Extracting New Facts in Knowledge Bases:-A matrix tri factorization approach

Tanmoy Mukherjee MUKHERJEE.TANMOY@GMAIL.COM
Singapore University of Technology and Design, Singapore

Vinay Pande VINAY.PANDE@RESEARCH.IIIT.AC.IN
International Institute of Information Technology Hyderabad, India

Stanley Kok STANLEYKOK@SUTD.EDU.SG
Singapore University of Technology and Design, Singapore

Abstract

Knowledge bases provide with the benefit of organizing knowledge in the *relational form* but suffer from incompleteness of new *entities* and *relationships*. Prior work on relation extraction has been focused on *supervised learning* techniques which are quite expensive. An alternative based on *distant supervision* has been of significant interest where one aligns records in the database with sentences of these records. A new line of work on embeddings of symbolic representations (Bordes et al., 2011) has shown promise. We introduce a Matrix tri factorization model which can find missing information in knowledge bases. Experiments show that we are able to query and find missing information from text and shows improvement over existing methods.

1. Introduction

Automatic text understanding has been a major challenge for scientists and AI researchers. As electronic media becomes more widely used, the amount of text in electronic form has also grown rapidly. A challenge for AI systems has been to gather,organize and make use of this massive amount of collected information. Organizing this data has wide applications in storing and indexing text for searching and retrieval, ranking of documents, classifying of documents, information extraction, question answering etc.

ICML workshop on Structured Learning: Inferring Graphs from Structured and Unstructured Inputs (SLG 2013).

There has been a recent interest in building large scale *Knowledge Bases* in the form of *multi-relational* graph data whose nodes represent entities and edges corresponds to relations. Multi-relational data plays a major role in areas such as recommendation systems, computational biology, social networks and has progressed into statistical relational learning (Getoor & Taskar, 2007). Knowledge Bases have been conceived for human-like reasoning due to the structure of the data. The nature and organization of the data aids in applications like entity resolution in NLP to image annotation in computer vision. Relations in knowledge bases can be represented as *triplets* of the form (*subject,predicate,object*) which are termed as *multi-relational graphs* (Bordes et al., 2013). Such data sources are also represented as 3-dimensional tensors where each dimension represents an adjacency matrix for a predicate.

They are quite popular in Semantic Web (Freebase,Opencyc, YAGO),and natural language processing (Wordnet).

The ability to represent complex and rich relationships makes these models quite popular for various tasks in language understanding. However they are often incomplete and have large dimensions with millions of entities and relations.

In this paper we propose a matrix tri factorization which can accurately learn to add facts in a knowledge base. We represent each entity (either subject or object) by a low dimensional vector that captures the uniqueness of facts.

1.1. Related Work

Multi-relational models have motivated applications such as collaborative filtering, link prediction in networks, and finding relations between entities. Relations can be similar or related and hence independently learning relationships for each model would be inefficient. Kemp et al. (Kemp et al., 2006) proposed a nonparametric Bayesian model called *Infinite Relational Model* (IRM) which discovers entities and possible set of relations between them. Sutskever et al. (Sutskever et al., 2009) proposed to train a factorized representation of relations which embeds a factorized representation of relation in a nonparametric Bayesian clustering framework. However these models have multiple embeddings per entity, which leads to bad generalization (Bordes et al., 2011). Paccanro et. al (Paccanaro & Hinton, 2001) proposed the *Linear Relational Embedding* where concepts were represented by distributed patterns of activity in neural networks. A natural extension to learning embeddings per entity is to formulate the problem as a matrix factorization problem. To learn multiple relationships consists of stacking multiple relations to be factorized by applying *tensor factorization* methods such as PARAFAC(Harshman & Lundy, 1994). *Collective matrix factorization* (Singh & Gordon, 2008) and RESCAL (Nickel et al., 2011) simultaneously factors several matrices, sharing parameters across factors when an entity shares multiple relations. RESCAL has shown to achieve state of the art performance on several relation datasets like YAGO (Nickel et al., 2012).

