Tools and Techniques for solving cold start Recommendation*

Nitin Mishra†
Rajeev Gandhi Memorial College of Engineering and Management
Nandyal, Andra pradesh, India
drnitinmishra10@gmail.com

Vimal Mishra‡
Director IERT Allahabad
Allahabad, Uttar Pradesh
vimal.mishra.upte@gmail.com

Saumya Chaturvedi§
saumyanmishra5@gmail.com

ABSTRACT

Recommendation Systems are very important systems that saves users time and resources by saving them from searching the bulk data. The best example is googling which searches and gives list of hundreds of pages. Therefore, a major challenge of Recommendation Systems can be how to make recommendations for a new user, that is called cold-start user problem in this papers we are trying to identify different kinds of cold start problems in Recommendation Systems. We are also trying to explore different types of solutions to these problems in last 10 years. are very important systems that saves users time and resources by saving them from searching the bulk data. The best example is googling which searches and gives list of hundreds of pages. Therefore, a major challenge of Recommender systems can be how to make recommendations for a new user, that is called cold-start user problem in this papers we are trying to identify different kinds of cold start problems recommender systems. We are also trying to explore different types of solutions to these problems in last 10 years.

CCS CONCEPTS

• **Information systems → Similarity measures**;

KEYWORDS

Recommendation System,Cold start problem, content based methods,collaborative filtering

ACM Reference Format:
Nitin Mishra, Vimal Mishra, and Saumya Chaturvedi. 2017. Tools and Techniques for solving cold start Recommendation. In *Proceedings of International Conference on Internet of Things and Machine Learning (IML2017)*. ACM, New York, NY, USA, Article 4, 6 pages. https://doi.org/http://dx.doi.org/10.1145/3109761.3109772

*Produces the permission block, and copyright information
†First Author
‡Incharge Guide of Research
§This author did documentation work.

IML2017, 17-18 October 2017, Liverpool, United Kingdom

ACM ISBN 978-1-4503-5243-7/17/10…$15.00
https://doi.org/http://dx.doi.org/10.1145/3109761.3109772

1 INTRODUCTION

The large volume of data makes accessing the useful information from the internet a difficult task. To solve this problem Recommender system were proposed. This is the system that recommends the user most probable items from large pool of items of interest. The important items that are recommended includes videos e.g YouTube, news e.g. newspapers websites, items in e-commerce websites, images etc. A social network is another area where recommendation algorithms are working is a popular way. Job websites are recommending jobseekers the best suitable jobs based on recommendation Algorithms. In the last few years a lot of recommendation algorithms have been proposed. We can divide them into three categories, first can be content based second can be collaborative filtering and third can be combination of other methods which are known as hybrid methods. The content based algorithms mainly focus on items content. They do not care about the users profiles for which these items are recommended. Popular applications can be newspaper websites like jagran.com. These websites make use of content based recommendation techniques to recommend items to its users. In Collaborative filtering techniques a profile of users is built on the basis of his past actions. Most popular Collaborative filtering methods are nearest neighbor search, singular value decomposition and matrix factorization technique. [29] For new items content based methods are better than collaborative filtering methods. So one should use methods that can combine the good properties of both the methods. The new method proposed is known as hybrid method which can recommend without knowing much of a users history. [6] Recommendation systems suffers many problems like scalability, Over Specialization problem, Lack of data, Changing data, Changing user preferences, unpredictable items, Sparsity and Cold-start problem.[24]

2 TYPES OF COLD START PROBLEM

Due to practical impact of this problem this area has attracted a number of researches in last 10 years. This problem is due to lack of data which causes our good working algorithm to fail in certain situations. [24] Cold start problem is one of the very crucial problem in recommendation system. The cold-start problem is of two types these are cold-start items and cold-start users [12]. Cold-start items Problem is caused by new items. For these new items we do not have sufficient ratings from users, so our recommender system has a problem. [1, 10**?**].

