Meta-path Selection for Extended Multi-Relational Matrix Factorization

Fatemeh Vahedian, Robin Burke, Bamshad Mobasher
Center for Web Intelligence, Depaul University, Chicago, IL 60604
{fvahedia,rburke,mobasher}@cs.depaul.edu

Abstract

Multi-relational matrix factorization is an effective technique for incorporating heterogenous data into prediction tasks, such as personalized recommendation. Recent research has extended the set of relations that can be applied within heterogeneous network settings by composing non-local relations using network meta-paths. One of the key problems in applying this technique is that the set of possible non-local relations is essentially unbounded. In this paper, we demonstrate that an information gain based technique for heuristic pruning of relations can enhance the performance of multi-relational matrix factorization recommenders.

Introduction

Multi-relational factorization is a learning technique applicable when there is a target relation to be learned in the context of multiple associated relations (Gantner et al. 2010; Drumond et al. 2014). For example, in predicting scientific collaborations, there may be co-authorship relations between individuals contributing to papers and articles, but there may also be relations of common membership in organizations, of common venues of publication, of citation, etc. The multi-relational approach allows the full variety of such relations to be applied to control the learning of the target relation.

Recent work has extended the locally-oriented multi-relational approach to include relations composed from multi-step typed network paths. For example, in the scientific publication area, it may be useful to include a relation from an author to the papers cited by papers written by his or her co-authors. This author-paper relation, which does not appear directly in the data, can be composed from existing relations: author-paper and paper-citation. Following research in heterogeneous networks (Sun and Han 2012), our approach conceives of each relation as a typed edge in a heterogeneous network, and the composition of these relations as the expansion of *meta-paths* within the network.

Researchers have shown that incorporating extended relations based on meta-paths improves the accuracy of recommender systems based on multi-relational matrix factorization(Vahedian, Burke, and Mobasher 2015). However, this approach raises additional computational questions: namely, how to limit the set of relations to consider and how to balance accuracy gains against training time. The set of meta-path expansions in a heterogeneous network is unbounded, but not all such relations contribute to improved performance. In addition, there are significant computational costs in incorporating a large number of relations.

In this paper, we show that a metric based on information gain can be used to prune relations before the multi-relational learning step and that pruned models exhibit improved recommendation accuracy.

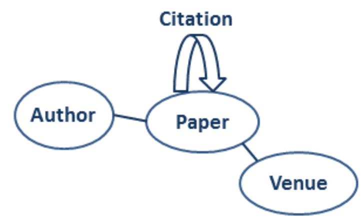

Figure 1: Network schema for the DBLP dataset

Meta-path-based relations

A heterogeneous network is a network with multiple types of nodes (for example, movies and actors) and multiple types of edges (for example, an "acted-in" relation between an actor and a movie, and a "directed-by" relation between a director and a movie). Edge types are defined by the types of nodes that they connect. Such networks are extremely common in social media applications: consider LinkedIn's users, employers, interest groups, educational institutions, job postings, posts, comments, etc. as just one example. A *network schema* is a high-level view of a heterogeneous network showing the node types and edge types.

Figure 1 shows the schema for the scientific publication network, DBLP, used in this paper with nodes for authors, papers, and venues where papers are published. There are edges between authors and papers they have written, and between papers and the venues in which they have appeared. In addition, there is a self-loop: citations in linking papers to each other.

A meta-path is a sequence of edges through the network schema – a sequence of edge types. Traversing a meta-path on a heterogeneous network means following all edges of a given type from a node to all possible successors. For example, consider a meta-path with a single edge type $author - paper$ within the fragment of the DBLP dataset shown in Figure 2. If we start with the author $R.B$ and follow this meta-path, we will arrive at a set of destination nodes: {P1, P2, P3}. In our examples, we typically denote an edge type with the initials of the beginning and ending node types. The meta-path in our example would be the ap meta-path. For clarity, we will use the letter c to refer to the paper destination node when that node is reached via a citation edge, as opposed to an author or venue edge.

