DRec:Multidomain Recommendation System for Social Community

Roshni K. Sorde
Department of CS & IT,
Dr. BabasahebAmbedkarMarathwad University
Aurangabad, Maharashtra,India
roshnisorde@gmail.com

Sachin N. Deshmukh
Department of CS & IT,
Dr. BabasahebAmbedkarMarathwad University
Aurangabad, Maharashtra,India
sndeshmukh@hotmail.com

Abstract—The Recommendation System is the software engines and the approaches for consideration proposal to the user which might be most probably matched with the liking of users. Usually, Recommendations system recommends on various fields like what items to buy, which movies to watch even the job recommendations, depending upon the users profile. Instinctively if the domains of users are captured and filtered out accordingly to recommend them will be a very useful idea. In this paper we will be discussing about the research done by us and the limitation of the system.We design a system for recommending domains in social network, using an explicit / offline data. We have tested it on two popular dataset namely Epinions and Ciao. The ratings of items are studied and performance measures are calculated with three different ways 1) MAP (Mean Average Precision) 2)F-measure and 3) nDCG (Normalized Discounted Cumulative Gain) . As well we have compared the results with 5 comparisons methods.All the techniques and methods are explained in paper.

Keywords-*Recommender Systems, Collaborative Filtering, Social Network, Probabilistic Topic Modeling, Epinions, Ciao*

*****

I. INTRODUCTION

All With the growth of Buying and selling of things Online, it became a big problem to find what the user is actually looking for. Search engines partially solved that problem, which has evolved a new branch and scope of system like Recommendation System [1].Recommender System are the software engines and approaches for providing suggestion of products to the user which might be most probably matched to the users choice. Usually the recommender system is a technology which filters out the information to envision in case a particular user will like a specific item; this is usually called as prediction problem, or to identify N set of items that will be of certain users interest called as Top N-recommendation problem [2]. From past few years the use of recommender System is being gradually increasing in various different applications, for instance application for recommending books, CDs and other products at different search engines like amazon.com , Netflix.com, ebay.com and so on. Even the Microsoft suggests many additional software's to user, to fix the bugs and so forth [3]. When a user downloads some software, a list of software is provided by the system. All the above examples would be result of diverse service, but all of them are categorized into a recommendation System, Identifying web-pages that will be of interest, or even implying backup ways of searching for information's [4]. There is huge number of algorithms used for making a personalized Recommender System, but out of them 2 algorithms became most popular they are Content Filtering from the item based filtering and Collaborative Filtering from the social community [5].They are used as the base for new era's recommender system[6].

The list of techniques which recommender system applies comes from other research domain such as Human Computer Interaction (HCI) or Information Retrieval (IR) [7]. However, most of these systems take in their core as algorithm that can be understand as a particular instance of a data mining (DM) technique. The data mining development consists of 3 steps, succession: Data Preprocessing Data Analysis and Result Interpretation [8][9].

The social research network is a very challenging and research oriented subject

II. BACKGROUND

The outcome of Recommender System is a list produced out of the two mainly two –collaborative filtering and Content filtering. Content based algorithm is built up solely at the time when a new profile of user is built [10]. All the information about the user's choice is stored in the users profile; from there the taste of the user is studied. In recommender system the taste is studied by combining the entire positively rated product in one group and then finding the maximum likelihood ratio of the ratio, and is suggested to the user [11].

Collaborative Filtering is one of the most research topics. The main concept is behind studying the social community and then deciding the user having the similar appreciation. If the users have similar choices and tastes then they fall in same category of choices. Even though the User has not rated the products but still it will be recommended to the user if they belong to the same category[12].

III. Method

A. *Probabilistic Methods*

Probabilistic item similarity functions discussed in *Computing Item Similarity*, several fully probabilistic formulations of collaborativefiltering have been proposed and gained some currency. Thesemethods generally aim to build probabilistic models of user behaviorand use those models to predict future behavior [13]. The core idea of probabilisticmethods is to compute either $P(i|u)$, the probability that user u will purchase or view item i, or the probability distribution $\mathbf{P}(ru,i|u)$over user u's rating of item i (and the related problem $E[ru,i]$, theexpected value of u's rating of i).