The situation of cold start problem happens when we do not know much about the user. In such situations because user is new to the system there is no previous interaction data of the user and system. So recommendation system is not able to make genuine

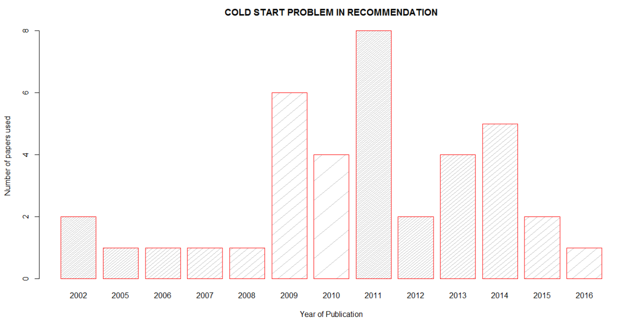

Figure 1: The plot shows the papers taken from each year from 2002-2016.

recommendations.[1]. Sometimes some recommendation system cannot start work. We have discussed content based recommender systems and collaborative filtering based recommender systems. Both of these systems suffer problems of cold start.[4]

Collaborative filtering needs previous records for good recommendation. But we have new items with no previous records and we can have new users without previous records.In collaborative filtering systems we have two kinds of problems. the problem which arises from items is known as cold start item problem and which arises from users is known as cold start user problem. [12] In content based system we can solve this problem in which historical ratings are not there. it does not mean content based systems to not suffer from cold start problem. they are also not unable to recommend new items to the user in absence of history of previous records of their users.[4]. The example can be the Recommendation of article of newspaper to new users. [23]

3 ABOUT THE PAPERS

This review paper has used 39 papers mostly from last 10 years of work. Two important papers are out of our 10-year boundary. All the papers are taken from journals and publications of reputation. Every paper discusses unique approach to solve cold start problem with different probabilities.

4 SOLUTIONS

The given below table gives the important methods used to solve the cold start problem in recommendation system in last 10 years. We have identified almost all the methods that were suitable to solve this problem to some extent.

4.1 Description of various methods

SN	Methods	Description
1	Predictive Feature based regression[18]	The resulting algorithms scale well on linear functions.[18]
2	presented *functional matrix factorization* (fMF)[39]	A decision tree is constructed which is then used for Recommendation.
3	Pure collaborative filtering model, Two way aspect model and naive Bayes Recommenders[27]	In this paper have used the method which is useful in all the three cases of cold start problem.The naive bayes method is helpful in solving cold start user problem and cold start item problem. The method which they have used makes the mapping of profiles flexible.The two aspects model which they have suggested is also useful in certain situations to solve the problem.
4	Representative-based matrix factorization[14]	They identified representative users and items. Then used matric factorization.
5	Interview process algorithm[30]	Algorithm asks questions from users to gather data and that data was used for recommendation.
6	Attribute aware matrix factorization model.[5]	In this method they have used attributes of entity (which is user or item) and used as latent pictures of matrix factorization model. then they have done training on Matrix factorization model by standard techniques. this method perform well on parameters of prediction accuracy as well as time taken in computation. This method is performing well in new item cold start problem.
7	context-aware semi-supervised co-training method named C-SEL[36]	In this paper Authors have proposed a supervised training method C-SEL. Then they have used matrix factorization to get user item context. For learning they have used a semi supervised training method. The algorithm first creates a weak prediction model from the existing data. Then these prediction are supervised by other predictions models. This method has advantage over other methods. First it does more accurate modelling. It supports supervised and semi supervised modelling technique to learn from unlabelled data.

