Representing and Reconstructing PhySH: Which Embedding Competent?

Xiaoli Chen
National Science Library
University of Chinese Academy of Sciences
Chinese Academy of Sciences
chenxl@mail.las.ac.cn

Zhixiong Zhang
National Science Library
Wuhan Library
University of Chinese Academy of Sciences
Chinese Academy of Sciences
Zhangzhx@mail.las.ac.cn

Abstract

In this paper, we conduct a comprehensive comparison of well-known embeddings' capability in capturing the hierarchical Physics knowledge. Several key findings are: (i) Poincaré embeddings do outperform if trained on PhySH, but it fails if trained on co-occurrence pairs which are extracted from raw text. (ii) No algorithm can properly learn hierarchies from the more realistic case of co-occurrence pairs, which contains more noisy relations other than hierarchical relations. (iii) Our statistic analysis of Poincaré embedding's representation of PhySH shows successful hierarchical representation share two characteristics: firstly, upper-level terms have a smaller semantic distance to root; secondly, upper-level hypernym-hyponym pairs should be further apart than lower-level hypernym-hyponym pairs.

1 Introduction

Concept hierarchy or taxonomy[1] is highly organized and expertly curated hierarchical hypernym-hyponym sets. How to effectively represent these terms with the hierarchical relation is the main hurdle for automatically taxonomy construction and other downstream applications.

Though embeddings have been taken for granted in most NLP pipelines, none of the previous work has fully explored which embeddings can capture hierarchical scientific knowledge. Even though Poincaré embedding is proved to have a better ability to capture hierarchical relations, it is learned based on existing WordNet hypernym-hyponym pairs. It is never been tested in the scientific domain. In this paper, we conduct a comprehensive comparison of well-known embeddings' performance in reconstructing Physical Subject Headings (PhySH) from raw APS datasets.

Our main contributions are mainly three-fold: Firstly, for the first time, we compare mainstream embeddings' capability to represent and reconstruct Physical Subject Headings (PhySH) both from raw text and PhySH. Secondly, our experiment shows Poincaré embedding is not sufficient for taxonomy induction from raw text. Thirdly, we explore the characteristics of successful representation of PhySH, which might be the inspiration for better taxonomy construction algorithms.

2 Related Work

Representations for Concept Hierarchy. Representations for concept hierarchy has been receiving quite growing interests in recent years (Kozareva et al., 2008; Carlson et al., 2010; Shen et al., 2018). It is the basis of automaticly taxonomy construction. In the survey study of (Wang et al., 2017), there are *Pattern-based* (Hearst, 1992; Wu et al., 2012; Kozareva and Hovy, 2010) methods and *distributional* (Navigli and Velardi, 2004; Luu et al., 2014; Padó and Lapata, 2003; Baroni and Lenci, 2010; Nguyen et al., 2017) methods use hand-crafted rule-based, co-occurrence features,syntactic features or graph features to learn representations of hierarchical pairs. They also apply pretrained neural laguage models such as Word2Vec (Mikolov et al., 2013), GloVe (Pennington et al., 2014).

Recently, Poincaré embedding (Nickel and Kiela, 2017) is proposed to better represent hierarchical relations. Following works like (Law et al., 2019) use Lorentzian distance to replace the Poincaré metric, (Dhingra et al., 2018) extends Poincaré embedding to apply in raw text with re-parameterization technique, (Leimeister and Wilson, 2019) and (Tifrea et al., 2019) introduce hyperbolic embeddings in word embeddings like Skipgram and GloVe. Effectively in reconstructing WordNet though, the Poincaré embedding is not quite perfect yet (De Sa et al., 2018). It has only been tested on WordNet re-

[1] In this paper, we use *taxonomy* and *concept hierarchy* as equal term.

