HINMF: A Matrix Factorization Method for Clustering in Heterogeneous Information Networks

Jialu Liu Jiawei Han
University of Illinois at Urbana-Champaign, Urbana, IL
{jliu64, hanj}@illinois.edu

Abstract

Non-negative matrix factorization (NMF) has become quite popular recently on the relational data due to its several nice properties and connection to probabilistic latent semantic analysis (PLSA). However, few algorithms take this route for the heterogeneous networks. In this paper we propose a novel clustering method for heterogeneous information networks by searching for a factorization that gives compatible clustering solutions across multiple sub-networks. Moreover, we develop an automatic weight learning strategy in order to balance the effects of different sub-networks brought to the consensus. Experimental results on real-world dataset demonstrate the effectiveness of our approach.

1 Introduction

Many real-world graph data can be represented as an information network, which composes of nodes interconnected with each other via meaningful links. It is also common that in real networks multiple types of nodes are connected by multiple types of links, forming *heterogeneous information networks* (HIN). For example, networked data extracted from bibliographic information contains multi-typed objects including authors, papers, terms and venues. Various types of links are formed between these different types of objects, e.g., author writes paper, venue publishes paper and paper is represented by a bag of terms.

Clustering on such heterogeneous information networks aims at partitioning the multi-typed objects into clusters such that nodes within the same cluster have more connections than those between different clusters. Efforts have been devoted to developing effective clustering approaches to characterize the heterogeneity property of information networks recently [Sun *et al.*, 2009; Deng *et al.*, 2011]. However, few algorithms are taking the route of matrix factorization which has been shown to be a highly effective approach for homogeneous or bipartite network data [Koren *et al.*, 2009; Xu *et al.*, 2003]. Some existing matrix factorization work for HIN is not designed directly for clustering. For example, [Wang *et al.*, 2011] proposed JNMF to solve the collaborative filtering task. The main challenge of applying the matrix factorization technique to HIN is how to limit the search of factorizations to those give meaningful and comparable clustering solutions across different types of nodes simultaneously.

In this paper, we propose a novel clustering approach based on non-negative matrix factorization (NMF) [Lee and Seung, 1999] for HIN following *star schema* defined in [Sun *et al.*, 2009]. Specifically, we follow the principle that the true clustering structure should be hidden as a common shared clustering solution for the *center type* and get propagated to other *attribute* types. To achieve this, we propose to formulate a regularized joint optimization framework to derive factorizations with the constraint that factors of the center type learnt from different sub-networks should be regularized toward a common consensus. Moreover, we develop an automatic weight learning strategy in order to balance the effects of these different sub-networks brought to the consensus.

2 Model Formulation

Assume that we are given a HIN comprised of bipartite networks between the shared center type and T attribute types[1]. Let $\{X^{(1)}, \ldots, X^{(t)}, \ldots, X^{(T)}\}$ denote the non-negative edge weight matrix for all sub-networks where each column of $X^{(t)}$ denotes a node of the center type and each row represents a node of the attribute type. For each bipartite graph $X^{(t)} \in \mathbb{R}^{M^{(t)} \times N}$, we wish to find low-ranked factorizations $U^{(t)} \in \mathbb{R}^{M^{(t)} \times K}$ and $V^{(t)} \in \mathbb{R}^{N \times K}$ such that $X^{(t)} \approx U^{(t)}(V^{(t)})^T$. Here $M^{(t)}$ is the number of nodes of the attribute type t, N is the number of nodes of the center type and K is the number of clusters we wish to discover.

One of the common reconstruction processes can be formulated as a set of *Frobenius norm* optimization problems for each individual sub-network, defined as:

$$\min_{U^{(t)}, V^{(t)}} ||X^{(t)} - U^{(t)} {V^{(t)}}^T||_F^2, \ s.t. \ U^{(t)} \geq 0, V^{(t)} \geq 0$$

where $||\cdot||_F$ is the Frobenius norm and $U^{(t)} \geq 0$, $V^{(t)} \geq 0$ represent the constraints that all the matrix elements are non-negative. However, such formulation assumes each sub-network to be independent and fails to model the HIN in a unified way.

[1] An HIN can be decomposed into a set of bipartite networks once it follows star schema.