Probabilistic Matrix Factorization

Probabilistic latent semantic analysis (PLSA, also called PLSI or probabilistic latent semantic indexing in the information retrieval literature) is a matrix factorization technique similar to singular value decomposition

The basis of PLSA is a probabilistic mixture model of user behavior, diagrammed with plate notation in Figure 1. PLSA decomposes the probability $P(i|u)$ by introducing a set Z of latent factors [14]. It assumes that the user selects an item to view or purchase by selecting a factor $z \in Z$ with probability $P(z|u)$ and then selecting an item with probability $P(i|z)$; $P(i|u)$ is therefore $\sum P(i|z)P(z|u)$.

This has the effect of representing users as a *mixture* of preference profiles or feature preferences and attributing item preference to the preference profile rather than directly to the users. The probabilities can be learned using approximation methods for Bayesian inference such as expectation maximization.

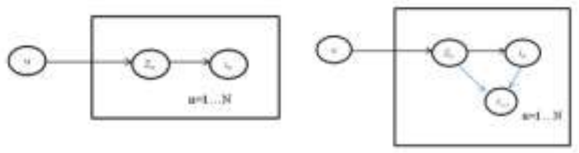

Figure 1: PLSI generative model a)model user purchase b) model real value rating varing by item.

The probabilities can be stored in matrices, so that the preference matrix **P** (where $pu,i= P(i|u)$) is decomposed into

$$P = \hat{}U\Sigma\hat{}TT \qquad (1)$$

$\hat{\mathbf{U}}$is the matrix of the mixtures of preference profiles for each user(so $\hat{}uu,z= P(z|u)$) and $\hat{\mathbf{T}}$is the matrix of preference profile probabilitiesof selecting various items. $\mathbf{\Sigma}$ is a diagonal matrix such that $\sigma z= P(z)$(the marginal probability or global weight of factor z) [15]. This factorizationallows prediction to be done meaningfully in unary domains byconsidering the probability that u will purchase i, in contrast to itemitem unary recommendation where the psuedo-predictions were only useful for ranking candidate items.

Let $\mathbf{R}k\in \mathrm{R}Nk \times Mk$ denote the rating matrix for the k-th domain, where k =1, . . . , K. Mk and Nkare the number of items and usersin each domain respectively. Let $\mathbf{P}k\in \mathrm{R}d\times Nk$and $\mathbf{Q}k\in \mathrm{R}d\times Mk$denote the latent feature matrices in k-th domain,with column $\mathbf{p}ki$and $\mathbf{q}kj$represent the latent feature vectorsof users and items respectively, where d denotes the dimensionof latent feature [16]. Adopting PMF model in differentdomains, the model is trained on rating data by minimizingthe square error:

$$\frac{1}{2}\sum_{i=1}^{N^k}\sum_{j=1}^{M^k} I_{ij}^k\,(R_{ij}^k - \left(p_i^k\right)^T q_j^k\,)^2 \;+\; \frac{\beta}{2}\sum_{i=1}^{N_k}||P_i^k||^2 \;+ \; j{=}1Mk||qjk||2\;) \qquad (2)$$

whereI_{ij}^kindicates the training data of user-item pairs belongedto domain k,||.||2 denotes the Frobenius norm tomake the solution more robust, and β is the regularizationcoefficient. One important difference between the PMF andour model is that we consider the training process acrosseach domain. Therefore, we have K objective functions intotal. The parameters $p_i^k$and $q_{\boldsymbol{j}}^{\boldsymbol{k}}$

in Eq.(2) can be minimized by Alternating Least Square(ALS) method, which

performs the following two updates alternatively.

First, optimizing Eq.(2) with respect to $p_i^k$for i = 1, 2, . . . Nk, in domain k and fixing all $q_{\boldsymbol{j}}^{\boldsymbol{k}}$leads to

$$p_i^k \;=\; (\,\sum_{j=1}^{M_k} I_{ij}^k\; q_j^k \left(q_j^k\right)^T \;+\; \beta\, I_d)^{-1}\left(\sum_{j=1}^{M_k} I_{\boldsymbol{ij}}^{\boldsymbol{k}}\, R_{ij}^k\, q_j^k\right) \quad (3)$$

Then, optimizing with respect to $q_{\boldsymbol{j}}^{\boldsymbol{k}}$for j = 1, 2, . . Mk, indomain k and fixing all p_i^kleads to

$$q_i^k \;=\; (\sum_{i=1}^{N_k} I_{ij}^k\; p_i^k \left(p_i^k\right)^T + \beta\, I_d)^{-1}(\sum_{i=1}^{N_k} I_{\boldsymbol{ij}}^{\boldsymbol{k}}\, R_{ij}^k\, p_i^k) \qquad (4)$$

