Framework for Incorporating Social Networks with Recommender Systems: An Implementation of Profiling Products

Usman Humayun Mirza
Department of Computer Engineering
College of Electrical and Mechanical Engineering,
National University of Sciences and Technology
H-12, Islamabad, Pakistan
usman572@gmail.com

Aqsa Bajwa
Department of Computer Engineering
College of Electrical and Mechanical Engineering,
National University of Sciences and Technology
H-12, Islamabad, Pakistan
aqsa.mbajwa@gmail.com

Usman Qamar
Department of Computer Engineering
College of Electrical and Mechanical Engineering,
National University of Sciences and Technology
H-12, Islamabad, Pakistan
usmanq@ceme.nust.edu.pk

ABSTRACT
Recommender systems in today's world have played a very important role in bridging the gap between the customers and the retailers. They have evolved the way retailers reach out to their potential customers and provide them with what they want. On the other hand, many service providers are exploring the social media to reach out to new customers and increase their presence on the internet. This research Endeavour encompasses the proposal and implementation of a framework for a recommender system that runs independent of any retailer's site. It mines product information from the internet and suggests a concept hierarchical clustering of the acquired data. An application has also been developed to give a visual outlook on the formed clusters and how the system will work in real life.

Keywords
Recommender systems, customer profiling, social media, user clustering, product clustering

1. INTRODUCTION
Recommender systems in two decades have become the most important tool in reaching out to the customer in the world of ecommerce. They have been successfully applied to enhance the quality of service for customers, and more importantly, to increase the sale of products and services[1]. The notion of a recommender system revolves around the concept of personalization. Personalization is all about tailoring the services around the preferences and behavior of the target customers. Customer profiling is a key factor in the process of personalization. It is a means of meeting the customer's needs more effectively and efficiently, making interactions faster and easier and, consequently, increasing customer satisfaction and the likelihood of repeat or similar visits [5]. However, the main issue with personalization is: how to provide personal recommendations based on a comprehensive knowledge of who customers are, how they behave, and how similar they are to other customers, and how to extract this knowledge from the available data and store it in customer profiles [2].

In this paper we argue that the way most recommender systems gather information about the users' needs to be upgraded. In most of the current recommender systems the main source of user data collection are registration forms, feedback forms and ratings. We believe that the information that individuals share on their social profiles nowadays is much more authentic than the feedback forms, ratings or the registrations forms they fill on different retails sites. Social media is the new hype of the current era and it should be fully exploited to learn about the potential customers and to explore new business ventures. In this paper we suggest such a system that runs independently of any specific retail site and gathers products information all across the internet. For user information, the systems suggest mining their social profiles across all social sites to gather information related to their interests, likes, dislikes, status updates, etc. Once the product and the user data have been collected, we suggest applying data sanitization and different clustering techniques on both the data to form a structured dataset of products and the users. Recommendations to users can then be made about their products of interests and likes.

2. BACKGROUND
The first recommender system was developed about two decades ago [6]. Early recommender systems helped users to sift through a large number of documents like Usenet news articles or Web pages. With the proliferation of e-

commerce, recommender systems have become a power business tool to enhance customers' capability to overcome the product information overload problem. Many recommender systems have employed the user's explicit feedback in the form of ratings [7][8][9][10][11]. For example, Syskills and Webert used the user ratings of a web page input data[12]. Some systems employed user ratings on movies as input to their system[8]. Some systems also have used implicit feedback obtained during the user's usage session. The system analyzes the system logs to find user's preferences[4].

3. RELATED WORK

Let's have a look at some systems whose architecture has inspired us to suggest our own.

3.1 Graph Model

Zan Huang, Wingyan Chung, and Hsinchun Chen demonstrated the workings of commonly used recommender systems through a graph model. In this model, the data is shown in graphical form in two layers - one layer comprising the data of the products and the second layer comprising the data of the users. In both layers the individual products and individual users are represented by the nodes of the graph. The similarity among the users is shown by weighted links between them and the same has been done for the products. The transactions between the user layer and the product layer are captured by the interlayer links. The transactions are basically user's click stream, browsing history, purchase history, and so on and so forth [4].