Algorithm 1 NMF for Heterogeneous Information Network

Input: HIN $\{X^{(1)}, \ldots, X^{(t)}, \ldots, X^{(T)}\}$, parameters α, number of clusters K

Output: Clustering on both center type and attribute types

Normalize each view $X^{(t)}$ such that $||X^{(t)}||_1 = 1$
Initialize $U^{(t)}$,$V^{(t)}$, V^* and $\beta^{(t)}$ $(1 \leq t \leq T)$
repeat
  for $t = 1$ **to** T **do**
    repeat
      Fixing V^*, $\beta^{(t)}$ and $V^{(t)}$, update $U^{(t)}$ by Eq. 4
      Compute $Q^{(t)}$ as in Eq. 2
      Normalize $U^{(t)}$ and $V^{(t)}$ as in Eq. 5
      Fixing V^*, $\beta^{(t)}$ and $U^{(t)}$, update $V^{(t)}$ by Eq. 6
    until Eq. 9 converges.
  end for
  Fixing $U^{(t)}$ and $V^{(t)}$ $(1 \leq t \leq T)$, update V^* and $\beta^{(t)}$ by Eqs. 7 and 8.
until Eq. 3 converges.
Cluster nodes of center type indicated by $\arg\max_k V^*_{j,k}$.
For each attribute type t, cluster nodes of this type indicated by $\arg\max_k U^{(t)}_{i,k} \sum_j V^*_{j,k}$. [3]

We instead propose a co-regularization framework by incorporating the disagreement between $V^{(t)}$ learnt from sub-networks and their consensus V^* as follows:

$$\min_{U^{(t)}s, V^{(t)}s, V^*} \sum_{t=1}^{T} \|X^{(t)} - U^{(t)}(V^{(t)})^T\|_F^2 + \alpha \sum_{t=1}^{T} \|V^{(t)}Q^{(t)} - V^*\|_F^2 \tag{1}$$
$$s.t.\ \forall 1 \leq t \leq T, U^{(t)} \geq 0, V^{(t)} \geq 0, V^* \geq 0$$

where $Q^{(t)}$ is a diagonal matrix defined as follows:

$$Q^{(t)} = Diag\left(\sum_{i=1}^{M^{(t)}} U^{(t)}_{i,1}, \sum_{i=1}^{M^{(t)}} U^{(t)}_{i,2}, \ldots, \sum_{i=1}^{M^{(t)}} U^{(t)}_{i,K}\right) \tag{2}$$

We multiply $V^{(t)}$ by $Q^{(t)}$ before computing the disagreement to ensure the comparability of $V^{(t)}$ learnt from different sub-networks. We notice that w.l.o.g. that $||X^{(t)}||_1 = 1$, [2]

$$||X^{(t)}||_1 \approx \sum_{k=1}^{K} ||U^{(t)}_{\cdot,k} \times Q^{(t)}_{k,k} / Q^{(t)}_{k,k} \sum_j V^{(t)}_{j,k}||_1 = ||V^{(t)}Q^{(t)}||_1.$$

Additionally, we use α as a fixed parameter tuning the weight between standard NMF reconstruction error and the disagreement term. Since both the "energies" of $X^{(t)}$ and $V^{(t)}Q^{(t)}$ are under control, this parameter is not much sensitive so throughout the experiment we set it to be 0.1.

Note that until now we treat all bipartite sub-networks of the whole HIN to be equally important. Take the relative weights of different sub-networks into consideration, we develop an automatic weight learning strategy via introducing a set of parameters $\beta^{(t)}$ into Eq. 1:

$$\min_{U^{(t)}s, V^{(t)}s, V^*, \beta^{(t)}s} \sum_{t=1}^{T} \beta^{(t)} \Big(\|X^{(t)} - U^{(t)}(V^{(t)})^T\|_F^2 + \alpha \|V^{(t)}Q^{(t)} - V^*\|_F^2] \Big) \tag{3}$$
$$s.t.\ \forall 1 \leq t \leq T, \quad U^{(t)} \geq 0, V^{(t)} \geq 0, V^* \geq 0,$$
$$\sum_t exp^{-\beta^{(t)}} = 1$$

The reason for the constraint represented in the exponential form instead of a naive linear combination is to avoid trivial solution that completely favours the sub-network with the least reconstruction error.

