Pairwise Tensor Factorization for learning new facts in Knowledge Bases

Tanmoy Mukherjee
IIIT Hyderabad
Hyderabad, India
mukherjee.tanmoy
@gmail.com

Vinay Pande
IIIT Hyderabad
Hyderabad, India
vinay.pande
@research.iiit.ac.in

Vasudeva Varma
IIIT Hyderabad
Hyderabad, India
vv@iiit.ac.in

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 approach based on *distant supervision* has been of significant interest where one aligns database records with sentences of these records. A new line of work on embeddings of symbolic representations [2] 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.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous; D.2.8 [**Software Engineering**]: Metrics—*complexity measures, performance measures*

General Terms

Theory

Keywords

ACM proceedings, LaTeX, text tagging

1. INTRODUCTION

Statistical relational learning analyzes the interaction between entities and their relationships and is an growing area in machine learning. Recent years has seen an increasing interest in analysis pairwise relational data that contains several entity types and multiple relations.

A challenge for AI systems has been to gather,organize and make use of massive amount of collected information. There has been a recent interest in building large scale *Knowledge Bases*(KB) which are *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 [9]. Manually created knowledge bases often lack information about entities and their properties, either due to missing information from the source used to create the knowledge base or because human annotators were unable to add facts due to imprecise knowledge. There has been a great deal of recent research focus on extracting knowledge from text [1, 8, 17] but little work has beed done on a problem of drawing inference from this imperfect extracted knowledge. For example, given the facts *PlayForTeam("Lionel Messi","Barcelona")* and *TeamPlaysInLeague("Barcelona","La Liga")* in knowledge base, can we infer and add new fact *PlayerPlaysInLeague("Lionel Messi","La Liga")* into the knowledge base. While existing systems like NELL[3] try to infer such facts by minning data from their knowledge bases and using limited set of rules, adding new fact takes much efforts in terms of creating new inference rules. While traditional logical inference techniques are too brittle to make inference, probablistic inference techniques [16] suffer from scalability issues. In this paper we introduce a model that can accurately learn to add additional facts to a database using only that database. We represent each entity by a low dimensional representation which captures local and global semantics [11].

Our model is flexible and allows us to query and find if certain entities that were not originally in the knowledge base are in certain relationships by exploring distributional word vectors. These vectors are learned by training a neural network model [11] using wikipedia and uses local and global properties. The word vectors thus are able to capture semantic and syntactic proeprties and allows to extend the database without parsing of any additional textual resources.

2. RELATED WORK

There is a vast amount of work done in extending knowledge bases using external corpora [8, 17]. However most techniques use a predefined, finite and fixed schema of relation types. Some text is labeled according to this schema, and this labeling is then used in supervised training for automatic extracting of relations. There has been an increased interest in learning with weak supervision where one aligns a database record with a sentence in which these records appear, effectively labeling the text which is further used to train a machine learning system [7, 13]. However this

Eleventh Workshop on Mining and Learning with Graphs. Chicago, Illinois, USA

	PRA		Pairwise Tensor Model	Neural Network [6]
Task	Paths	p@100	p@100	p@100
athletePlaysForTeam	125	0.46	0.42	0.48
from their knowledge bases athletePlaysInLeague	15	0.84	0.86	0.85
athletePlaysSport	34	0.78	**0.82**	0.80
stadiumLocatedInCity	18	0.62	**0.68**	0.65
teamHomeStadium	66	0.48	0.52	0.52
teamPlaysInCity	29	0.86	**0.92**	0.9
teamPlaysInLeague	36	0.70	**0.80**	0.78
companiesHeadquarteredIn	42	0.54	0.58	0.56
publicationJournalist	25	0.70	0.74	0.73
teamWonTrophy	56	0.50	**0.55**	0.52
worksFor	62	0.60	**0.68**	0.64

Table 1: Evaluation on Nell knowledge base

method relies on the availibility of an existing database with a fixed schema. There has been minimal work done in drawing reliable inference in noisy knowledge bases. The closest work has been in using path ranking algorithm [12] where the authors showed that a soft inference procedure based on combination of constrained, weighted random walk through a knowledge base is able to reliably infer new facts or belifs.

3. TENSOR FACTORIZATION MODEL

Factorization models for tensors are studied in several fields. The most general model is the Tucker Decomposition [19] which has been studies in problems in tag recommendation [18, 14]. A special case of Tucker Decomposition is Canonical Decomposition (CD) [4] also known as PARAFAC [10]. An issue with CD is that it considers only the three wise interactions however we would like to consider interactions between the enitities and their relations. We consider the pairwise tensor model (PITF) [15] which considers various interactions between the two inputs.

We compare with randomly initialized word vectors, pretrained word vectors with 100 dimensional word vectors from the model of Collobert[6]. Using wikipedia text, the neural network model learns word vectors using the local context and the global document context. The resulting word vectors thus capture syntactic and semantic information.

Let $e_1, e_2 \in \mathcal{R}^d$ be the vector representation of two entities. PITF simply models the two way interactions between entities and its relationships. We compute a score $f(e_1, R, e_2)$ by factorizing in two way interactions

$$f(e_1, R, e_2) = \sum_{i=1}^{d}(e_1^T W_R^{[1:k]}) + \sum_{i=1}^{d}(e_1^T e_2) \qquad (1)$$

We define $W^{[1:k]} \in \mathcal{R}^{d \times d \times k}$ as a tensor.

Since we are interested in ranking or finding missing entities, the pairwise model is able to model the interactions between the entities as well as between the entity and relationship.

