

USAAR-WLV: Hypernym Generation with Deep Neural Nets

Liling Tan, Rohit Gupta^α and Josef van Genabith^β

Universität des Saarlandes / Campus A2.2, Saarbrücken, Germany
University of Wolverhampton^α / Wulfruna Street, Wolverhampton, UK
Deutsches Forschungszentrum für Künstliche Intelligenz^β /
Stuhlsatzenhausweg, Saarbrücken, Germany
alvations@gmail.com, r.gupta@wlv.ac.uk,
josef.van_genabith@dfki.de

Abstract

This paper describes the USAAR-WLV taxonomy induction system that participated in the Taxonomy Extraction Evaluation task of SemEval-2015. We extend prior work on using vector space word embedding models for hypernym-hyponym extraction by simplifying the means to extract a projection matrix that transforms any hyponym to its hypernym. This is done by making use of function words, which are usually overlooked in vector space approaches to NLP. Our system performs best in the chemical domain and has achieved competitive results in the overall evaluations.

1 Introduction

Traditionally, broad-coverage semantic taxonomies such as CYC (Lenat, 1995) and WordNet ontology (Miller, 1995) have been manually created with much effort and yet they suffer from coverage sparsity. This motivated the move towards unsupervised approaches to extract structured relational knowledge from texts (Lin and Pantel, 2001; Snow et al., 2006; Velardi et al., 2013).¹

Previous work in taxonomy extraction focused on rule-based, clustering and graph-based approaches. Although vector space approaches are popular in current NLP researches, ontology induction studies have yet to catch on the frenzy. Fu et al. (2014) proposed a vector space approach to hypernym-hyponym identification using word embeddings that

¹For the rest of the paper, *taxonomy* and *ontology* will be used interchangeably to refer to a hierarchically structure that organizes a list of concepts.

trains a projection matrix² that converts a hyponym vector to its hypernym. However, their approach requires an existing hypernym-hyponym pairs for training before discovering new pairs.

Our system submitted to the SemEval-2015 taxonomy building task is most similar to the approach by Fu et al. (2014) in using word embeddings projections to identify hypernym-hyponym pairs. As opposed to previous method our method does not requires prior taxonomical knowledge.

Instead of training a projection matrix, we capitalize on the fact that hypernym-hyponym pair often occurs in a sentence with an ‘*is a*’ phrase, e.g. “*The goldfish (Carassius auratus auratus) is a freshwater fish*”.³ Intuitively, if we single-tokenize the ‘*is a*’ phrase prior to training a vector space, we can make use of the vector that represents the phrase in capturing a hypernym-hyponym pair as such the multiplication of $v(\text{goldfish})$ and $v(\text{is-a})$ will be similar to the cross product $v(\text{fish})$ ($v(\text{goldfish}) \times v(\text{is-a}) \approx v(\text{fish})$).

There is little or no previous work that manipulates non-content word vectors in vector space models studies for natural language processing. Often, non-content words⁴ were implicitly incorporated into the vector space models by means of syntactic frames (Sarmiento et al., 2009) or syntactic parses (Thater et al., 2010).

Our main contribution for ontological induction

²In this case, the projection matrix is a vector space feature function.

³From <http://en.wikipedia.org/wiki/Goldfish>.

⁴Words that are not noun (entities/arguments), verbs (predicates), adjectives or adverbs (adjuncts).

using vector space models are primarily (i) the use of non-content word vectors and (ii) simplifying a previously complex process of learning a hypernym-hyponym transition matrix. The implementation of our ontological induction approach is open-sourced and available on our GitHub repository.⁵

1.1 Task Definition

Similar to Fountain and Lapata (2012), the SemEval-2015 Taxonomy Extraction Evaluation (TaxEval) task addresses taxonomy learning without the term discovery step, i.e. the terms for which to create the taxonomy are given (Bordea et al., 2015). The focus is on creating the hypernym-hyponym relations.

In the TaxEval task, taxonomies are evaluated through comparison with gold standard taxonomies. There is no training corpus provided by the organisers of the task and the participating systems are to generate hyper-hyponyms pairs using a list of terms from four different domains, viz. chemicals, equipment, food and science.

The gold standards used in evaluation are the *ChEBI ontology* for the chemical domain (Degtarenko et al., 2008), the *Material Handling Equipment taxonomy*⁶ for the equipment domain, the *Google product taxonomy*⁷ for the food domain and the *Taxonomy of Fields and their Different Sub-fields*⁸ for the science domain. In addition, all four domains are also evaluated against the sub-hierarchies from the WordNet ontology that subsumes the Suggested Upper Merged Ontology (Pease et al., 2002).

2 Related Work

There are a variety of methods used in taxonomy induction. They can be broadly categorized as (i) pattern/rule based, (ii) clustering based, (iii) graph based and (iv) vector space approaches.

⁵<https://github.com/alvations/USAAR-SemEval-2015/tree/master/task17-USAAR-WLV>

⁶<http://www.ise.ncsu.edu/kay/mhetax/index.htm>

⁷<http://www.google.com/basepages/producttype/taxonomy.en-US.txt>

⁸http://sites.nationalacademies.org/PGA/Resdoc/PGA_044522

2.1 Pattern/Rule Based Approaches

Hearst (1992) first introduced ontology learning by exploiting lexico-syntactic patterns that explicitly links a hypernym to its hyponym, e.g. “*X and other Ys*” and “*Ys such as X*”. These patterns could be manually constructed (Berland and Charniak, 1999; Kozareva et al., 2008) or automatically bootstrapped (Girju, 2003).