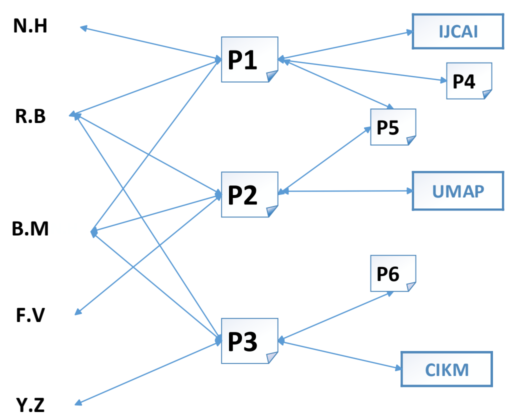

Figure 2: DBLP dataset meta-path example

With the meta-path formalism, it is possible to examine a wide variety of relations. For example, the $author - paper - author - paper$ ($apap$) meta-path encodes a relation between an author and a paper, based on shared co-authors. The relation includes all papers written by any of the original author's co-authors. It can be viewed as a kind of profile expansion based on connections through the network.

Prior work has demonstrated that meta-paths of various lengths could contribute to a multi-component weighted hybrid in a variety of network settings (Burke and Vahedian 2013; Burke, Vahedian, and Mobasher 2014; Vahedian and Burke 2014; Vahedian 2014). However, as should be clear from the discussion above, there is no inherent limit to the number of meta-paths that can be constructed. Limiting the set of relations is a key goal and was explored in the context of weighted hybrid recommendation in (Burke, Vahedian, and Mobasher 2014).

Multi-relational matrix factorization

Multi-relational matrix factorization computes latent factors used to map users to items that they will prefer, but instead of using only rating profiles, as in standard factorization schemes such as (Koren 2008), it allows for the creation of additional latent factors based on other *auxiliary relations* (Drumond et al. 2014). (Gantner et al. 2010) uses a movie recommendation domain in which user-movie is the target relation (the system is predicting movies that users will prefer) and the auxiliary relations are movie-genre, movie-director, movie-actors, and movie-credits (a superset of actor, director and other key roles, such as cinematographer).

In the multi-relational matrix factorization model DMF from (Gantner et al. 2010), different latent feature models are defined for each relation. Parameters are learned from the factorization process in such a way that they are optimized for the best performance on each relation individually. The DMF model associates one latent feature vector model with each relation r. Different feature matrices (shown as $\Phi_{t,r}$ in the equation below) are associated with each relation r for different target relations. The DMF loss function decomposes over the target relation and each component can be optimized independently of each other. For example, the parameter learning model for the UM relation as a target relation with R auxiliary relations is calculated as

$$(\varphi^*(A), \Phi^*_{AB}) = argmin \\ L_{AB}(D_{AB}, \hat{y}_{AB}(.; \varphi_{AB}(A), \Phi_{AB}))+ \\ \sum_{i=1}^{R} \alpha_{AB,MP_i} L_{MP_i}(D_{MP_i}, \hat{y}_{AB,MP_i}(.; \varphi_{AB}(U), \Phi_{AB,MP_i})) \tag{1}$$

where φ is the set of model parameters and $y_{AB}(.; \varphi)$ is the prediction model for relation AB parameterized with φ. In this model the functions $\hat{y}_{AB,MP_i}$ form the auxiliary reconstruction of relation MP_i when the AB relation is the target relation. L_{AB} is the loss function and $\alpha_{AB,x}$ the importance of relation x when AB is predicted such that $0 < \alpha_{a,b} < 1$ and $\alpha_{a,a} = 1$. Each MP_i is an auxiliary relation and R denotes the number of such relations. Essentially, the model is optimizing the AB relation while treating the other relations as regularization parameters.

The CATSMF model (Drumond et al. 2014) aims to improve the efficiency of the DMF model when applied to multiple targets. Since the DMF model must learn parameters for each relation individually, the number of parameters to be learned grows by a factor of number of relations in the network. In order to deal with this problem, CATSMF limits the parameters needed for the auxiliary relations by coupling them together. It also enables the learning of interactions between the different auxiliary relations.

