

# Robust Multi-Relational Clustering via $\ell_1$ -Norm Symmetric Nonnegative Matrix Factorization

**Kai Liu**

Colorado school of Mines  
Department of EECS  
Golden, Colorado 80401  
kaliu@mines.edu

**Hua Wang**

Colorado school of Mines  
Department of EECS  
Golden, Colorado 80401  
HUAWANGCS@gmail.com

## Abstract

In this paper, we propose an  $\ell_1$ -norm Symmetric Nonnegative Matrix Tri-Factorization ( $\ell_1$  S-NMTF) framework to cluster multi-type relational data by utilizing their interrelatedness. Due to introducing the  $\ell_1$ -norm distances in our new objective function, the proposed approach is robust against noise and outliers, which are inherent in multi-relational data. We also derive the solution algorithm and rigorously analyze its correctness and convergence. The promising experimental results of the algorithm applied to text clustering on IMDB dataset validate the proposed approach.

## 1 Introduction

Traditional clustering aims to partition data points into several groups, such that the data points in the same group can share some commonalities whilst those from different groups are dissimilar. With the recent progresses of Internet and computational technologies, data have started to appear in much richer structures. To be more specific, in many real-world problems a pair of object can be related in several different ways, which inevitably complicates the problem and calls for new clustering algorithms for better understanding to the data. To address this new challenge, Wang *et al.* (Wang et al., 2011c; Wang et al., 2011d) proposed nonnegative matrix factorization (NMF) (Lee and Seung, 1999) based computational algorithms that have successfully solved the problems.

Due to its mathematical elegance and its equivalence to  $K$ -means clustering and spectral clustering (Ding et al., 2005), NMF (Lee and Seung, 1999) has been broadly studied in recent years and successfully solved a variety of practical problems in data mining and machine learning, such as those

in computer vision (Wang et al., 2011b), bioinformatics (Wang et al., 2013), natural language understanding (Wang et al., 2011a), to name a few. Compared to many traditional clustering methods, such as  $K$ -means clustering, NMF has better mathematical interpretation, which usually lead to improved accuracy on clustering (Ding et al., 2010). Traditional clustering algorithms concentrate on dealing with homogeneous data, in which all the data belong to one single type (Wang et al., 2011d). To deal with the richer data structures in modern real-world applications, symmetric Nonnegative Matrix Tri-Factorization (NMTF) (Wang et al., 2011c) have demonstrated its effectiveness on simultaneous clustering of multi-type relational data by utilizing the interrelatedness among different data types.

Traditional NMF algorithms routinely use the least square error functions, which are notably known to be sensitive against outliers (Kong et al., 2011). However, at the era of big data outliers are inevitable due to the ever increasing data sizes. As a result, developing a more robust NMF model for multi-relational data clustering has become more and more important. In this paper, we further develop the symmetric NMF clustering model proposed in (Wang et al., 2011c) by using the  $\ell_1$ -norm distances, such that our new clustering model is more robust against outliers, which is of particular importance in multi-relational data.

## 2 Robust Multi-Relational Clustering via $\ell_1$ -Norm Symmetric NMTF (S-NMTF)

In this section, we first introduce the backgrounds to use symmetric NMF to cluster multi-relational data. Then we develop our new  $\ell_1$ -norm symmetric NMF model for better robustness against outlying data. The solution algorithm to our new model will be proposed and analyzed in the next section.

**Notations.** In this paper, we use upper case letters to denote matrices. Given a matrix  $M$ , its en-

try at the  $i$ -th row and  $j$ -th column is denoted as  $M_{(ij)}$ . The Frobenius norm of a matrix  $M$  is denoted as  $\|M\|_F = \left(\sum_i \sum_j M_{(ij)}^2\right)^{1/2}$  and its  $\ell_1$ -norm is denoted as  $\|M\|_1 = \sum_i \sum_j |M_{(ij)}|$ .

