the adaptive weighting and grouping and the re recommender module. The clustering modules are used to group the products and the groups for effective recommendation. The adaptive weighting module is used to give more preference to the items or the products which are rated and preferred by similar users. The recommender module gives the ultimate recommendations to the users.

3.1.1 Item Grouping (Product-Product Clustering) Module

Products or the items can be grouped based on the attributes like category of the product, cost, comfort level, color, geographic region where it can be best used and the sales history of the products. This will help in effective identification of the candidate products for recommendation. Closest Centroidal clustering is used in identifying the product versus product. The items will be marked with the same group ID and during processing they will be joined into one file.

3.1.2 User Grouping (User-User Clustering) Module

Users are grouped into varying communities based on the information obtained from their profile. The profile of the users are crawled and the preprocessing is done to extract information like age, location, profession, interests, the type of friends, and the purchase and the browsing histories of the user. Closest Centroidal clustering is used for the clustering. The recommendation domain can be narrowed down drastically by means of this grouping. Multi-track transcend approach has been considered in each individual profile of the user so as to enrich the join mode at verifiable grouping of the friend's suggestion about the product.

3.1.3 Product-User Clustering Module

Product-User grouping can be done to identify the products which are being purchased by the users. The potential customers can be identified effectively by this cluster. This can be further used in the product and user grouping. This is a vital clustering and can be represented by a bipartite graph. Closest Centroidal clustering is used in identifying the product versus user grouping based on the ranking of the products by the user and the user attributes based on the survey conducted. Apart from the standard local matching using sparse density verification, competing weight functions are embedded into the adaptive weighting approach, which uses slanted item grouping.

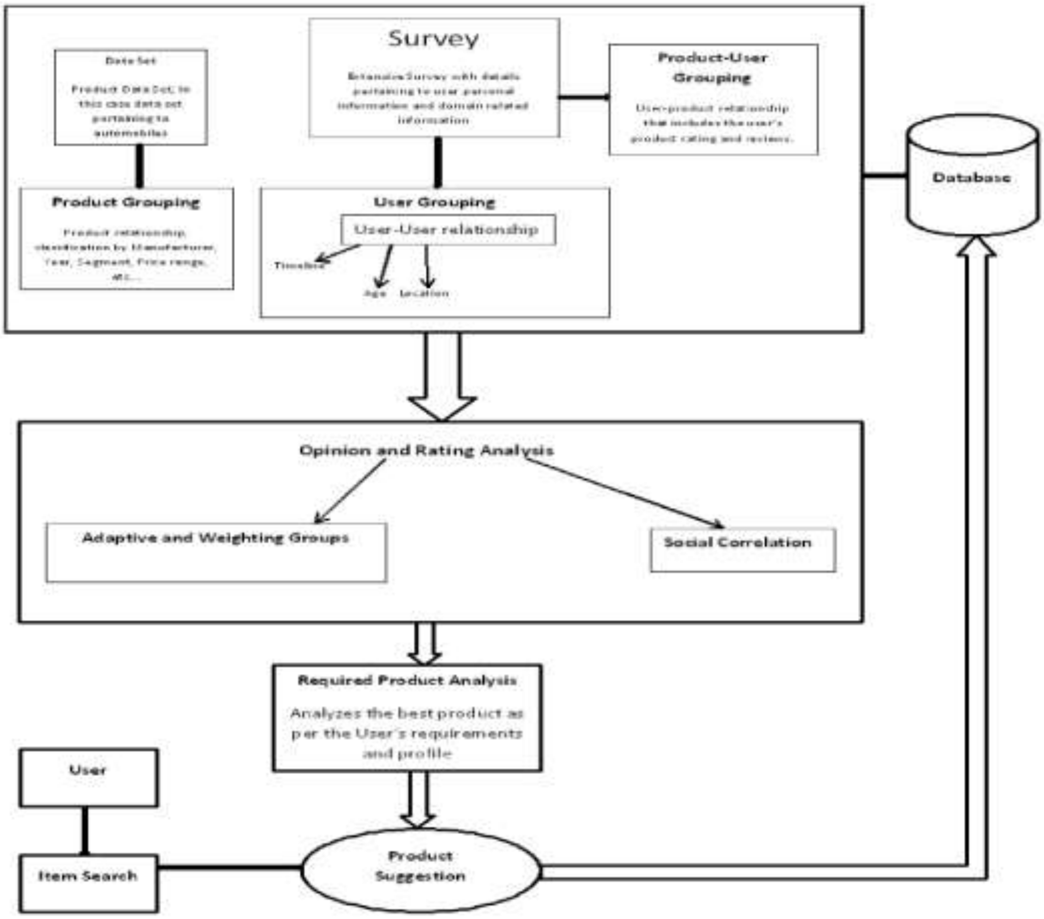


Fig. 1: Recommender System

3.1.4 Adaptive Weighting and Grouping (Dominating) Product Recognizer

Nearest neighbor analysis that is used to compare the product quality and other similar attributes and validating the quality of the product based on the attributes collected from the datasets is represented in Fig. 2, sparse density verification is used to obtain reliable findings of product with high efficiency, these different weight functions are tested on a large set of grouping technique of the disparity pairs. This also uses slanted item grouping. Final outcome of this module is a clear segregated layer between the dominated items and suppressed items. The dominating attributes would contribute more towards the rating based recommendation than the other attributes. So we may consider these relevant or dominating attributes first without losing the quality of the recommendation. This component is one of the highlights of the recommendation system. If we consider all the attributes of the items, the system may have to handle a large number of data as each product is supposed to have a number of attributes and it will vary from product to product. The relevance of the attributes also varies from product to product. To avoid the overhead of including more number of attributes is overcome by splitting the attributes into dominant and suppressed attributes and then considering the dominant attributes for the recommendation. The suppressed attributed or the less relevant attributes will be considered for further refining the recommendation, if needed, at any instance of time.

