Simultaneous Clustering of Multi-Type Relational Data via Symmetric Nonnegative Matrix Tri-Factorization

Hua Wang
Department of Computer
Science and Engineering
University of Texas at
Arlington, TX 76019, USA
huawangcs@gmail.com

Heng Huang
Department of Computer
Science and Engineering
University of Texas at
Arlington, TX 76019, USA
heng@uta.edu

Chris Ding
Department of Computer
Science and Engineering
University of Texas at
Arlington, TX 76019, USA
chqding@uta.edu

ABSTRACT

The rapid growth of Internet and modern technologies has brought data involving objects of multiple types that are related to each other, called as *multi-type relational data*. Traditional clustering methods for single-type data rarely work well on them, which calls for more advanced clustering techniques to deal with multiple types of data simultaneously to utilize their interrelatedness. A major challenge in developing simultaneous clustering methods is how to effectively use all available information contained in a multi-type relational data set including *inter-type* and *intra-type* relationships. In this paper, we propose a Symmetric Nonnegative Matrix Tri-Factorization (S-NMTF) framework to cluster multi-type relational data at the same time. The proposed S-NMTF approach employs NMTF to simultaneously cluster different types of data using their inter-type relationships, and incorporate the intra-type information through manifold regularization. In order to deal with the symmetric usage of the factor matrix in S-NMTF, we present a new generic matrix inequality to derive the solution algorithm, which involves a fourth-order matrix polynomial, in a principled way. Promising experimental results have validated the proposed approach.

Categories and Subject Descriptors

H.3.3 [**Information Storage And Retrieval**]: Information Search and Retrieval—*Clustering*; I.2.6 [**Artificial Intelligence**]: Learning

General Terms

Algorithms, Experimentation, Performance

Keywords

Clustering, Multi-Type Relational Data, Web, Nonnegative Matrix Tri-Factorization

1. INTRODUCTION

Clustering aims to partition a set of objects into groups or clusters such that objects in a same cluster are similar while those in

different clusters are dissimilar. Most traditional clustering algorithms concentrate on dealing with *homogeneous* data, in which all the objects are of one single type. Recently, the rapid progress of Internet and computational technologies have brought data much richer in structure, involving objects of multiple types that are related to each other. For example, in a Web search system in Figure 1, we have four different types of data objects including words, Web pages, search queries and Web users. Each of the four types of entities have their own attributes. Meanwhile different types of data are also interrelated to each other by various means, *e.g.*, Web pages and words are related via co-occurrences. Due to their high heterogeneity and close interrelationships, these data are called as *multi-type relational data* [10]. The information contained in a multi-type relational data set typically appears in two forms as follows:

- **Inter-type relationships** characterize the relations between data objects from different types, such as the co-occurrences between data objects from different types, as illustrated by the green solid lines in Figure 1.
- **Intra-type relationships** characterize the native relations between objects within one data type, *e.g.*, as illustrated by the blue dashed lines in Figure 1, the internet hyperlinks among Web pages, the pairwise affinities among users that are induced from user attributes, *etc*.

Inter-type relationships appear exclusively in multi-type relational data, while intra-type information, due to its homogeneity, appears in both multi-type relational data and traditional single-type data.

The rich structures of multi-type relational data provide a potential opportunity to improve the clustering accuracy, which, however, also present a new challenge on how to take advantage of all available information. In this paper, we tackle this new, yet important, problem to simultaneously cluster multiple types of relational data. Our goal is to make full use of the both forms of information of a multi-type relational data set.

1.1 Difficulties of Traditional Clustering Methods on Multi-Type Relational Data

Because each type of a multi-type relational data set can be viewed as homogeneous data, one may argue that clustering heterogeneous data can be solved using traditional methods to cluster each type of objects independently. However, this may not work well.

First, various data types in a multi-type relational data set are interrelated to each other. Clustering each data type independently will lose these interaction information, such as that conveyed by internet hyper-links or co-occurrences, which are essential to gain a full understanding of the Web data.

Second, in practice the clustering of a data type may not be the one based on the intra-type relationships (*i.e.*, the similarity of fea-
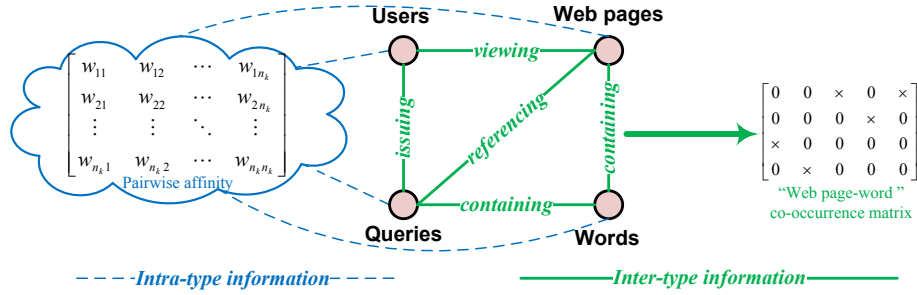
Figure 1: A Web search system is an example of multi-type relational data, involving (1) *Inter-type relationships* (green solid lines): the relations between objects in different data types. (2) *Intra-type relationships* (blue dashed lines): the relations between different objects in a same data type.

tures) but the one based on the inter-type relationships between this data type and others. For example, in the Web search system in Figure 1, instead of the similarity between Web pages themselves, people are often interested in the relationships between user groups and Web page styles.