There is no inherent restriction on the types of relations that can be incorporated as auxiliary relations. This creates an attractive opportunity to integrate extended relations built from meta-paths. However, like other factorization techniques, DMF and CATSMF work best over sparse relations. When relations are dense (as happens with longer meta-paths), the learning process slows greatly: dense relations add many more calculations and also many more constraints. Incorporating additional relations has a similar effect – this problem being the key motivation for the development of CATSMF. It is also the case that the indiscriminate addition of relations may lead to overfitting and decreased accuracy.

Controlling meta-path generation

As we have seen, relation generation is in principle unbounded: nodes and edges can be revisited, as seen in a relation like $author-paper-author-paper$. In addition to the problems of factorization for dense relations and overfitting, there are significant computational costs in generating numerous relations in large networks. Still, prior research has shown that the inclusion of some longer meta-paths can significantly improve recommendation accuracy. It is therefore important to control this process – ideally, we would like to be able to estimate in advance what relations are likely to make a substantial contribution and include only those components.

To achieve this goal, we use an information gain computation for each meta-path to estimate the amount of information that is available through it. Intuitively, the idea is that each incorporated relation should add information to the recommendation calculation. If it is redundant or overly noisy, then it can be omitted from the model.

Consider a relation AB created by following some meta-path that begins with A and ends with B, and containing an arbitrary number of composed relations in between. We wish to compute the information to be gained starting from A and arriving at B via this meta-path. We will represent this value as the information gain $IG(A, B)$ and compute it as follows: $IG(A, B) = H(A) - H(A|B)$ where $H(A)$ is entropy of dimension A and $H(A|B)$ is the conditional entropy of A given B.

$H(A)$ is the entropy of entity type A. It is a function of the probability of the items in A: $H(A) = -\sum_i p(a_i)log(p(a_i))$. To associate probabilities with each node a_i in A, we use the property of graphs that the probability of encountering a node on a random walk is, in the limit, proportional to its degree. Therefore, the probability of node a_i relative to other nodes in dimension A is calculated as $p(a_i) = Degree(a_i)/\sum_{n\in A} Degree(n)$

Conditional entropy measures the uncertainty of one dimension, given another dimension. Considering an AB projection of a network, we make use of a two-dimensional matrix to calculate the probability of dimension B given A ($P(B|A)$) as follows. The probability of reaching node b in dimension B is calculated as the fraction of meta-path expansions from node a leading to node b out of all possible expansions from node a reaching some node in B. The conditional probability is therefore calculated as

$$P(b|a) = \frac{\#path(a \to b)}{\sum_{n\in B} \#path(a \to n)}$$

For example, consider the $author - paper - citation - paper - citation$ meta-path and the AC projection relation derived from it. We calculate the overall entropy of the author dimension using the degrees of the author nodes. Then we calculate all meta-path expansions for the meta-path and count how many times, for each author, a particular paper is encountered by following this expansion. With this information, we can calculate $H(A|C)$ for this meta-path. Note that this value could be quite different than what would be calculated for a different meta-path connecting the same entities: $author - paper - citation$, for example. If the values for $H(A)$ and $H(A|C)$ are roughly the same, then $IG(A, C)$ will be near zero, suggesting that this particular meta-path does not add much information beyond what is already contained in A.

Normalized information gain

Normalization of the information gain measure is essential because we are comparing relations with different starting and ending entity types. We compute the normalized information gain (NIG) value for two entity types in the network A and B, as the information gain value is normalized by the square root of the product of the entropies of the two dimensions: $NIG(A, B) = IG(A, B)/\sqrt{H(A)H(B)}$

Dataset

DBLP is a database of bibliographic information on computer science journals and proceedings. We augmented this with additional citation data[1], and explored two different recommendation tasks: venue recommendation and citation recommendation.

As a preliminary step, we removed papers with fewer than 15 citations. The nodes in this network are author, paper, and venue,[2] and the direct links used for this dataset are $author - paper$, $paper - venue$ and $paper - citation$ as seen in Figure 1.

The venue recommendation task involves recommending venues of possible publication to authors. In order to build this recommendation model, we generate author profiles based on the $author - paper - venue$ (apv) meta-path, then select an author who has published in at least five different venues.