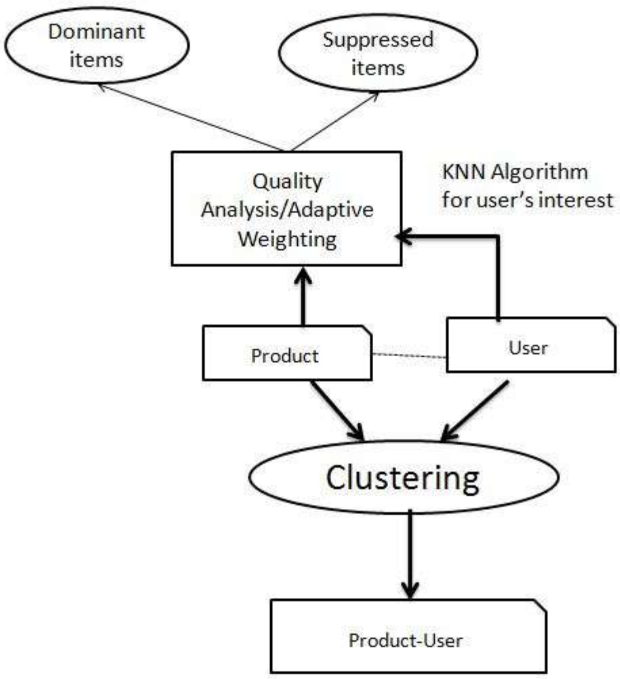


Fig. 2: Adaptive Weighting and Grouping

3.1.5 Search System User Specific Recommender

Recommender systems is a suggestion generation engine comprising of an information filtering technique that predicts 'rating' or 'preference' that the user would give to an item [4]. The basic idea of Collaborative Filtering is used in the system. But it is enhanced with the adaptive weighting and grouping method. This method splits the attributes into dominant and less relevant attributes. Now, the recommender system will find out the recommended products from the candidate products based on the dominant attributes. The suppressed attributes may be considered at any point of time, if needed. This may reduce the computation overhead required. The rating of the friends and the overall crowd sourcing opinion (rating) is taken into consideration which gives the recommended search results for the user. Recommendation techniques for rating prediction is done by collaborative filtering where systems recommend items based on the ratings given by the similar users with similar preferences. K - Nearest Neighbor algorithm is used to find user with similar preferences. Collaborative Filtering based technique primarily depends on the user ratings of the items. We also consider the opinions or the comments of the user's friends for the recommendation. The opinions may influence the recommendation. So we have taken that also into account while doing the recommendation. This has added to the accuracy of the system also. Accuracy and diversity are enhanced by recommending fewer items in the final result. This is represented in the Fig. 3.

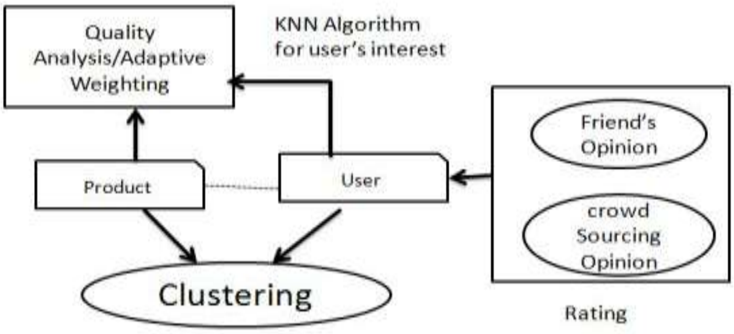


Fig. 3: Search System User Specific Recommender

3.1.6 User Co-relation Recommendation Search

KNN Algorithm is used for identifying the users' interests and the search criteria, based on his preferences and similarity with other users. The concept of correlation between the users and their opinions are being considered for making the recommendation. The rating of the user's friends about the particular product and the similar products are also being considered for the suggestion to the user in this recommendation search. It is illustrated in the Fig. 4.

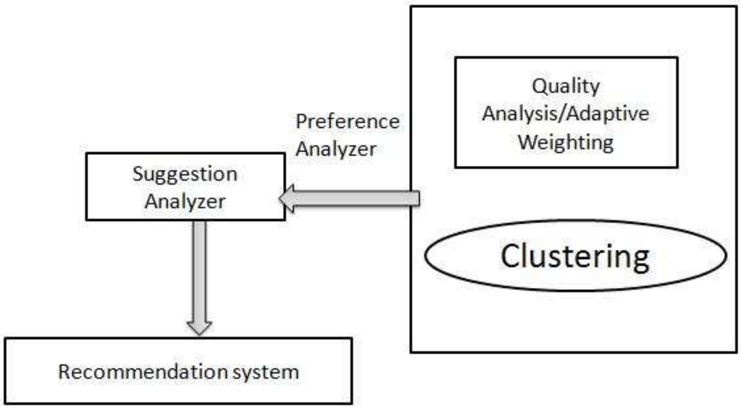


Fig. 4: User Co-relation Recommendation Search

4. RECOMMENDATION SYSTEM IMPLEMENTATION

The system is implemented for the recommendation of the cars. The users are asked to rate the items they purchased. Now the users are grouped based on their interests and preferences. Likewise, the products are also grouped based on the attributes of the products like the category of the product, cost, color, comfort level, geographic regioin where the car can be best used and the sales history of the product. There is another grouping bet ween the products and the users. This is represented by a bipartite graph. This graph has the users and the products as the nodes and the edges represents the rating the user offers for a car. If the user purchases a car it is regarded as a rating. These are further used for the future predictions of the cars for which the user has not done any rating. This prediction makes the recommendation to be more accurate.