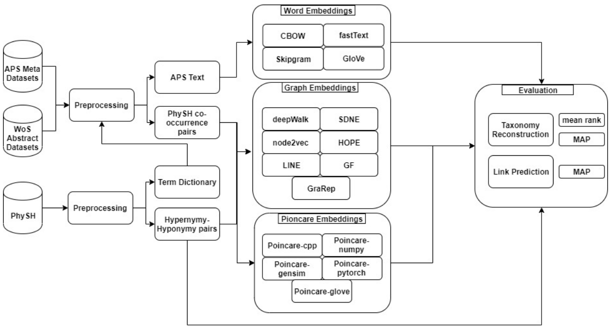

Figure 1: Evaluation Pipeline

construction and Hyperlex entailment (Nickel and Kiela, 2017). Whether it is an effective tool in representing hierarchical relations from raw domain text need to be further explored.

Embeddings Analysis. With the fast pacing of text representation technology, it is also important to revisit existing embedding methods for different downstream tasks. Several previous works have explicitly done this work based on their unique perspectives.Gladkova et al. (2016) explores GloVe's ability to encode different morphological and semantic relations. (Zuccon et al., 2015) analyze word embeddings for information retrieval. Nooralahzadeh et al. (2019) compared COW and Skipgram by using Gensim implementation (Řehůřek and Sojka, 2010) with several different hyper-parameters settings and different domain corpus. Sanchez and Riedel (2017) explored different datasets in evaluation hypernyms identification by using GloVe. (Lastra-DÃ■az et al., 2019) surveys main word embeddings for word similarity.

Despite the above-mentioned work, there is still a missing part describing which embedding is the optimal choice for taxonomy induction. In this paper, we design our evaluation pipeline to choose the optimal embedding scheme for taxonomy learning and construction. In our paper, we consider two perspectives to represent and construct concept hierarchy: (i) Learn and construct from raw texts by word embeddings; (ii) Learn and construct from extracted co-occurrence pairs from raw texts by graph embeddings and Poincaré embeddings.

3 Method

In our pipeline (Figure 1), we follow three steps: raw text and PhySH preprocessing; learn various embeddings with different hyperparameters; evaluate embeddings by reconstructing PhySH.

We evaluate the following embeddings:

	Model Name	Metric	Dimensions 5	10	20	50	100	200
Word Embeddings	GloVe	*mean rank*	2168·30	2568·93	2237·33	2142·94	2188·83	2271·32
		MAP	0·18	0·05	0·06	0·07	0·06	0·06
	COW	*mean rank*	2883·64	3196·44	2937·09	3162·85	1894·02	3096·82
		MAP	0·69	0·63	0·70	0·63	**0·72**	0·64
	Skipgram	*mean rank*	2595·87	3939·61	3091·23	2732·08	3683·61	2893·45
		MAP	0·68	0·60	0·67	0·68	0·63	0·70
	fastText	*mean rank*	2461·97	3004·28	2903·78	3391·16	3456·85	2493·87
		MAP	0·67	0·67	0·69	0·66	0·59	0·65
Graph Embeddings	deepWalk	*mean rank*	**244·03**	469·47	624·21	726·95	780·47	811·30
		MAP	0·18	0·05	0·06	0·05	0·05	0·05
	GF	*mean rank*	1189·78	1003·12	916·56	825·82	682·73	629·40
		MAP	0·01	0·01	0·01	0·01	0·02	0·03
	GraRep	*mean rank*	676·07	944·67	849·28	813·40	828·20	840·75
		MAP	0·05	0·01	0·02	0·03	0·03	0·03
	HOPE	*mean rank*	-	749·61	776·90	803·42	838·26	874·51
		MAP	-	0·12	0·11	0·10	0·08	0·06
	LINE	*mean rank*	387·65	360·36	459·23	432·79	423·59	425·38
		MAP	0·07	0·06	0·06	0·05	0·06	0·08
	node2vec	*mean rank*	490·53	458·11	462·65	459·00	453·37	450·80
		MAP	0·02	0·03	0·04	0·04	0·04	0·04
	SDNE	*mean rank*	917·78	836·11	823·17	960·99	931·31	991·00
		MAP	0·04	0·10	0·10	0·02	0·04	0·02
Poincaré Embeddings	Pioncare Gensim	*mean rank*	765·08	734·58	747·20	750·99	739·38	745·25
		MAP	0·03	0·03	0·03	0·03	0·03	0·03
	Pioncare Cpp	*mean rank*	438·84	428·82	441·85	449·64	452·56	457·11
		MAP	0·06	0·09	0·09	0·09	0·09	0·09
	Pioncare Numpy	*mean rank*	935·95	880·16	861·51	874·15	892·52	879·84
		MAP	0·01	0·02	0·02	0·02	0·02	0·02
	Pioncare Pytorch	*mean rank*	1169·85	1151·57	1167·01	1164·53	1169·49	1165·13
		MAP	0·08	0·08	0·08	0·08	0·08	0·08
	Pioncare GloVe	*mean rank*	1268·48	1263·33	1250·31	1169·00	1165·30	1003·67
		MAP	0·01	0·01	0·01	0·03	0·04	0·06