## 2.1 Problem Formalization

$K$ -type relational data set can be denoted as  $\chi = \{\chi_1, \chi_2, \dots, \chi_K\}$ , where  $\chi_k = \{x_1^k, x_2^k, \dots, x_{n_k}^k\}$  represents the data set of  $k$ -th type. Suppose we are given a set of relationship matrices  $\{R_{kl} \in \mathbb{R}^{n_k \times n_l}\}_{(1 \leq k \leq K, 1 \leq l \leq K)}$  between different types of data objects, then we have  $R_{kl} = R_{lk}^T$ . Our goal is to simultaneously partition the data objects in  $\chi_1, \chi_2, \dots, \chi_K$  into  $c_1, c_2, \dots, c_K$  disjoint clusters respectively.

## 2.2 Our objective

To cluster multi-relation data, symmetric NMF has been taken advantage that solves the following optimization problem (Wang et al., 2008):

$$\begin{aligned} \min J = & \sum_{1 \leq k < l \leq K} \|R_{kl} - G_k S_{kl} G_l^T\|_F^2, \\ \text{s.t. } & G_k \geq 0, \forall 1 \leq k \leq K. \end{aligned} \quad (1)$$

It has also been shown that solving the above equation is equivalent to solve (Long et al., 2006):

$$\min J = \|R - GSG^T\|_F^2, \quad \text{s.t. } G \geq 0, \quad (2)$$

in which

$$\begin{aligned} R = & \begin{bmatrix} 0^{n_1 \times n_1} & R_{12}^{n_1 \times n_2} & \dots & R_{1K}^{n_1 \times n_K} \\ R_{21}^{n_2 \times n_1} & 0^{n_2 \times n_2} & \dots & R_{2K}^{n_2 \times n_K} \\ \vdots & \vdots & \ddots & \vdots \\ R_{K1}^{n_K \times n_1} & R_{K2}^{n_K \times n_2} & \dots & 0^{n_K \times n_K} \end{bmatrix}, \\ G = & \begin{bmatrix} G_1^{n_1 \times c_1} & 0^{n_1 \times c_2} & \dots & 0^{n_1 \times c_K} \\ 0^{n_2 \times c_1} & G_2^{n_2 \times c_2} & \dots & 0^{n_2 \times c_K} \\ \vdots & \vdots & \ddots & \vdots \\ 0^{n_K \times c_1} & 0^{n_K \times c_2} & \dots & G_K^{n_K \times c_K} \end{bmatrix}, \\ S = & \begin{bmatrix} 0^{c_1 \times c_1} & S_{12}^{c_1 \times c_2} & \dots & S_{1K}^{c_1 \times c_K} \\ S_{21}^{c_2 \times c_1} & 0^{c_2 \times c_2} & \dots & S_{2K}^{c_2 \times c_K} \\ \vdots & \vdots & \ddots & \vdots \\ S_{K1}^{c_K \times c_1} & S_{K2}^{c_K \times c_2} & \dots & 0^{c_K \times c_K} \end{bmatrix}, \end{aligned} \quad (3)$$

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

Despite its successfulness of the method proposed in (Wang et al., 2011c) in multi-relational data clustering, the objectives in Equations (1—2) use the squared  $\ell_2$ -norm distances to measure the matrix approximation errors, which, though, are prone to outliers. As a result, the clustering results could be heavily dominated by outlying data points with large approximation errors (Kong et al., 2011; Nie et al., 2011; Wang et al., 2014). To improve the robustness of the clustering model, following prior works (Kong et al., 2011; Nie et al., 2011; Wang et al., 2014) we propose to use the following  $\ell_1$ -norm symmetric NMTF model for multi-relational data clustering:

$$\min J = \|R - GSG^T\|_1 \quad \text{s.t. } G \geq 0, \quad (4)$$

In this new formulation, the approximation errors are measured by the  $\ell_1$ -norm distances, which are expected to be more insensitive to outlying data points. As shown in Figure 1, when there exist outliers in the input data, traditional squared Frobenius-norm NMF are inclined to cluster incorrectly, while the  $\ell_1$ -norm NMF are more robust and can cluster more accurately.

