ISSN: 0975-8585

Research Journal of Pharmaceutical, Biological and Chemical Sciences

Recommendation System with Demographic Attributes for Fake Review Identification

Sivaramakrishnan N*, Subramaniyaswamy V.

School of Computing, SASTRA University, Thanjavur, Tamil Nadu, India.

ABSTRACT

Recommendation systems are a type of information filtering system. It is a tool which helps in decision making. Traditional system provides personalized ratings without considering user preferences, thus fails to meet the requirements. Preference based recommendation system presents a personalized service recommendation list. Keyword aware service recommendation approach is used to provide recommendations based on key terms. Collaborative filtering approach is applied to generate recommendations based on user queries and responses. We propose a recommendation system which is based on a demographic attribute namely age which eases the recommendation. Reviews or feedback of the product helps in promoting or demoting the product. Certain reviewers are paid or getting incentives by the online sellers in order to get the product promoted. Fake reviews have to be differentiated. Finally, experiments are conducted on real-world data sets, and the outcome significantly improves the recommendation results with demographic attributes over existing approaches.

Keywords: Recommender system, collaborative filtering, demographics, fake reviews, behavioural patterns.

Corresponding author

ISSN: 0975-8585

INTRODUCTION

Recommender System can be defined as a system that produces individualized recommendation as output in a large space of possible options. Currently, these systems are incorporating social information. They were initially based on

- Content-based filtering
- Collaborative filtering
- Hybrid Filtering

Content based approach recommends items that are similar to items the user preferred or queried in the past. It relies on product features and textual item descriptions. Collaborative filtering approach may consider user's social environment. It recommends items based on the opinions of other customers who have similar tastes or preferences as like the active user. Collaborative recommender system can either memory based or model based. Hybrid recommender system is the combination of content based filtering and collaborative filtering recommender system.

Recommendation systems help the user to choose the product without much difficulty. In this modern society everyone is going for online-shopping as they are running out of time. Eventually there arises a necessity having recommendation generation engine (as depicted in Figure.1) in those sites which simplifies the job of user. Customer satisfaction is the main focus of business. Nowadays there prevails popularity by price, people options in the e-shopping website. Henceforth recommendation system has to keep track of user preferences and multi-criteria ratings. These voluminous data needs the distributed and parallel computing concepts for the scalability problem in a big data environment. A wide area of research is still in action to improve the recommendation systems and services. Thus it leads to the sales and business profits.

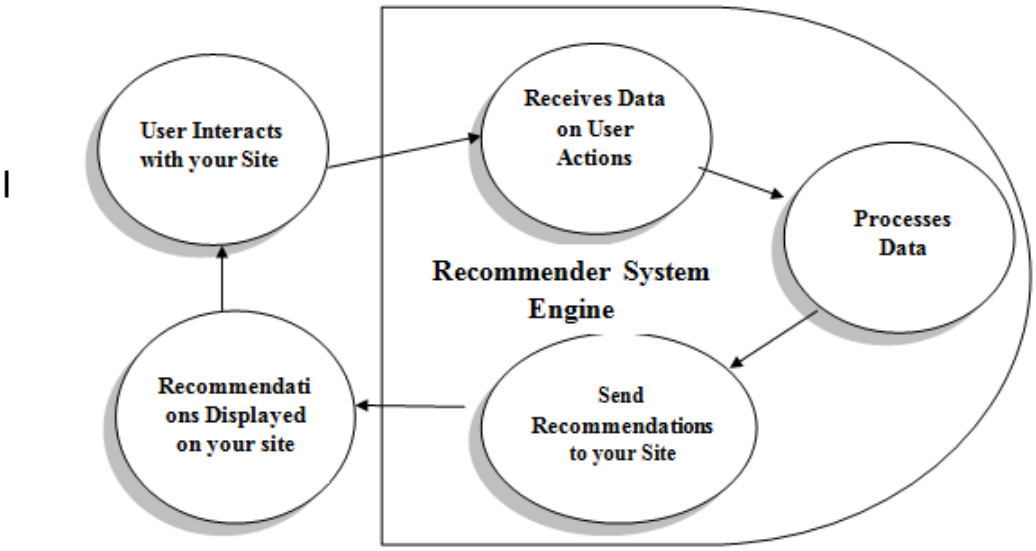

Figure 1: Recommendation Engine Process

Recommendation system is an information filtering system. In the future, they will use implicit, local and personal information from the Internet of things. There are pervasive applications of recommendation systems with the provocation of personalized recommendations and services.