The second recommendation task we explored is citation recommendation. For this model, we assume that an author has written a paper (or is considering writing a paper) and the task is recommending a list of relevant citations. The target relation for this recommendation task is the $paper - citation$ (pc) path.

Experiments and methodology

For each dataset, the target relation was randomly partitioned into 80% training and 20% test data. Relations were generated from the training data starting with the direct relations used in the basic DMF model and adding two-step and three-step meta-paths starting from the first entity of the target relation. These are DMF, DMF2 and DMF3 in our figures.

We built multi-relational factorization models for each collection of relations using the implementations of CATSMF and DMF made available by the authors of (Drumond et al. 2014). [3] This implementation is self-contained and requires no external parameter setting other choosing an optimization criterion. We chose Bayesian Personalized

[1]Citation-network V1 from http://aminer.org/billboard/citation

[2]We intend to include title and abstract information in future research.

[3]http://ismll.de/catsmf/mrFac.tar.gz

Ranking as the optimization criterion (BPR-opt), as described in (Drumond et al. 2014). For all recommendation models, we evaluated recall and precision on recommendation lists of length one through ten.

In addition to the DMF models generated based on meta-paths, we calculated the normalized information gain measure to find the most informative meta-paths for those models. In order to test normalized information gain as a pruning heuristic for DMF, the relations found to be least informative in terms of NIG were removed from the most inclusive DMF model for each task. Figure 3 shows the meta-paths included in each model for the venue recommendation task. The striped areas indicate relations successively pruned from the DMF-IG1, DMF-IG2, and DMF-IG3 models. Figure 4 shows the meta-paths included for citation recommendation.

Note that for the purpose of this paper we did not generate CATSMF models based on extended meta-paths. In our preliminary work, we found the CATSMF model using two step paths did not show any accuracy improvement. Due to the increased density of the two-step and three-step models, CATSMF is also extremely slow in computing models using these relations. So, we do not view CATSMF as a practical alternative for constructing models using extended meta-paths.

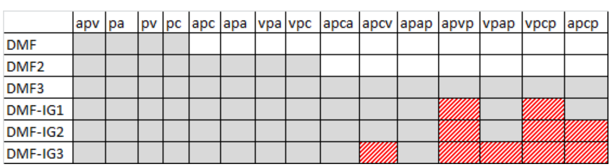

Figure 3: Vnue recommendation: relations and factorization models

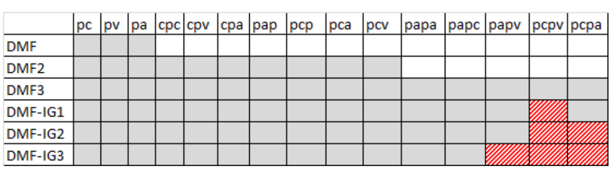

Figure 4: Citation recommendation: relations and factorization models

Results and discussion

The results from the datasets confirms earlier findings that including relations derived from extended meta-paths in heterogeneous networks enhances the performance of recommendation models in both recall and precision. We also see that we can, with some reliability, estimate the importance of each relation using the normalized information gain metric. Removing paths with the lowest information gain does not harm the performance of the factorized model and can improve performance.

Venue recommendation

Figure 6 shows precision and recall results for the four algorithm variants shown in Figure 3. The important finding is that the versions of the DMF algorithm that incorporate both two-step and three-step meta-paths demonstrate improvements in both precision and recall, with the best performance found in the model incorporating two-step paths, DMF2. Surprisingly, the model incorporating three-step extended relations (DMF3) shows a major decrease in performance compared to DMF2, although still showing better recall and precisions than the original DMF and CATSMF models. One reason for the decline in performance could be the redundancy of the relations encountered in the three-step paths. For example, the venue profile generated based on $venue-paper-author-paper$ and $venue-paper-citation-paper$ could be very similar because both of them generate a VP relation which may reflect a similar set of papers for each venues. We are still exploring the cause of this finding.