Table 1: PhySH reconstruction from APS datasets, with word embeddings trained on raw text, graph embeddings and Poincaré embeddings trained on co-occurrence of PhySH terms in raw text. We only include each embedding's optimal result in the table.

- Word embeddings: CBOW and Skipgram (Mikolov et al., 2013), fastText (Joulin et al., 2017), GloVe(Pennington et al., 2014)[2].
- Graph embeddings: deepWalk (Perozzi et al., 2014), node2vec (Grover and Leskovec, 2016), LINE (Tang et al., 2015), LLE (Roweis and Saul, 2000), HOPE (Ou et al., 2016), GF(Ahmed et al., 2013), SDNE(Wang et al., 2016)[3].
- Poincaré embeddings: Poincaré-gensim[4], Poincaré-cpp [5], Poincaré-pytorch[6], Poincaré-numpy[7], Poincaré-glove[8] (Tifrea et al., 2019).

.

Word embeddings are trained on *title* and *abstract* of APS publications. The PhySH terms' embedding vectors will be extracted for taxonomy

[2]CBOW, Skipgram and fastText are trained by https://github.com/NIHOPA/word2vec_pipeline. GloVe is trained by https://github.com/stanfordnlp/GloVe

[3]Graph embeddings are implemented by OpenNE repository https://github.com/thunlp/OpenNE

[4]https://radimrehurek.com/gensim/models/poincare.html

[5]https://github.com/TatsuyaShirakawa/poincare-embedding.git

[6]https://github.com/facebookresearch/poincare-embeddings

[7]https://github.com/nishnik/poincare_embeddings.git

[8]https://github.com/alex-tifrea/poincare_glove

	Model Name	Metric	Dimensions					
			5	**10**	**20**	**50**	**100**	**200**
Graph Embeddings	deepWalk	*mean rank*	357·26	496·36	546·02	537·64	525·74	519·82
		MAP	0·22	0·19	0·21	0·22	0·22	0·23
	GF	*mean rank*	277·24	125·89	50·67	8·90	2·93	9·79
		MAP	0·10	0·35	0·58	0·65	0·66	0·66
	GraRep	*mean rank*	-	78·87	34·45	22·19	13·45	82·18
		MAP	-	0·49	0·53	0·56	0·58	0·57
	HOPE	*mean rank*	-	561·03	758·32	691·95	615·45	515·47
		MAP	-	0·64	0·47	0·43	0·43	0·45
	LINE	*mean rank*	489·49	344·14	141·35	34·32	15·84	10·00
		MAP	0·04	0·07	0·23	0·52	0·60	0·62
	node2vec	*mean rank*	265·65	264·94	265·69	264·81	269·52	265·20
		MAP	0·33	0·34	0·35	0·35	0·34	0·35
	SDNE	*mean rank*	72·58	33·18	478·27	517·55	512·10	492·46
		MAP	0·37	0·54	0·12	0·04	0·02	0·02
Poincaré Embeddings	Pioncare Gensim	*mean rank*	8·08	6·58	7·04	7·43	6·63	6·20
		MAP	0·61	0·61	0·62	0·61	0·62	0·61
	Pioncare Cpp	*mean rank*	12·04	11·74	8·12	6·75	8·17	6·95
		MAP	0·61	0·61	0·62	0·62	0·62	0·62
	Pioncare Numpy	*mean rank*	382·52	291·56	272·80	232·12	249·01	247·75
		MAP	0·46	0·53	0·56	0·58	0·58	0·59
	Pioncare Pytorch	*mean rank*	3·83	3·22	2·88	**2·61**	2·80	2·82
		MAP	0·93	**0·94**	**0·94**	**0·94**	**0·94**	**0·94**