Last, but not least, heterogeneous data contain different types of objects. Processing and interpreting them in one single way may not be appropriate and presents a major challenge. Ad hoc integration or normalization (*e.g.*, concatenating different features into a vector of fixed length) rarely works.

As a result, recent research has advanced swiftly from simple clustering of one type of data to simultaneous clustering of multiple types of data, for both two data types (pairwise co-clustering) [3, 6, 7, 16] and multiple (more than two) data types (high-order co-clustering) [10, 12, 15]. Through simultaneous clustering, one can discover the hidden global structures in the heterogeneous data, which seamlessly integrates multiple data types to provide a better picture of the underlying data distribution.

1.2 Our Contributions

Given the inter-type relationships and the intra-type information, we present a Symmetric Nonnegative Matrix Tri-Factorization (S-NMTF) framework for simultaneous clustering of multi-type relational data. In our new approach, we use NMTF to simultaneously cluster different types of data upon their inter-type relationship matrices. Meanwhile, users are allowed to provide optional intra-type information for different types of data in form of pairwise affinity, which is incorporated as manifold regularization to NMTF. Promising empirical results in extensive experiments validate the proposed method. We summarize our contributions as following.

1. We present a simple, yet effective, framework to tackle the complicated problem of simultaneous clustering of multi-type relational data, which aims to fully exploit all available information.

2. Existing NMF algorithms usually deal with asymmetric (rectangle data-feature) matrix factorization [4, 6, 12–14] and optimize one factor matrix at a time, which only involve second-order matrix polynomials. When dealing with symmetric matrix factorization, a fourth-order matrix polynomial is involved, which is much harder to be dealt with and heuristics are usually used in existing works. In this paper, as an important theoretical contribution, we present a new generic matrix inequality (in Lemma 4), such that the difficulty is tackled in a principled way.

2. SIMULTANEOUSLY CLUSTER MULTI-TYPE RELATIONAL DATA BY S-NMTF

In this section, we first briefly review NMTF based co-clustering of two-type relational data, from which we gradually develop the proposed S-NMTF approach for simultaneous clustering of multi-

type relational data. Our approach employs both inter-type relationships as well as intra-type information.

2.1 Problem Formalization

We first formalize the problem of simultaneous clustering of multi-type relational data.

We denote a K-type relational data set as $\mathcal{X} = \{\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_K\}$, where $\mathcal{X}_k = \{\mathbf{x}_1^k, \mathbf{x}_2^k, \ldots, \mathbf{x}_{n_k}^k\}$ represents the data objects of the k-th type. Suppose that we are given a set of relationship matrices $\left\{ R_{kl} \in \Re^{n_k \times n_l} \right\}_{(1 \le k \le K, 1 \le l \le K)}$ between different types of data objects, and we have $R_{kl} = R_{lk}^T$. Our goal is to simultaneously partition the data objects in $\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_K$ into $c_1, c_2, \ldots, c_K$ disjoint clusters respectively.

Most existing simultaneous clustering algorithms only rely on inter-type relationships, *i.e.*, R_{ij}. However, in practice we may also have intra-type information for each type of data. For the k-th type of data, for example, we usually have pairwise affinities $W_k \in \Re^{n_k \times n_k}$ between the objects of the type. In this study, we aim to make full use of both inter-type relationships R_{kl} (in Section 2.3) and intra-type information W_k (in Section 2.4).

Throughout this paper, we denote $A_{(ij)}$ as the entry at the i-th row and j-th column of a matrix A.

2.2 A Brief Review of Co-Clustering of Two-Type Relational Data Using NMTF

The close connection between Nonnegative Matrix Factorization (NMF) and clustering [4] provides the potential theory to develop co-clustering methods. Ding *et al.* proposed to use NMTF [6] to simultaneously cluster the rows and columns of an input nonnegative relationship matrix R_{12} by decomposing it into three nonnegative factor matrices, which minimizes the following objective:

$$J_1 = \|R_{12} - G_1 S_{12} G_2^T\|^2, \ s.t. \ G_1 \ge 0, G_2 \ge 0, S_{12} \ge 0, \quad (1)$$

where $\| \cdot \|$ denotes the Frobenius norm of a matrix, $G_1 \in \Re_+^{n_1 \times c_1}$ and $G_2 \in \Re_+^{n_2 \times c_2}$ are the cluster indicator matrices for $\mathcal{X}_1$ and $\mathcal{X}_2$ respectively, and $S_{12} \in \Re_+^{c_1 \times c_2}$ absorbs the different scales of R_{12}, G_1 and G_2. Note that, the original NMF problem [9] requires R_{12} to be nonnegative. In co-clustering scenarios, this constraint (thereby the nonnegativity constraint on S_{12}), however, can be relaxed [5], which leads to the semi-NMTF problem to minimize:

$$J_2 = \|R_{12} - G_1 S_{12} G_2^T\|^2, \quad s.t. \quad G_1 \ge 0, G_2 \ge 0 \ . \quad (2)$$