---

### Algorithm 1: Algorithm to solve $\ell_1$ -norm S-NMTF

---

**Data:** Relationship matrices:  $\{R_{ij}\}_{1 \leq i < j \leq K}$

**Result:** Factor matrices:  $\{G_k\}_{1 \leq k \leq K}$

1. Construct  $R, G, S$

2. Initialize  $G$  as in (Ding et al., 2006).

**repeat**

3. Construct diagonal matrix  $D$ , where

$$D(i, i) = \frac{\sum |R - GSG^T|_i}{\|R - GSG^T\|_i^2}.$$

4. Compute

$$S = (G^T G)^{-1} G^T R G (G^T G)^{-1}.$$

5. Update

$$G_{(ij)} \leftarrow G_{(ij)} \left[ \frac{(RDGS)_{(ij)}}{(GSG^T DGS)_{(ij)}} \right]^{\frac{1}{4}}.$$

**until Converges**

---

## 3 Algorithm to Solve $\ell_1$ -Norm S-NMTF and its analysis

The computational algorithm for the proposed  $\ell_1$ -norm S-NMTF approach is summarized in Algorithm 1 (Due to space limit, the derivation of the algorithm is skipped and will be provided in our journal version of the paper). Upon solution, the

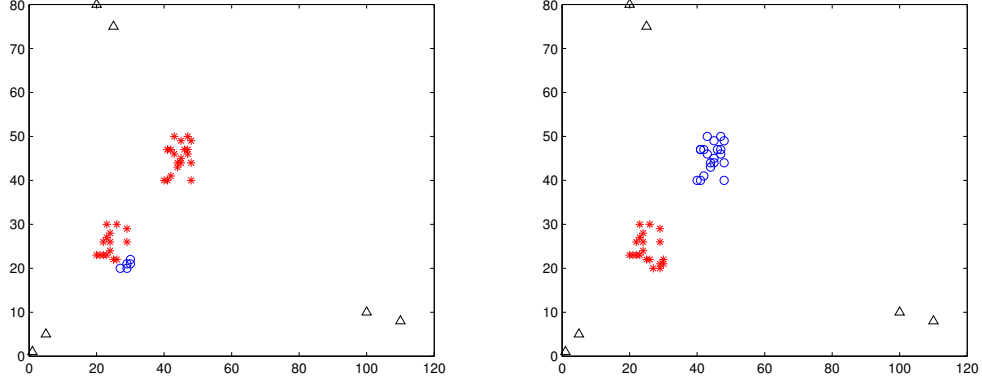


Figure 1: Clustering data in two clusters with some outliers (represented as triangle). **Left:** Clustering performance by using traditional squared Frobenius-norm NMF algorithm. **Right:** Clustering performance by using the proposed  $\ell_1$ -norm NMF algorithm.

final cluster labels are obtained from the resulted  $G_k$ .

The following theorems guarantee the correctness of Algorithm 1 (Due to space limit, the derivation of the algorithm is skipped and will be provided in our journal version of the paper).

**Theorem 3.1** *If the updating rules of  $G$  and  $S$  in Algorithm 1 converges, the final solution satisfies the KKT optimal condition.*

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

The following lemmas and theorem guarantee the convergence of Algorithm 1 (Due to space limit, the derivation of the algorithm is skipped and will be provided in our journal version of the paper).

**Lemma 3.2** (Lee and Seung, 1999)  *$Z(h, h')$  is an auxiliary function of  $F(h)$  if the conditions  $Z(h, h') \geq F(h)$  and  $Z(h, h') = F(h)$  are satisfied.*

**Lemma 3.3** (Lee and Seung, 1999) *If  $Z$  is an auxiliary function for  $F$ , then  $F$  is non-increasing under the update  $h^{(t+1)} = \arg \min_h Z(h, h')$ .*

**Theorem 3.4** *Let*

$$J(G) = \text{tr}(-2RDGSG^T + GSG^T DGSG^T), \quad (5)$$

then the following function

$$\begin{aligned} Z(G, G') = & -2 \sum_{ijkl} G'_{(ji)} S_{(jk)} G'_{(kl)} D_{(ll)} R_{(li)} (1 + \log \frac{G_{(ji)} G_{(kl)}}{G'_{(ji)} G'_{(kl)}}) \\ & + \sum_{ij} (G' S G'^T D G' S)_{(ij)} \frac{G_{(ij)}^4}{G'^3_{(ij)}} \end{aligned} \quad (6)$$

is an auxiliary function of  $J(G)$ . Furthermore, it is a convex function in  $G$  and its global minimum is

$$G_{(ik)} = G_{(ik)} \left[ \frac{(RDGS)_{(ik)}}{(GSG^T DGS)_{(ik)}} \right]^{\frac{1}{4}} \quad (7)$$

Based on the property of auxiliary function and convex function, by updating  $G$ , we can always get the optimal solution to the object function, thus determining the final cluster label.

