Tensor based Relations Ranking for Multi-relational Collective Classification

Chao Han†, Qingyao Wu†, Michael K. Ng§, Jiezhang Cao†, Mingkui Tan†, Jian Chen†
†*School of Software Engineering, South China University of Technology, Guangzhou, China*
§*Department of Mathematics, Hong Kong Baptist University, Hong Kong, China*
Email: hanchaoc@gmail.com, qyw@scut.edu.cn, mng@math.hkbu.edu.hk,
caojiezhang@gmail.com, mingkuitan@scut.edu.cn, ellachen@scut.edu.cn

***Abstract*—In this paper, we study relations ranking and object classification for multi-relational data where objects are interconnected by multiple relations. The relations among objects should be exploited for achieving a good classification. While most existing approaches exploit either by directly counting the number of connections among objects or by learning the weight of each relation from labeled data only. In this paper, we propose an algorithm, TensorRrCc, which is able to determine the ranking of relations and the labels of objects simultaneously. Our basic idea is that highly ranked relations within a class should play more important roles in object classification, and class membership information is important for determining a ranking quality over the relations w.r.t. a specific learning task. TensorRrCc implements the idea by modeling a Markov chain on transition probability graphs from connection and feature information with both labeled and unlabeled objects and propagates the ranking scores of relations and relevant classes of objects. An iterative progress is proposed to solve a set of tensor equations to obtain the stationary distribution of relations and objects. We compared our algorithm with current collective classification algorithms on two real-world data sets and the experimental results show the superiority of our method.**

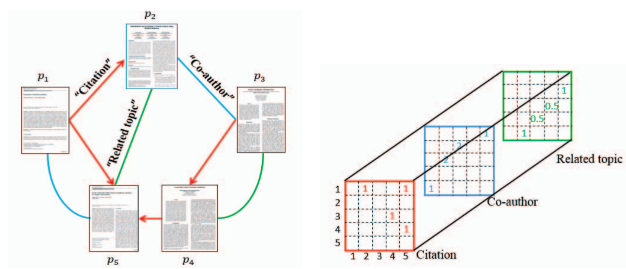
***Keywords*-relations ranking, classification, tensor, multi-relational data**

I. Introduction

Collective classification aims to exploit linkage information among objects whose class labels are correlated for improving classification accuracy. It is an important and intense research problem in last decade which can assist in many applications of data mining, e.g., recommending specific items for individuals, discovering communities in social networks, and detecting fraud in communication networks. A variety of collective classification approaches have been explored in the literature [1], [2], but most of them focus on single-relational data. However, in many real-world applications, objects are involved in multiple types of relations which are complex, various, and discriminative. The relations between one pair of objects commonly indicate different semantic representation. To be more specific, we show in Figure 1(a) an example of a bibliographic network over research papers. This example contains three relations, such as "Citation", "Co-author", and "Related Topics".

*The first two authors contributed to this work equally. Correspondence should be addressed to M. Tan and J. Chen.

Recently, some researchers (see [3], [4]) have considered the multi-relational collective classification. Kong et al. [3] captures the subtlety of different types of relations by counting the number of connections for each kind of relation respectively. This method is straightforward and easy to implement, but the link counting may not provide additional discriminative power to enhance the classification. Other studies [4], [5] try to weight the importance of relations by training the weight parameters from labeled data. Relation knowledge obtained in this way is generally helpful but with high risks of overfitting. In this paper, we proposed



(a) (b)

Figure 1. An example of bibliographic network. (a) Examples of multiple types of connections; (b) A tensor representation of the data.

a tensor-based relations ranking and collective classification algorithm, TensorRrCc, to determine the importance of relations and predict class labels for objects simultaneously. Our intuition is that highly ranked relations within a class should play more important roles in object classification, and class membership information is also important for determining a ranking quality over the relations. Specifically, we use tensor to represent the multiple relations among objects. As shown in Figure 1(b), a three-dimensional tensor is used to represent the relational data in Figure 1(a). Note that, each two-dimensional slice of the tensor represents an adjacency matrix for one relation. After that, we model the problem as a Markov chain on a set of transition probability graphs from both connection and feature information of both labeled and unlabeled data and propagate the ranking scores of relations and relevant classes of objects on these graphs.