Recommender systems use a broad range of techniques from information retrieval, statistics, machine learning and data mining to search for similarities among items and customer preferences.

Research in development of Recommender Systems expands to following areas:

- Product / Service Recommendation

- Tourism/Travel recommendation
- E-Commerce recommendation
- Web search recommendation
- Music recommendation
- TV shows recommendation
- Movie recommendation
- Library recommendation

Recommender system does not rely on product transaction record.

LITERATURE REVIEW

Context aware Social Information retrieval

Generating recommendation includes different weighting methods which are based on different text feature selection strategies. The objective of strategies is to analyze the relation between an aspect's frequencies. The linear regression model (LRM) and the probabilistic regression model (PRM) applied for finding recommendations [1].

Latent Semantic Analysis approach (LSA) method which is used to find similar documents where utilization of the user's context for recommending a different number of items [2]

Three social factors-personal interest, interpersonal interest similarity, and interpersonal influence are fuse into a unified personalized recommendation model based on probabilistic matrix factorization [4].

Context-aware Personal Information Retrieval (CPIR) algorithm has been proposed which considers both the participatory and implicit-topical properties of the context to improve the retrieval performance. Addressing problem is to capture the topical diversity among the collection of participating user PWIs (Personal Web Information) using any of the Clustering Techniques [5].

Presenting a novel approach using information extraction and summarization method to improve the quality of the reply for users' queries in social network [6].

Information Filtering Framework

Unified Collaborative and Content-Based Web Service Recommendation proposes a novel approach considers simultaneously both rating data (e.g., QoS) and semantic content data (e.g., functionalities) of web services using a probabilistic generative model [7].

Hybrid filtering mechanism is proposed to eliminate irrelevant or less relevant results which combine content-based filtering and collaborative filtering.

Novel algorithm called intensity-based contraction (IC) aiming at reducing the level of sparseness of the data. Intensity-based contraction clustering (ICC) algorithm to demonstrate the efficiency [8] .

Applying data fusion can be helpful for improving effectiveness of item recommendation. The popularity-based item novelty model to measure the novelty of recommendations. Semantic analysis can be used to detect and merge synonymous tags and to examine temporal feature to improve the recommendation performance [9].

As depicted in figure 2, we have to apply various filtering methods such as content based, collaborative filtering and hybrid filtering methods and evaluate their performance. Original dataset can be divided into training and testing datasets. Combinations of many filtering methods are called as Information Filtering Framework.