8	n-dimensional Markov random field[21]	In this paper authors have proposed Markov random field method. in this method that attributes such as age occupation genre or release year is considered important . We create Markov random profile of users and items. The latent profile determines the attitude of the user to an item: if the latent profiles of both users and items are very similar, then the response is very strong, otherwise weak.This method is very effective in solving all the three types of cold start problems.
9	Collaborative filtering method prediction[29]	This method uses collaborative filtering of other users for prediction of new users. this method works well when lot of related users are found for new user all items. But it fails when a unique kind of user or item arives.
10	Combination of well-known classifiers[6]	Uses combination of well-known classifiers to classify the new item and users.
11	Filter bots[19]	Proposed filterbots to generate ratings. The ratings were based on user and item attributes. Different filterbots were proposed.
12	Pairwise Regression Model[18]	In this method user and items were taken as pair. Outer product of the attributes of user, item pair generates the ratings. Regression based approach is used on this.
13	Regression based matrix factorization model[2]	This approach is based on a model that predict response as a multiplicative function of row and column latent features that are estimated through regressions. Then accurate scalable fitting methods are used to solve the problem of cold start.
14	A monte carlo Expectation maximization Algorithm[24]	Uses famous monte carlo algorithm variant to solve cold start problem.
15	Topic model(fLDA)[3]	In this method the system is proposed which runs the content of item as words and learn topics over words and creates latent profile of the item. then this model is optimised by MCEM method. experimental results performed in this paper shows that this method performs better then other models in the item cold start problem. The disadvantage of this approach is that due to the methods used it becomes very slow. The method is very time taking also.
16	Attribute Mapping model[3]	It is based on bayesian personalized rankings.
17	Bayesian personalized Ranking[22]	In this method implicit feedback is taken for items from users. Then they had suggested criteria BPR from maximum posterior estimator. This criteria generates optimal personalized rankings. then they have proposed a learning algorithm which maximizes these rankings.
18	Functional matrix factorization[39]	In this paper they have suggested approach of matrix factorization. Matrix factorization technique construction interview together initial data from the user. Then this method construct a tree which changes not according to the answer given by the user.
19	Multiple Question Decision tree Model[39]	AThis approach is an extension of previous work discussed in 18th here also they create a multiple question decision tree then this is compared with Matrix factorization model. this model decides which note to move according to the response chosen by the user. This is a very good approach but only problem with this is that user sometimes do not take interest in giving genuine information. They are in a hurry and they choose anything which biases the recommender systems. .
20	Semi Supervised Discriminative selection model[30]	This method also takes interview and takes input from the user. but the process of creation of interview is different. this method uses tree like methods discussed in 19 and 20. The users are directed towards the child nodes of the tree according to their answer and its weight represented in their method.This method can detect emotions playing a role in selection of options.

21	Active Learning method[37]	In this paper they have discussed semi-supervised discriminative selection model. they claim that this model mines important valuable items to construct the query. this method also utilizes semi-supervised scheme with graph regularization and expert guidance. In their experiment this method outperforms top and measures of cold start recommendation.
22	Local Collective Embedding[7]	This paper they are using friend recommender method. They are using to step process to make more accurate recommendations. In the first step on the basis of similarity, a list of possible friends is generated. In the second step they use the assumption that friend of friend is also a friend. using this approach they do co-clustering from the previous list and generate the final list of friends. Using their choices cold start problem is solved.
23	Transfer Learning[9, 13, 26, 34, 35]	As the name suggest this method is associated with transfer of Information and learning from that information from multiple sources. We have information about users and we have information about data. in most of the solutions which are using this method we try to get both user and item knowledge in egg from axillary data sources through principled matrix based learning framework from heterogeneous data.
24	Transfer Knowledge [8, 16] (without overlapping)	In this method also We take knowledge from other domains. But in this method overlapping is not done. This method also gives good results as discussed in the papers.
25	Collective Knowledge[11]	In this method data from different collaborative filtering domains(System, data and Temporal) is taken and shared using different knowledge transfer styles (Rating-Pattern, Latent feature and Domain Correlating Styles)
26	Rating matrix generative model for effective cross domain filtering.[30]	This method finds relation between different rating matrixes by finding cluster level rating matrix. This can be further extended to a cluster level rating model.
27	Tag-induced Cross-Domain Collaborative Filtering (TagCDCF)[11, 38]	This method makes use of your user generated tags to attain similarity between different domains. We do collaborative filtering using these tags in different domains. the advantage of this method is that we do not need exact users and items in different domains. hence we can solve the cold start problem even if we are not able to find the same user or item in other domains.
28	Combination of Classification and similarity methods[28]	In this method widely known classification algorithms in combination with similarity techniques and prediction mechanisms provide the necessary means for retrieving recommendations
29	Error reflected models[1, 10]	In this methods models are generated from explicit ratings and then collaborative filtering recommender systems model are applied. This method predict the ratings and then identifies the error in each predicted rating. using this error information and precomputed models is collectively known as error reflected model. They have used this error reflected models to make new predictions.
30	Enhanced content-based algorithm using social networking (ECSN)[1, 17]	This method makes use of content based algorithm and social networking techniques . We can get information of similar users using social networking techniques. then we can use content based algorithms or on the items created by them. The prediction ratings generated using this method solve the cold start problem.