The main contributions of this paper are as follows.

- We propose an algorithm, TensorRrCc, that integrates relation ranking and object classification, allowing them

2374-8486/17 $31.00 © 2017 IEEE
DOI 10.1109/ICDM.2017.112

to mutually enhance each other.

- We develop solve a set of tensor equations in Markov chain model to obtain a stationary distribution as evaluation scores for classification and ranking.
- By building a tensor-based Markov chain model, our algorithm can effectively make use of both labeled and unlabeled data information to boost the classification and ranking performance.

The rest of the paper is organized as follows. The related work is introduced in Section 2. The proposed methodology is detailed in Section 3. The experimental results are presented in Section 4. Conclusions are given in Section 5.

II. Related Work

Collective Classification. Various collective classification approaches have been developed to perform learning on relational data [1], [6]. ICA [1] is one of the most well-known algorithms which transforms the relational information into a feature vector by counting the labels of connected nodes. Macskassy [7] developed a classifier, wvRN+RL, to estimate the distribution probability of labels for each node through iteratively computing neighboring labels. A few researchers have recently examined the semi-supervised collective classification task focusing on the less labeled data. In [2], the authors examine the performance of many semi-supervised variants, including ICA.

Heterogeneous Network Learning. Ji et al. [8] proposed a RankClass algorithm to use the ranking of heterogeneous objects to perform the various objects classification. Eswaran et al. [9] proposed ZooBP to perform on heterogeneous graphs. These works consider multiple types of nodes and relations while we focus on only one type of node. Kong et al. [3] proposed an approach for solving such problem, which employs the meta-path method to transform a heterogeneous network to multiple relations which are considered as a sequence of feature vectors by aggregating the label information of neighbors via each relation. This method is used as a comparison in the experiment.

Tensor-based Multi-relational Learning. There has been a growing interest in tensor methods for multi-relational learning, partially due to their natural representation of multi-relational data. These approaches have been applied successfully in many applications, such as community discovery [10], link prediction [11], and ranking [12]. More recently, Ng et al. [13] proposed a framework, MultiRank, to seek the stationary probability distributions of a set of tensors for a ranking problem in multi-relational data. Later on, this approach is employed for computing the hub, authority scores of objects, and the relevance scores of relations [14], image retrieval [15] and discovering the community structure [16]. Different from these approaches, we focus on the problem of object classification and relations ranking in which the objects attribute and the labels information of each entity are taken into account.

III. Methodology

We use a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ to represent multi-relational data set, where the node set $\mathcal{V}$ denotes a set of objects, $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ denotes the set of links between the nodes in $\mathcal{V}$. Each object i is represented by a feature vector $\boldsymbol{f}_i \in \mathbb{R}^d$ and is associated with one or several labels. Suppose there are n objects, m relations, and q possible labels. The task of this paper is to predict the class labels of unlabeled objects accurately as much as possible and give the important relations associated with each class label.

The basic idea of our work is to consider a random walk in multi-relational data via semantic connections and feature-based correlations, where the multi-relational data is represented by a tensor. A Markov chain model is used to solve a set of tensor equations to obtain the stationary probabilities of objects and relations for each label, which is the estimation score for determining the ranking of relations and labels of objects. By building a tensor-based Markov chain model, our algorithm can effectively make use of both labeled and unlabeled data information to boost the classification and ranking performance.

A. Tensor Representation

We use tensor to represent the multiple relations among objects [13]. We call $\boldsymbol{\mathcal{A}}= (a_{i,j,k})$ a real $(2,1)$th order $(n \times m)$-dimensional tensor, where $i, j = 1, 2, \ldots, n$, and $k = 1, 2, \ldots, m$. We refer (i, j) to be the indices for objects and k to be the indices for relations. Specifically, if an object i is linked to an object j through the relation k, then $a_{i,j,k}$ is set to 1. $\boldsymbol{\mathcal{A}}$ is a nonnegative tensor for $a_{i,j,k} \geq 0, \forall i, j, k$.