## 4 Experiments Result

In this section, We test our proposed algorithm on IMDB dataset by using its inter-type relationship information.

### 4.1 Data set

We use the dataset from **ACL-IMDB** provided by (Maas et al., 2011). In this dataset, there is a sub-training set of 25000 highly polar movie reviews, in which positive and negative comments come up with one half(12500) each. The dataset also includes the following two important files: the content of each comment and the corresponding *URL*

where each comment comes from. There are also some other files but not related with the experiment we conduct, thus we skip them.

## 4.2 Experiments settings

In our experiment, we set the multi-type data as 3 types: author, comment and word. As it is discussed in the 3rd part, there are three relationships we need to find, which correspond to three matrices we need to construct the multi-type data matrix:comment-author, comment-word and author-word. By making use of the *URL* of every comment, we can find the author who posts the corresponding comment, thus we can build the author-comment matrix. Since each comment with content is given by the dataset file, we could therefore construct the matrix of comment-word, and the author-word matrix is the product of author-comment matrix and comment-word matrix.

We could find the first 1500 authors who post comments most, since the comments from the same person are more likely to have some correlations, such as similar sentence structures, same words and etc. We also rule out the stop-words since they may disturb the clustering and they are meaningless to the property of comments. To make our experiments to be more persuasive, we also add some noise to the three relationship matrices with a ratio of 25 percentage(1/5 in amplitude). By randomly choosing 500 authors from 1500, we could generate many sub-datasets to conduct our experiments.

## 4.3 Experiments Results

We compare the performance of our proposed  $\ell_1$ -norm S-NMTF algorithm with other methods such as P-NMF, Frobenius norm S-NMTF, traditional NMF and  $K$ -means clustering. For simplicity, we only compare the clustering accuracy of comment-word matrix since its label (positive or negative) is fixed(the grounding label), thus could be compared with the clustering results by using the clustering algorithms.

Table 1 shows that when the data is pure, in many cases(more than the listed),  $\ell_1$ -norm S-NMTF approach has better performance than others

Table 2 illustrates the situation when some noise is added to the data, it is easy to find that  $\ell_1$ -norm S-NMTF algorithm is the best in terms of clustering accuracy. This meets our analysis in our Motivation part.

Alg	L11	L22	PMF	NMF	Kms
set 1	<b>0.578</b>	0.528	0.554	0.510	0.504
set 2	<b>0.583</b>	0.556	0.551	0.521	0.521
set 3	<b>0.584</b>	0.559	0.555	0.501	0.501
set 4	<b>0.551</b>	0.522	0.502	0.527	0.506
set 5	<b>0.566</b>	0.534	0.506	0.529	0.531
set 6	<b>0.558</b>	0.517	0.510	0.535	0.526

Table 1: Clustering Accuracy with Pure Data.

Alg	L11	L22	PMF	NMF	Kms
sub 1	<b>0.586</b>	0.545	0.508	0.530	0.530
sub 2	<b>0.575</b>	0.535	0.540	0.518	0.532
sub 3	<b>0.567</b>	0.528	0.520	0.500	0.500
sub 4	<b>0.574</b>	0.533	0.525	0.500	0.500
sub 5	<b>0.574</b>	0.537	0.530	0.519	0.518
sub 6	<b>0.556</b>	0.525	0.524	0.504	0.505

Table 2: Clustering Accuracy with Noise.

Careful examination in Table 3 reveals the fact that  $\ell_1$ -norm S-NMTF algorithm performs more robust than any other algorithm. Though the clustering accuracy of  $\ell_1$ -norm S-NMTF decreases when noise exists, still it reduces the least among the five algorithms. This result convincingly demonstrates the robustness of our proposed  $\ell_1$ -norm S-NMTF method.