We computed the normalized information gain for all the components as shown in Figure 5. The 3-step meta-paths do show lower NIG values except for $apap$ and $apca$. The $apap$ meta-path is the author-paper relation in which the author is linked to all the paper written by his or her co-authors, regardless of his or her own authorship. It is logical that the venues related to these papers would be of interest to the target author. The $apca$ relation is an author-author relation in which the author is related to the authors of papers that he or she has cited. This also makes sense in this recommendation context, although it is interesting that this relation is more informative than the one formed by the $apcv$ meta-path, which more directly links cited papers to venues.

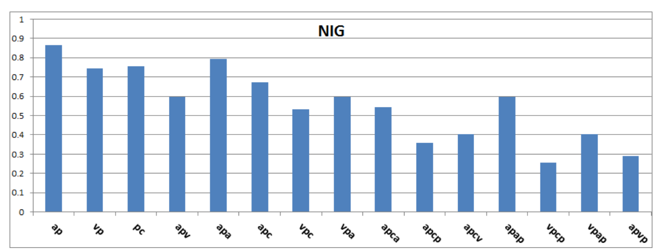

Figure 5: Venue recommendation: NIG value for relations

To test meta-path selection using NIG, we generated three different models, as shown in Figure 3, with different levels of pruning: the relations corresponding to the striped cells were filtered out. The best performing model is DMF-IG3 which removes 5 of the three-step relations. DMF-IG1 in which only two relations are removed performs similarly to DMF2 and shows enhancement over DMF3, which includes all meta-paths. The NIG measure is effective here in filtering out longer meta-paths, without losing accuracy. The DMF-IG2 shows only a slight decrease in performance compared to DMF2.

Citation recommendation

The precision and recall curve for citation recommendation is shown in Figure 8 for the five algorithm variants

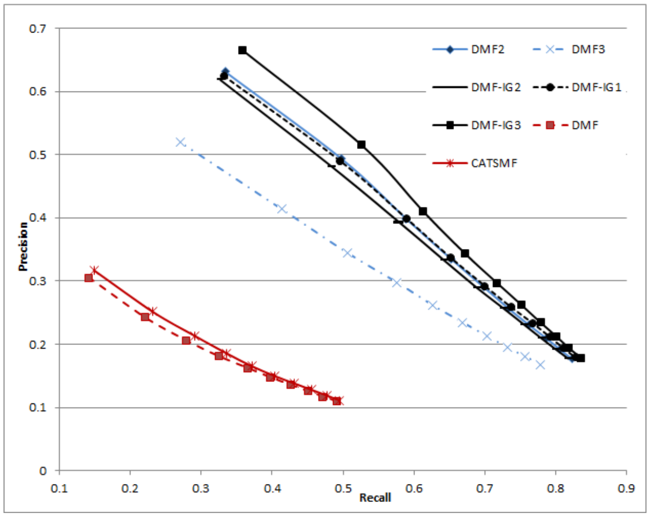

Figure 6: Venue recommendation: recall vs. precision

shown in Figure 4 including twelve meta-path driven relations. The same pattern as venue recommendation can be seen: the DMF2 model including two-step meta-paths is superior. However, the decrease in performance for DMF3 is small compared to the venue recommendation case.

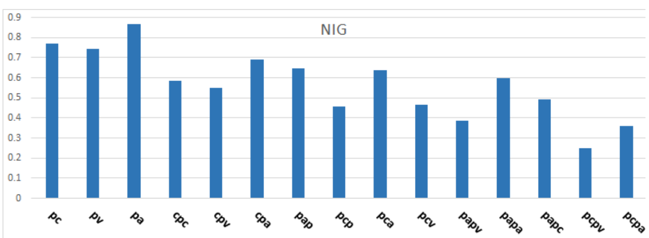

Figure 7: Citation recommendation: NIG value for relations

Figure 7 shows the distribution of NIG values for the different relations used in this task. Of the 3-step relations, $papa$ and $papc$ have the higher information gain values. The $papc$ meta-path is retrieving papers cited by the papers of co-authors. It is reasonable to expect these will be good recommendations for citations. The $papa$ suggests using authors one step removed in the co-authorship graph: co-authors of co-authors. This suggests a certain amount of intermural co-citation behavior, which is also to be expected.