Table 2: PhySH reconstruction from PhySH hypernym-hyponym pairs. Since there is no context infomation, word embeddings are not applicable here.

reconstruction. Graph embeddings and Poincaré embeddings are trained on the co-occurrence of PhySH terms in each of APS publications. As with (Nickel and Kiela, 2017), we also train graph embeddings and Poincaré embeddings on PhySH hypernym-hyponym pairs.

Taxonomy Reconstruction: We follow (Nickel and Kiela, 2017) to reconstruct taxonomy based on embedding vectors. For each embedding vector in Poincaré disk space, which is denoted as $\mathcal{B}^d = \{x \in R^d, \|x\|_2 \leq 1\}$. The norm of each vector can measure the radius of each vector, while the hyperbolic distance can measure the closeness of two vectors. The closest two are assigned as hypernym-hyponym pairs. The hyperbolic distance of two vector points $u, v \in \mathcal{B}^d$ is calculated follow as (Nickel and Kiela, 2017).

$$d_H(u, v) = arcosh\left(1 + 2 * \frac{\|u - v\|^2}{(1 - \|u\|^2)(1 - \|v\|^2)}\right) \quad (1)$$

The distance could only tell how semantically close are the node pairs (u, v). But which one is the parent node is not answered. One property that makes hyperbolic space outstanding for the hierarchical structure is that the hyperbolic disc area and circle length grow exponentially with its radius. Node with the smaller norm is the higher-level term.

4 Evaluation and Results

4.1 Evaluation Datasets

APS (American Physical Society) has made available their publications data for researchers with the total number of $661,209$ articles and citations and dates back to 1893[9]. We utilize the article metadata datasets. PhySH (Physics Subject Headings) is the Physics concept hierarchy. It is used to organize publications of APS. It is open-source on Github[10]. APS metadata datasets only contain *title* field, we retrieve *abstract* from Web of Sciences database[11].

4.2 Evaluation Metrics

mean rank and *MAP* metrics are used to measure taxonomy reconstruction performance. *mean rank* is calculated for each node's distance of ground truth children against all other nodes. *MAP* is the mean average precision at the threshold of each correctly retrieved child.

$$\text{mean\_rank}(u) = \frac{sp(u)}{sp(u) + lp(u)} \in [0, 1] \quad (2)$$

$lp(u)$ is the furthest length from node u to its descendants. $sp(u)$ is the shortest length from node u to root node. The optimal embedding should score a low *mean rank* and a high *MAP*.

4.3 PhySH Reconstruction Evaluation

PhySH Reconstruction From Raw Text. In this experiment, we extract co-occurrence of PhySH terms in each APS publication. The graph embeddings are trained on the co-occurrence graph. Poincaré embeddings are trained on the noisy co-occurrence pairs. Word embeddings are trained on APS publication raw texts. PhySH terms' representation vectors are extracted from word vectors in the postprocessing step.