3.2 Improvisation on the Graph Model

Lekha G. Rao and Siddharth C. Ravi KanthRao suggested a model in concept that adds a new dimension to the way the recommender systems work in general. It adds the use of social media to gather user information so that the recommendations can be made based on the user's interests. It suggested an Open Id/ Single Sign so that the retailer system can have an access to the customer's social profiles across different social sites. It improvises the graph model in such a way that it replaces the user layer with users interest and identity descriptors. Nodes represent interests and identities [3].

4. PROPOSED ARCHITECTURE

The proposed model in fig. 1 takes layered approach from the graph model and incorporates the user information from the social media as described in the later system. However, in the proposed architecture, the second layer will consist of users and each entity will define a group of users and not a single user. The users will be clustered together on the basis of their similarities with one another and the information about these similarities will be mined from social medium. The product layer of the proposed architecture will also be clustered on the basis of the similarities between the products. The relevant user cluster is then mapped onto the related/ required product cluster. That means that every user in one cluster will be displayed all products grouped in the product cluster which is linked with the user cluster. This link is established on the basis of user information mined from their social profiles. By using an effective clustering technique in both layers we will not need weighted links among the users or products that represent the similarities, because the strongly similar users and products are already clustered together and each such cluster is represented by a single node in the two layers.

Another main characteristic of the model is that it represents an independently working recommender system that mines information about the products as well as the users from the internet. It is not a retailer site specific system. When the users sign in, the information about their likes and dislikes, interests and hobbies is mined from the social networking sites APIs and on the basis of this information, the users are assigned particular cluster. This clustering represents all users with similar interests in one cluster, e.g., all users with interests in book reading are part of one cluster, all users interested in music are clustered in another, and so on and so forth. Similarly, the product layer is also clustered in the same way. This similarity index can be as broad as the diversity in the data allows and can be as compact as the implementer's wish.

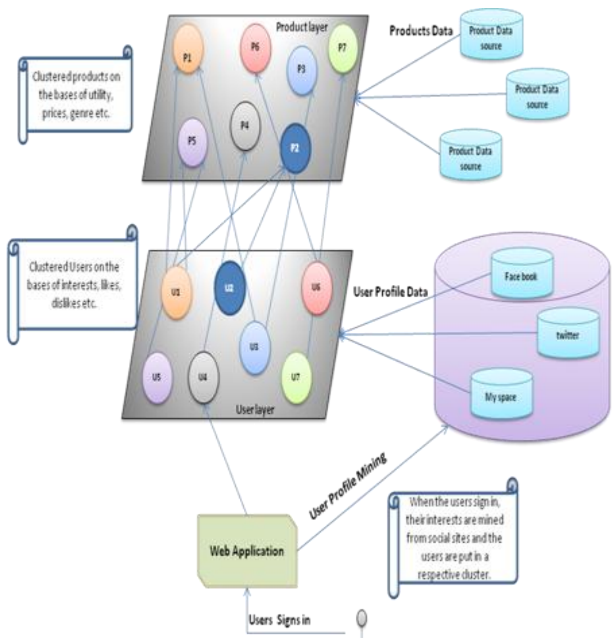

Fig 1. Proposed Architecture

4.1 Implementation of Products Layer

The data set chosen for the implementation of the product's layer in proposed architecture (Fig 1) consisted of jobs data. For credibility of the jobs data, LinkedIn was selected for mining open jobs information. Once the data was extracted, concept hierarchical clustering technique was used for data clustering.

4.2 Data Acquisition and Preparation

LinkedIn is a business oriented social networking service. LinkedIn API was used for job searching. Jobs data from Oct 2013 to Aug 2014 were extracted in unstructured form.