We assume that any two objects in multi-relational data can be connected via some relations, so $\boldsymbol{\mathcal{A}}$ is irreducible. As we would like to determine the stable probability distributions of both objects and relations simultaneously in multi-relational data, irreducibility is a reasonable assumption that we will use in the following analysis and discussion. It is clear that when $\boldsymbol{\mathcal{A}}$ is irreducible, the two corresponding tensors $\boldsymbol{\mathcal{O}}$ and $\boldsymbol{\mathcal{R}}$ are also irreducible.

In order to perform semi-supervised learning, we construct a transition probability graph for a Markov chain of all the labeled and unlabeled objects, and then make use of the idea in topic-sensitive PageRank [17] and random walk with restart [18] to propagate the label information from labeled data to unlabeled data.

In the graph, the Markov transition probabilities $\boldsymbol{\mathcal{O}}= (o_{i,j,k})$ and $\boldsymbol{\mathcal{R}}= (r_{i,j,k})$ w.r.t. objects and relations can be obtained by normalizing the entries of $\boldsymbol{\mathcal{A}}$ as follows:

$$o_{i,j,k} = \frac{a_{i,j,k}}{\sum\limits_{i=1}^{n} a_{i,j,k}}, \quad i = 1, \ldots, n,$$

$$r_{i,j,k} = \frac{a_{i,j,k}}{\sum\limits_{k=1}^{m} a_{i,j,k}}, \quad k = 1, \ldots, m.$$

Here, $o_{i,j,k}$ can be interpreted as the probability of visiting the ith object by given that jth object is currently visited and the kth relation is used, and $r_{i,j,k}$ can be interpreted as the probability of using the kth relation given that ith object is visited from the jth object. Let $X_t = [X_t = 1, \ldots, X_t = n]$ and $Z_t = [Z_t = 1, \ldots, Z_t = m]$ be the random variables referring to visiting any particular object and using any particular relation respectively at the time t. The transition probabilities are written as follows:

$$o_{i,j,k} = P\left[X_t = i \,|X_{t-1} = j, Z_t = k\right],$$

$$r_{i,j,k} = P\left[Z_t = k \,|X_t = i, X_{t-1} = j\right].$$

If there is a dangling node ($a_{i,j,k}$ is equal to 0 for all $1 \le i \le n$ [19]), the values of $o_{i,j,k}$ can be set to $1/n$ (an equal chance to visit any object). Similarly, $r_{i,j,k}$ can be set to $1/m$ (an equal chance to use any relation), if $a_{i,j,k}$ is equal to 0 for all $1 \le k \le m$. We call them transition probability tensors which are analog of transition probability matrices in Markov chains [20].

B. Transition Probabilities from Connections

Let $\boldsymbol{x}$ be a column vector of length n and $\boldsymbol{z}$ be a column vector of length m. Let $\boldsymbol{\mathcal{A}xz}$ be a vector in $\mathbb{R}^n$ such that

$$(\boldsymbol{\mathcal{A}xz})_i = \sum_{j=1}^{n}\sum_{k=1}^{m} a_{i,j,k} x_j z_k, \;\; i = 1, \ldots, n.$$

Similarly, $\boldsymbol{\mathcal{A}xx}$ is a vector in $\mathbb{R}^m$ such that

$$(\boldsymbol{\mathcal{A}xx})_k = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{i,j,k} x_i x_j, \;\; k = 1, \ldots, m.$$