4.1. User Level Mapping

This component module represents the way in which the users are grouped based on their similarities of attributes. Clustering is being done here in such a way that the users are mapped to each other based on the similarities of their location, gender, license holder, type of driving, hobbies, brand and color. The users are grouped into clusters based on similarities and the common attributes as represented in Fig. 5. This user level mapping is done to narrow down the search domain to be considered for recommendation. Also, it is useful in forming the community of users based on their interests or preferences. This will lead to more effective and efficient recommendation.

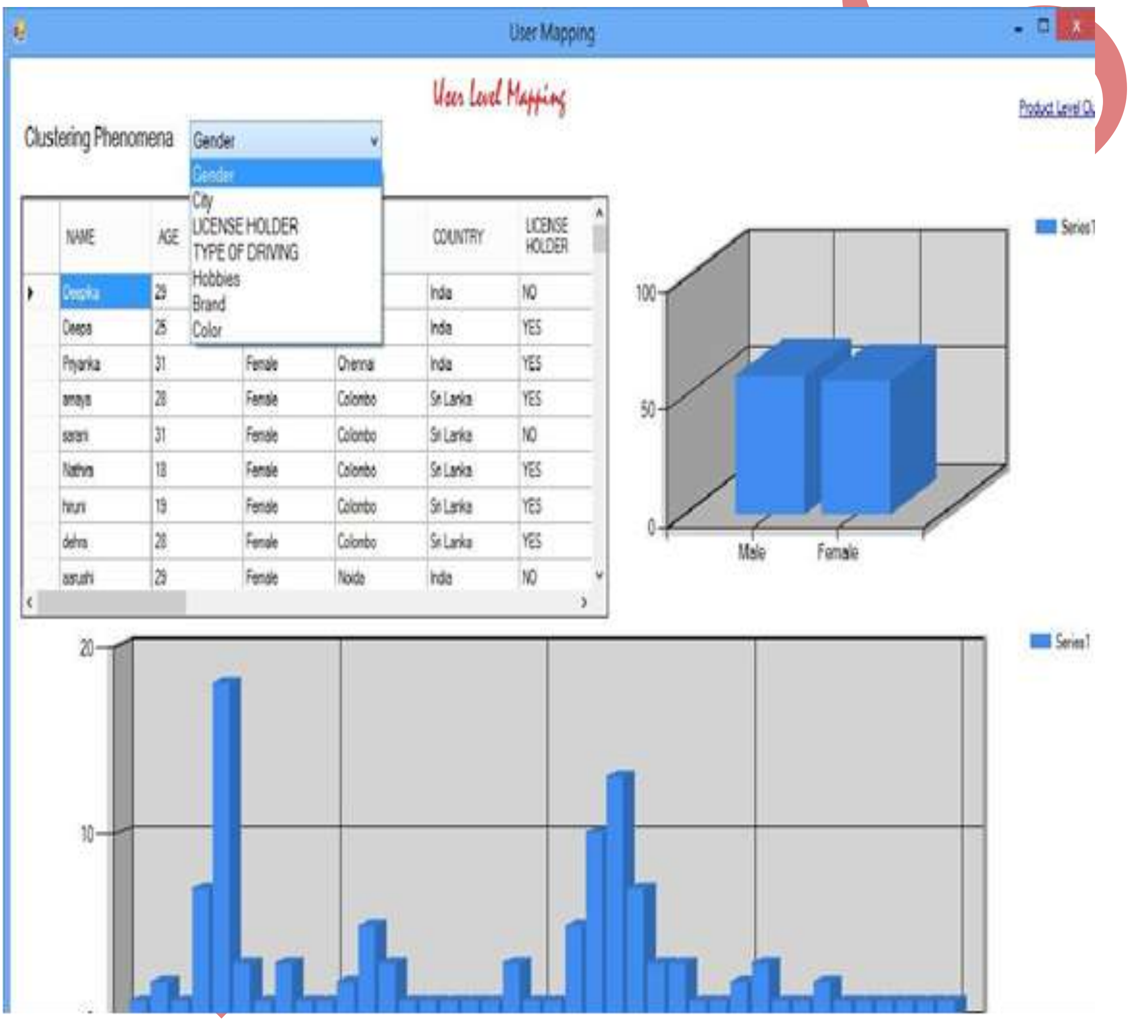


Fig. 5: User level mapping

4.2. Product Level Mapping

This module represents the way in which the product data set has been clustered. First based on the closest centroid classifier, the decider and the non-decider columns are obtained and then, based on the decider columns the products are grouped into the respective clusters. This is again similar to the case with dominant and suppressed attributes. And also based on the peak RPM, price and horsepower of the cars in each make they have been clustered into dominant and suppressed products which are represented as cluster 1 and cluster 2 in Fig. 6.