To solve the above optimization problem, we propose an iterative optimization algorithm. The following two steps are iterated until convergence: (1) fix V^* and $\beta^{(t)}$, minimize Eq. 3 over $U^{(t)}$ and $V^{(t)}$, and (2) fix $U^{(t)}$ and $V^{(t)}$, minimize Eq. 3 over V^* and $\beta^{(t)}$. This two-step procedure is summarized in Algorithm 1. The iterative procedure converges to a local minimum. Here we give the equations for $U^{(t)}$, $V^{(t)}$, V^* and $\beta^{(t)}$ respectively:

$$U^{(t)}_{i,k} \leftarrow U^{(t)}_{i,k} \frac{(X^{(t)}V^{(t)})_{i,k} + \alpha \sum_{j=1}^{N} V^{(t)}_{j,k} V^*_{j,k}}{(U^{(t)}{V^{(t)}}^T V^{(t)})_{i,k} + \alpha \sum_{i=1}^{M^{(t)}} U^{(t)}_{i,k} \sum_{j=1}^{N} {V^{(t)}_{j,k}}^2} \tag{4}$$

$$U^{(t)} \leftarrow U^{(t)}{Q^{(t)}}^{-1},\ V^{(t)} \leftarrow V^{(t)}Q^{(t)} \tag{5}$$

$$V^{(t)}_{j,k} \leftarrow V^{(t)}_{j,k} \frac{({X^{(t)}}^T U^{(t)})_{j,k} + \alpha V^*_{j,k}}{(V^{(t)}{U^{(t)}}^T U^{(t)})_{j,k} + \alpha V^{(t)}_{j,k}} \tag{6}$$

$$V^* \leftarrow \frac{\sum_{t=1}^{T} \beta^{(t)} V^{(t)} Q^{(t)}}{\sum_{t=1}^{T} \beta^{(t)}} \geq 0 \tag{7}$$

$$\beta^{(t)} \leftarrow -\log \frac{RE^{(t)}}{\sum_t RE^{(t)}} \tag{8}$$

where $RE^{(t)}$ represents the reconstruction error for the bipartite sub-network related to attribute type t:

$$\|X^{(t)} - U^{(t)}(V^{(t)})^T\|_F^2 + \alpha \|V^{(t)}Q^{(t)} - V^*\|_F^2 \tag{9}$$

3 Experiment

In this experiment, we use a subset of the DBLP records that belong to four research areas: artificial intelligence, information retrieval, data mining and database. It contains 4023 authors, 20 venues and 11771 unique terms in total. The first sub-network records the number of papers each author publishes in different venues and the second sub-network contains all the terms each author writes in the abstract of his/her papers with stopwords removed. To demonstrate how the clustering performance can be improved by the proposed approach, we have compared with the following algorithms:

- A-V: We report the clustering performance after running NMF on the author-venue sub-network. We normalize U and V after convergence following [Xu *et al.*, 2003].
- A-T. It is similar to A-V but we turn to use the author-term sub-network.

[2] $||\cdot||_1$ represents the summation of all elements.

[3] This step is inspired from the connection of NMF and PLSA as studied in a multi-view NMF algorithm [Liu *et al.*, 2013].

Table 1: Clustering performance on DBLP dataset (%)

Method	AC(%)		NMI(%)	
	Author	Venue	Author	Venue
A-V	92.35	**100.0**	77.12	**100.0**
A-T	77.24	-	47.28	-
NetClus	90.86	**100.0**	73.51	**100.0**
HINMF	**94.07**	**100.0**	**80.67**	**100.0**

- NetClus [Sun *et al.*, 2009]. It is a rank-based algorithm proposed recently to integrate ranking and clustering together in HIN with star schema.
- HINMF. Our proposed method in this paper.

The clustering results regarding authors and venues are evaluated by comparing the obtained label using clustering algorithms with that provided by the dataset. The accuracy (AC) and normalized mutual information (NMI) [Xu *et al.*, 2003] are used to measure the performance.

Table 1 shows the clustering performance of different algorithms on this dataset. As we can see, HINMF outperforms its competitors and successfully integrates the complementary information from different sub-networks. NetClus was originally designed for the paper-centered HIN and we here apply it directly to this author-centered schema, which could be the reason it fails to perform better than the NMF running on single author-venue sub-network.