Given two transition probability tensors $\boldsymbol{\mathcal{O}}$ and $\boldsymbol{\mathcal{R}}$, we study the following probabilities:

$$P\left[X_t = i\right] = \sum_{j=1}^{n}\sum_{k=1}^{m} o_{i,j,k} \times P\left[X_{t-1} = j, Z_t = k\right], \quad (1)$$

$$P\left[Z_t = k\right] = \sum_{i=1}^{n}\sum_{j=1}^{n} r_{i,j,k} \times P\left[X_t = i, X_{t-1} = j\right], \quad (2)$$

where $P\left[X_{t-1} = j, Z_t = k\right]$ is the joint probability distribution of X_{t-1} and Z_t, and $P\left[X_t = i, X_{t-1} = j\right]$ is the joint probability of X_t and X_{t-1}. Here we employ a product form of individual probability distributions for joint probability distributions. Using this assumptions, (1) and (2) become

$$P\left[X_t = i\right] = \sum_{j=1}^{n}\sum_{k=1}^{m} o_{i,j,k} \times P\left[X_{t-1} = j\right] P\left[Z_t = k\right],$$

$$P\left[Z_t = k\right] = \sum_{i=1}^{n}\sum_{j=1}^{n} r_{i,j,k} \times P\left[X_t = i\right] P\left[X_{t-1} = j\right].$$

Here, we consider to achieve a stationary distribution of objects, denoted by $\bar{\boldsymbol{x}} = [\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n]^T$ with $\sum_{i=1}^{n} \bar{x}_i = 1$ and a stationary probability distribution of relations, denoted by $\bar{\boldsymbol{z}} = [\bar{z}_1, \bar{z}_2, \ldots, \bar{z}_m]^T$ with $\sum_{k=1}^{m} \bar{z}_k = 1$, where

$$\bar{x}_i = \lim_{t\to\infty} P\left[X_t = i\right], \quad \text{and} \quad \bar{z}_k = \lim_{t\to\infty} P[Z_t = k]$$

for $1 \le i \le n,$ and $1 \le k \le m$.

Using the above equations, we have

$$\bar{x}_i = \sum_{j=1}^{n}\sum_{k=1}^{m} o_{i,j,k}\bar{x}_j\bar{z}_k, i = 1, 2..., n, \quad (3)$$

$$\bar{z}_k = \sum_{i=1}^{n}\sum_{j=1}^{n} r_{i,j,k}\bar{x}_i\bar{x}_j, k = 1, 2, ..., m. \quad (4)$$

Formally, under the tensor operations for (3) and (4), we compute the stationary probabilities of the objects and relations by solving the following tensor equations:

$$\bar{\boldsymbol{x}} = \boldsymbol{\mathcal{O}}\bar{\boldsymbol{x}}\bar{\boldsymbol{z}}, \quad (5)$$

$$\bar{\boldsymbol{z}} = \boldsymbol{\mathcal{R}}\bar{\boldsymbol{x}}^2. \quad (6)$$

C. Transition Probabilities from Nodes

Two objects with similar attributes indicate that they belong to one class. Hence, the feature-based correlations among objects can be regarded as the transition probability. Here, we use a common correlation measure, cosine similarity, to construct the transition probability matrix.

For objects i and j, the cosine similarity is given by

$$\cos\left(\boldsymbol{f}_i, \boldsymbol{f}_j\right) = \frac{\boldsymbol{f}_i \cdot \boldsymbol{f}_j}{\|\boldsymbol{f}_i\| \, \|\boldsymbol{f}_j\|}.$$

We construct an n-by-n matrix $\boldsymbol{C} = (c_{i,j})$ where $c_{i,j} = cos(\boldsymbol{f}_i, \boldsymbol{f}_j)$ to store the cosine similarity between objects i and j. For the entries of each column sum of the transition probability matrix equal one, we obtain the transition probability matrix $\boldsymbol{W}$ by normalizing $\boldsymbol{C}$ with respect to each column, $\sum_{i=1}^{n} w_{ij} = 1, j = 1, 2, ..., n$.

We obtain the stable probability distribution of objects from their attributes by solving the following equation:

$$\bar{\boldsymbol{x}} = \boldsymbol{W}\bar{\boldsymbol{x}}, \quad (7)$$

with $\sum_{i=1}^{n} \bar{x}_i = 1$.