In order to avoid overfitting on test data, we use the weighted-regularization

$\sum_{j=1}^{M_k} m_{vj}^{v}\;\left|q_j^k\right|2$ and$\sum_{i=1}^{N_k} n_{ui}^k\;\left|p_{ji}^k\right|2$ instead of the original regularization terms in Eq.(2) in experiment, where $m_{vj}^k$and $n_{ui}^k ui$ denote the number of ratings of item $vj$and user uiin Dk, respectively.

IV. EXPERIMENTS

A. *Dataset:*

In this paper we have examine on two datasets which are multidomain along with the ratings and trustnetworkEpinions[1] and Ciao[2]. Both of them are most popular consumer review websites where users not only provide reviews to the products they know very well and also maintain the trustlinks of the product with the users.

The version of the two datasets used in this study is published by the authors of including data records until Oct 2012. The dataset used are for various research purposed by the Researchers and still been used.

TABLE I. STATISTISC OF THE DATASETS

Features	Epinions	Ciao
Rating Table	416645	280134
Trust network	10869	79693
No of Users	10844	7375
No of product id (items)	162952	106435
Category id(domain)	27	28
Rating	5	5
Avg rating per user	38.42	37.98
Avg Rating per item	2.71	2.63

B. *Performance Measures:*

So far, we have followed a common practice ondatasets to evaluate prediction accuracy by the MAP (Mean Average Precision), F-measure and nDCG (normalized discounted cumulative gain) which are a commonly known classical measure [18]. For each users a precision and recall is calculated and at ranked position j, for preference j, the average precision is calculated by

$$AP(u) = \frac{\sum_{j=1}^{n} prec\ (j) \times pref\ (j)}{no\ .ofpreferreditems} \qquad (5)$$

$$MAP = \frac{\sum_{u \epsilon U}\ AP(u)}{|U|} \qquad (6)$$

F-measure is calculated with the help of precision and recall value for the top-n list

$$F1 = \frac{2 \times precision\ \times recall}{precision\ + recall} \qquad (7)$$

And for calculation of nDCG, the pref(j) is also used.

$$nDCG = \frac{1}{IDCG} \times \sum_{j=1}^{n} \frac{2^{pref\ (j)-1}}{log_{2}(j+1)} \qquad (8)$$

However the IDCG is found with the help of perfect ranking algorithm, ie.nDCG give priority to top ranking entities. More the values of his measuring models indicate the better he recommendation system[19].

C. *Comparisions*

Here we have compared our project with two variants and method of system and 3 other baseline methods to demonstrate the effectiveness of each part our system

- D'Rec with single class Insingleclass model, we know that each domain is liked by different users here we have calculated the value of any randomly selected domain and its value, considering the single class.
- D'Rec with multiple class in this model we have considered all the domain and categories the user like for.as all the user are multi facet it means that the user can like other domains too.
- Probabilistic Matrix Factorization (PMF)PMF virtually is a low rank matrix factorization model and assumes that a user generates a rating for an item by adding Gaussian noise to the inner product Rij= (pi)Tqj, where pi ∈Rd and qj∈Rd associate with latent factor vector of user and item [20].
- PMF with Domains (PMF-D). This model takes multiple domains information of items into consideration, so the PMF-D treats different domains independently but has N users in all domains.
- Multiclass Co-Clustering (MCoC) [21]. This method proposes a framework to extend traditional CF by dividing users and items into multiple subgroups. Different with our framework, it views this allocation procedure as a Multiclass Co-Clustering problem.

V. RESULT AND ANALYSIS.

TABLE II. TPERFORMANCE COMPARISONS OF TOP-N RECOMMENDATION ON EPINIONS IN TERMS OF MAP, F1, NDCG.