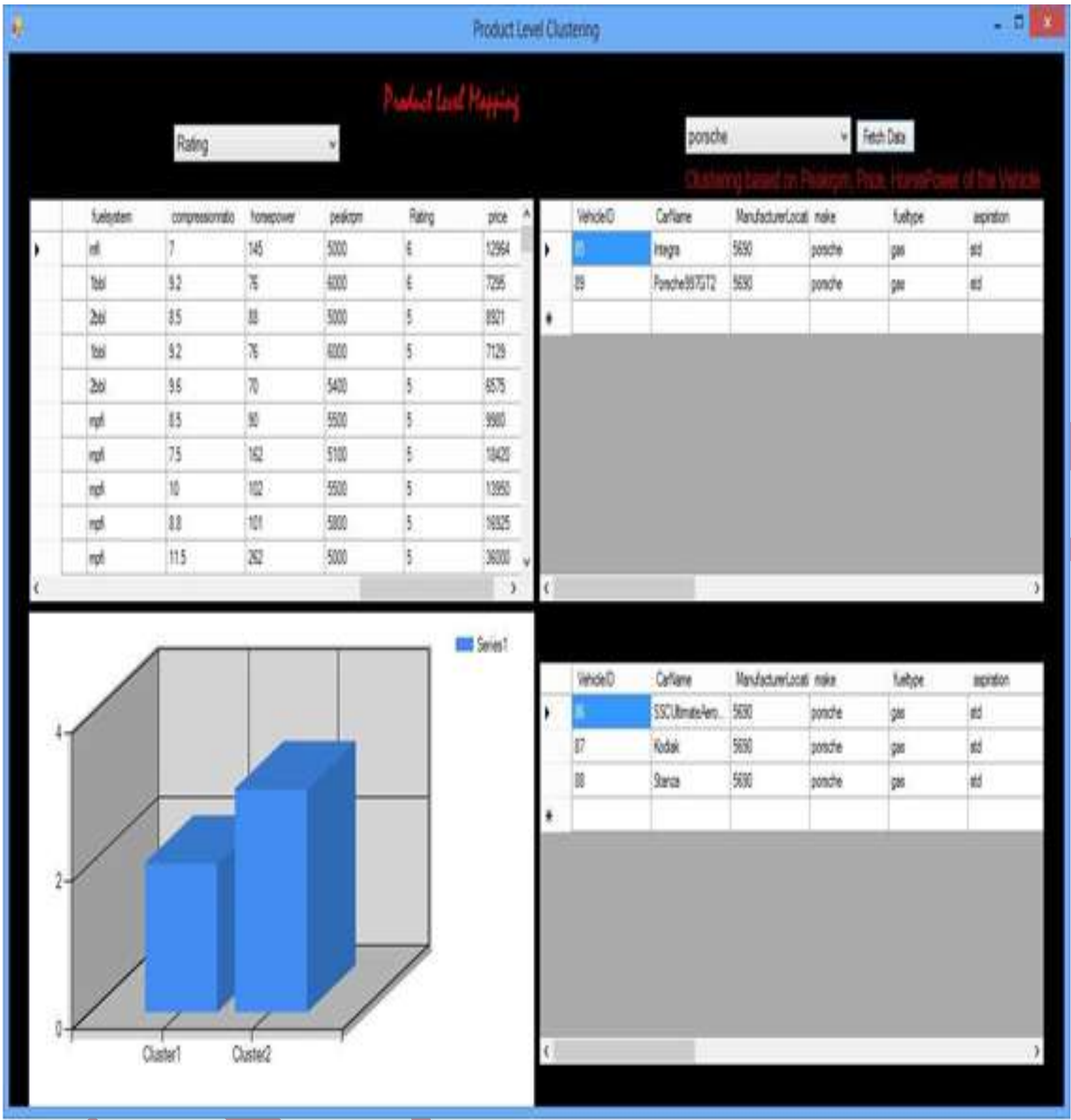


Fig. 6: Product level mapping

4.3. User-Product Level Clustering

In this module the users have been related to the products. It also analyses how each user has rated the product and how clusters are formed based on his ratings. After identification of the users interests in the products that have been given to him to rate association rule is used to map the clusters in such a way that the rating of each car that the user has given can be associated to the other and we can come up with an efficient cluster to which the user may belong to. Fig. 7 represents how first level, second level and third level user product clustering is done.

