

Neural Latent Relational Analysis to Capture Lexical Semantic Relations in a Vector Space

Koki Washio^{*†1} and Tsuneaki Kato^{*2}

^{*}Department of Language and Information Sciences, The University of Tokyo

[†]RIKEN Center for Advanced Intelligence Project

{¹kokiwashio@g.ecc, ²kato@boz.c}.u-tokyo.ac.jp

Abstract

Capturing the semantic relations of words in a vector space contributes to many natural language processing tasks. One promising approach exploits lexico-syntactic patterns as features of word pairs. In this paper, we propose a novel model of this pattern-based approach, neural latent relational analysis (NLRA). NLRA can generalize co-occurrences of word pairs and lexico-syntactic patterns, and obtain embeddings of the word pairs that do not co-occur. This overcomes the critical data sparseness problem encountered in previous pattern-based models. Our experimental results on measuring relational similarity demonstrate that NLRA outperforms the previous pattern-based models. In addition, when combined with a vector offset model, NLRA achieves a performance comparable to that of the state-of-the-art model that exploits additional semantic relational data.

1 Introduction

Information on the semantic relations of words is important for many natural language processing tasks, such as recognizing textual entailment, discourse classification, and question answering. There are two main approaches to obtain the distributed relational representations of word pairs.

One is the vector offset method (Mikolov et al., 2013a,b). This approach represents word pairs as the vector offsets of their word embeddings. Another approach exploits lexico-syntactic patterns to obtain word pair representations. As a pioneer work, Turney (2005) introduced latent relational analysis (LRA), based on the *latent relation hypothesis*. It states that word pairs that co-occur in similar lexico-syntactic patterns tend to have similar semantic relations (Turney, 2008b; Turney and Pantel, 2010). LRA is expected to complement

the vector offset model because word embeddings do not contain information on lexico-syntactic patterns that connect word pairs in a corpus (Shwartz et al., 2016).

However, LRA cannot obtain the representations of word pairs that do not co-occur in a corpus. Even with a large corpus, observing a co-occurrence of all semantically related word pairs is nearly impossible because of Zipf's law, which states that most content words rarely occur. This data sparseness problem is a major bottleneck of pattern-based models such as LRA.

In this paper, we propose neural latent relational analysis (NLRA) to solve that data sparseness problem. NLRA unsupervisedly learns the embeddings of target word pairs and co-occurring patterns from a corpus. In addition, it jointly learns the mapping from the word embedding space to the word-pair embedding space. By this mapping, NLRA can generalize the co-occurrences of word pairs and patterns, and obtain the relational embeddings for arbitrary word pairs even if they do not co-occur in the corpus.

Our experimental results on the task of measuring relational similarity show that NLRA significantly outperforms LRA, and it can also capture semantic relations of word pairs without co-occurrences. Moreover, we show that combining NLRA and the vector offset model improves the performance and leads to competitive results to those of the state-of-the-art method that exploits additional semantic relational data.

2 Background

2.1 Vector Offset Model

The vector offset model (Mikolov et al., 2013a,b; Levy and Goldberg, 2014) obtains word embeddings from a corpus and represents each word pair (a, b) as the vector offset of their embedding as

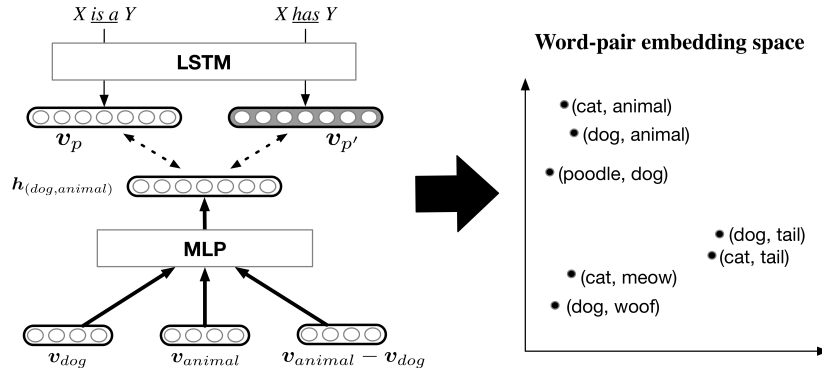


Figure 1: An illustration of NLRA

follows:

$$\mathbf{v}_{(a,b)} = \mathbf{v}_b - \mathbf{v}_a \quad (1)$$

where \mathbf{v}_a and \mathbf{v}_b are the word embeddings of a and b respectively.