ISSN: 0975-8585

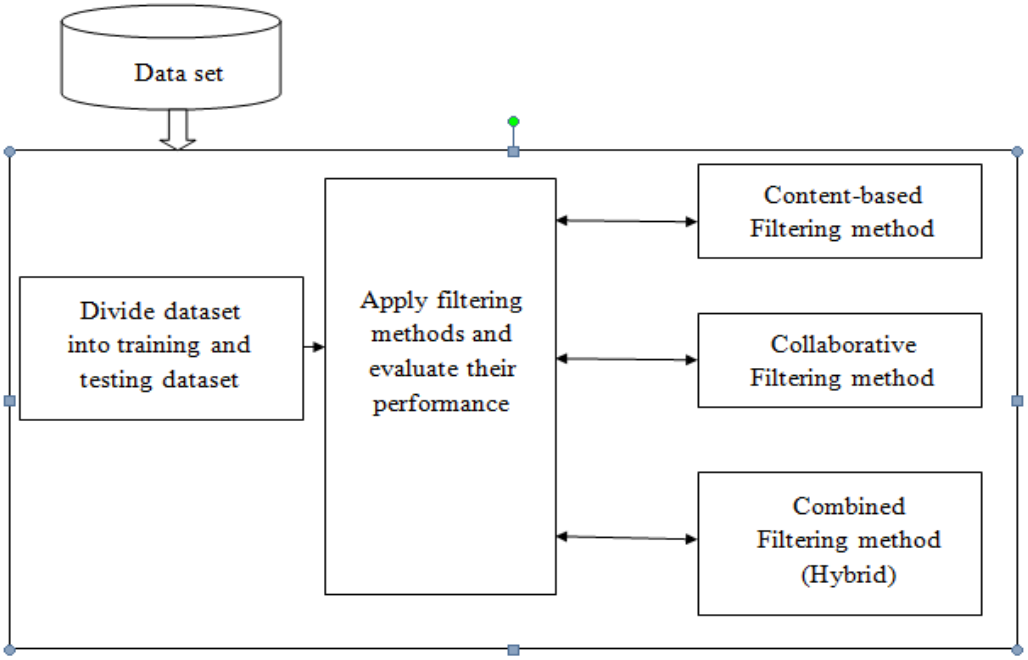

Figure 2: Information Filtering Framework

PROPOSED WORK

Demographic based Recommender system

A hybrid algorithm with demographic attributes is proposed and applied. The demographic correlation is calculated by the nearest neighborhood item. Collaborative Filtering algorithm is enhanced with the application of demographic attributes. Reviews are filtered based on their preferences and demographic data. The information extracted from demographic attributes helps in the generation of accurate predictions in the recommendation systems[12].

Recommendations are generated based on the location information of the user. Considering the real time application such as mobile searching where the context information is also obtained from the user's mobile devices using Bayesian networks reflecting user preferences [15].

Recommendation methods

Current recommendation methods are mainly of three categories: collaborative, content-based and hybrid approach. Collaborative approach filters the users having similar preferences of the product. Content-based considers the past preferences of the user. Hybrid approach is a mixture of both collaborative and content-based. In our paper CF algorithm is used which is further interpreted as user-based CF approach and item-based CF approach. A keyword aware service recommendation system which splits the positive and negative reviews based on the feedback has been proposed [10]. The reviews are extracted from the previous user ratings.

Approach and steps

Keywords can be interpreted as positive and negative and stored in a stack. If the input of the user is mapped with the words in the stack then it is marked as positive or negative review based on its meaning. User-based collaborative filtering algorithm is widely applied in many recommendation systems. CF algorithm filters recommendation based on the similar preferences of the previous and active user [10]. It is the powerful technique as it improves the efficiency and quality of service.

ISSN: 0975-8585

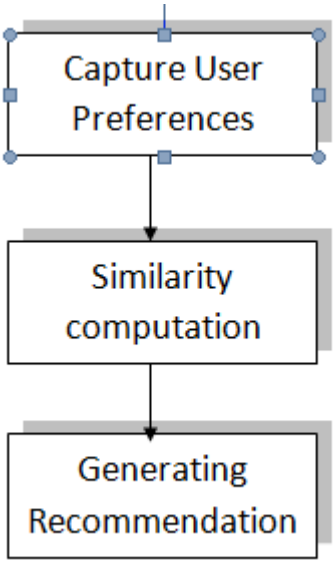

Figure 3: Flow of User-based CF algorithm

Capturing preferences of the user:

Preferences of the active and the previous user are formalized into different sets. The user gives reviews by choosing keywords from the keyword-candidate list and the domain thesaurus. Keyword-candidate list contains the list of keywords for reviewing. Domain thesaurus groups the set of keywords having similar meaning.

a) Pre-processing: This is done before the keyword extraction process. Primary HTML tags, stop words, script tags, comment tags, delimiters, inclusive content has to be removed from the review with the help of keyword stripping algorithm.

b) Keyword extraction: The reviews are formulated into the specific keyword set based on the keyword candidate list and the domain thesaurus. If the keyword is present in that thesaurus, then it is extracted into the preference keyword set. The repetition of the keyword is recorded and it is useful for calculating the preference keyword weight age factor [3].

Similarity computation:

If the reviews of the previous user in par with the reviews of the active user, then similarity computation is done. Similarity is distinguished as approximate similarity computation and exact similarity computation.

Approximate similarity computation:

It is applied for the measurement of binary or non-binary variables. Here the weightage factor of the keyword is taken into account. It is based on Pearson correlation which is nothing but the linear correlation of previous and active user keyword set. It gives a positive or a negative value based on the correlation.

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}}$$

ISSN: 0975-8585

Exact similarity computation:

It is a cosine based approach and preferably applied in Analytic Hierarchy Process (AHP) approach. The weightage factor is calculated by the function term frequency/inverse frequency. Finally the similarity of the cosine function is calculated.

Generating personalised ratings and recommendation:

After the process of similarity computation, personalised ratings of each service are calculated. The similarity computation is compared with a standard threshold value to filter the previous user reviews. Ranking of product is also done based on the personalised ratings. Finally a list of ratings is presented to the active user. A product which has more ratings is highly recommended to the user [10].