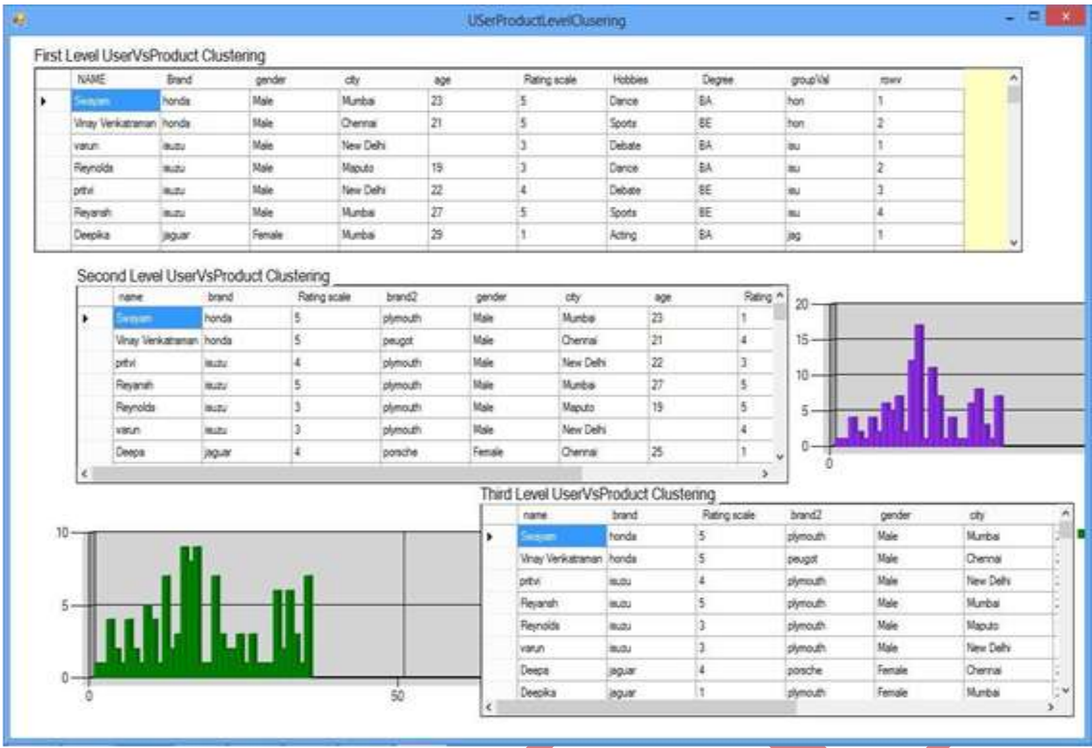


Fig. 7: User-Product level clustering

4.4. User-User Clustering

In this module, the users are grouped based on their similarities. Users who belong to the same cluster can be selected and chosen. For example, we may want people who belong to the same age and location and thus can be retrieved. The kind of a combination that is needed in a group the users in and get the similar users to get a cluster in such a way that the users can find the users who are similar to them can be decided. The admin here choses the necessary cluster he wants to provide for the final output so that the users can choose the similar users from that cluster. Either a single attribute or multiple attributes can be selected as clustering criteria.In the above example Age, Gender and City have been selected as clustering attributes, as represented by Fig. 8.

4.5. Search Data-User Preference

As illustrated in Fig. 9, the user can make a search with a keyword of his choice. His search is being recorded every time he gives a new keyword and he is shown a set of products relative to the keyword he has entered for the search criteria. The user can also search based on his preference which will give him a result based on how he has rated the products in the past. Then there are two filter criteria, the fuel and the the rating from which the user can choose the desired field to refine his search. The user can further choose the option Enable more recommendation that will allow him to choose if he wants to choose from the first cluster or the second cluster or both. Here the first cluster refers to the dominant items and the second cluster refers to the suppressed items or he can have a varied list from both the given clusters.