In order to train this model we minimize the following objective function

$$J(W, E) = \sum_{i=1}^{N}\sum_{c=1}^{C} max(0, 1 - f(e_1^{(i)}, R^{(i)}, e_2^{(i)} + f(e_1^i, R^{(i)}, e_{neg}))) \qquad (2)$$

where N is the number of training examples and we score the correct triple higher than a negative triple. We choose a $e_n eg$ a corrputed entity by picking a random entity and create a corrputed triplet. The model is trained by using stochastic gradient descent similar to the training procedure as [2]

4. EXPERIMENTS

In our experiments, we follow similar experiments as in [5] for Wordnet. There are a total of 38,966 entities and 11 relations. We use 112,581 triples for training and 10,544 for test. We consider the wordnet relationships - has instance, type of,member meronym, member holonym, part of, has part, subordinate instance of, domain region, synset domain region, similar to, and domain topic.

For each triplet (e_1, R, e), we compute the score $f(e_1, R, e)$ for all other entities in the knowledge base. We sort the results in descending order and report the top ranked entity as correct entity. Our pairwise tensor model was able to obtain a score of 25% while the neural models achieve 10.6%.

We also report experiments on NELL knowledge base and compare with the path ranking algorithm (PRA) [12]. Results are as shown in Table 1. From Table 1, our approach performs significantly better than PRA and neural network approaches. Specifically, Our approach performs better for relations where direct inferencing is very difficult.

4.1 Classification

In this experiment we ask the model whether a set of triples is true or not. With the help of a vocabulary of semantic word vectors, we can query whether certain relationships hold or not even for entities that were not originally in Wordnet.

We use a development set to find a threshold for each relation such that $f(e_1, R, e_2) \geq T_R$ the relation holds or else it is false. As explained in the training stage we randomly switch entities and relations from correctly testing triplets. We observe an accuracy of 78% with semantic word vectors. In contrast the model by Collobert [6] acheived 66.7%. There is decrese in performance if the entities were initialized randomly.

5. CONCLUSION

In this paper we use the *Pairwise Tensor Model* where we model pariwise relations between the entity and its relationships. Similar to [5] this model has better performance for ranking as well as for predicting unseen relationships between entities. It enables the extension of a knowledge base without external textual corpora.

6. REFERENCES

[1] E. Agichtein and L. Gravano. Snowball: extracting relations from large plain-text collections. In *Proceedings of the fifth ACM conference on Digital libraries*, DL '00, pages 85–94, New York, NY, USA, 2000. ACM.

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

[3] A. Carlson, J. Betteridge, B. Kisiel, B. Settles, E. R. Hruschka Jr, and T. M. Mitchell. Toward an architecture for never-ending language learning. In *AAAI*, 2010.

[4] J. Carroll and J.-J. Chang. Analysis of individual differences in multidimensional scaling via an n-way generalization of âĂIJeckart-youngâĂİ decomposition. *Psychometrika*, 35(3):283–319, September 1970.

[5] D. Chen, R. Socher, C. D. Manning, and A. Y. Ng. Learning new facts from knowledge bases with neural tensor networks and semantic word vectors. *CoRR*, abs/1301.3618, 2013.

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

[7] M. Craven and J. Kumlien. Constructing biological knowledge bases by extracting information from text sources. pages 77–86. AAAI Press, 1999.

[8] O. Etzioni, M. Cafarella, D. Downey, A.-M. Popescu, T. Shaked, S. Soderland, D. S. Weld, and A. Yates. Unsupervised named-entity extraction from the web: an experimental study. *Artif. Intell.*, 165(1):91–134, June 2005.

[9] L. Getoor and B. Taskar. *Introduction to Statistical Relational Learning*. The MIT Press, 2007.

[10] R. Harshman. Foundations of the parafac procedure: Models and conditions for an "explanatory" multi-modal factor analysis. *UCLA Working Papers in Phonetics*, 16, 1970.

[11] E. H. Huang, R. Socher, C. D. Manning, and A. Y. Ng. Improving word representations via global context and multiple word prototypes. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers - Volume 1*, ACL '12, pages 873–882, Stroudsburg, PA, USA, 2012. Association for Computational Linguistics.

[12] N. Lao, T. Mitchell, and W. W. Cohen. Random walk inference and learning in a large scale knowledge base. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '11, pages 529–539, Stroudsburg, PA, USA, 2011. Association for Computational Linguistics.

[13] M. Mintz, S. Bills, R. Snow, and D. Jurafsky. Distant supervision for relation extraction without labeled data. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2 - Volume 2*, ACL '09, pages 1003–1011, Stroudsburg, PA, USA, 2009. Association for Computational Linguistics.

[14] S. Rendle, L. Balby Marinho, A. Nanopoulos, and L. Schmidt-Thieme. Learning optimal ranking with tensor factorization for tag recommendation. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '09, pages 727–736, New York, NY, USA, 2009. ACM.

[15] S. Rendle and L. Schmidt-Thieme. Pairwise interaction tensor factorization for personalized tag recommendation. In *Proceedings of the third ACM international conference on Web search and data mining*, WSDM '10, pages 81–90, New York, NY, USA, 2010. ACM.

[16] M. Richardson and P. Domingos. Markov logic networks. *Mach. Learn.*, 62(1-2):107–136, Feb. 2006.

[17] R. Snow, D. Jurafsky, and A. Y. Ng. Semantic taxonomy induction from heterogenous evidence. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, ACL-44, pages 801–808, Stroudsburg, PA, USA, 2006. Association for Computational Linguistics.

[18] P. Symeonidis, A. Nanopoulos, and Y. Manolopoulos. Tag recommendations based on tensor dimensionality reduction. In *Proceedings of the 2008 ACM conference on Recommender systems*, RecSys '08, pages 43–50, New York, NY, USA, 2008. ACM.

[19] L. Tucker. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31(3):279–311, September 1966.