Simultaneous clustering on $\mathcal{X}_1$ and $\mathcal{X}_2$ is then achieved by solving Eq. (2). Because the rows of the resulted G_k ($k \in \{1, 2\}$) (with normalization) can be interpreted as the posterior probability for clustering on $\mathcal{X}_k$ [4, 6], the cluster label of $\mathbf{x}_i^k$ is obtained by

$$l\left(\mathbf{x}_i^k\right) = \arg \max_j G_{k(ij)} \ . \quad (3)$$

2.3 Simultaneous Clustering of Multi-Type Relational Data Using Inter-Type Relationships

A natural generalization of the co-clustering objective in Eq. (2) to simultaneous clustering of multi-type relational data is to solve the following optimization problem [10, 12]:

$$\min J_3 = \sum_{1 \le k < l \le K} \| R_{kl} - G_k S_{kl} G_l^T \|^2, \tag{4}$$
$$s.t. \quad G_k \ge 0, \ \forall \ 1 \le k \le K \ .$$

However, it is not straightforward to solve Eq. (4) by generalizing existing iterative multiplicative NMTF solution algorithms. Motivated by [3] that deals with bipartite graphs, we propose to solve the optimization problem in Eq. (4) by solving an equivalent Symmetric NMTF problem.

We first introduce the following useful lemma.

LEMMA 1. *The optimization problem in Eq.* (2) *can be equivalently solved by the following S-NMTF problem:*

$$\min J_4 = \| R - GSG^T \|^2, \quad s.t. \quad G \ge 0, \tag{5}$$

in which

$$R = \begin{bmatrix} 0^{n_1 \times n_1} & R_{12}^{n_1 \times n_2} \\ R_{21}^{n_2 \times n_1} & 0^{n_2 \times n_2} \end{bmatrix},$$

$$G = \begin{bmatrix} G_1^{n_1 \times c_1} & 0^{n_1 \times c_2} \\ 0^{n_2 \times c_1} & G_2^{n_2 \times c_2} \end{bmatrix}, \quad S = \begin{bmatrix} 0^{c_1 \times c_1} & S_{12}^{c_1 \times c_2} \\ S_{21}^{c_2 \times c_1} & 0^{c_2 \times c_2} \end{bmatrix}, \tag{6}$$

where the superscripts denote the matrix sizes, and $R_{21} = R_{12}^T$, $S_{21} = S_{12}^T$. $0^{n_1 \times n_1}$ *is a matrix with all zero entries of size* $n_1 \times n_1$.

PROOF. Following the definitions of R, G and S, we can derive

$$\| R - GSG^T \|^2 = \left\| \begin{bmatrix} 0 & R_{12} \\ R_{12}^T & 0 \end{bmatrix} - \begin{bmatrix} 0 & G_1 S_{12} G_2^T \\ G_2 S_{12}^T G_1^T & 0 \end{bmatrix} \right\|^2$$
$$= 2 \| R_{12} - G_1 S_{12} G_2^T \|,$$

which proves the lemma. $\square$

Based upon Lemma 1, we have the following theorem.

THEOREM 1. *It is equivalent to solve Eq.* (4) *and to solve*

$$\min J_5 = \| R - GSG^T \|, \quad s.t. \quad G \ge 0, \tag{7}$$

in which

$$R = \begin{bmatrix} 0^{n_1 \times n_1} & R_{12}^{n_1 \times n_2} & \cdots & R_{1K}^{n_1 \times n_K} \\ R_{21}^{n_2 \times n_1} & 0^{n_2 \times n_2} & \cdots & R_{2K}^{n_2 \times n_K} \\ \vdots & \vdots & \ddots & \vdots \\ R_{K1}^{n_K \times n_1} & R_{K2}^{n_K \times n_2} & \cdots & 0^{n_K \times n_K} \end{bmatrix},$$

$$G = \begin{bmatrix} G_1^{n_1 \times c_1} & 0^{n_1 \times c_2} & \cdots & 0^{n_1 \times c_K} \\ 0^{n_2 \times c_1} & G_2^{n_2 \times c_2} & \cdots & 0^{n_2 \times c_K} \\ \vdots & \vdots & \ddots & \vdots \\ 0^{n_K \times c_1} & 0^{n_K \times c_2} & \cdots & G_K^{n_K \times c_K} \end{bmatrix}, \tag{8}$$

$$S = \begin{bmatrix} 0^{c_1 \times c_1} & S_{12}^{c_1 \times c_2} & \cdots & S_{1K}^{c_1 \times c_K} \\ S_{21}^{c_2 \times c_1} & 0^{c_2 \times c_2} & \cdots & S_{2K}^{c_2 \times c_K} \\ \vdots & \vdots & \ddots & \vdots \\ S_{K1}^{c_K \times c_1} & S_{K2}^{c_K \times c_2} & \cdots & 0^{c_K \times c_K} \end{bmatrix},$$

where $R_{ji} = R_{ij}^T$ *and* $S_{ij} = S_{ji}^T$.