Besides the evaluation on authors and venues, we list the top ten words for each cluster k by sorting $U_{i,k}^{(2)}$ in Table 2. Similar to Line 15 in Algorithm. 1, this is inspired from the connection of NMF and PLSA [Liu *et al.*, 2013].

Table 2: Top 10 words in different clusters.

Cluster 1	Cluster 2	Cluster 3	Cluster 4
learning	retrieval	mining	data
based	information	data	database
knowledge	web	clustering	query
problem	search	based	queries
model	query	patterns	xml
algorithm	based	frequent	system
approach	document	large	databases
systems	text	efficient	systems
system	language	databases	based
reasoning	model	classification	processing

We have also conducted experiments on parameter α as shown in Fig. 1. It can be observed that the performance is not much sensitive with respect to different values of α. Thus through the experiment, we set it to be 0.1.

For β, Fig. 2 shows its variation w.r.t. number of iterations. It is quite interesting that initially $\beta^{(1)}$ related to author-venue sub-network is slightly larger than $\beta^{(2)}$ and the former soon decreases significantly. A possible reason is that during the first several iterations, factorizations learnt on author-term sub-network get trapped in the local optimum and results in a large reconstruction error. By later incorporating the knowledge from author-venue sub-network, it gets out of that local minimum and succeeds in searching for better solutions in terms of the reconstruction error.

Fig. 3 shows the convergence curve together with its performance. The black solid line shows the value of the objective function and the red dashed line indicates the accuracy.

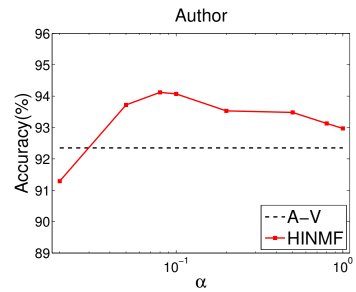

Figure 1: Performance of HINMF w.r.t. varying α

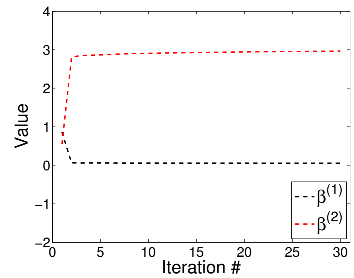

Figure 2: Value of β w.r.t. increasing iterations

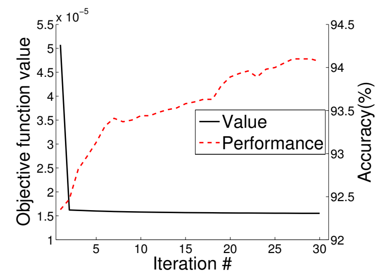

Figure 3: Convergence and corresponding performance curve

4 Acknowledgement

This work was supported in part by the U.S. National Science Foundation grants IIS-0905215, U.S. Army Research Laboratory under Cooperative Agreement No. W911NF-09-2-0053 (NS-CTA).

References

[Deng *et al.*, 2011] H. Deng, J. Han, B. Zhao, Y. Yu, and C. Lin. Probabilistic topic models with biased propagation on heterogeneous information networks. In *KDD*, pages 1271–1279, 2011.

[Koren *et al.*, 2009] Y. Koren, R. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, 2009.

[Lee and Seung, 1999] D. Lee and S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, 1999.

[Liu *et al.*, 2013] J. Liu, C. Wang, J. Gao, and J. Han. Multi-view clustering via joint nonnegative matrix factorization. In *SDM*, 2013.

[Sun *et al.*, 2009] Y. Sun, Y. Yu, and J. Han. Ranking-based clustering of heterogeneous information networks with star network schema. In *KDD*, pages 797–806, 2009.

[Wang *et al.*, 2011] F. Wang, T. Li, X. Wang, S. Zhu, and C. Ding. Community discovery using nonnegative matrix factorization. *Data Mining and Knowledge Discovery*, 22(3):493–521, 2011.

[Xu *et al.*, 2003] W. Xu, X. Liu, and Y. Gong. Document clustering based on non-negative matrix factorization. *SIGIR*, pages 267–273, 2003.