This method regards relational information as the change in multiple topological dimensions from one word to the other in the word embedding space (Zhila et al., 2013). Meanwhile, it does not contain the information of lexico-syntactic patterns that were shown to capture complementary information with word embeddings in previous studies on the lexical semantic relation detection (Levy et al., 2015; Schwartz et al., 2016).

2.2 Latent Relational Analysis

LRA takes a set of word pairs as input and generates the distributed representations of those word pairs based on their co-occurring patterns.

Given target word pairs $W = \{(a_1, b_1), \dots, (a_n, b_n)\}$, LRA constructs a list of lexico-syntactic patterns that connect those pairs, such as *is a* or *in the*, from the corpus for each word pair. Then, those patterns are generalized by replacing any or all or none of the intervening words with wildcards. As a feature selection, the generalized patterns generated from many word pairs are used as features. We define the set of these target feature patterns as $C = \{p_1, \dots, p_m\}$. Then, the $2n \times 2m$ matrix M is constructed. The rows of M correspond to pairs (a_i, b_i) and reversed pairs (b_i, a_i) . The columns of M correspond to patterns Xp_iY and swapped patterns Yp_iX , where X and Y are the slots for the words of the word pairs. The value of M_{ij} represents the strength of the association between the corresponding word pair and pattern, which is calculated using weighting methods such as positive pointwise mutual information

(PPMI). After these processes, the singular value decomposition (SVD) is applied to M , and the vector $\mathbf{v}_{(a,b)}$ is assigned to each word pair (a, b) .

Although pattern-based approaches such as LRA have achieved promising results in some semantic relational tasks (Turney, 2008a,b), they have a crucial problem that a co-occurrence of all semantically related word pairs cannot be observed because of Zipf’s law, which states that the frequency distribution of words has a long tail. In other words, most words occur very rarely (Hanks, 2009). For the word pairs without co-occurrences, LRA cannot obtain their vector representations.

3 Neural Latent Relational Analysis

We introduce NLRA, based on the latent relation hypothesis. NLRA represents the target word pairs and lexico-syntactic patterns as embeddings. Similar to the skip-gram model (Mikolov et al., 2013a), NLRA updates those representations unsupervisedly, such that the inner products of the word pairs and patterns in which they co-occur in a corpus have high values. Through this learning, the word pairs that co-occur in similar patterns have similar embeddings. Moreover, NLRA can generalize the co-occurrences of the word pairs and patterns by constructing the embeddings of the word pairs from their word embeddings, thus solving the data sparseness problem of word co-occurrences. Therefore, NLRA can provide representations that capture the information of lexico-syntactic patterns even for the word pairs that do not co-occur in a sentence.

Figure 1 is an illustration of our model. NLRA encodes a word pair (a, b) into a dense vector as follows:

$$\mathbf{h}_{(a,b)} = MLP([\mathbf{v}_a; \mathbf{v}_b; \mathbf{v}_b - \mathbf{v}_a]) \quad (2)$$

where $[v_a; v_b; v_b - v_a]$ is the concatenation of the word embeddings of a and b and their vector offsets; *MLP* is a multilayer perceptron with nonlinear activation functions.

A pattern p is a sequence of the words w_1, \dots, w_k . The sequence of the corresponding word embeddings w_1, \dots, w_k are encoded using long short-term memory (LSTM) (Hochreiter and Schmidhuber, 1997). Then, the final output vector v_p is used as the pattern embedding.

For unsupervised learning, we use the negative sampling objective (Mikolov et al., 2013a). Given a set of observed triples $(a, b, p) \in D$, where a and b are words such that $(a, b) \in W$, or $(b, a) \in W$ and p is a co-occurring pattern from a corpus, the objective is as follows:

$$L = \sum_{(a,b,p) \in D} \log \sigma(v_p \cdot h_{(a,b)}) + \sum_{(a,b,p') \in D'} \log \sigma(-v_{p'} \cdot h_{(a,b)}) \quad (3)$$

where D' is a set of randomly generated negative samples and σ is the sigmoid function. We sampled 10 negative patterns for each word pair. This objective is maximized using the stochastic gradient descent.