The proof of Theorem 1 can be easily obtained by generalizing the proof of Lemma 1 to multi-type relational data.

Theorem 1 presents a general framework via S-NMTF to simultaneously cluster multi-type relational data using the mutual relationship matrices. However, the symmetric usage of the factor matrix G in NMTF leads to a fourth-order matrix polynomial when deriving the iterative algorithm, which is harder to be dealt with compared to the second-order matrix polynomials involved in standard (asymmetric) NMTF. In Section 3, we will introduce a new matrix inequality in Lemma 4 to solve this problem.

2.4 Incorporating Intra-Type Information via Graph Regularization

The optimization objectives in Eq. (2) and Eq. (7) only involve the inter-type relationships of a multi-type relational data set, whereas the intra-type information, though often available, are not used. We incorporate them through Laplacian regularization [2, 7].

For a two-type relational data set, given the intra-type information in form of the pairwise affinity matrices W_1 and W_2 for $\mathcal{X}_1$ and $\mathcal{X}_2$ respectively, we can incorporate them into Eq. (2) as following:

$$J_6 = \| R_{12} - G_1 S_{12} G_2^T \|^2 + 2\lambda \left[\mathbf{tr} \left(G_1^T L_1 G_1 \right) + \mathbf{tr} \left(G_2^T L_2 G_2 \right) \right],$$
$$s.t. \quad G_1 \ge 0, \ G_2 \ge 0, \tag{9}$$

where D_k is the diagonal degree matrix with $D_{k(ii)} = \sum_j W_{k(ij)}$, and $L_k = D_k - W_k$ is the corresponding graph Laplacian. Because L_k is the discrete approximation of the Laplace-Beltrami operator on the underlying data manifold [1], the last two terms reflects the label smoothness of the two types of data points. The smoother the data labels are with respect to the underlying data manifolds, the smaller their values will be.

Using R, S and G defined in Eq. (8), and denote

$$W = \begin{bmatrix} W_1^{n_1 \times n_1} & 0^{n_1 \times n_2} & \cdots & 0^{n_1 \times n_K} \\ 0^{n_2 \times n_1} & W_2^{n_2 \times n_2} & \cdots & 0^{n_2 \times n_K} \\ \vdots & \vdots & \ddots & \vdots \\ 0^{n_K \times n_1} & 0^{n_K \times n_2} & \cdots & W_K^{n_K \times n_K} \end{bmatrix}, \tag{10}$$

we approach simultaneous clustering on multi-type relational data by solving the following optimization problem:

$$\min J_{\text{S-NMTF}} = \| R - GSG^T \|^2 + 2\lambda \, \mathbf{tr} \left[G^T \left(D - W \right) G \right],$$
$$s.t. \quad G \ge 0, \tag{11}$$

where D is the diagonal degree matrix with $D_{(ii)} = \sum_j W_{(ij)}$. We call Eq. (11) as our proposed S-NMTF approach. As can be seen, the complicated problem of simultaneous clustering of multi-type relational data is finally modeled by a very simple mathematical formulation, which is the first contribution of this work.

3. ALGORITHM FOR S-NMTF

The computational algorithm for the proposed S-NMTF approach is listed in Algorithm 1. Upon solution, the final cluster labels are obtained from the resulted G_k using Eq. (3).

The main challenge to derive Algorithm 1 is the fourth-order matrix polynomial incurred by the symmetric usage of G in Eq. (11). Existing works [4, 6, 12] tackle this difficulty in an intuitive way: solve the problem as a standard NMF problem with two different factor matrices and set them as same in the solution. In this work, as a theoretical contribution, we present a new matrix inequality in Lemma 4 to tackle this difficulty in a principled way.

Algorithm 1: Algorithm to solve S-NMTF in Eq. (11)

   Data: Relationship matrices: $\{R_{ij}\}_{1\le i<j\le K}$
   Unsupervised pairwise affinity matrices: $\{W_k\}_{1\le k\le K}$
   Result: Factor matrices: $\{G_k\}_{1\le k\le K}$
   1. Construct R, G, S as in Eq. (8), and W as in Eq. (10).
   2. Initialize G as in [6].
   repeat
       3. Compute $S = \left(G^T G\right)^{-1} G^T R G \left(G^T G\right)^{-1}$.
       4. Update $G_{ij} \leftarrow G_{ij} \left[\dfrac{(RGS + \lambda WG)_{ij}}{\left(GSG^T GS + \lambda DG\right)_{ij}} \right]^{\frac{1}{4}}$.
   until *Converges*

3.1 Correctness of the Algorithm

The following theorem guarantees the correctness of Algorithm 1.