In this paper, we explore non-negative tri factorization for simultaneously clustering multiple types of entities.

2. Non Negative Matrix Tri Factorization

In this section we briefly review NMF and NTMF. In general NMF factorizes input non-negative matrix X into two non-negative matrices

$$X \approx FG^{\top} \tag{1}$$

In this paper we consider the following non negative 3-factor decomposition introduced in (Ding et al., 2006)

$$X \approx FSG^{\top} \tag{2}$$

where $F^{\top}F = I$ and $G^{\top}G = I$

We consider data set $\mathcal{X} = \{\mathcal{X}_1, \mathcal{X}_2, ..., \mathcal{X}_K\}$, where $\mathcal{X}_K = \{\mathbf{x_1^k}, \mathbf{x_2^k}, ..., \mathbf{x_{n_k}^k}\}$ represent data object of k type. We are given a set of relationship matrices $\{R_{kl} \in \mathcal{R}^{n_k \times n_l}\}_{(1 \le k \le K, 1 \le l \le K)}$. Our task is to cluster the dataset $\mathcal{X}$ into different clusters maintaining pairwise affinity. We construct R,G,S and W following (Wang et al., 2008)

Algorithm 1 Algorithm to solve GNNTF

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 and W
2. Initialize G .

repeat
3. Compute $S = (G^T G)^{-1} G^T R G (G^T G)^{-1}$
4. Update $G_{ij} \leftarrow G_{ij}[\frac{(RGS + \lambda WG)_{ij}}{(GSG^T SG + \lambda DG)_{ij}}]$

until Converges

2.1. Objective Function

Ding et. al(Ding et al., 2006) proposed NMTF to simultaneously cluster rows and columns of an input non-negative matrix by decomposing in three non-negative factor matrices.

$$X \approx G_1 S G_2^T \tag{3}$$

We write the objective function where the input is the relationship matrix R_{12}

$$J_1 = ||R_{12} - G_1 S_{12} G_2^T|| \quad s.t, G_1 \ge 0, G_2 \ge 0, S_{12} \ge 0 \tag{4}$$

where $||.||$ denotes the Frobenius norm of the matrix.

The objective function only incorporates the inter type relationships. For intra type relationships, we incorporate Laplacian regularization (Cai et al., 2011) and rewrite Eqn(2) as:

$$J_2 = ||R_{12} - G_1 S_{12} G_2^T||^2 + 2\lambda[\mathrm{Tr}(G_1^T L_1 G_1) + \mathrm{Tr}(G_2^T L_2 G_2)] \tag{5}$$

where $L_k = D_k - W_k$ is the corresponding graph Laplacian, D_k is the diagonal degree matrix with $D_{ii} = \sum_j W_{ij}$. Simultaneous clustering on $\mathcal{X}_1$ and $\mathcal{X}_2$ is then achieved by solving Equation (2), the cluster label $\mathbf{x_i^k}$ is obtained by:

$$l(\mathbf{x_i^k}) = \arg\max_{\mathbf{j}} \mathbf{G_{k(ij)}}$$

We rewrite equation (5) as

$$\begin{array}{ll} \text{minimize} & J_{NNTF} = ||R - GSG^T|| + 2\lambda \, \mathrm{Tr}[G^T L G] \\ \text{s.t} & G \ge 0, S \ge 0 \end{array} \tag{6}$$

Equation (6) is our major objective function.