Table 1 is the performance of PhySH reconstruction by learning representation from raw APS datasets[12]. None of the embeddings get the best result in both metrics. Word embeddings like CBOW achieve better *MAP*, while graph embeddings like deepWalk outperform in *mean rank*. Poincaré embeddings did not show any superior. Learn PhySH

[9]https://journals.aps.org/datasets

[10]https://github.com/physh-org/PhySH

[11]Web of Science is a commercial database of Clarivate Analytics, it can be accessed by most universities and institutions

[12]We experiment each embedding with different hyperparameters by grid search, we present the optimal performance of each embedding in the tables.

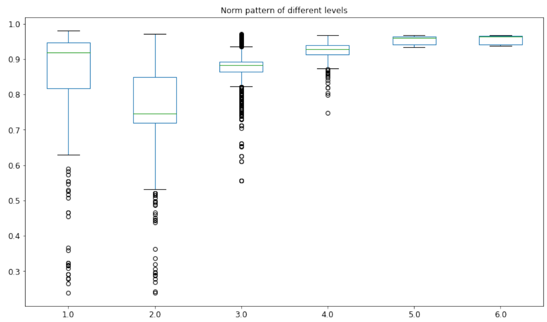

Figure 2: Norm of Poincaré embedding vector at different taxonomy levels

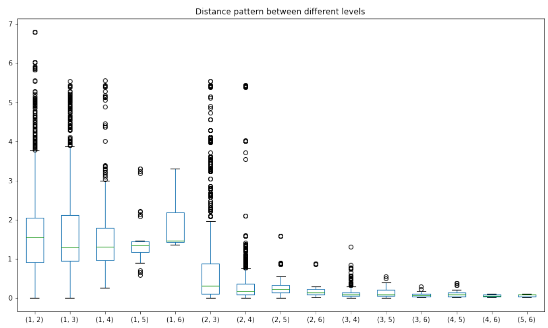

Figure 3: Distance of Poincaré embedding vector accross different taxonomy levels

from noisy co-occurrence pairs are much more complicated than the mammal tree of the WordNet described in the origin paper(Nickel and Kiela, 2017). We can conclude Poincaré embeddings are not sufficient for learning and representing from the co-occurrence pairs.

PhySH Reconstruction From PhySH. Table 2 is the performance of PhySH reconstruction by learning representation from PhySH hypernym-hyponym pairs. The graph embeddings are trained on the PhySH hypernym-hyponym graph. Poincaré embeddings are trained on the PhySH hypernym-hyponym pairs.

In this experiment, Poincaré's official implementation *Poincaré-Pytorch* wins with far better results than other algorithms. This is because Poincaré is trained with the loss function designed to learn hierarchies, while graph embeddings are trained to learn from neighbors and global graph structure. However, *GF* at dimension 100 and *LINE* at dimension 200 also get very good performance.

4.4 The Hierarchical Characteristics of PhySH Poincaré embedding

If we understand the successful representation characteristics of taxonomy hierarchical relations, it will be the help of taxonomy construction. We will analyze what are the hierarchical characteristics of PhySH preserved by Poincaré embeddings in this section.

In Figure 2, we visualize how the norm value varies in different PhySH level. There is a clear pattern from taxonomy level 2 to level 6: lower-level terms have bigger norm values. It means lower terms are further from the root term. The pace of the decrease of the norm in lower levels seems to decelerate, which needs to be further validated. However, the norm of level 1 terms is rather distributed, which we think is the points where Poincaré embedding fails.

In Figure 3, we compare the distance of terms over different PhySH levels. The ancestor nodes are further than parent nodes. For each node, its distance to the child is smaller than the distance to parent, and the distance to the child is nearly half as the distance to parent. These patterns are important for a successful representation of taxonomy.

5 Conclusion and Future Work

we compare word embeddings, graph embeddings, and Poincaré embeddings by reconstructing PhySH. We consider two scenario case: reconstructing from raw texts and reconstructing from existing PhySH. The experiment shows even though Poincaré embeddings far outweigh other embeddings in reconstructing PhySH from PhySH, it is also not competent as other embeddings in reconstructing PhySH from raw APS texts.