THEOREM 2. *If the updating rules of G and S in Algorithm 1 converges, the final solution satisfies the KKT optimal condition.*

PROOF. For the objective in Eq. (11), following the standard theory of constrained optimization, we introduce the Lagrangian multipliers 4Λ and minimize the Lagrangian function as follows:

$$L\left(G, S\right) = J_{\text{S-NMTF}} + \textbf{tr}\left(4\Lambda G^T\right), \qquad (12)$$

which gives

$$\frac{\partial L}{\partial G} = -4RGS + 4GSG^T GS - 4\lambda WG + 4DG + 4\Lambda, \quad (13)$$

$$\frac{\partial L}{\partial S} = -2G^T RG + 2G^T GSG^T G. \qquad (14)$$

Fixing G, letting $\partial L/\partial S = 0$, from Eq. (14) we obtain

$$S = \left(G^T G\right)^{-1} G^T R G \left(G^T G\right)^{-1}. \qquad (15)$$

The KKT complementary condition for the nonnegativity of G_{ij} gives $G_{ij}\Lambda_{ij} = 0$. Setting $\partial L/\partial G = 0$, from Eq. (13) we have:

$$\left(-4RGS + 4GSG^T GS - 4\lambda WG + 4DG\right)_{ij} G_{ij} = 0. \quad (16)$$

This is the fixed point relationships that the solution must satisfy.

At convergence, $G^{(\infty)} = G^{(t+1)} = G^{(t)}$, thus we can derive

$$\left(-4RGS + 4GSG^T GS - 4\lambda WG + 4DG\right)_{ij} G_{ij}^4 = 0,$$

which is identical to Eq. (16) and proves Theorem 2. □

3.2 Convergence of the Algorithm

Now, we analyze the convergence of Algorithm 1 using the auxiliary function approach [9].

LEMMA 2. *[9] $Z\left(h, h'\right)$ is an auxiliary function of $F\left(h\right)$ if the conditions $Z\left(h, h'\right) \ge F\left(h\right)$ and $Z\left(h, h'\right) = F\left(h\right)$ are satisfied. [9] If Z is an auxiliary function for F, then F is non-increasing under the update $h^{(t+1)} = \arg\min_h Z\left(h, h'\right)$.*

LEMMA 3. *[6] For any matrices $A \in \Re_+^{n\times n}$, $B \in \Re_+^{k\times k}$, $S \in \Re_+^{n\times k}$ and $S' \in \Re_+^{n\times k}$, and A and B are symmetric, the following inequality holds:*

$$\sum_{ip} \frac{(AS'B)_{ip} S_{ip}^2}{S'_{ip}} \ge \textbf{tr}\left(S^T ASB\right). \qquad (17)$$

As one of our contribution, we prove the following generic matrix inequality to analyze objective functions involving 4-th order matrix polynomials, such as that our S-NMTF objective in Eq. (11).

LEMMA 4. *For any nonnegative symmetric matrices $A \in \Re_+^{k\times k}$ and $B \in \Re_+^{k\times k}$, for $H \in \Re_+^{n\times k}$ the following inequality holds:*

$$\textbf{tr}\left(HAH^T HBH^T\right) \le$$
$$\sum_{ik}\left(\frac{H'AH'^T H'B + H'BH'^T H'A}{2}\right)_{ik} \frac{H_{ik}^4}{H_{ik}'^3}. \qquad (18)$$

PROOF. Let $H_{ik} = H'_{ik}u_{ik}$. The 1st term in RHS of Eq. (18) is

$$\sum_{ik}(H'AH'^T H'B)_{ik}\frac{H_{ik}^4}{H_{ik}'^3} = \sum_{ijkrpq} H'_{jr}A_{rk}H'_{ik}H'_{ip}B_{pq}H'_{jq}u_{jq}^4$$

Now, switching indexes: $i \Leftrightarrow j, p \Leftrightarrow q, r \Leftrightarrow k$, we obtain

$$\sum_{ik}(H'AH'^T H'B)_{ik}\frac{H_{ik}^4}{H_{ik}'^3} = \sum_{ijkrpq} H'_{ik}A_{kr}H'_{jr}H'_{jq}B_{qp}H'_{ip}u_{ip}^4$$

The 2nd term in RHS of Eq. (18) is

$$\sum_{ik}(H'BH'^T H'A)_{ik}\frac{H_{ik}^4}{H_{ik}'^3} = \sum_{ijkrpq} H'_{ip}B_{pq}H'_{jq}H'_{jr}A_{rk}H'_{ik}u_{ik}^4.$$

Now, switching indexes: $i \Leftrightarrow j, p \Leftrightarrow q, r \Leftrightarrow k$, we obtain

$$\sum_{ik}(H'BH'^T H'A)_{ik}\frac{H_{ik}^4}{H_{ik}'^3} = \sum_{ijkrpq} H'_{jq}B_{qp}H'_{ip}H'_{ik}A_{kr}H'_{jr}u_{jr}^4$$