After unsupervised learning, we can obtain word pair representations $v_{(a,b)}$ as follows:

$$v_{(a,b)} = [h_{(a,b)}; h_{(b,a)}] \quad (4)$$

4 Evaluation

4.1 Dataset

In our evaluation, we used the SemEval-2012 Task 2 dataset (Jurgens et al., 2012) for the task of measuring relational similarity. This dataset contains a collection of 79 fine-grained semantic relations. For each relation, there are a few prototypical word pairs and a set of several dozen target word pairs. The task is to rank the target pairs based on the extent to which they exhibit the relation. In our experiment, we calculated the score of a target word pair with the average cosine similarity between it and each prototypical word pair. The models are evaluated in terms of the MaxDiff accuracy and Spearman’s correlation. Following previous works (Rink and Harabagiu, 2012; Zhila et al., 2013), we used the test set that includes 69 semantic relations to evaluate the performance.

4.2 Baselines

VecOff. We used the 300-dimensional pre-trained GloVe (Pennington et al., 2014)¹ and represented word pairs as described in Section 2.1.

LRA. We implemented LRA as described in Section 2.2. We set W as the lemmatized word pairs of the dataset. We used the English Wikipedia as a corpus. For each word pair, we searched for patterns of from one to three words. When searching for patterns, the left word and right word adjacent to the patterns were lemmatized to ignore their inflections. Following (Turney, 2008b), we selected C as the top $20|W|$ generalized patterns. Then, M was constructed using PPMI weighting, and its dimensionality was reduced to 300 using SVD.

4.3 Our methods

NLRA. For each word pair in the dataset, co-occurring patterns were extracted from the same corpus in the same manner as with LRA, resulting in D . For word embeddings, we used the same pre-trained GloVe as VecOff. These embeddings were updated during the training. For *MLP*, we used three affine transformations followed by the batch normalization (Ioffe and Szegedy, 2015) and tanh activation. The size of each hidden layer of the MLP was 300. To encode the patterns, we used LSTM with the 300-dimensional hidden state. The objective was optimized by AdaGrad (Duchi et al., 2011) (whose learning rate was 0.01). We trained the model for 50 epochs.

NLRA+VecOff. This method combines NLRA and VecOff by averaging their score for a target word pair.

4.4 Result and Analysis

Table 1 displays the overall result.

NLRA vs. LRA

First, NLRA outperformed LRA in terms of both the average accuracy and correlation. These differences were statistically significant ($p < 0.01$) with the paired t-test. These results indicate that generalizing patterns with LSTM is better than by using wildcards. Moreover, NLRA can successfully calculate the relational similarity for the word pairs that do not co-occur in the corpus. Table 2 shows an example of the Reference–Express relation, where the middle-score pair *handshake:cordiality*

¹<https://nlp.stanford.edu/projects/glove/>

Relation	Accuracy				Correlation			
	VecOff	LRA	NLRA	NLRA+VecOff	VecOff	LRA	NLRA	NLRA+VecOff
Class-Inclusion	0.543	0.485	0.533	0.56	0.487	0.427	0.622	0.611
Part-Whole	0.45	0.427	0.465	0.488	0.304	0.282	0.38	0.395
Similar	0.414	0.346	0.412	0.436	0.267	0.123	0.271	0.315
Contrast	0.343	0.349	0.377	0.374	0.108	0.065	0.092	0.124
Attribute	0.462	0.414	0.447	0.486	0.406	0.299	0.367	0.456
Non-Attribute	0.39	0.366	0.369	0.381	0.217	0.16	0.125	0.174
Case Relations	0.468	0.438	0.536	0.558	0.391	0.291	0.553	0.544
Cause Purpose	0.444	0.471	0.448	0.485	0.345	0.387	0.397	0.454
Space-Time	0.5	0.428	0.516	0.525	0.424	0.31	0.489	0.493
Reference	0.441	0.447	0.449	0.465	0.297	0.346	0.404	0.378
Average	0.443	0.415	0.453	0.475	0.321	0.246	0.36	0.391

Table 1: Average MaxDiff accuracy and Spearman’s correlation of each major relation group.

Pair	Human	LRA	NLRA
laugh:happiness	50	0.217	0.578
nod:agreement	46	0.245	0.347
tears:sadness	44	0.381	0.483
...
scream:terror	26	0.396	0.417
handshake:cordiality	24	0 (no pattern)	0.34
lie:dishonesty	16	0.206	0.394
...
discourse:relationship	-60	0.331	0.275
friendliness:wink	-68	0 (no pattern)	0.26

Table 2: The scores assigned by humans, LRA, and NLRA for the Reference-Express relation. The pairs are sorted in descending order according to the human score.

and the low-score pair *friendliness:wink* have no co-occurring pattern. In these cases, LRA could not obtain the representations of those word pairs nor correctly assign the score. By contrast, NLRA could accomplish both because it could generalize the co-occurrences of word pairs and patterns.