The number of jobs being collected was 3,923. The attributes associated with each job included LinkedIn job id, company id, job status, job position, job link, job type, job experience level, job poster id, job first name, job last name, job function, job description, job short description, job posted date, job expiry date, job location description, job id, job company name.

4.3 Concept Hierarchical Clustering

Since data acquired through LinkedIn was textual data and non-numeric, therefore, it needed careful analysis as to which type of clustering technique should be used for grouping similar jobs together. Fig. 2 shows the framework overview which was used for job clustering. After careful analysis and study of data, the following candidate attributes were shortlisted:

- Job – Field/ Domain : e.g., IT, HR etc
- Job Location: city/ country of job.
- Job Title/ Position: e.g., for managers, assistant manager, etc.
- Career Level/ Experience level: The career path like entry level, Executive, Associate, Mid Senior level and Director. All these positions require different years to experience.

A hierarchy is defined by specifying the total ordering among these attributes at the schema level. We can now easily specify explicit groupings for a small portion of intermediate-level data. The system generates the attribute ordering so as to construct a meaningful concept hierarchy based on the number of distinct attribute values. The attribute with the most distinct values is placed at the lowest level of the hierarchy.

Jobs{} < Job Position, Location < Job Experience Level < Job Function/ Field

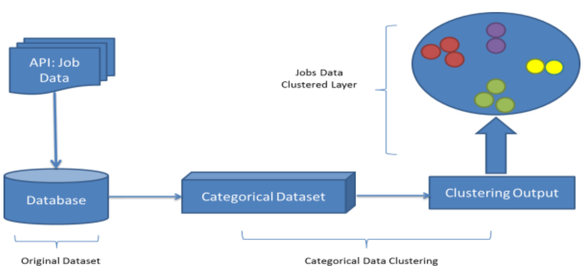

Fig 2: Proposed Clustering Framework

Categorical attributes have a finite (but possibly large) number of distinct values, with no ordering among the values. Therefore, an ordering is derived out of available attributes. Fig. 3 shows how hierarchical logic in the recommender system for jobs is implemented.

Level 1: Job Function/ Field = {Job Experience level, Job Position/ Title, Job Location}

Level 2: Job Experience level = {Job Position/ Title, Job Location}

Level 3: Job title, location and company come at the third level of hierarchy in our implementation.

To specify the level in which an attribute is included in concept hierarchy is dependent on a few exception and constraints which are:-

- An attribute value should be mandatory (must value)
- An attribute value with least distinct value without having nil/ zero value will be placed above other attributes in hierarchy.
- The attribute with most number of distinct values will be placed lower in order.

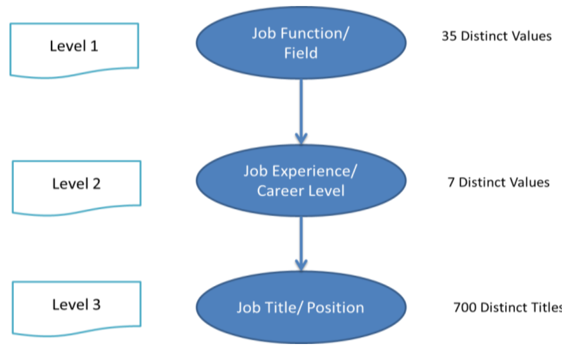

Fig 3. Concept Hierarchy Generation

5. RESULTS

5.1 Level 1 Clustering

Table 1 shows 35 clusters are formed on data of 3923 LinkedIn jobs. There are 35 distinct values for job function attribute. Hence, 35 subset of clustered jobs based on the field are created.

Job Function Description	Number of Records Clustered
Finance	217
Engineering	596
Supply Chain	95
Information Technology	429
Administrative	74
Other	194
Advertising	9
Human Resources	207
Legal	22
Art/Creative	54
Manufacturing	66
Distribution	6
Strategy/Planning	11
Management	165
Production	25
Sales	588
Purchasing	7
Business Development	112
Product Management	7
Quality Assurance	21
Project Management	92
Consulting	302
Marketing	111
Writing/Editing	30
Education	79
Research	19
Training	22
Public Relations	7

Health Care Provider	90
Customer Service	74
General Business	20
Science	11
Design	76
Accounting/Auditing	59
Analyst	26
Total Records	**3923**

Table 1: Clustering on Job Function/ Field

5.2 Level 2 Clustering

Table 2 shows 7 clusters are formed from the data when job experience level is kept for clustering of jobs.