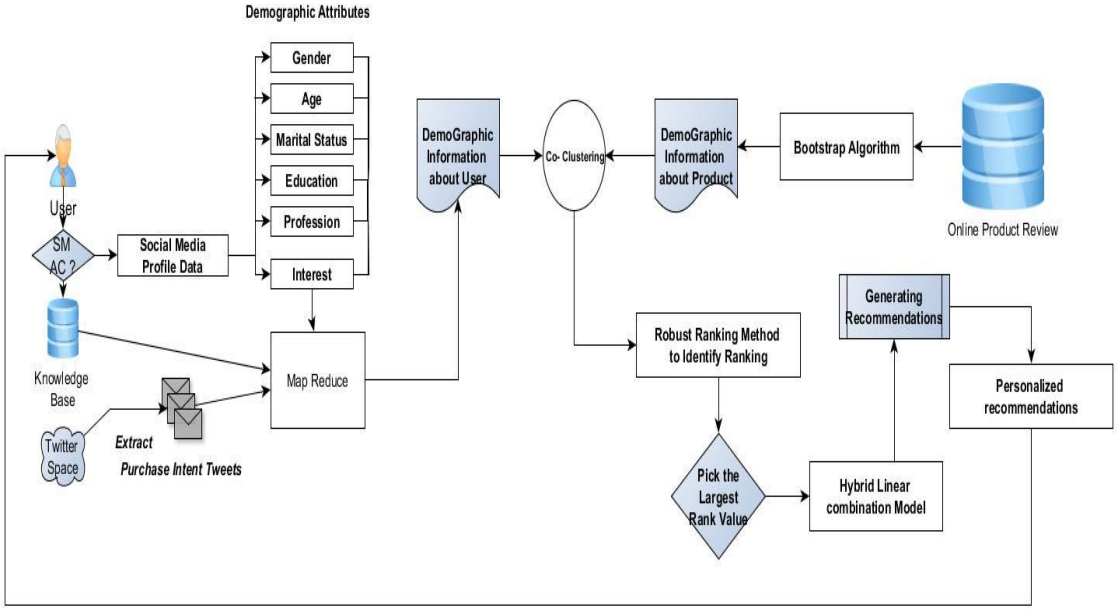

Figure 4: Recommendation systems with demographic attributes

As mentioned in the figure-4 the Product details are extracted from online product review which is available in many of the ecommerce websites. Similarly user preferences and their details are extracted from online public profile that is collected from online public profile which is available in social media micro blogs. Bootstrap algorithm used to learn the pattern and extracting demographic phrases from online product review. Using this algorithm we can extract top frequent patterns using apriori algorithm. Jaccard Coefficient similarity measure method used to find the matching between two extracted phrases or keywords. Bayesian classification algorithm used to predict outcome of interest based on training and testing datasets. Dirichlet smoothing method used to eliminate meaningless data or noisy data. Robust Ranking method applied for point wise, pair wise and list wise approach. Demographic method applied to identify tweet contents that contain purchase intents.

Demographic based recommendation system

Advancement of technologies leads to the upcoming of demographic recommendation systems. In this case, recommendations are based on demographic attributes such as age, gender, location, education etc. The information is filtered based on the demographic data [14]. This type of recommendation is popular as it reduces the ambiguity of users. For example: people in coastal regions prefer tough woolen clothes or thick coats. Those who are in hot places like Arab countries prefer silk or cotton wears based on their weather conditions. This is an example for location-based recommendation system. Demographic data collected from user profiles is **symbolized** as vectors and similarities between those profiles are also computed. Demographic correlations are defined by the similarity of two items.

ISSN: 0975-8585

Analysis of age-based recommendation system

A literature for mobile-recommendation systems is done based on the age of the user. User preferences may vary demographically in terms of age, location etc. This analysis as depicted in table 1 - mobile recommendations represents an approximate estimation of user based on their age.

Table 1. Pivot value computation for brand usage

Mobile Brand Name	< 15	16 – 25	26 - 35	36 – 45	46 - 55	>55
Lava	0.00%	0.95%	1.96%	5.26%	0.00%	0.00%
LG	0.00%	0.00%	5.91%	0.00%	0.00%	0.00%
Blackberry	0.00%	0.00%	9.80%	5.26%	0.00%	2.38%
Apple	0.00%	2.88%	7.84%	10.53%	0.00%	4.76%
Sony	0.00%	15.24%	11.76%	5.26%	0.00%	0.00%
Option 15	0.00%	1.90%	0.00%	0.00%	0.00%	0.00%
Mi	0.00%	1.90%	3.92%	0.00%	2.13%	2.38%
Asus	1.85%	5.71%	0.00%	0.00%	2.13%	0.00%
XOLO	0.00%	0.95%	0.00%	0.00%	2.13%	0.00%
Moto	12.96%	9.52%	3.92%	0.00%	4.26%	2.39%
HTC	0.00%	8.57%	9.80%	0.00%	4.26%	0.00%
Micromax	0.00%	11.43%	7.84%	10.53%	6.38%	16.67%
Karbonn	0.00%	0.95%	0.00%	0.00%	8.51%	0.00%
Nokia	5.56%	12.38%	7.84%	21.05%	34.04%	35.71%
Samsung	79.63%	27.62%	29.41%	42.11%	36.16%	35.71%
Grand Total	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