Three models are built using the meta-path filtering technique. DMF-IG1, which filters out only one relation, performs exactly the same as DMF3. Despite the lower NIG values, removing additional low-information-gain meta-paths causes a slight decline in performance compared to DMF3. All variants are still significantly improved over both the DMF and CATSMF models.

Related Work

Recommender systems based on complex networks have been studied extensively in recent years (Durão and Dolog 2009; Song, Zhang, and Giles 2011)See the multi-network approach of (Chen et al. 2012), the layered approach of (Kazienko, Musial, and Kajdanowicz 2011), and the linear weighted hybrid model (Gemmell et al. 2012; Burke, Vahedian, and Mobasher 2014).

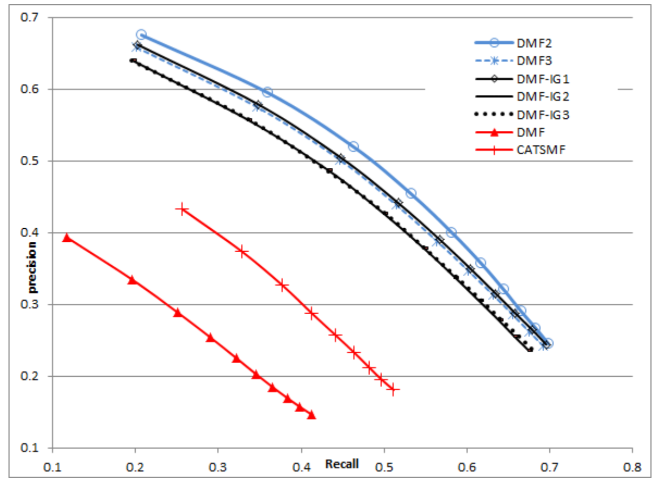

Figure 8: Citation recommendation: recall vs. precision

Although there is a great deal of research in link prediction (Kunegis and Lommatzsch 2009; Benchettara, Kanawati, and Rouveirol 2010), we find this work less compelling as a basis for recommender systems. Link prediction research does not typically take a user-centered view of the task, and considerations such as diversity and personalization are absent. Also, these approaches are typically designed for homogeneous networks incorporating a single relation type. Note in particular that the venue recommendation task explored here cannot be formulated as a link prediction problem because there are no edges between authors and venues in the original network.

In addition to network-oriented techniques, a separate thread of research has developed in multi-relational matrix factorization to make predictions for highly correlated data. Singh and Gordon (Singh and Gordon 2008) proposed collective matrix factorization, as a model of pairwise relation data. Coupled matrix factorization and tensor factorization can extend the multi-relational model to deal with higher arity relations as shown in (Acar, Kolda, and Dunlavy 2011).

Normalized mutual information has been studied as a measure for feature selection in different areas (Fleuret 2004; Peng, Long, and Ding 2005). In the recommendation domain, it has been used to predict the contribution of recommendation components in a weighted hybrid recommendation model (Vahedian and Burke 2014).

Conclusion

Recommendation using multi-relational matrix factorization in networked data can be enhanced through in the inclusion of extended relations derived from meta-path expansions. However, experiments have also shown that including longer meta-paths is computationally expensive in both path generation and factorization and may, at times, decrease the performance of the final recommendation model. We have demonstrated here that meta-path selection using normalized information gain enables us to choose relations with the greatest likelihood of improving the model's recommendations. Our results on two recommendation tasks wtihin the DBLP dataset show that it is possible to prune the relations in a multi-relational model without losing significant accuracy, and in some cases, with improved accuracy.[4]

While the normalized information gain measure has proved useful, it is clear that there are aspects of the relation pruning problem that it does not address. The information gain associated with a meta-path is a function of the network structure alone. However, we know that the recommendation task has a strong influence on what relations will be useful. Also, normalized information gain treats each meta-path as independent, when clearly this is not the case. In our future work, we will examine variations of this metric to take recommendation task and relation redundancy into account.

Acknowledgments

This work was supported in part by the National Science Foundation under Grant No.IIS-1423368 (Multi-dimensional Recommendation in Complex Heterogeneous Networks).