Alg	L11	L22	PMF	NMF	Kms
s.1(P)	<b>0.547</b>	0.546	0.521	0.546	0.546
s.1(N)	<b>0.546</b>	0.525	0.516	0.540	0.545
s.2(P)	<b>0.543</b>	0.543	0.534	0.543	0.543
s.2(N)	<b>0.543</b>	0.539	0.531	0.513	0.531
s.3(P)	<b>0.536</b>	0.536	0.524	0.536	0.536
s.3(N)	<b>0.536</b>	0.534	0.522	0.517	0.508

Table 3: Clustering Accuracy Contrast.

## 5 Conclusion

In this paper, we presented an  $\ell_1$ -norm Symmetric Nonnegative Matrix Tri-Factorization Framework to cluster multi-type relational data simultaneously. Our proposed approach clusters different types of data, using its inter-type relationship by transforming the original problem into a symmetric NMTF problem. We also presented an auxiliary function and high order matrix inequality to derive the solution algorithm. The proposed algorithm not only makes use of the rich data struc-

ture to improve the clustering accuracy, but also remains robust when there is noise and outliers. Experimental results demonstrate the potential usage and advantage of  $\ell_1$ -norm S-NMTF in clustering especially when there are outliers, which is in accordance with our theory analysis.

## References

- C. Ding, X. He, and H.D. Simon. 2005. On the equivalence of nonnegative matrix factorization and spectral clustering. In *SDM*.
- C. Ding, T. Li, W. Peng, and H. Park. 2006. Orthogonal nonnegative matrix t-factorizations for clustering. In *SIGKDD*.
- C. Ding, T. Li, and M.I. Jordan. 2010. Convex and semi-nonnegative matrix factorizations. *IEEE TPAMI*, 32(1):45–55.
- Deguang Kong, Chris Ding, and Heng Huang. 2011. Robust nonnegative matrix factorization using  $\ell_2$ -norm. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, pages 673–682. ACM.
- D.D. Lee and H.S. Seung. 1999. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791.
- Bo Long, Zhongfei Mark Zhang, Xiaoyun Wu, and Philip S Yu. 2006. Spectral clustering for multi-type relational data. In *Proceedings of the 23rd international conference on Machine learning*, pages 585–592. ACM.
- Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, and Christopher Potts. 2011. Learning word vectors for sentiment analysis. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 142–150, Portland, Oregon, USA, June. Association for Computational Linguistics.
- Feiping Nie, Heng Huang, Chris Ding, Dijun Luo, and Hua Wang. 2011. Robust principal component analysis with non-greedy  $\ell_1$ -norm maximization. In *IJCAI Proceedings-International Joint Conference on Artificial Intelligence*, volume 22, page 1433. Cite-seer.
- F. Wang, T. Li, and C. Zhang. 2008. Semi-supervised clustering via matrix factorization. In *SDM*.
- H. Wang, H. Huang, F. Nie, and C. Ding. 2011a. Cross-language web page classification via dual knowledge transfer using nonnegative matrix tri-factorization. In *SIGIR*.
- H. Wang, F. Nie, H. Huang, and C. Ding. 2011b. Dyadic transfer learning for cross-domain image classification. In *ICCV*.
- Hua Wang, Heng Huang, and Chris Ding. 2011c. Simultaneous clustering of multi-type relational data via symmetric nonnegative matrix tri-factorization. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, pages 279–284. ACM.
- Hua Wang, Feiping Nie, Heng Huang, and Chris Ding. 2011d. Nonnegative matrix tri-factorization based high-order co-clustering and its fast implementation. In *Data Mining (ICDM), 2011 IEEE 11th International Conference on*, pages 774–783. IEEE.
- Hua Wang, Heng Huang, Chris Ding, and Feiping Nie. 2013. Predicting protein–protein interactions from multimodal biological data sources via nonnegative matrix tri-factorization. *Journal of Computational Biology*, 20(4):344–358.
- Hua Wang, Feiping Nie, and Heng Huang. 2014. Robust distance metric learning via simultaneous  $\ell_1$ -norm minimization and maximization. In *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, pages 1836–1844.