A careful examination of the RHS of the above four equations shows that they are identical except u^4 terms. Thus, the RHS of Eq. (18) is

$$\sum_{ijkrpq} H'_{ip}B_{pq}H'_{jq}H'_{jr}A_{rk}H'_{ik}\frac{u_{ik}^4 + u_{jr}^4 + u_{jq}^4 + u_{ip}^4}{4}. \quad (19)$$

The LHS of Eq. (18) is $\sum_{ijkrpq} H'_{ip}B_{pq}H'_{jq}H'_{jr}A_{rk}H'_{ik}u_{ik}u_{jr}u_{jq}u_{ip}$. For any $a, b, c, d > 0$, we have $a^4 + b^4 + c^4 + d^4 \ge 2(a^2b^2 + c^2d^2) \ge 4(ab)(cd)$, thus $u_{ik}u_{jr}u_{jq}u_{ip} \le \left(u_{ik}^4 + u_{jr}^4 + u_{jq}^4 + u_{ip}^4\right)/4$, which proves Lemma 4. □

Based upon the above lemmas, we prove the following theorem.

THEOREM 3. *Let*

$$J_{\text{S-NMTF}}\left(G\right) = \textbf{tr}\left(-2RGSG^T + GSG^T GSG^T \right.$$
$$\left. +2\Lambda G^T DG - 2\lambda G^T WG\right), \qquad (20)$$

then the following function

$$Z\left(G, G'\right) = -2\sum_{ijkl} G'_{ji}S_{jk}G'_{kl}R_{li}\left(1 + \log\frac{G_{ji}G_{kl}}{G'_{ji}G'_{kl}}\right)$$
$$+ \sum_{ij}\left(G'SG'^T G'S\right)_{ij}\frac{G_{ij}^4}{G_{ij}'^3} + \sum_{ij}\left(DG'\Lambda\right)_{ij}\frac{G_{ij}^4 + G_{ij}'^4}{G_{ij}'^3}$$
$$- 2\lambda\sum_{ijk} G'_{ji}W_{jk}G'_{ki}\left(1 + \log\frac{G_{ji}G_{ki}}{G'_{ji}G'_{ki}}\right)$$

is an auxiliary function of $J_{\text{S-NMTF}}\left(G\right)$. Furthermore, it is a convex function in G and its global minimum is

$$G_{ik} = G_{ik}\left[\frac{(RGS + \lambda WG)_{ik}}{(GSG^T GS + DG\Lambda)_{ik}}\right]^{\frac{1}{4}}. \qquad (21)$$

PROOF. By applying Lemma 4, we have

$$\mathbf{tr}\left(GSG^TGSG^T\right) \leq \sum_{ij}\left(G'SG'^TG'S\right)_{ij}\frac{G_{ij}^4}{G_{ij}'^3}.$$

Because of Lemma 3 and the inequality of $2ab < a^2 + b^2$, we have

$$\mathbf{tr}\left(\Lambda G^TDG\right) \leq \sum_{ij}\left(DG'\Lambda\right)_{ij}\frac{G_{ij}^2}{G_{ij}'} \leq \sum_{ij}\left(DG'\Lambda\right)_{ij}\frac{G_{ij}^4 + G_{ij}'^4}{G_{ij}'^3}.$$

Because $z \leq 1 + \log z, \forall z > 0$, we have

$$\mathbf{tr}\left(RGSG^T\right) \geq \sum_{ijkl}G'_{ji}S_{jk}G'_{kl}R_{li}\left(1 + \log\frac{G_{ji}G_{kl}}{G'_{ij}G'_{kl}}\right).$$

$$\mathbf{tr}\left(G^TWG\right) \geq \sum_{ijk}G'_{ji}W_{jk}G'_{ki}\left(1 + \log\frac{G_{ji}G_{ki}}{G'_{ji}G'_{ki}}\right).$$

Summing over all these bounds, we get $Z(G,G')$, which clearly satisfies (1) $Z(G,G') \geq J_{\text{S-NMTF}}(G)$ and (2) $Z(G,G) = J_{\text{S-NMTF}}(G)$. Therefore, $Z(G,G')$ is an auxiliary function of $J_{\text{S-NMTF}}(G)$.

Following the same derivations as in [5–7, 12, 13], it can be verified that the Hessian matrix of $Z(G,G')$ is a positive definite diagonal matrix (we skip the derivations due to space limit). Thus $Z(G,G')$ is a convex function of G. We can obtain the global minimum of $Z(G,G')$ by setting $\partial Z(G,G')/\partial G_{ij} = 0$ and solving for G, from which we can get Eq. (21). This completes the proof of Theorem 3. □

THEOREM 4. *Updating G using the rule in Algorithm 1 monotonically decreases $J_{\text{S-NMTF}}(G)$ in Eq. (20).*

PROOF. According to Lemma 2 and Theorem 3, we can get that $J_{\text{S-NMTF}}(G^0) = Z(G^0, G^0) \geq Z(G^1, G^0) \geq J_{\text{S-NMTF}}(G^1) \dots$. Thus $J_{\text{S-NMTF}}(G)$ is monotonically decreasing. □

Because of Eqs. (14–15), in each iteration step of Algorithm 1 $S^{(t+1)} = \arg\min_{S^{(t)}} J_{\text{S-NMTF}}\left(S^{(t)}, G^{(t)}\right)$, which, together with Theorem 3 and 4, guarantees the convergence of Algorithm 1, because $J(G)$ in Eq. (11) is obviously lower bounded by 0.