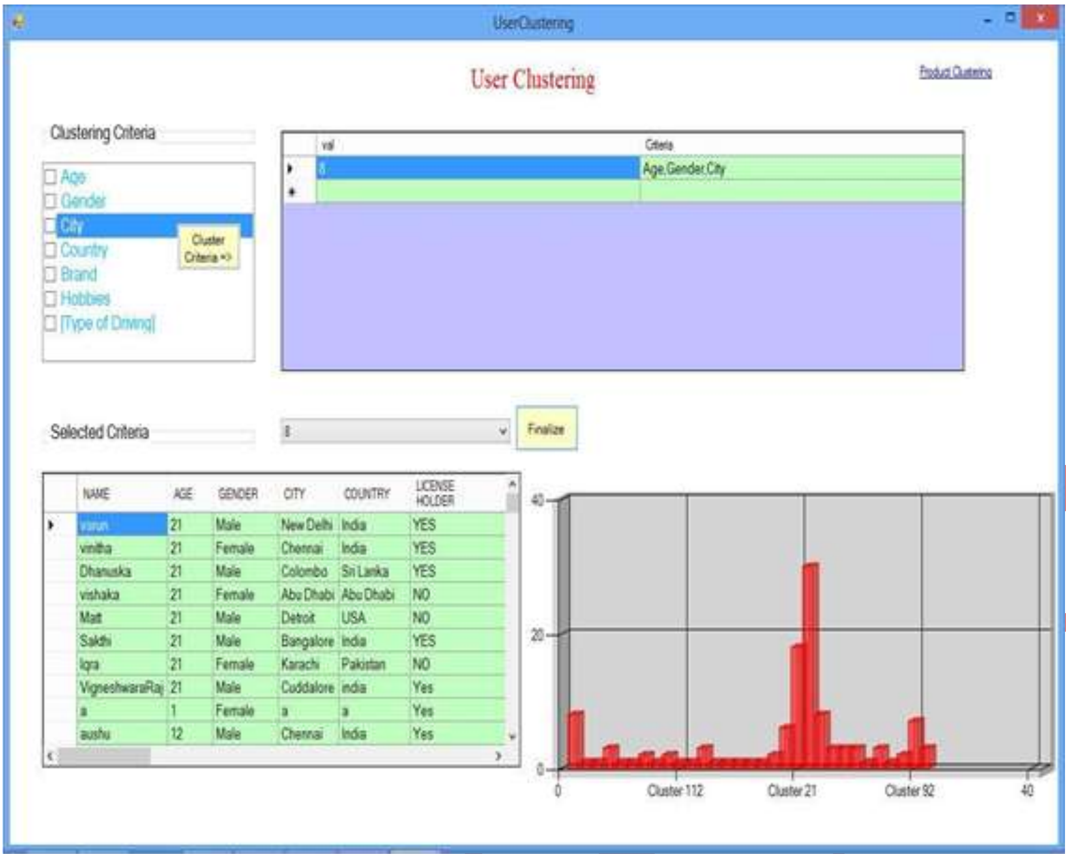


Fig. 8: User Clustering

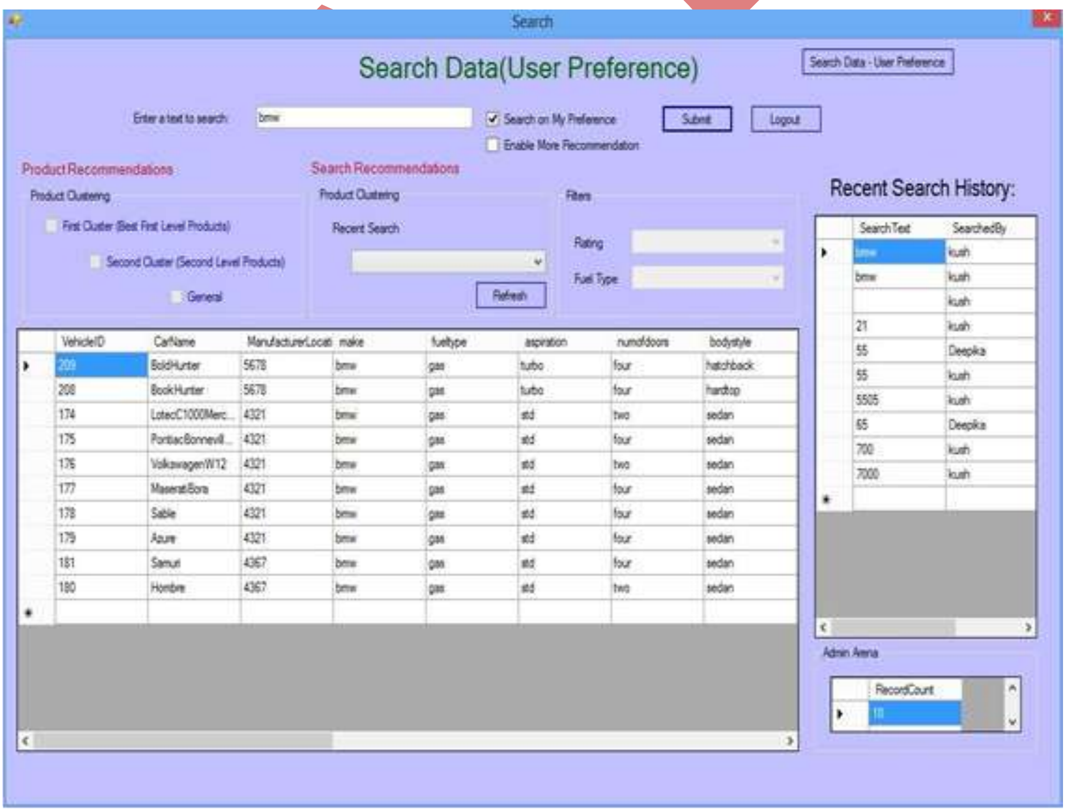


Fig. 9: Search Data-User Preference

4.6. Search Data-Relative User Preference

Incorporation of a final filter based on price is done. Most of the searches, when related to the searching of cars have been associated with price as a filter and thus the final search has been done in such a way that price has been included as a filter and the cars closest to the price selected have only been displayed in the search results. Finally relative users are taken in the category specified by the admin in the user-user clustering module and the users in that cluster are shown to the user as represented in Fig. 10. The user can select the user he wants and also choose the keyword that he has searched for and get his results. The relative users results are appended to the actual search of the user. The top five searches performed by the user in all the searches that he has made in all his history have been taken into consideration and the final recommendation of the product is given to the user.