D. The TensorRRCc Algorithm

We start a random walker from the given labeled nodes. The walker iteratively visits the neighboring nodes with the transition probabilities $\boldsymbol{\mathcal{O}}$ and $\boldsymbol{W}$ given in (5) and (7). At each step, it has probability $\alpha (0 < \alpha < 1)$ to return the label information of labeled nodes. A weighting parameter γ is used to scale the ratio of walking in $\boldsymbol{\mathcal{O}}$ and $\boldsymbol{W}$. We use $\beta = \gamma \times (1 - \alpha)$ to simplify the description of the equation. The walker with stationary probabilities will finally stay at different nodes. Formally, these stationary probabilities are computed using the following equation:

$$\bar{\boldsymbol{x}} = (1-\alpha-\beta)\boldsymbol{\mathcal{O}}\bar{\boldsymbol{x}}\bar{\boldsymbol{z}} + \beta \boldsymbol{W}\bar{\boldsymbol{x}} + \alpha \boldsymbol{l}, \quad (8)$$

where $\boldsymbol{l}$ is an assigned probability distribution vector of size n referring to the labeled object in the current label c. To construct $\boldsymbol{l}$, one simple way is to use a uniform distribution of the objects with the class label $c(c = 1, 2, \ldots, q)$. More precisely,

$$[\boldsymbol{l}]_i = \begin{cases} 1/n_c, & \text{if } c \in Y_i; \\ 0 \quad , & \text{otherwise.} \end{cases} \quad (9)$$

where n_c is the number of objects associated with the label c in the labeled data set, Y_i is the label set of object i.

In this paper, We present an iterative algorithm to solve the tensor equation in (8) and (6) simultaneously. After finite iterations, we can obtain the stationary probability distribution of objects and relations for each class label respectively. We can know the top several relations strongly related with each label based on the probability distribution of relations. And the labels of unlabeled data can be predicted based on the probability distribution of objects.

Algorithm 1 The TensorRRCC Algorithm.

Input: $\boldsymbol{\mathcal{O}}$, $\boldsymbol{\mathcal{R}}$, $\boldsymbol{W}$, $\boldsymbol{l}$, $\boldsymbol{x}_0$, $\boldsymbol{z}_0$; Parameters: α, β, ε;

```
1: repeat
2:    set t = t + 1;
3:    x_t = (1 − α − β) O x_{t−1} z_{t−1} + β W x_{t−1} + α l;
4:    z_t = R x_t^2;
5: until ||x_t − x_{t−1}|| + ||z_t − z_{t−1}|| < ε;
```

IV. Experimental Results

In this section, we exhibit that our proposed algorithm can give a ranking of relations for each class label and predict for unlabeled examples effectively. We experiment on two data sets which are extracted from DBLP[1] and ACM[2] digit libraries respectively. Their characteristics (number of nodes, features, relations, labels, network density, and label cardinality) are summarized in Table I.

Table I
SUMMARY OF DATA SETS

Data set	#N	#F	#R	#L	ND	LC
DBLP	4057	8898	20	4	1.96×10^{-1}	1
ACM	1484	4067	1629	13	1.2×10^{-7}	1.16

In order to demonstrate the performance of our algorithm, we compare with the following methods on two data sets.

- **HCC**: Kong et al. [3] proposed this method for heterogeneous network classification, in which the meta path-based linkage among objects can be viewed as multiple relations.
- **HCC-SS**: We employ a semi-supervised approach semi-ICA [2] to replace the base classifier ICA in HCC.
- **wvRN+RL** [7]. It is a collective classification method which transfers content and structure information to relationship among objects respectively. Hence, it can solve multi-relational classification problem.
- **EMR**: C. Preisach & L. Schmidi-Thieme [21] use an ensemble to combine multiple relations while ignoring their difference. We train an ICA classifier for each relation which votes for the prediction.
- **ICA**: It is commonly used as a comparison in collective classification [1]. For multiple relations, we aggregate them all into one relation for employing this algorithm to show the necessity of multi-relational setting.