We further demystify what is the success of Poincaré embeddings in reconstructing PhySH from PhySH. The future work would be how to design a powerful taxonomy induction algorithm which could benifit from the characteristics of our paper.

Acknowledgments

The authors thank anonymous reviewers for their helpful comments that improved the manuscript. The authors also thank American Physical Society for providing access to the APS Datasets.

This work is supported in part by the National Natural Science Foundation of China under contracts No 71950003, and the project "Design and Research on a Next Generation of Open Knowledge Services System and Key Technologies" (2019XM55).

6 References

References

Amr Ahmed, Nino Shervashidze, Shravan Narayanamurthy, Vanja Josifovski, and Alexander J. Smola. 2013. Distributed large-scale natural graph factorization. In *Proceedings of the 22Nd International Conference on World Wide Web*, WWW '13, pages 37–48, New York, NY, USA. ACM.

Marco Baroni and Alessandro Lenci. 2010. Distributional memory: A general framework for corpus-based semantics. *Computational Linguistics*, 36(4):673–721.

Andrew Carlson, Justin Betteridge, Richard C. Wang, Estevam R. Hruschka, Jr., and Tom M. Mitchell. 2010. Coupled semi-supervised learning for information extraction. In *Proceedings of the Third ACM International Conference on Web Search and Data Mining*, WSDM '10, pages 101–110, New York, NY, USA. ACM.

Christopher De Sa, Albert Gu, Christopher RÃl', and Frederic Sala. 2018. Representation tradeoffs for hyperbolic embeddings. *Proceedings of machine learning research*, 80.

Bhuwan Dhingra, Christopher Shallue, Mohammad Norouzi, Andrew Dai, and George Dahl. 2018. Embedding text in hyperbolic spaces. In *Proceedings of the Twelfth Workshop on Graph-Based Methods for Natural Language Processing (TextGraphs-12)*, pages 59–69, New Orleans, Louisiana, USA. Association for Computational Linguistics.

Anna Gladkova, Aleksandr Drozd, and Satoshi Matsuoka. 2016. Analogy-based detection of morphological and semantic relations with word embeddings: what works and what doesn't. In *Proceedings of the NAACL Student Research Workshop*, pages 8–15, San Diego, California. Association for Computational Linguistics.

Aditya Grover and Jure Leskovec. 2016. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

Marti A. Hearst. 1992. Automatic acquisition of hyponyms from large text corpora. In *COLING 1992 Volume 2: The 15th International Conference on Computational Linguistics*.

Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. 2017. Bag of tricks for efficient text classification. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 427–431. Association for Computational Linguistics.

Zornitsa Kozareva and Eduard Hovy. 2010. A semi-supervised method to learn and construct taxonomies using the web. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 1110–1118, Cambridge, MA. Association for Computational Linguistics.

Zornitsa Kozareva, Ellen Riloff, and Eduard Hovy. 2008. Semantic class learning from the web with hyponym pattern linkage graphs. In *Proceedings of ACL-08: HLT*, pages 1048–1056, Columbus, Ohio. Association for Computational Linguistics.

Juan J. Lastra-DÃ■az, Josu Goikoetxea, Mohamed Ali Hadj Taieb, Ana GarcÃ■a-Serrano, Mohamed Ben Aouicha, and Eneko Agirre. 2019. A reproducible survey on word embeddings and ontology-based methods for word similarity: Linear combinations outperform the state of the art. *Engineering Applications of Artificial Intelligence*, 85:645 – 665.

Marc T Law, Jake Snell, and Richard S Zemel. 2019. Lorentzian distance learning.

Matthias Leimeister and Benjamin J. Wilson. 2019. Skip-gram word embeddings in hyperbolic space.

Anh Tuan Luu, Jung-jae Kim, and See Kiong Ng. 2014. Taxonomy construction using syntactic contextual evidence. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 810–819, Doha, Qatar. Association for Computational Linguistics.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. 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.