4. EMPIRICAL STUDIES

In this section, we experimentally evaluate the proposed S-NMTF approach. Our new method has only one parameter λ in Eq. (11). Empirically, we set it as $\lambda = 0.01$.

4.1 Co-Clustering of Two-Type Relational Data

We first evaluate the proposed method on two-type relational data, the simplest multi-type relational data. We use the following three data sets: Newsgroup4 data set, WebKB4 data set, WebACE data set, whose details can be found in [7].

4.1.1 Experimental Settings

For co-clustering of two-type relational data, we need two inputs from a data set: the relationship matrix between the two types of data, and the pairwise affinity matrices for each type of data. We obtain the relationship matrices directly from the testing data sets. We construct neighborhood graphs from the both sides of the relationship matrix following [7] to obtain the pairwise affinity matrices for the both types of the entities, where, following [7], the neighborhood size is set as 10.

We compare the proposed S-NMTF approach against two related clustering methods that combine NMF and Laplacian regularization: graph regularized NMF (GNMF) [2] method and dual regularized co-clustering (DRCC) [7] method. These two methods are

Table 1: Clustering accuracy in co-clustering tasks.

Methods	Newsgroup4	WebKB4	WebACE
GNMF	0.889	0.731	0.513
DRCC	0.931	0.738	0.568
PMF	0.923	0.727	0.566
CKMeans	0.643	0.594	0.477
S-NMTF	**0.937**	**0.783**	**0.613**

Table 2: Normalized mutual information in co-clustering tasks.

Methods	Newsgroup4	WebKB4	WebACE
GNMF	0.716	0.462	0.618
DRCC	0.782	0.491	0.629
PMF	0.758	0.488	0.602
CKMeans	0.612	0.375	0.491
S-NMTF	**0.795**	**0.554**	**0.643**

largely same, except that the former uses two-factor factorization and imposes Laplacian regularization on one side factor, while the latter uses the three-factor factorization and imposes Laplacian regularizations on the both side factors. Following [7], two graphs are built on both feature side and data point side for DRCC method, and its parameters are set as $\lambda = \mu$ which is same as in [7].

In addition, we also compare our approach against the following two methods: penalized matrix factorization (PMF) [12] method and constrained K-means (CKmeans) [11] method.

To evaluate the clustering results, we adopt two standard measures widely used for clustering [2]: clustering accuracy and normalized mutual information (NMI).

4.1.2 Experimental Results

The clustering performance measured by clustering accuracy and normalized mutual information of the compared methods on the three data sets are listed in Table 1 and Table 2. The results show that the proposed S-NMTF approach consistently outperforms the other methods, sometimes very significantly, which demonstrate the effectiveness of the proposed S-NMTF approach in co-clustering of two-type relational data.

4.2 Simultaneous Clustering of Multi-Type Relational Web Data

Finally we evaluate the proposed S-NMTF approach for simultaneous clustering on multi-type relational data by using both inter-type relationships and intra-type information, which is the ultimate goal of this work.

4.2.1 Data Set

For the data set and experimental settings in this subsection, we largely follow those in [12] (We acknowledge the author of [12] for sharing the data).

We use a data set sampled from the Bulletin Board Systems (BBS) in [8]. In a BBS system, the users first register IDs. Using their IDs, the users can read others' published messages and leave their own messages. The whole system consists of many discussion fields, each of which contains many boards with similar themes. The boards are named to reflect the contents of the articles in them [8]. Once an ID posts a new article (initial article) on one board, the others can show their opinions by replying the initial article using reply articles. The initial article and reply articles

Table 3: A subset of data sampled from a BBS data set [8].

Field name	Board name
Computer Science	C++ Builder
Computer Science	Delphi
Computer Science	Database
Sports	Basketball
Sports	Volleyball
Sports	Badminton

Table 4: The F1 measure of the four compared algorithms.

	MLSA	SRC	PMF	S-NMTF
$d = 3$	0.712	0.731	0.795	**0.832**
$d = 5$	0.756	0.634	0.815	**0.839**
$d = 7$	0.711	0.621	0.780	**0.811**
$d = 9$	0.699	0.482	0.734	**0.759**

constitute a topic. Each board contains many topics. Each topic connects with several IDs through articles.