31	Community based information .[25]	In this paper they have used community information about the missing user in the system to solve cold start problem of recommendation system. The limitation of this method is that it cannot be used for every system. External data needs to be accessed and needs permission sometimes.This method was suggested in 1997 but it is still used today for solving cold start problem.
32	Heuristic Similarity measure [4]	This method uses similarity with existing users and use the data of similar users to solve the problem.
33	Content based recommendation[20]	This method uses content based where there is problem of collaborative filtering cold start is there.
34	Cold start problem with sparse ratings[15]	They have used combination of clustering and unique algorithm designed by them to solve this problem
35	Deep Learning[32]	In this paper they have suggested deep learning based method to solve cold start problem in recommendation system. They have merged deep learning with collaborative filtering.
36	Bibliometrics and Fuzzy Linguistic Modeling[31]	A fuzzy linguistic approach that using bibliometrics aids to soft or remove the necessity of interaction of users providing them with personalized profiles built beforehand, thus reducing the cold start problem.
37	Ranker Ensemble Method[33]	Making use of interaction profiles and content data, a hybrid of ranking algorithms is proposed to optimize the required objectives. This solution was able to achieve 8th position during A/B testing, and some of the ideas can be useful in a more complex ensemble.

5 CONCLUSION

From the above review we can reach the conclusion that many techniques can be useful for solving cold start problem in recommendation system. some of the techniques can be used in some special domains While the other techniques can be extended to many domains. Some technique important are collaborative filtering Matrix factorization and generating latent profiles of users. Some models which are useful for solving a linear regression[2] and tree based models[24]. In recent years some techniques using deep learning fuzzy linguistic Modelling and ranker Ensembler method are also producing very good results. Choice of method for solving cold start problem depends on the domain, the time which we you can give recommendation system to solve, and the kind of business you are running. After considering all these points we can choose one of the method or combination of other methods to solve cold start problem in any system.

REFERENCES

[1] Gediminas Adomavicius and Alexander Tuzhilin. 2005. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE transactions on knowledge and data engineering* 17, 6 (2005), 734–749.

[2] Deepak Agarwal and Bee-Chung Chen. 2009. Regression-based Latent Factor Models. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '09)*. ACM, New York, NY, USA, 19–28. https://doi.org/10.1145/1557019.1557029

[3] Deepak Agarwal and Bee-Chung Chen. 2010. fLDA: Matrix Factorization Through Latent Dirichlet Allocation. In *Proceedings of the Third ACM International Conference on Web Search and Data Mining (WSDM '10)*. ACM, New York, NY, USA, 91–100. https://doi.org/10.1145/1718487.1718499

[4] Hyung Jun Ahn. 2008. A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem. *Information Sciences* 178, 1 (2008), 37–51.

[5] Zeno Gantner, Lucas Drumond, Christoph Freudenthaler, Steffen Rendle, and Lars Schmidt-Thieme. 2010. Learning Attribute-to-Feature Mappings for Cold-Start Recommendations. In *Proceedings of the 2010 IEEE International Conference on Data Mining (ICDM '10)*. IEEE Computer Society, Washington, DC, USA, 176–185. https://doi.org/10.1109/ICDM.2010.129

[6] Asela Gunawardana and Christopher Meek. 2009. A Unified Approach to Building Hybrid Recommender Systems. In *Proceedings of the Third ACM Conference on Recommender Systems (RecSys '09)*. ACM, New York, NY, USA, 117–124. https://doi.org/10.1145/1639714.1639735

[7] Neil Houlsby, José Miguel Hernández-Lobato, and Zoubin Ghahramani. 2014. Cold-start active learning with robust ordinal matrix factorization. In *International Conference on Machine Learning*. 766–774.

[8] Shangrong Huang, Jian Zhang, Shiyang Lu, and Xian-Sheng Hua. 2015. Social Friend Recommendation Based on Network Correlation and Feature Co-Clustering. In *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval (ICMR '15)*. ACM, New York, NY, USA, 315–322. https://doi.org/10.1145/2671188.2749325

[9] Meng Jiang, Peng Cui, Fei Wang, Qiang Yang, Wenwu Zhu, and Shiqiang Yang. 2012. Social recommendation across multiple relational domains. In *Proceedings of the 21st ACM international conference on Information and knowledge management*. ACM, 1422–1431.