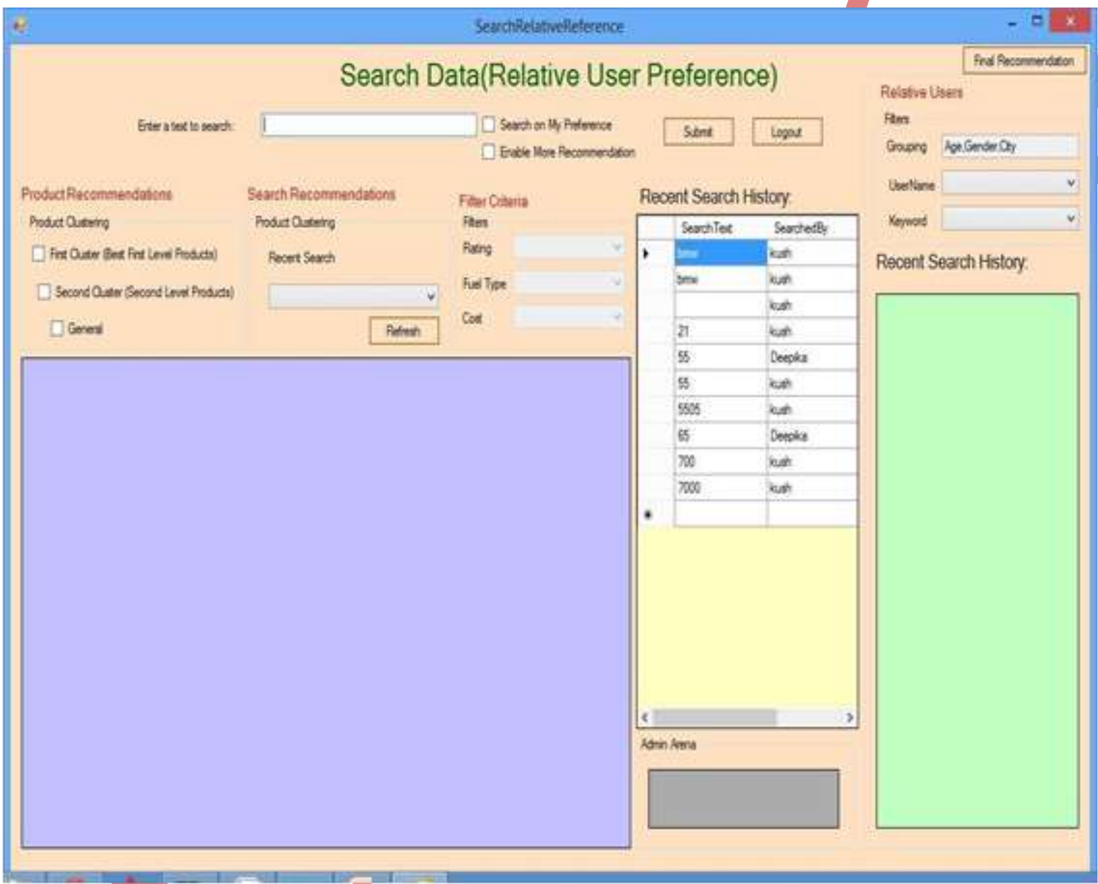


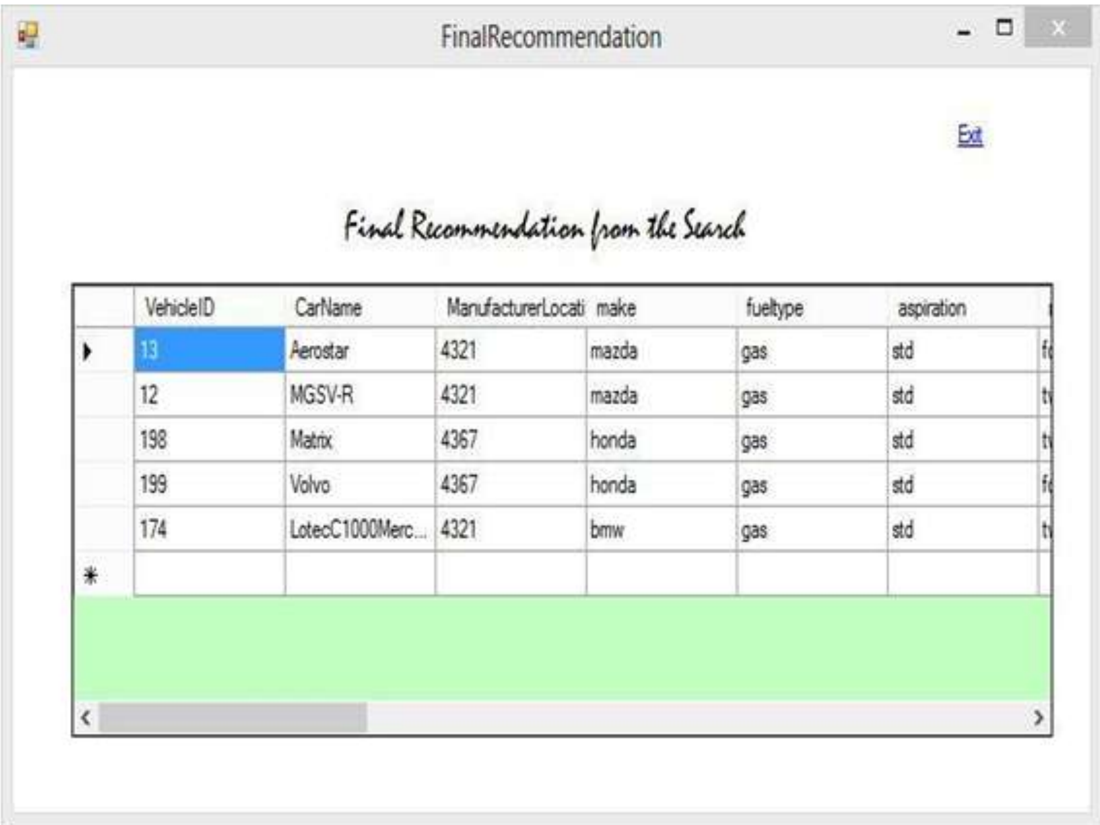
Fig. 10: Search data-Relative User Preference

5. EXPERIMENTAL EVALUATION AND ANALYSIS

An exhaustive analysis was done on the results obtained from the recommender system by using the input dataset collected from the survey conducted among the potential customers and the product- review dataset collected from the website www.toptenreviews.com. Both the datasets were large enough to give good results for the recommender system. The detailed discussion of the results with the analysis is given in the next section.

5.1. Results

The final output of the project is an optimal recommendation of the cars to the user. Top five search results are represented by the output. Factors such as the users profile, users ratings of the cars, users preference, and the search history of the users with similar attributes to user currently searching to give an optimal recommendation to the user, the one who is searching, have been taken into consideration. The final ordering is however done on the price preference of the user that is what price range he prefers to buy the car at as represented by Fig. 11.



	VehicleID	CarName	ManufacturerLocati	make	fueltype	aspiration
▸	13	Aerostar	4321	mazda	gas	std
	12	MGSV-R	4321	mazda	gas	std
	198	Matrix	4367	honda	gas	std
	199	Volvo	4367	honda	gas	std
	174	LotecC1000Merc...	4321	bmw	gas	std
*						

Fig. 11:. Final Recommendation

5.2. Performance Evaluation

In the existing systems, the number of initial search results (or the pool of data from where the final result is got) is much smaller as it does not take into account a lot of factors. On the other hand the proposed system takes into account search history and search results of other similar users and adds those results to the search pool thereby increasing the diversity of items. This is represented in Fig. 12, where y-axis is number of items and x-axis is the system used. However accuracy of recommended items is not affected as, the data added to the pool is relevant to the user ensuring that the final shortlisted items are of highest relevance. By listing out the option of which users search history to choose from and also which keyword, it ensures that there not only is a trade-off between accuracy and diversity but also an optimal result in both accuracy and diversity. The search results obtained are not only more in number when compared to the existing system but is also more efficient as the users can get info of the products searched by their friends and the search of the friends is also embedded into the search of the user. Also we take into consideration the users search and his ratings into account and order the final output based on the price limit given by the user thus making it a more refined search. Thus the final recommendation given to the user comprises of all the ratings given by him at the start, the search results of the relative users given by the admin and finally the filters for search that help him to select what he wants and the price at which he wants to buy the car at.