Methods		PMF	PMF-D	MCoC	DRec-S	DRec-M
N=5	MAP	0.2309	0.7821	0.3647	**2.6983**	1.8552
	F1	0.7648	0.9648	0.5014	**1.1648**	0.5014
	nDCG	0.1999	0.2301	**0.7000**	0.0012	0.6000
N=10	MAP	0.2291	0.7811	0.3603	**4.4450**	1.2289
	F1	0.8710	1.0716	0.4974	**1.2710**	0.4945
	nDCG	0.1199	0.2291	**0.7000**	8.2007E-4	0.6000
N=15	MAP	0.2319	0.7831	0.3600	**5.2627**	1.0535E-4
	F1	0.8935	1.0935	0.4970	**1.2935**	0.4975
	nDCG	0.1999	0.2319	**0.7000**	7.167E-4	0.6000

Performance of Epinions: Table 2 shows the experimental results on epinions dataset with different evaluation metrices: MAP,F-measure, nDCG, when we vary no of returned items n=5,10,15. For the three variants of DRec methods, we pick the value of regularization parameter λ=100 then we empirically set the number to nearest neighbors as 5, the value of newton parameter τ =0.01.From table 2 DRec-Single gives best result in case of MAP and F Measure and Multi co clustering gives best result for nDCG under the evaluation condition. As we compare the PMF and PMF with Domains we can clearly see the domains give better result than plain PMF it shows that's the multiple domains do benefits on recommendation task. DRec which allocates users into their interested domains by user clustering and topic mining, outperforms PMF-D, which simply assumes users are belonged to all domains

Performance on Ciao: Ciao is smaller dataset as compared to the the above datatset with few ratings.in case of ciao the multi co clustering, and PMF with Domain gives good result as shown in table 3. Whereas the DRec Single and Multiple performances degrades.Although the Drec with single domain is somewhat god if the number of domains are increased.

TABLE III. PERFORMANCE COMPARISONS OF TOP-N RECOMMENDATION ON CIAO IN TERMS OF MAP, F1, NDCG.

Methods		PMF	PMF-D	MCoC	DRec-S	DRec-M
N=5	MAP	0.5176	0.7826	**1.3586**	0.0499	0.8424
	F1	**1.1648**	0.9648	0.4998	0.8983	0.0941
	nDCG	0.6000	**0.7648**	0.7000	1.3156E-4	0.0024
N=10	MAP	0.5692	0.7811	**1.7586**	0.0134	0.8262
	F1	**1.2710**	1.0710	0.4973	1.2599	0.0265
	nDCG	0.6000	**0.8710**	0.7000	9.0583D-5	0.0072
N=15	MAP	0.2319	0.7831	**2.0517**	0.0242	0.8392
	F1	1.2935	1.0935	0.4875	**1.3179**	0.0471
	nDCG	0.6000	**0.8352**	0.7000	7,8308E-5	0.0045

VI. RELATED WORK

In this section we have studied the existing recommendation systems that have used collaborative filteringtechnique. CF is categorized into two types memory based and model based [22, 23]. Memory based technique is easy to implement as well as widely used by many system but it has certain problems like it has limited scalability for large dataset and works poorly in sparse data. Giving poor result.

In contrast the model based approach uses the machine learning and mathematical concepts. There are numerous model based approaches like PCA, Clustering, PLSA, Bayesian, sparse etc. the latent factorization modeling technique give good results for the sparse data, and is one of the most popular algorithm which gives best result [24].

Recently the users are not considered as a single entity for finding out their interest .instead it is studied that even though user are multifaceted their choices can be predicted with the help of social network such as flipkart, ciao etc. user can have similar taste with other person, and can even depend on the choice of their friend and family in the network. Various algorithms are their which uses trustnetwork for recommending items to the user [25]. This solves the sparsity problems of the dataset.

VII. ConClusion

The architecture for DRec is highly modular and enables using various algorithms under the business knowledge layer.

We have designed an interface for entering business rules that can be used for explicit user feedback. For entering the strength of the rules we have introduced the expectancy concept, which is a way of representing condense of both positive and negative rules. We have analyzed methods for combining multiple rules that can be applied to a single entity or entity pairs.

The design concepts were verified by implementing a prototype that was adapted to datasets from various domains. We have used 3 ways of measuring recommendation accuracy and used it for our prototype implementation.

ACKNOWLEDGMENT

The review has been partially supported by Department of Computer Science and IT Dr. BabasahebAmbedkarMarathwada University, Aurangabad. The views expressed here are those of the authors only.

REFERENCES

[1] Karypis, George. "Evaluation of item-based top-n recommendation algorithms."Proceedings of the tenth international conference on Information and knowledge management.ACM, 2001.