Job Experience/ Expertise level	Number of Records
Entry level	211
Director	110
Not Applicable	1707
Executive	24
Internship	48
Mid-Senior level	1261
Associate	562

Table 2: Number of Jobs Clustered on Job Experience Basis

Statistics of job Clusters formed on the basis of level 1 and level 2 is shown in Table 3. Clustered by the proposed approach leaves no instance un-clustered. Result is that all the records are clustered in a way suitable for integration with recommender system or with any other layer, e.g., User Profile Layer.

Level	Number of Clusters	Clustered Instances	Not clustered Instances	Result
1	35	3923	0	100%
2	245	3923	0	100%

Table 3: Job Function – Human Resource

6. CONCLUSION

Recommender systems have changed the way people find products, information, and even other people. The technology behind recommender systems has evolved over the past 20 years into a rich collection of tools that enable the practitioner or researcher to develop effective recommenders. Our main contribution in this work was the idea of promoting an enterprise free system that effectively incorporates the information on the social media with the recommender systems. The clustering technique was specifically designed to cater for the non numeric data which itself brought many challenges with it. No tool is available for verifying categorical data clustering and the result varies with respect to the type of data that is under study, therefore, the selected records were physically worked through. It was seen that 99% of the jobs were correctly found in the cluster that contained similar jobs. All in all, the architecture successfully supported the new dimension and the clustering technique used for the LinkedIn jobs data generated fruitful results.

REFERENCES

[1] Choochart Haruechaiyasak, Chatchawal Tipnoe, Sarawoot Kongyoung, Chaianun Damrongrat, Niran Angkawattanawit, "A Dynamic Framework for Maintaining Customer Profiles in E-Commerce Recommender Systems"

[2] GediminasAdomavicius, AlexanderTuzhilin, "Using Data Mining Methods to Build Customer Profiles"

[3] Lekha G. Rao, Siddharth C. Ravi KanthRao, "Connecting the Dots: Retailer, User and Social Sites", 2011 IEEE

[4] Zan Huang, "A Graph Model for E-Commerce Recommender Systems", February 2003

[5] Bin Zhang, Zhonglin Lin, Wenjun Yin, Jin Dong, Minghua Zhu, "Establishing & Refinement of Customer Profile using Transactional Data",2008 IEEE

[6] Goldberg, Nichols, Oki & Terry, "Using collaborative filtering to weave an information tapestry". 1992 Communications of ACM, 35(12), 61-70.

[7] Balabanovic, & Shoham, , "Fab: Content-based, collaborative recommendation" 1997. Communications of ACM 40(3), 66-72

[8] Basu, Hirsh, Cohen, "Recommendation as classification: Using social and content-based information in recommendation". In proceedings of the 15th National Conference on Artificial Intelligence, pages (714-720), Madison, Wisconsin, July 1998.

[9] Claypool, Gokhale, Miranda, Murnikov, Netes, & Sartin, "Combining content-based and collaborative filters in an online newspaper". In proceedings of the ACM SIGIR workshop on Recommender Systems, Berkeley, California, August 1999.

[10] Condliff, lewis, Madigan, & Posse, "Bayesian mixed-effects models for recommender systems". In proceedings of the ACM SIGIR Workshop on Recommender Systems, Berkeley, California, August 1999.

[11] Mooney & Roy, “Content-based book recommending using learning for text categorization". In proceedings of the Fifth ACM Conference on digital libraries, pages (195-204), San Antonio, TX, June 2000;

[12] Pazzani & Billsus, "Learning and revising user profiles: The identification of interesting web sites" 1997.