[10] Heung-Nam Kim, Abdulmotaleb El-Saddik, and Geun-Sik Jo. 2011. Collaborative error-reflected models for cold-start recommender systems. *Decision Support Systems* 51, 3 (2011), 519–531.

[11] Bin Li, Qiang Yang, and Xiangyang Xue. 2009. Transfer Learning for Collaborative Filtering via a Rating-matrix Generative Model. In *Proceedings of the 26th Annual International Conference on Machine Learning (ICML '09)*. ACM, New York, NY, USA, 617–624. https://doi.org/10.1145/1553374.1553454

[12] Blerina Lika, Kostas Kolomvatsos, and Stathes Hadjiefthymiades. 2014. Facing the cold start problem in recommender systems. *Expert Systems with Applications* 41, 4 (2014), 2065–2073.

[13] Jixiong Liu, Jiakun Shi, Wanling Cai, Bo Liu, Weike Pan, Qiang Yang, and Zhong Ming. 2017. Transfer Learning from APP Domain to News Domain for Dual Cold-Start Recommendation. In *CEUR Workshop Proceedings*. 38.

[14] Nathan N. Liu, Xiangrui Meng, Chao Liu, and Qiang Yang. 2011. Wisdom of the Better Few: Cold Start Recommendation via Representative Based Rating Elicitation. In *Proceedings of the Fifth ACM Conference on Recommender Systems (RecSys '11)*. ACM, New York, NY, USA, 37–44. https://doi.org/10.1145/2043932.2043943

[15] Nitin Mishra, Saumya Chaturvedi, Vimal Mishra, Rahul Srivastava, and Pratibha Bargah. 2017. Solving Sparsity Problem in Rating-Based Movie Recommendation System. In *Computational Intelligence in Data Mining*. Springer, 111–117.

[16] Orly Moreno, Bracha Shapira, Lior Rokach, and Guy Shani. 2012. TALMUD: Transfer Learning for Multiple Domains. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management (CIKM '12)*. ACM, New York, NY, USA, 425–434. https://doi.org/10.1145/2396761.2396817

[17] Rahul Srivastava nitin mishra, saumya chaturvedi. 2014. Research Challenges in Recommendation System. In *Proceedings of the CONIAPS (CONIAPS '14)*. IIIT Jabalpur, Jabalpur MP, India, 39–46.

[18] Seung-Taek Park and Wei Chu. 2009. Pairwise Preference Regression for Cold-start Recommendation. In *Proceedings of the Third ACM Conference on Recommender Systems (RecSys '09)*. ACM, New York, NY, USA, 21–28. https://doi.org/10.1145/1639714.1639720

[19] Seung-Taek Park, David Pennock, Omid Madani, Nathan Good, and Dennis DeCoste. 2006. NaïVe Filterbots for Robust Cold-start Recommendations. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '06)*. ACM, New York, NY, USA, 699–705. https://doi.org/10.1145/1150402.1150490

[20] Michael J Pazzani and Daniel Billsus. 2007. Content-based recommendation systems. In *The adaptive web*. Springer, 325–341.

[21] Furong Peng, Jianfeng Lu, Yongli Wang, Richard Yi-Da Xu, Chao Ma, and Jingyu Yang. 2016. N-dimensional Markov random field prior for cold-start recommendation. *Neurocomputing* 191 (2016), 187 – 199. https://doi.org/10.1016/j.neucom.2015.12.099

[22] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian personalized ranking from implicit feedback. In *Proceedings of the twenty-fifth conference on uncertainty in artificial intelligence*. AUAI Press, 452–461.

[23] Francesco Ricci, Lior Rokach, and Bracha Shapira. 2011. Introduction to recommender systems handbook. In *Recommender systems handbook*. Springer, 1–35.

[24] Yu Rong, Xiao Wen, and Hong Cheng. 2014. A Monte Carlo algorithm for cold start recommendation. In *Proceedings of the 23rd international conference on World wide web*. ACM, 327–336.