Datasets		Latent Factor Model	RESCAL	MRC	SME	Our Approach
Kinships	Area under PR curve	0.946 ± 0.005	0.95	0.84	0.907 ± 0.008	0.96
	Log-likelihood	-0.029 ± 0.001	N/A	-0.045 ± 0.002	N/A	-0.028 ± 0.001
UMLS	Area under PR curve	0.990 ± 0.003	0.98	0.98	0.983 ± 0.003	0.990
	Log-likelihood	-0.002 ± 0.0003	N/A	-0.004 ± 0.001	N/A	-0.002 ± 0.001
Nations	Area under PR curve	0.909 ± 0.009	0.84	0.75	0.883 ± 0.02	0.93
	Log-likelihood	-0.202 ± 0.008	N/A	-0.311 ± 0.022	N/A	-0.198 ± 0.008

Table 1. Comparisons of the performance obtained by our approach and RESCAL,MRC,SME and Latent Factor Model

Dataset	Train. size	Test size	Labeled	Symbols
WordNet	216017	5000	No	synsets
ConceptNet	11332	0	No	lemmas
Wikipedia	1498298	0	No	lemmas
Extended WordNet	786105	5000	Yes	lemmas+synsets
Unambig. Wikipedia	981841	0	Yes	lemmas+synsets

Table 2. Multiple data sources used for learning representations of lemmas and synsets. Labeled indicates when triplets consist of text lemmas for which the corresponding synsets are known.

We are omitting details of the optimization algorithm and the optimization procedure. For further details we refer readers to (Ding et al., 2006; Cai et al., 2011; Wang et al., 2008)

3. Experiments

3.1. Data

We perform experiments on standard multi type relational data and standard heterogeneous source of knowledge from web (Bordes et al., 2011). For multirelational data we compare our approach with standard tensor factorization datasets (Jenatton et al., 2012) and also with (Nickel et al., 2012; Kok & Domingos, 2007; Bordes et al., 2011) as shown in Table 1. For experimental purpose we construct neighborhood graph from the relationship matrix (Cai et al., 2011) to obtain pairwise affinity matrices for both types of entities. We set it to 10. Similar to (Chen et al., 2013) we initialize our word vectors with randomly initialized word vectors and pre-trained vectors from the unsupervised model of (Collobert & Weston, 2008) and also from Brown clusters (Brown et al., 1992).

From Table 1 we can see that our results are comparable with (Jenatton et al., 2012)

3.2. Ranking and Classification

Chen et. al (Chen et al., 2013) computes a score for each triplet for all other entities in the knowledge base $e \in E$ and sort them based on a descending order. Questions relating to triples could have multiple answers and hence we report the percentage of times i.e recall with higher numbers showing correct entity has been correctly estimated. Our model obtains a ranking recall score of 15.3 % while Chen et. al (Chen et al., 2013) obtains 20.9 %.

For the task of correctly estimating if a relation is true or not we achieve an accuracy of 72% while the Neural Tensor model achieves 75%. The Hadamard model and the similarity model (Chen et al., 2013) achieves 66.7 % and 51.6 % respectively.

4. Conclusion

We introduced a Matrix Tri Factorization model with pairwise constraints for predicting relations in Knowledge Bases. This paper is intended to explore efficient matrix factorization method for predicting or extracting relations. A line of direction which we are interested is similar to Universal schema for Open IE (Riedel et al., 2013) In future, we will try to integrate Linked Open Data similar to (Nickel et al., 2012)

References

Bordes, A., Weston, J., Collobert, R., and Bengio, Y. Learning structured embeddings of knowledge bases. In *AAAI*, 2011.

Bordes, Antoine, Glorot, Xavier, Weston, Jason, and Bengio, Yoshua. A semantic matching energy function for learning with multi-relational data. *Ma-*

chine Learning: Special Issue on Learning Semantics, 2013.

Brown, Peter F., deSouza, Peter V., Mercer, Robert L., Pietra, Vincent J. Della, and Lai, Jenifer C. Class-based n-gram models of natural language. *Computational Linguistics*, 18:467–479, 1992.

Cai, Deng, He, Xiaofei, Han, Jiawei, and Huang, Thomas S. Graph regularized non-negative matrix factorization for data representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(8):1548–1560, 2011.