[2] Asanov, Daniar. "Algorithms and methods in recommender systems." Berlin Institute of Technology, Berlin, Germany (2011)

[3] Survey Paper on Recommendation System Muktakohar,ChhaviRana Department of computer science and Engg U.I.E.T, MDU,Rohtak .

[4] Gunawardana, Asela, and Guy Shani. "A survey of accuracy evaluation metrics of recommendation tasks." The Journal of Machine Learning Research 10 (2009): 2935-2962.

[5] Cover, T.,and Hart, P., Nearest neighbor pattern classification. Information Theory, IEEE Transactions on, 13(1):21–27, 1967

[6] Miranda, Tim, et al. "Combining content-based and collaborative filters in an online newspaper." In Proceedings of ACM SIGIR Workshop on Recommender Systems. 1999.

[7] Almazro, Dhoha, et al. "A survey paper on recommender systems." arXiv preprint arXiv:1006.5278 (2010).

[8] Pazzani, Michael J., and Daniel Billsus. "Content-based recommendation systems."The adaptive web.Springer Berlin Heidelberg, 2007.325-341.

[9] Sarwar, Badrul, et al. "Item-based collaborative filtering recommendation algorithms." Proceedings of the 10th international conference on World Wide Web.ACM, 2001.

[10] Das, A. S., Datar, M., Garg, A., &Rajaram, S. (2007, May). Google news personalization: scalable online collaborative filtering. In Proceedings of the 16th international conference on World Wide Web (pp. 271-280).ACM.

[11] Van Meteren, Robin, and Maarten Van Someren."Using content-based filtering for recommendation." Proceedings of the Machine Learning in the New Information Age: MLnet/ECML2000 Workshop. 2000.

[12] Ungar, L. H., and Foster, D. P. (1998) Clustering Methods for Collaborative Filtering. In Workshop on Recommender Systems at the 15th National Conference on Artificial Intelligence.

[13] Su, Xiaoyuan, and Taghi M. Khoshgoftaar. "A survey of collaborative filtering techniques." Advances in artificial intelligence 2009 (2009): 4.

[14] Massarani, Leonardo C. "Content-indexing search system and method providing search results consistent with content filtering and blocking policies implemented in a blocking engine." U.S. Patent No. 6,336,117. 1 Jan. 2002.

[15] Y. Koren, R. Bell, and C. Volinsky, “Matrix Factorization Techniques for Recommender Systems,” Computer, vol. 42, no. 8, pp. 30–37, Aug. 2009.

[16] Zhang, Xi, et al. Toprec: domain-specific recommendation through community topic mining in social network. Proceedings of the 22nd international conference on World Wide Web. International World Wide Web Conferences Steering Committee, 2013.

[17] S. J. Nowlan and G. E. Hinton, “Simplifying Neural Networks by Soft Weight-sharing,” Neural Comput., vol. 4, no. 4, pp. 473–493, Jul. 1992.

[18] A. Mnih and R. Salakhutdinov, “Probabilistic matrix factorization,” in Advances in neural information processing systems, 2007, pp. 1257–1264.

[19] K. Goldberg, T. Roeder, D. Gupta, and C. Perkins, “Eigentaste: A constant time collaborative filtering algorithm,” Information Retrieval, vol. 4, no. 2, pp. 133–151, 2001.

[20] B. Xu, J. Bu, C. Chen, and D. Cai. An exploration of improving collaborative recommender systems via user-item subgroups. In Proceedings of the 21st international conference on World Wide Web, pages 21–30, 2012.

[21] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl. Item-based collaborative filtering recommendation algorithms. In Proceedings of the 10th international conference on World Wide Web, pages 285–295, 2001.

[22] J. L. Herlocker, J. A. Konstan, A. Borchers, and J. Riedl.An algorithmic framework for performing collaborative filtering. In Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval, pages 230–237, 1999.

[23] X. Su and T. M. Khoshgoftaar. A survey of collaborative filtering techniques. Advance in Artificial Intelligence, pages 4:2–4:2, 2009.

[24] J. Tang, H. Gao, and H. Liu. mtrust: discerning multi-faceted trust in a connected world. In Proceedings of the fifth ACM international conference on Web search and data mining, pages 93–102, 2012.

[25] Y. Zhang, B. Cao, and D.-Y. Yeung. Multi-domain collaborative filtering. In Proceedings of the Twenty-Sixth conference on Uncertainty in artificial intelligence, pages 725–732, 2010.