[25] Shaghayegh Sahebi and William W Cohen. 1997. Community-based recommendations: a solution to the cold start problem. *Proceedings of WOODSTOCK'97* (1997).

[26] Martin Saveski and Amin Mantrach. 2014. Item Cold-start Recommendations: Learning Local Collective Embeddings. In *Proceedings of the 8th ACM Conference on Recommender Systems (RecSys '14)*. ACM, New York, NY, USA, 89–96. https://doi.org/10.1145/2645710.2645751

[27] Andrew I. Schein, Alexandrin Popescul, Lyle H. Ungar, and David M. Pennock. 2002. Methods and Metrics for Cold-start Recommendations. In *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '02)*. ACM, New York, NY, USA, 253–260. https://doi.org/10.1145/564376.564421

[28] Yue Shi, Martha Larson, and Alan Hanjalic. 2011. Tags as bridges between domains: Improving recommendation with tag-induced cross-domain collaborative filtering. In *International Conference on User Modeling, Adaptation, and Personalization*. Springer, 305–316.

[29] Xiaoyuan Su and Taghi M. Khoshgoftaar. 2009. A Survey of Collaborative Filtering Techniques. *Adv. in Artif. Intell.* 2009, Article 4 (Jan. 2009), 1 pages. https://doi.org/10.1155/2009/421425

[30] Mingxuan Sun, Fuxin Li, Joonseok Lee, Ke Zhou, Guy Lebanon, and Hongyuan Zha. 2013. Learning Multiple-question Decision Trees for Cold-start Recommendation. In *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining (WSDM '13)*. ACM, New York, NY, USA, 445–454. https://doi.org/10.1145/2433396.2433451

[31] Alvaro Tejeda-Lorente, Juan Bernabé-Moreno, Carlos Porcel, and Enrique Herrera-Viedma. 2017. Using Bibliometrics and Fuzzy Linguistic Modeling to Deal with Cold Start in Recommender Systems for Digital Libraries. In *Advances in Fuzzy Logic and Technology 2017*. Springer, 393–404.

[32] Jian Wei, Jianhua He, Kai Chen, Yi Zhou, and Zuoyin Tang. 2017. Collaborative filtering and deep learning based recommendation system for cold start items. *Expert Systems with Applications* 69 (2017), 29–39.

[33] Murat Yagci and Fikret Gurgen. 2017. A Ranker Ensemble for Multi-objective Job Recommendation in an Item Cold Start Setting. In *Proceedings of the Recommender Systems Challenge 2017*. ACM, 2.

[34] Ming Yan, Jitao Sang, Tao Mei, and Changsheng Xu. 2013. Friend transfer: Cold-start friend recommendation with cross-platform transfer learning of social knowledge. In *Multimedia and Expo (ICME), 2013 IEEE International Conference on*. IEEE, 1–6.

[35] Ming Yan, Jitao Sang, and Changsheng Xu. 2015. Unified youtube video recommendation via cross-network collaboration. In *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval*. ACM, 19–26.

[36] Mi Zhang, Jie Tang, Xuchen Zhang, and Xiangyang Xue. 2014. Addressing Cold Start in Recommender Systems: A Semi-supervised Co-training Algorithm. In *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '14)*. ACM, New York, NY, USA, 73–82. https://doi.org/10.1145/2600428.2609599

[37] Xi Zhang, Jian Cheng, Ting Yuan, Biao Niu, and Hanqing Lu. 2013. Semi-supervised Discriminative Preference Elicitation for Cold-start Recommendation. In *Proceedings of the 22Nd ACM International Conference on Information & Knowledge Management (CIKM '13)*. ACM, New York, NY, USA, 1813–1816. https://doi.org/10.1145/2505515.2507869

[38] Zi-Ke Zhang, Chuang Liu, Yi-Cheng Zhang, and Tao Zhou. 2010. Solving the cold-start problem in recommender systems with social tags. *EPL (Europhysics Letters)* 92, 2 (2010), 28002.

[39] Ke Zhou, Shuang-Hong Yang, and Hongyuan Zha. 2011. Functional Matrix Factorizations for Cold-start Recommendation. In *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '11)*. ACM, New York, NY, USA, 315–324. https://doi.org/10.1145/2009916.2009961