Chen, Danqi, Socher, Richard, Manning, Christopher D., and Ng, Andrew Y. Learning new facts from knowledge bases with neural tensor networks and semantic word vectors. *CoRR*, abs/1301.3618, 2013.

Collobert, R. and Weston, J. A unified architecture for natural language processing: Deep neural networks with multitask learning. In *International Conference on Machine Learning, ICML*, 2008.

Ding, Chris, Li, Tao, Peng, Wei, and Park, Haesun. Orthogonal nonnegative matrix t-factorizations for clustering. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '06, pp. 126–135, New York, NY, USA, 2006. ACM. ISBN 1-59593-339-5. doi: 10.1145/1150402.1150420. URL http://doi.acm.org/10.1145/1150402.1150420.

Getoor, Lise and Taskar, Benjamin. *Introduction to Statistical Relational Learning*. The MIT Press, 2007.

Harshman, Richard A. and Lundy, Margaret E. Parafac: Parallel factor analysis. *Computational Statistics & Data Analysis*, 18(1):39 – 72, 1994. ISSN 0167-9473. doi: 10.1016/0167-9473(94)90132-5.

Jenatton, Rodolphe, Le Roux, Nicolas, Bordes, Antoine, and Obozinski, Guillaume. A latent factor model for highly multi-relational data. In *NIPS 2012 - Neural Information Processing Systems*, Lake Tahoe, Nevada, United States, 2012. URL http://hal.inria.fr/hal-00776335.

Kemp, Charles, Tenenbaum, Joshua B., Griffiths, Thomas L., Yamada, Takeshi, and Ueda, Naonori. Learning systems of concepts with an infinite relational model. In *Proceedings of the 21st national conference on Artificial intelligence - Volume 1*, AAAI'06, pp. 381–388. AAAI Press, 2006. ISBN 978-1-57735-281-5. URL http://dl.acm.org/citation.cfm?id=1597538.1597600.

Kok, Stanley and Domingos, Pedro. Statistical predicate invention. In *In Z. Ghahramani (Ed.), Proceedings of the 24th annual international conference on machine learning (ICML-2007*, pp. 433–440, 2007.

Nickel, Maximilian, Tresp, Volker, and Kriegel, Hans-Peter. A three-way model for collective learning on multi-relational data. In Getoor, Lise and Scheffer, Tobias (eds.), *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, ICML '11, pp. 809–816, New York, NY, USA, June 2011. ACM. ISBN 978-1-4503-0619-5.

Nickel, Maximilian, Tresp, Volker, and Kriegel, Hans-Peter. Factorizing yago: scalable machine learning for linked data. In *Proceedings of the 21st international conference on World Wide Web*, WWW '12, pp. 271–280, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1229-5. doi: 10.1145/2187836.2187874. URL http://doi.acm.org/10.1145/2187836.2187874.

Paccanaro, Alberto and Hinton, Geoffrey E. Learning distributed representations of concepts using linear relational embedding. *IEEE Trans. Knowl. Data Eng.*, 13(2):232–244, 2001.

Riedel, Sebastian, Yao, Limin, Marlin, Benjamin M., and McCallum, Andrew. Relation extraction with matrix factorization and universal schemas. In *Joint Human Language Technology Conference/Annual Meeting of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL '13)*, June 2013.

Singh, Ajit P. and Gordon, Geoffrey J. Relational learning via collective matrix factorization. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '08, pp. 650–658, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-193-4. doi: 10.1145/1401890.1401969. URL http://doi.acm.org/10.1145/1401890.1401969.

Sutskever, Ilya, Salakhutdinov, Ruslan, and Tenenbaum, Josh. Modelling relational data using bayesian clustered tensor factorization. *Advances in neural information processing systems*, 22:1821–1828, 2009.

Wang, Fei, Li, Tao, and Zhang, Changshui. Semi-supervised clustering via matrix factorization, 2008.