We use a subset of the BBS data in [8], in which several boards are sampled from several discussion fields. In each board, 80 topics are sampled randomly. The names of the fields and boards that we use are listed in Table 3. The user IDs related to these topics and boards are found out. Then the tensor is constructed by the co-occurrence of these three data types.

4.2.2 Experimental Settings

In the experiments, there exist three data types: topics ($\mathcal{X}_1$), user IDs ($\mathcal{X}_2$) and boards ($\mathcal{X}_3$). The topic-user matrix (R_{12}) is constructed with the number of articles each user posted in each topics with TF-IDF normalization. The topic-board matrix (R_{13}) is constructed such that if a topic belongs to a board, then the corresponding entry of R_{13} is 1. R_{23} is constructed such that if the user had posted any articles on that board, then the corresponding element of R_{23} is set to 1. Finally the elements of R_{23} are also normalized using TF-IDF scheme.

We only use the pairwise affinity matrices W_1 and W_2 for $\mathcal{X}_1$ and $\mathcal{X}_2$, which are constructed using R_{12} in a same way as in Section 4.1. We set $W_3 = I$ to simulate the case in real applications when the intra-type information for $\mathcal{X}_3$ is not available.

Besides our approach, the results of applying the Spectral Relational Clustering (SRC) [10] method and Multiple Latent Semantic Analysis (MLSA) [15] method, and PMF method are also included for comparison. All these three methods are for simultaneous clustering on multi-type relational data.

The evaluation metric is the F1 score computed using the clustering results on topics, the ground truth of which is set to be the classes corresponding to the field names they belong to.

4.2.3 Experimental Results

The experimental results are shown in Table 4, in which the value of d represent different number of clusters. From the table we can clearly see the superiority of the proposed S-NMTF approach. Besides, we also see that PMF method is always better than the other two methods. By examining the implementation details, we can see that these results are consistent with the information used by these methods: MLSA and SRC methods only use the inter-type relationships, while our S-NMTF approach additionally utilizes the intra-type information.

5. CONCLUSIONS

In this paper, we presented a general Symmetric Nonnegative Matrix Tri-Factorization (S-NMTF) framework to simultaneously cluster multi-type relational data. Our approach clusters different types of data at the same time using inter-type relationships by transforming the original problem into a symmetric NMF problem, into which we can optionally incorporate the intra-type information. In order to deal with the symmetric usage of the factor matrix in S-NMTF, we presented a new generic matrix inequality to derive the solution algorithm, which involves a fourth-order matrix polynomial, in a principled way. Extensive empirical studies in evaluating various aspects of our approach have demonstrated encouraging results, which validate the usefulness of the proposed S-NMTF approach.

Acknowledgment.

6. REFERENCES

[1] M. Belkin, P. Niyogi, and V. Sindhwani. On manifold regularization. In *AISTAT*, 2005.

[2] D. Cai, X. He, J. Han, and T. S. Huang. Graph regularized non-negative matrix factorization for data representation. *IEEE TPAMI*, 99, 2010.

[3] I. Dhillon. Co-clustering documents and words using bipartite spectral graph partitioning. In *SIGKDD*, 2001.

[4] C. Ding, X. He, and H. Simon. On the equivalence of nonnegative matrix factorization and spectral clustering. In *SDM*, 2005.

[5] C. Ding, T. Li, and M. Jordan. Convex and semi-nonnegative matrix factorizations. *IEEE TPAMI*, 32(1):45–55, 2010.

[6] C. Ding, T. Li, W. Peng, and H. Park. Orthogonal nonnegative matrix t-factorizations for clustering. In *SIGKDD*, 2006.

[7] Q. Gu and J. Zhou. Co-clustering on manifolds. In *SIGKDD*, 2009.

[8] Z. Kou and C. Zhang. Reply networks on a bulletin board system. *Physical Review E*, 67(3):36117, 2003.

[9] D. Lee and H. Seung. Algorithms for non-negative matrix factorization. In *NIPS*, 2001.

[10] B. Long, Z. Zhang, X. Wu, and Y. P. Spectral clustering for multi-type relational data. In *ICML*, 2006.

[11] K. Wagstaff, C. Cardie, S. Rogers, and S. Schrödl. Constrained k-means clustering with background knowledge. In *ICML*, 2001.

[12] F. Wang, T. Li, and C. Zhang. Semi-supervised clustering via matrix factorization. In *SDM*, 2008.

[13] H. Wang, H. Huang, F. Nie, and C. Ding. Cross-language web page classification via dual knowledge transfer using nonnegative matrix tri-factorization. In *SIGIR*, 2011.

[14] H. Wang, F. Nie, H. Huang, and C. Ding. Dyadic transfer learning for cross-domain image classification. In *ICCV*, 2011.

[15] X. Wang, J. Sun, Z. Chen, and C. Zhai. Latent semantic analysis for multiple-type interrelated data objects. In *SIGIR*, 2006.

[16] H. Zha, X. He, C. Ding, H. Simon, and M. Gu. Bipartite graph partitioning and data clustering. In *CIKM*, 2001.