Analysis of mobile phone survey is done and based on that percentage of users in each brand is calculated. Pivot table presents the percentage of user accurately. A graph is plot which brings out the exact percentage of users who prefer mobile brands according to their age.

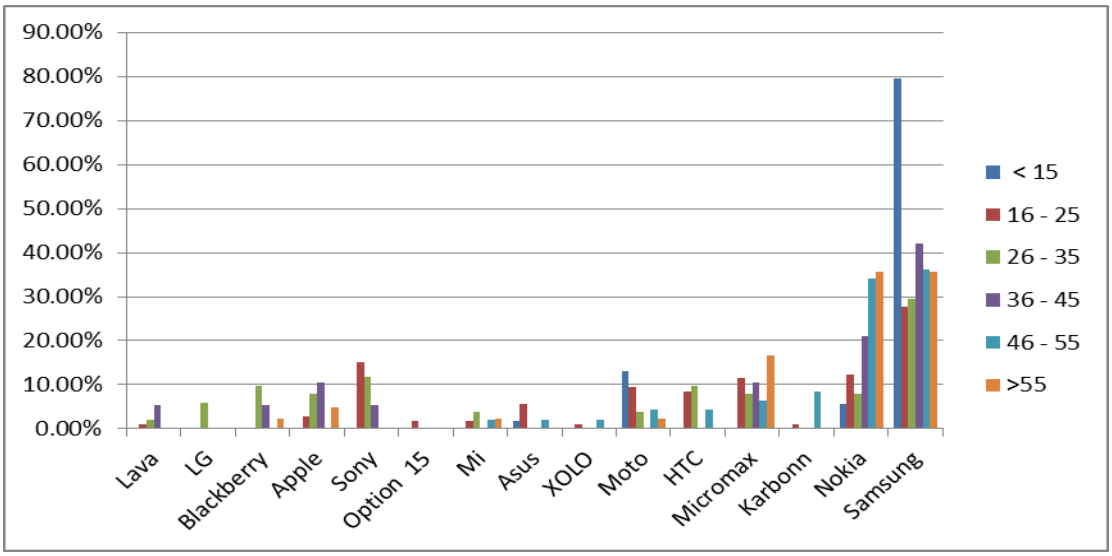

Figure 5: Analysis of mobile phone brand usage based on age

As depicted in the figure 5, it is quite clear that the people who are under the age of 25 generally prefers smart phones with upgraded features such as more memory, latest android OS and high camera resolution. People who are above the age of 50 prefer featured (button-type) mobiles. Thus there is significant relationship between age of the user and mobile brand preferences.

Avoiding Fake reviews:

Reviews or feedback of the product helps in promoting or demoting the product. Reviews can be singleton or a group review. Nowadays not all the reviews are genuine. Certain reviewers are paid or getting incentives by the online sellers in order to get the product promoted.

Behavioural Patterns:

There exists a necessity to differentiate fake reviews and genuine reviews. There are some basic ideas to differentiate those reviews: a) Customer has to read at-least twenty reviews before estimating the product. b) Fake reviewers use more adjectives and adverbs or exaggerating words with many exclamation points namely: “awesome!!!! Highly recommended!!!” c) Fake reviewers also use first-person subjects such as “I” and “Me’’ d) A genuine reviewer does not use the full name of a product. He might also mention some of the negative reviews about the product. e) Avoid extreme reviews and reviews which has more superlatives like greatest or worst”. f) Avoid reviews by lame usernames such as sales person, angry customer, review-girl unknown user etc, Skip those anonymous comments.