NLRA+VecOff vs. Other Models

Second, NLRA+VecOff outperformed the other models. These differences were statistically significant (the correlation difference between NLRA+Vecoff and NLRA: $p < 0.05$; the other differences: $p < 0.01$). These results indicate that lexico-syntactic patterns and the vector offset of word embeddings capture complementary information for measuring relational similarity. This is inconsistent with the findings of [Zhila et al. \(2013\)](#). That work combined heterogeneous models, such as the vector offset model, pattern-based model, etc., and stated that the pattern-based model was less significant than the vector offset model, based on their ablation study. We believe that this was because their pattern-based model did not generalize patterns with wildcards nor select useful features. Their pattern-based model seemed to suffer from sparse feature space. In our experiment, NLRA helped VecOff, for example, for the Part-Whole relation, Cause Purpose rela-

Model	Accuracy	Correlation
Rink and Harabagiu (2012)	0.394	0.229
Mikolov et al. (2013b)	0.418	0.275
Levy and Goldberg (2014)	0.452	-
Zhila et al. (2013)	0.452	0.353
Iacobacci et al. (2015)	0.459	0.358
Turney (2013)	0.472	0.408
VecOff	0.443	0.321
LRA	0.415	0.264
NLRA	0.453	0.36
NLRA+VecOff	0.475	0.391

Table 3: Published results of other models on the SemEval2012 Task 2 dataset.

tion, and Space-Time relation, where there seemed to be prototypical patterns indicating those relations. Meanwhile, VecOff helped NLRA for the Attribute relation, where the relational patterns seemed to be diverse. These results showed that the combined model is robust.

4.5 Comparison to other systems

We compared the results of our models to other published results. Table 3 displays those results. [Rink and Harabagiu \(2012\)](#) is the pattern-based model with naive Bayes. [Mikolov et al. \(2013b\)](#), [Levy and Goldberg \(2014\)](#), and [Iacobacci et al. \(2015\)](#) are the vector offset models. [Zhila et al. \(2013\)](#) is the model composed of various features. [Turney \(2013\)](#) extracts the statistical features of two word pairs from a word-context co-occurrence matrix and trains the classifier with additional semantic relational data to assign a relational similarity for two word pairs.

NLRA+VecOff achieved a competitive performance to the state-of-the-art method of [Turney \(2013\)](#). Note that our method learns unsupervisedly and does not exploit additional resources, and the method of [Turney \(2013\)](#) cannot obtain the distributed representation of word pairs.

A work similar to ours, [Bollegala et al. \(2015\)](#), represented lexico-syntactic patterns as the vector offset of co-occurring word pairs and updated the

vector offsets of word pairs such that word pairs that co-occur in similar patterns have similar offsets. They evaluated their model on all 79 semantic relations of the dataset and achieved 0.449 accuracy. In their setting, NLRA+VecOff achieved 0.47 accuracy, outperforming their model.

5 Related Work

5.1 Word Pairs and Co-occurring Patterns

Hearst (1992) detected the hypernymy relation of word pairs from a corpus using several handcrafted lexico-syntactic patterns. Turney and Littman (2005) used 64 handcrafted lexico-syntactic patterns as features of word pairs to represent word pairs as vectors. To obtain word-pair embeddings, Turney (2005) extended the method of Turney and Littman (2005) as LRA. Our work is a neural extension of LRA.

Washio and Kato (2018) proposed the method similar to ours in lexical semantic relation detection. Their neural method modeled the co-occurrences of word pairs and dependency paths connecting two words to alleviate the data sparseness problem of pattern-based lexical semantic relation detection. While they assigned randomly initialized embeddings to each dependency path, our work encodes co-occurring patterns with LSTM for better generalization. Jameel et al. (2018) embedded word pairs with the context words occurring around word pairs instead of lexico-syntactic patterns. Their method cannot obtain embeddings of word pairs that do not co-occur in a corpus because they directly assigned embeddings to word pairs. By contrast, NLRA can obtain embeddings for those word pairs.

In another research area, relation extraction, several works have explored an idea similar to the latent relation hypothesis (Riedel et al., 2013; Toutanova et al., 2015; Verga et al., 2017). They factorized a matrix of entity pairs and co-occurring patterns, while they focused on named entity pairs instead of word pairs and did not consider co-occurrence frequencies.