Roberto Navigli and Paola Velardi. 2004. Learning domain ontologies from document warehouses and dedicated web sites. *Comput. Linguist.*, 30(2):151–179.

Kim Anh Nguyen, Maximilian KÃűeper, Sabine Schulte im Walde, and Ngoc Thang Vu. 2017. Hierarchical Embeddings for Hypernymy Detection and Directionality. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 233–243, Copenhagen, Denmark.

Maximilian Nickel and Douwe Kiela. 2017. Poincaré embeddings for learning hierarchical representations. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 6341–6350. Curran Associates, Inc.

Farhad Nooralahzadeh, Lilja Øvrelid, and Jan Tore Lønning. 2019. Evaluation of domain-specific word embeddings using knowledge resources. *LREC 2018 - 11th International Conference on Language Resources and Evaluation*, pages 1438–1445.

Mingdong Ou, Peng Cui, Jian Pei, Ziwei Zhang, and Wenwu Zhu. 2016. Asymmetric transitivity preserving graph embedding. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1105–1114. ACM.

Sebastian Padó and Mirella Lapata. 2003. Constructing semantic space models from parsed corpora. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, pages 128–135, Sapporo, Japan. Association for Computational Linguistics.

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, Doha, Qatar. Association for Computational Linguistics.

Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '14, pages 701–710, New York, NY, USA. ACM.

Radim Řehůřek and Petr Sojka. 2010. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pages 45–50, Valletta, Malta. ELRA. http://is.muni.cz/publication/884893/en.

Sam T. Roweis and Lawrence K. Saul. 2000. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326.

Ivan Sanchez and Sebastian Riedel. 2017. How well can we predict hypernyms from word embeddings? a dataset-centric analysis. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 401–407, Valencia, Spain. Association for Computational Linguistics.

Jiaming Shen, Zeqiu Wu, Dongming Lei, Chao Zhang, Xiang Ren, Michelle T. Vanni, Brian M. Sadler, and Jiawei Han. 2018. Hiexpan: Task-guided taxonomy construction by hierarchical tree expansion. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '18, pages 2180–2189, New York, NY, USA. ACM.

Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei. 2015. Line: Large-scale information network embedding. In *WWW*. ACM.

Alexandru Tifrea, Gary Becigneul, and Octavian-Eugen Ganea. 2019. Poincare glove: Hyperbolic word embeddings. In *International Conference on Learning Representations*.

Chengyu Wang, Xiaofeng He, and Aoying Zhou. 2017. A Short Survey on Taxonomy Learning from Text Corpora: Issues, Resources and Recent Advances. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 1992, pages 1190–1203, Stroudsburg, PA, USA. Association for Computational Linguistics.

Daixin Wang, Peng Cui, and Wenwu Zhu. 2016. Structural deep network embedding. In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 1225–1234, New York, NY, USA. ACM.

Julie Weeds, David Weir, and Diana McCarthy. 2004. Characterising measures of lexical distributional similarity. In *Proceedings of the 20th International Conference on Computational Linguistics*, COLING '04, Stroudsburg, PA, USA. Association for Computational Linguistics.

Wentao Wu, Hongsong Li, Haixun Wang, and Kenny Q. Zhu. 2012. Probase: A probabilistic taxonomy for text understanding. In *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data*, SIGMOD '12, pages 481–492, New York, NY, USA. ACM.

Ichiro Yamada, Kentaro Torisawa, Jun'ichi Kazama, Kow Kuroda, Masaki Murata, Stijn De Saeger, Francis Bond, and Asuka Sumida. 2009. Hypernym discovery based on distributional similarity and hierarchical structures. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 929–937, Singapore. Association for Computational Linguistics.

Guido Zuccon, Bevan Koopman, Peter Bruza, and Leif Azzopardi. 2015. Integrating and evaluating neural word embeddings in information retrieval. In *Proceedings of the 20th Australasian Document Computing Symposium*, ADCS '15, pages 12:1–12:8, New York, NY, USA. ACM.