Techniques and Algorithm:

Spam reviews are of many types and they vary based on their comments such as misleading, non-review and brand-only reviews [14]. Misleading reviews are reviews which has the link of other related product. Brand-only reviews only promote their brand values. Non-review refers to random texts, irrelevant questions and answers. Many spam review detection techniques such as Factor graph model, Rating consistency check, GSRank, Behavioural footprint, temporal pattern discovery model are available to overcome the misleading reviews and brand-only reviews. We can detect fake reviews based on labelling the reviews, spam indicators, supervised and unsupervised model. There are some techniques which keeps track of fake reviewer behavioural patterns and comments. Reviews play a vital in estimating and deciding about the product. It serves as a catalogue and it helps in filtering of products and services.

CONCLUSIONS AND FUTURE WORK

In this paper, we have proposed demographic recommendation system which is used to indicate user’s preferences and user based collaborative filtering algorithm is adopted to generate appropriate demographic attribute based recommendations. A mixture of Collaborative Filtering approach and content-based approach is termed as hybrid algorithm that helps in overcoming the defects of collaborative filtering approach. We have shown that recommendations can be formulated as a learning to rank problem which takes as input parameters derived from both product and user demographics. In addition, we have conducted a user study to gauge the performance of our system in a real-world deployment. The results have also shown that our system is more effective in generating recommendation results better matching users’ preferences. We believe our study has profound influence on both research and industrial communities. An extension of demographic recommendation system is also applied into various research areas. Identification of fake reviews has become a serious issue in online shopping. Innovative techniques and algorithm to differentiate the fake review over the genuine review are still being developed.

REFERENCES

[1] Chen G,Chen L, Augmenting service recommender systems by incorporating contextual opinions from user reviews,User Modeling and User-Adapted Interaction 2015; 25(3),295-329.

[2] Alhamid MF, Rawashdeh M, Al Osman H, Hossain MS, El Saddik, Towards context-sensitive collaborative media recommender system,Multimedia Tools And Applications 2015;74(24), 11399-11428.

[3] Shunmei Meng, Wanchun Dou, Xuyun Zhang, Jinjun Chen, KASR:A Keyword-Aware Service Recommendation Method on Map reduce for Big Data Applications,IEEE Transactions on Parallel and Distributed systems 2014;25(12), 3221-3231.

[4] Qian X , Feng H , Zhao G , Mei T, Personalized recommendation combining user interest and social circle,IEEE Transactions on Knowledge and Data Engineering 2014;26(7), 1763-1777.

[5] Han X, Wei W, Miao C, Mei J P, Song H, Context-aware personal information retrieval from multiple social networks,IEEE Computational Intelligence Magazine 2014; 9(2),18-28.

[6] Peng M , Gao B, Zhu J, Huang J, Yuan M , Li F, High quality information extraction and query-oriented summarization for automatic query-reply in social network, Expert Systems with Applications 2016 ; 44, 92-101.

[7] Yao L,Sheng QZ ,Ngu AHH, Yu J Segev A, Unified collaborative and content-based web service recommendation, IEEE Transactions on Services Computing 2015;8(3), 453-466.

[8] Cui B , Jin H, Liu Z, Deng J, Improved collaborative filtering with intensity-based contraction Journal of Ambient Intelligence and Humanized Computing 2015; 6(5), 661-674.

[9] Wu H, Pei Y , Li B , Kang Z , Liu X , Li H, Item recommendation in collaborative tagging systems via heuristic data fusion, Knowledge-Based Systems 2015; 75,124-140.

[10] Adomavicius G,Tuzhilin A, Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions, IEEE Trans. on Knowledge and Data Engineering 2005; 17(6), 734-749.

[11] Linden G, Smith B, York J, Amazon.com Recommendations: Item-to-Item Collaborative Filtering, IEEE Internet Computing 2003; 7(1), 76-80, 2003.

[12] Vozalis M, Margaritis KG, Collaborative Filtering Enhanced by Demographic Correlation, Proceeding of the AIAI Symposium on Professional Practice in AI, part of the 18th World Computer Congress 2004, Toulouse, France.

[13] Safiek Mokhlis, Consumer Choice Criteria in Mobile Phone Selection: An Investigation of Malaysian University Students, International Review of Social Sciences and Humanities 2012; 2(2), 203-212.

[14] Arjun Mukherjee, Abhinav Kumar, Bing Liu ,Junhui Wang, Meichun Hsu, Malu Castellanos, Riddhiman Ghosh, Spotting opinion spammers using behavioral footprints Proceedings of the 19th ACM 2013, New York, NY, USA.

[15] Moon-Hee Park, Jin-Hyuk Hong, Location-based recommendation system using Bayesian user preference model in mobile devices, International Conference on Ubiquitous Intelligence and Computing, 2007; 1130-1139.