References

Acar, E.; Kolda, T. G.; and Dunlavy, D. M. 2011. All-at-once optimization for coupled matrix and tensor factorizations. *CoRR* abs/1105.3422.

Benchettara, N.; Kanawati, R.; and Rouveirol, C. 2010. A supervised machine learning link prediction approach for academic collaboration recommendation. In *Proceedings of the fourth ACM conference on Recommender systems*, 253–256. ACM.

Burke, R., and Vahedian, F. 2013. Social web recommendation using metapaths. In *Proceedings of the 5th Workshop on Recommender Systems and the Social Web*.

Burke, R. D.; Vahedian, F.; and Mobasher, B. 2014. Hybrid recommendation in heterogeneous networks. In *User Modeling, Adaptation, and Personalization - 22nd International Conference, UMAP 2014, Aalborg, Denmark, July 7-11, 2014. Proceedings*, 49–60.

Chen, J.; Chen, G.; Zhang, H. L.; Huang, J.; and Zhao, G. 2012. Social recommendation based on multi-relational analysis. In *International Conference on Web Intelligence and Intelligent Agent Technology*, 471–477.

Drumond, L. R.; Diaz-Aviles, E.; Schmidt-Thieme, L.; and Nejdl, W. 2014. Optimizing multi-relational factorization models for multiple target relations. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, CIKM '14, 191–200. New York, NY, USA: ACM.

[4]We have replicated these findings with other datasets, but have not included these results for reasons of space.

Durão, F. A., and Dolog, P. 2009. A personalized tag-based recommendation in social web systems. In *Proceedings of International Workshop on Adaptation and Personalization for Web 2.0*.

Fleuret, F. 2004. Fast binary feature selection with conditional mutual information. *J. Mach. Learn. Res.* 5:1531–1555.

Gantner, Z.; Drumond, L.; Freudenthaler, C.; Rendle, S.; and Schmidt-Thieme, L. 2010. Learning attribute-to-feature mappings for cold-start recommendations. In *Data Mining (ICDM), 2010 IEEE 10th International Conference on*, 176–185. IEEE.

Gemmell, J.; Schimoler, T.; Mobasher, B.; and Burke, R. 2012. Resource recommendation in social annotation systems: A linear-weighted hybrid approach. *Journal of Computer and System Sciences* 78(4):1160–1174.

Kazienko, P.; Musial, K.; and Kajdanowicz, T. 2011. Multidimensional social network in the social recommender system. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on* 41(4):746–759.

Koren, Y. 2008. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, 426–434. ACM.

Kunegis, J., and Lommatzsch, A. 2009. Learning spectral graph transformations for link prediction. In *Proceedings of the 26th Annual International Conference on Machine Learning*, ICML '09, 561–568. New York, NY, USA: ACM.

Peng, H.; Long, F.; and Ding, C. 2005. Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* 27(8):1226–1238.

Singh, A. P., and Gordon, G. J. 2008. Relational learning via collective matrix factorization. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '08, 650–658. New York, NY, USA: ACM.

Song, Y.; Zhang, L.; and Giles, C. L. 2011. Automatic tag recommendation algorithms for social recommender systems. *ACM Transactions on the Web* 5(1):4.

Sun, Y., and Han, J. 2012. *Mining Heterogeneous Information Networks: Principles and Methodologies*. Synthesis Lectures on Data Mining and Knowledge Discovery. Morgan & Claypool Publishers.

Vahedian, F., and Burke, R. D. 2014. Predicting component utilities for linear-weighted hybrid recommendation. In *Proceedings of the 6th Workshop on Recommender Systems and the Social Web*.

Vahedian, F.; Burke, R. D.; and Mobasher, B. 2015. Network-based extension of multi-relational factorization models. In *Poster Proceedings of the 9th ACM Conference on Recommender Systems, RecSys 2015, Vienna, Austria, September 16, 2015.*

Vahedian, F. 2014. Weighted hybrid recommendation for heterogeneous networks. In *Eighth ACM Conference on Recommender Systems, RecSys '14*, 429–432.