We conduct experiments on both data sets by randomly picking up $\{10, 20, 30, 40, 50, 60, 70, 80, 90\}\%$ of the examples as the training data, and the remaining for testing.

A. Experiment on DBLP

In this section, we test our algorithm on data set DBLP which is reported by J. Ming et al. [22]. DBLP contains publication from 20 computer science conferences on four research areas: database (DB), data mining (DM), artificial intelligence (AI), and information retrieval (IR), and each of them contains five conferences, see Table II. For each author, a bag-of-words representation of all the paper titles published by the author is regarded as its content attribute. For relations among authors, each conference is regarded as one relation, and two authors are correlated through one of them if they have published papers on the corresponding conference. Each author is assigned with a class label indicating his/her research area. The task of this experiment is to give rankings of each conference relations in a research area and assign labels for the authors based on their content and relational information.

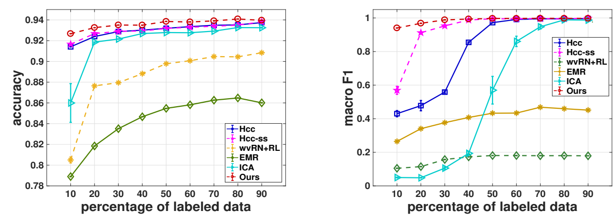
Table II
RANKINGS OF RELATIONS IN EACH RESEARCH AREA

DB		DM		AI		IR	
conf.	rk.	conf.	rk.	conf.	rk.	conf.	rk.
VLDB	1	KDD	1	IJCAI	1	SIGIR	1
SIGMOD	2	ICDM	2	AAAI	2	CIKM	2
ICDE	3	PAKDD	3	ICML	3	ECIR	3
EDBT	4	SDM	4	ECML	6	WWW	4
PODS	6	PKDD	6	CVPR	11	WSDM	19

Results and discussion. Figure 2(a) shows the accuracy results on DBLP. We can see from Figure 2(a) that: i) Our algorithm always results in best accuracy. The wvRN+RL considers the feature attribute of examples as one of the relations which are not competitive with our methods for the weak effect of content information. ii) Our proposed tensor representation method outperforms the EMR and ICA approaches, which both neglect the difference among relations. This suggests the superiority of tensor-based representation

[1]http://dblp.uni-trier.de/
[2]http://dl.acm.org/

idea in learning the multi-relational data. iii) Both ICA and wvRN+RL without considering the semi-supervised learning mechanism suffer a performance degradation when there are less than 20% labeled data. Our method achieves significant improvement against these baselines in this case.



(a) DBLP (b) ACM

Figure 2. Performance on Datasets W.R.T. Incremental Labeled Data.

B. Experiment on ACM

In this section, we test our algorithm on ACM dataset which is extracted from the ACM digital library with KDD conference from 1999 to 2010 and SIGIR conference from 2000 to 2010. Each publication contains the title, keywords, authors, concepts, conferences, citation, published year, and index terms. The index terms of papers are given by ACM based on the ACM CCS[3]. The feature attribute of each publication is represented by a bag-of-word vector from its title. Other information can be organized six relations, i.e., authors, concepts, conferences, keywords, published year and citations. The task of this experiment is to predict the ACM index terms for publications based on their representative and relational information.

Results and discussion. For each publication may have more than one index term, we use the multi-label metric, macro F1, to measure the performance of all algorithms on ACM. Figure 2(b) shows that our algorithm outperforms or equals other methods with different percentage of labeled data. As the labeled data are less than 40%, our algorithm has a great superiority. The EMR and wvRN+RL algorithms perform terribly, which is consistent with the result on DBLP because they treat all relations equally.

In figure 3, we show the probability distribution of relations in each class label. The height and color of each bar indicate the probability of the corresponding relation in current label. We can see the probability distribution of relations in each class label has a slight difference, but the rankings are basically consistent. It shows that the ”concept” and ”conference” relations are more important than others in the classification process, which means the two relations can provide enough and highly accurate connected information (less isolated examples and most connected examples with similar label sets).