5.2 Relation to Knowledge Graph Embedding

Knowledge graph embedding (KGE) embeds entities and relations in knowledge graph (KG), where entities and relations corresponds to nodes and edges respectively (Nickel et al., 2011; Bordes et al., 2013; Socher et al., 2013; Wang et al., 2014;

Lin et al., 2015; Yang et al., 2015; Nickel et al., 2016; Trouillon et al., 2016; Liu et al., 2017; Wang et al., 2017; Ishihara et al., 2018). By considering words and lexico-syntactic patterns as nodes and edges, respectively, a corpus can be viewed as a graph, i.e., corpus graph (CG). Thus, NLRA can be regarded as corpus graph embedding (CGE) models based on the latent relation hypothesis.

Although KGE models can be easily applied to CG, several differences exist between KG and CG. First, the nodes and edges of CG are (sequences of) linguistic expressions, such as tokens, lemmas, phrases, etc. Thus, the nodes and edges of CG might exhibit compositionality and ambiguity, while KG does not have those properties. Second, the edges of CG have weights based on co-occurrence frequencies unlike the edges of KG. Finally, CG might have a large number of edges types while the number of KG edges is at most several thousands. An interesting research direction is exploring models suitable for CGE to capture the property of linguistic expressions and their relations in the embedding space.

6 Conclusion

We presented NLRA, which learns the distributed representation of word pairs capturing semantic relational information through co-occurring patterns encoded by LSTM. This model jointly learns the mapping from the word embedding space into the word-pair embedding space to generalize co-occurrences of word pairs and patterns. Our experiment on measuring relational similarity demonstrated that NLRA outperforms LRA and can successfully solve the data sparseness problem of word co-occurrences, which is a major bottleneck in pattern-based approaches. Moreover, combining the vector offset model and NLRA yielded competitive performance to the state-of-the-art method, though our method relied only on unsupervised learning. This combined model exploits the complementary information of lexico-syntactic patterns and word embeddings.

In our future work, we will apply word-pair embeddings from NLRA to various downstream tasks related to lexical relational information.

Acknowledgments

This work was supported by JSPS KAKENHI Grant numbers JP17H01831. We thank Satoshi Sekine and Kentaro Inui for helpful discussion.

References

- Danushka Bollegala, Takanori Maehara, and Ken-ichi Kawarabayashi. 2015. Embedding semantic relations into word representations. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1222–1228.
- Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 2787–2795. Curran Associates, Inc.
- John Duchi, Elad Hazan, and Yoram Singer. 2011. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, 12(Jul):2121–2159.
- Patrick Hanks. 2009. The impact of corpora on dictionaries. In Paul Baker, editor, *Contemporary Corpus Linguistics*, pages 214–236. Continuum, London, Great Britain.
- Marti A. Hearst. 1992. Automatic acquisition of hyponyms from large text corpora. In *COLING 1992 Volume 2: The 15th International Conference on Computational Linguistics*.
- Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation*, 9(8):1735–1780.
- Ignacio Iacobacci, Mohammad Taher Pilehvar, and Roberto Navigli. 2015. Senseembed: Learning sense embeddings for word and relational similarity. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 95–105. Association for Computational Linguistics.
- Sergey Ioffe and Christian Szegedy. 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*, pages 448–456.
- Takahiro Ishihara, Katsuhiko Hayashi, Hitoshi Manabe, Masashi Shimbo, and Masaaki Nagata. 2018. Neural tensor networks with diagonal slice matrices. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 506–515. Association for Computational Linguistics.
- Shoaib Jameel, Zied Bouraoui, and Steven Schockaert. 2018. Unsupervised learning of distributional relation vectors. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 23–33. Association for Computational Linguistics.
- David Jurgens, Saif Mohammad, Peter Turney, and Keith Holyoak. 2012. Semeval-2012 task 2: Measuring degrees of relational similarity. In **SEM 2012: The First Joint Conference on Lexical and Computational Semantics – Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012)*, pages 356–364. Association for Computational Linguistics.
- Omer Levy and Yoav Goldberg. 2014. Linguistic regularities in sparse and explicit word representations. In *Proceedings of the Eighteenth Conference on Computational Natural Language Learning*, pages 171–180. Association for Computational Linguistics.
- Omer Levy, Steffen Remus, Chris Biemann, and Ido Dagan. 2015. Do supervised distributional methods really learn lexical inference relations? In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 970–976. Association for Computational Linguistics.
- Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. 2015. Learning entity and relation embeddings for knowledge graph completion. In *AAAI*, volume 15, pages 2181–2187.
- Hanxiao Liu, Yuexin Wu, and Yiming Yang. 2017. Analogical inference for multi-relational embeddings. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 2168–2178, International Convention Centre, Sydney, Australia. PMLR.
- Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013a. Distributed representations of words and phrases and their compositionality. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2, NIPS’13*, pages 3111–3119, USA. Curran Associates Inc.
- Tomas Mikolov, Wen-tau Yih, and Geoffrey Zweig. 2013b. Linguistic regularities in continuous space word representations. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 746–751. Association for Computational Linguistics.
- Maximilian Nickel, Lorenzo Rosasco, Tomaso A Poggio, et al. 2016. Holographic embeddings of knowledge graphs. In *AAAI*, volume 2, pages 3–2.
- Maximilian Nickel, Volker Tresp, and Hans-Peter Kriegel. 2011. A three-way model for collective learning on multi-relational data. In *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, ICML ’11, pages 809–816, New York, NY, USA. ACM.

- Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543. Association for Computational Linguistics.
- Sebastian Riedel, Limin Yao, Andrew McCallum, and Benjamin M. Marlin. 2013. Relation extraction with matrix factorization and universal schemas. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 74–84. Association for Computational Linguistics.
- Bryan Rink and Sanda Harabagiu. 2012. Utd: Determining relational similarity using lexical patterns. In **SEM 2012: The First Joint Conference on Lexical and Computational Semantics – Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012)*, pages 413–418. Association for Computational Linguistics.
- Vered Shwartz, Yoav Goldberg, and Ido Dagan. 2016. Improving hypernymy detection with an integrated path-based and distributional method. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2389–2398, Berlin, Germany. Association for Computational Linguistics.
- Richard Socher, Danqi Chen, Christopher D Manning, and Andrew Ng. 2013. Reasoning with neural tensor networks for knowledge base completion. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 926–934. Curran Associates, Inc.
- Kristina Toutanova, Danqi Chen, Patrick Pantel, Hoi-fung Poon, Pallavi Choudhury, and Michael Gamon. 2015. Representing text for joint embedding of text and knowledge bases. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1499–1509. Association for Computational Linguistics.
- Tho Trouillon, Johannes Welbl, Sebastian Riedel, Eric Gaussier, and Guillaume Bouchard. 2016. Complex embeddings for simple link prediction. In *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 2071–2080, New York, New York, USA. PMLR.
- Peter Turney. 2008a. A uniform approach to analogies, synonyms, antonyms, and associations. In *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*, pages 905–912. Coling 2008 Organizing Committee.
- Peter D. Turney. 2005. Measuring semantic similarity by latent relational analysis. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence*, pages 1136–1141.
- Peter D Turney. 2008b. The latent relation mapping engine: Algorithm and experiments. *Journal of Artificial Intelligence Research*, 33:615–655.
- Peter D. Turney. 2013. Distributional semantics beyond words: Supervised learning of analogy and paraphrase. *Transactions of the Association for Computational Linguistics*, 1:353–366.
- Peter D Turney and Michael L Littman. 2005. Corpus-based learning of analogies and semantic relations. *Machine Learning*, 60(1-3):251–278.
- Peter D Turney and Patrick Pantel. 2010. From frequency to meaning: Vector space models of semantics. *Journal of artificial intelligence research*, 37:141–188.
- Patrick Verga, Arvind Neelakantan, and Andrew McCallum. 2017. Generalizing to unseen entities and entity pairs with row-less universal schema. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 613–622. Association for Computational Linguistics.
- Quan Wang, Zhendong Mao, Bin Wang, and Li Guo. 2017. Knowledge graph embedding: A survey of approaches and applications. *IEEE Transactions on Knowledge and Data Engineering*, 29(12):2724–2743.
- Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. 2014. Knowledge graph embedding by translating on hyperplanes. In *AAAI*, volume 14, pages 1112–1119.
- Koki Washio and Tsuneaki Kato. 2018. Filling missing paths: Modeling co-occurrences of word pairs and dependency paths for recognizing lexical semantic relations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1123–1133. Association for Computational Linguistics.
- Bishan Yang, Wen tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. 2015. Embedding entities and relations for learning and inference in knowledge bases. In *International Conference on Learning Representations (ICLR)*, pages 809–816.
- Alisa Zhila, Wen-tau Yih, Christopher Meek, Geoffrey Zweig, and Tomas Mikolov. 2013. Combining heterogeneous models for measuring relational similarity. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1000–1009. Association for Computational Linguistics.