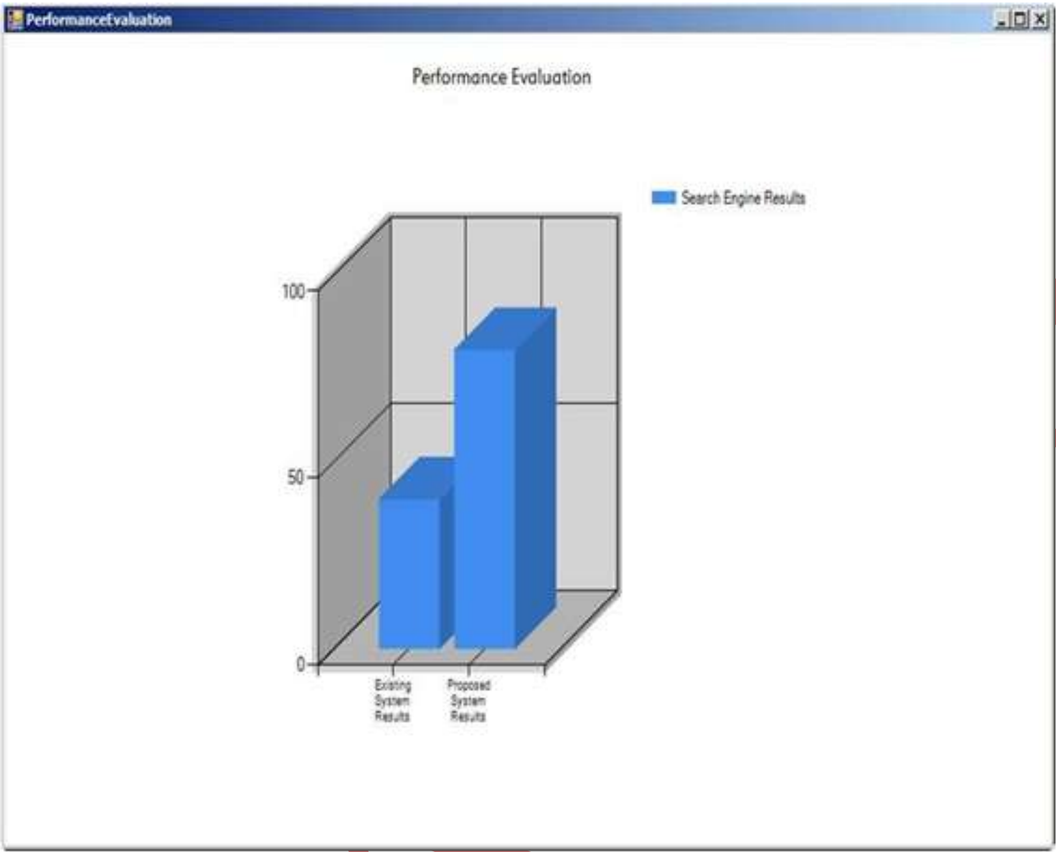


Fig. 12: Performance Evaluation

6. CONCLUSION

A novel method of mining has been incorporated to the system so as to thrive the suggestion engine more effective for every individual user. This system efficiently overcomes the drawbacks of existing systems by proposing a more generalized rating of the product that has been gathered in our system to suggest to any user according to his choice. The resulting product has been considered specifically according to each user's individual opinion about the product and the crowd sourcing opinion. According to the rating of each of these mentioned factors an optimal suggestion for the search has been made to the user. A proper correlation for each of the users with the related users based on their similarities of attributes and thus defining the search results of what each user with similarities has also searched for similar to the product and also appending the search results of the similar users with the current user thus providing a social correlation factor among the users in a very efficient manner has been established. That once a set of results are found to be compatible with the users interests, they are sorted according to price before being finally displayed, as a result ensuring the user gets value for money especially in the current economic scenario.

7. ACKNOWLEDGEMENTS

The authors would like to express sincere gratitute to all our collegues and friends who have rendered their inevitable help for bringing this work a successful one.

REFERENCES

[1] Freddy Chong Tat Chua, Hady W. Lauw, and Ee-Peng Lim, Generative models for item adoptions using social correlation, IEEE Transactions on Knowledge and Data Engineering, vol. 25, num. 9, 2013.

[2] D. Kempe, J. Kleinberg, and E. Tardos, Maximizing the spread of influence through a social network, Proc. 9th ACM SIGKDD International Conference Knowledge Discovery and Data Mining (KDD), pp. 137–146. ACM, 2003.

[3] Gediminas Adomavicius and Young Ok Kwon, Improving aggregate recommendation diversity using ranking-based techniques, IEEE Transaction on Knowledge and Data Engineering, vol. 24, num. 5, pp. 896–911, 2012.

[4] Joseph Konstan Badrul Sarwar, George Karypis and John Riedl, Item-based collaborative filtering recommendation algorithms, IEEE Transaction on Knowledge and Data Engineering, pp. 519– 530, 2013.

[5] Z. Y. Dingqi Yang, Daqing Zhang and Z. Wang., A sentiment-enhanced personalized location recommendation system, ACM Hypertext, 2013.

[6] Y. Koren, Factorization meets the neighborhood: a multifaceted collaborative filtering model, Proc. 14th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 426–434. ACM, 2008.

[7] Aggarwal, C. C., Wolf, J. L., Wu, K., and Yu, P.S., Horting Hatches an Egg:A New Graph-theoretic Approach to Collaborative Filtering, Proc. ACM KDD'99 Conference. San Diego, CA. pp. 201-212, 1999.

[8] Billsus, D., and Pazzani, M. J., Learning Collaborative Information Filters, Proc. ICML '98. pp. 46-53. 1998.

Biographical Notes

Mr. Madhu K. P. is presently pursuing Ph. D. in the Department of Computer Science and Engineering (Specialization in Location Based Recommender Systems), Anna University, Chennai, India.

Dr. D. Manjula is working as a Professor in the Department of Computer Science and Engineering, Anna University, Chennai, India.