[3]The ACM Computing Classification Systems: dl.acm.org/ccs/ccs.cfm.

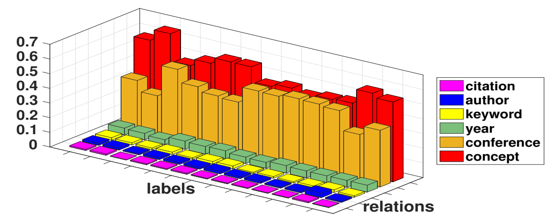
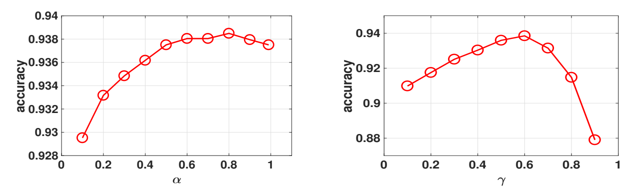


Figure 3. Relations Ranking of each label on ACM.

C. Sensitivity Study

The TensorRRCC algorithm has two essential parameters α and γ (β). We test the performance of the algorithm on DBLP when α varies from 0.1 to 0.99 and γ varies from 0.1 to 0.9 respectively. From Figure 4(a), we see, in general, the accuracy firstly increases and goes down varying with α increasing. It gets best when $\alpha = 0.8$ on DBLP and we set $\alpha = 0.8$ as the default value in all experiments. Figure 4(b) shows the accuracy variance while γ increases. We can see the algorithm performs best when $\gamma = 0.6$ on DBLP, and we set $\gamma = 0.6$ as the default value. While for ACM dataset, we set $\gamma = 0$ to only consider the relational information which is good enough compared with other algorithms.

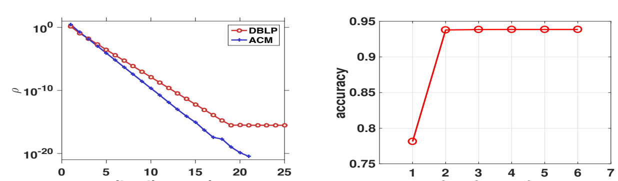


(a) α (b) γ

Figure 4. Sensitivity of parameters on DBLP.

D. Convergence study

We discuss the convergence and the performance of the TensorRRCC algorithm w.r.t. the iteration number. In Figure 5(a), we can see the algorithm gets convergence after more than 20 iterations on both datasets. The subfigure (b) shows the accuracy on DBLP varying with the iteration number, and it gets stable after 3 iterations.



(a) convergence (b) accuracy

Figure 5. Convergence of TensorRRCC on Datasets.

V. Conclusion

In this paper, we proposed a tensor based algorithm, TensorRrCc, for relations ranking and collective classification in multi-relational data. We represented the multi-relational data as a three-way tensor and introduced a Markov chain based scheme to determine the relations ranking and object labels simultaneously based on the stable probability distribution, which was obtained by using an iterative procedure to solve the equations. Experimental results on two real-world data sets demonstrated that the effectiveness of the proposed algorithm to rank the multiple relations and better performance than compared methods used in this paper.

Acknowledgment

This work was supported by National Natural Science Foundation of China (NSFC) under Grants 61502177 and 61602185, Recruitment Program for Young Professionals, Fundamental Research Funds for the Central Universities under Grants D2172500 and D2172480, Special Planning Project of Guangdong Province under Grant 609055894069, Guangdong Provincial Scientific and Technological Funds under Grants 2017B090901008 and 2017A010101011, CCF-Tencent Open Research Fund, RGC GRF HKBU12306616 and CRF C1007-15G.

References

[1] P. Sen, G. Namata, M. Bilgic, L. Getoor, B. Galligher, and T. Eliassi-Rad, "Collective classification in network data," *AI magazine*, vol. 29, no. 3, p. 93, 2008.

[2] L. K. McDowell and D. W. Aha, "Semi-supervised collective classification via hybrid label regularization," in *ICML*, 2012.

[3] X. Kong, P. S. Yu, Y. Ding, and D. J. Wild, "Meta path-based collective classification in heterogeneous information networks," in *Proc. of the 21st ACM Int. Conf. on Information and Dnowledge Management*, 2012, pp. 1567–1571.

[4] P. Vijayan, S. Subramanian, and B. Ravindran, "Multi-label collective classification in multi-attribute multi-relational network data," in *IEEE/ACM Int. Conf. on Advances in Social Networks Analysis and Mining*, 2014, pp. 509–514.

[5] X. Shi, J.-F. Paiement, D. Grangier, and S. Y. Philip, "Learning from heterogeneous sources via gradient boosting consensus." in *SDM*, 2012, pp. 224–235.

[6] H. Eldardiry and J. Neville, "Across-model collective ensemble classification," in *AAAI*, 2011.

[7] S. A. Macskassy, "Improving learning in networked data by combining explicit and mined links," in *AAAI*, vol. 22, 2007, pp. 590–595.

[8] M. Ji, J. Han, and M. Danilevsky, "Ranking-based classification of heterogeneous information networks," in *Proc. of the 17th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, 2011, pp. 1298–1306.

[9] D. Eswaran, S. Günnemann, C. Faloutsos, D. Makhija, and M. Kumar, "Zoobp: Belief propagation for heterogeneous networks," *Proceedings of the VLDB Endowment*, vol. 10, no. 5, 2017.

[10] K. Maruhashi, F. Guo, and C. Faloutsos, "Multiaspectforensics: Pattern mining on large-scale heterogeneous networks with tensor analysis," in *Advances in Social Networks Analysis and Mining (ASONAM), 2011 Int. Conf. on*, 2011, pp. 203–210.

[11] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko, "Translating embeddings for modeling multi-relational data," in *Advances in Neural Information Processing Systems*, 2013, pp. 2787–2795.

[12] R. Jenatton, N. L. Roux, A. Bordes, and G. R. Obozinski, "A latent factor model for highly multi-relational data," in *Advances in Neural Information Processing Systems*, 2012, pp. 3167–3175.

[13] M. K.-P. Ng, X. Li, and Y. Ye, "Multirank: co-ranking for objects and relations in multi-relational data," in *Proc. of the 17th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, 2011, pp. 1217–1225.

[14] X. Li, M. K. Ng, and Y. Ye, "Har: Hub, authority and relevance scores in multi-relational data for query search." in *SDM*, 2012, pp. 141–152.

[15] X. Li, Y. Ye, and M. K. Ng, "Multivcrank with applications to image retrieval," *IEEE Trans. on Image Processing*, vol. 25, no. 3, pp. 1396–1409, 2016.

[16] X. Li, M. K. Ng, and Y. Ye, "Multicomm: Finding community structure in multi-dimensional networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 4, pp. 929–941, 2014.

[17] T. H. Haveliwala, "Topic-sensitive pagerank: A context-sensitive ranking algorithm for web search," *IEEE Trans. on knowledge and Data Engineering*, vol. 15, no. 4, pp. 784–796, 2003.

[18] H. Tong, C. Faloutsos, and J.-Y. Pan, "Random walk with restart: fast solutions and applications," *Knowledge and Information Systems*, vol. 14, no. 3, pp. 327–346, 2008.

[19] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: Bringing order to the web." Stanford InfoLab, Tech. Rep., 1999.

[20] S. M. Ross, *Introduction to probability models*. Academic press, 2014.

[21] C. Preisach and L. Schmidt-Thieme, "Ensembles of relational classifiers," *Knowledge and Information Systems*, vol. 14, no. 3, pp. 249–272, 2008.

[22] M. Ji, Y. Sun, M. Danilevsky, J. Han, and J. Gao, "Graph regularized transductive classification on heterogeneous information networks," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2010, pp. 570–586.