These methods rely on surface-level patterns and incorrect items are frequently extracted because of parsing errors, polysemy, idiomatic expressions, etc.

2.2 Clustering Approaches

Clustering based approaches are mostly used to discover hypernym (is-a) and synonym (is-like) relations. For instance, to induce synonyms, Lin (1998) clustered words based on the amount of information needed to state the commonality between two words.⁹

Contrary to most bottom-up clustering approaches for taxonomy induction (Caraballo, 2001; Lin, 1998), Pantel and Ravichandran (2004) introduced a top-down approach, assigning the hypernyms to clusters using co-occurrence statistics and then pruning the cluster by recalculating the pairwise similarity between every hyponym pair within the cluster.

2.3 Graph-based Approaches

In graph theory (Biggs et al., 1976), similar ideas are conceived with a different jargon. In graph notation, *nodes/vertices* form the atom units of the graph and nodes are connected by directed *edges*. A *graph*, unlike an ontology, regards the hierarchical structure of a taxonomy as a by-product of the individual pairs of *nodes* connected by a directed *edges*. In this regard, a single *root* node is not guaranteed and to produce a tree-like structure.

Disregarding the overall hierarchical structure, the crux of graph induction focuses on the different techniques of edge weighting between individual node pairs and graph pruning or edge collapsing (Kozareva and Hovy, 2010; Navigli et al., 2011; Fountain and Lapata, 2012; Tuan et al., 2014).

⁹Commonly known as Lin information content measure.

2.4 Vector Space Approaches

Semantic knowledge can be thought of as a two-dimensional vector space where each word is represented as a point and semantic association is indicated by word proximity. The vector space representation for each word is constructed from the distribution of words across context, such that words with similar meaning are found close to each other in the space (Mitchell and Lapata, 2010; Tan, 2013).

Although vector space models have been used widely in other NLP tasks, ontology/taxonomy inducing using vector space models has not been popular. It is only since the recent advancement in neural nets and word embeddings that vector space models are gaining ground for ontology induction and relation extraction (Saxe et al., 2013; Khashabi, 2013).

3 Methodology

This section provides a brief overview of our system’s approach to taxonomy induction. The full system is released as open-source and contains documentation with additional implementation details.¹⁰

3.1 Projecting a Hyponym to its Hypernym with Transition Matrix

Fu et al. (2014) discovered that hypernym-hyponyms pairs have similar semantic properties as the linguistics regularities discussed in Mikolov et al. (2013b). For instance: $v(\text{shrimp}) - v(\text{prawn}) \approx v(\text{fish}) - v(\text{goldfish})$.

Intuitively, the assumption is that all words can be projected to their hypernyms based on a transition matrix. That is, given a word x and its hypernym y , a transition matrix Φ exists such that $y = \Phi x$, e.g. $v(\text{goldfish}) = \Phi \times v(\text{fish})$.

Fu et al. proposed two projection approaches to identify hypernym-hyponym pairs, (i) uniform linear projection where Φ is the same for all words and Φ is learnt by minimizing the mean squared error of $\|\Phi x - y\|$ across all word-pairs (i.e. a domain independent Φ) and (ii) piecewise linear projection that learns a separate projection for different word clusters (i.e. a domain dependent Φ , where a taxonomy’s domain is bounded by its terms’ cluster(s)). In both

¹⁰<https://github.com/alvations/USAAR-SemEval-2015/blob/master/task17-USAAR-WLV/README.md>

projections, hypernym-hyponym pairs are required to train the transition matrix Φ .

3.2 Inducing a Hypernym with *is-a* Vector

Instead of learning a supervised transition matrix Φ , we propose a simpler unsupervised approach where we learn a vector for the phrase “*is-a*”. We single-tokenize the adjacent “is” and “a” tokens and learn the word embeddings with *is-a* forming part of the vocabulary in the input matrix.

Effectively, we hypothesize that Φ can be replaced by the “*is-a*” vector. To achieve the piecewise projection effects of Φ , we trained a different deep neural net model for each TaxEval domain and assume that the “*is-a*” scales automatically across domains. For instance, the multiplication of the $v(\text{tiramisu})$ and the $v(\text{is-a}_{\text{food}})$ vectors yields a proxy vector and we consider the top ten word vectors that are most similar to this proxy vector as the possible hypernyms, i.e. $v(\text{tiramisu}) \times v(\text{is-a}_{\text{food}}) \approx v(\text{cake})$.

4 Experimental Setup

4.1 Training Data

There is no specified training corpus released for the SemEval-2015 TaxEval task. To produce a domain specific corpus for each of the given domains in the task, we used the Wikipedia dump and preprocessed it using WikiExtractor¹¹ and then extracted documents that contain the terms for each domain individually.

We trained a skip-gram model phrasal word2vec neural net (Mikolov et al., 2013a) using gensim (Řehůřek and Sojka, 2010). The neural nets were trained for 100 epochs with a window size of 5 for all words in the corpus.¹²

4.2 Evaluation Metrics

For the TaxEval task, the multi-faceted evaluation scheme presented in Navigli (2013) was adopted to compare the overall structure of the taxonomy against a gold standard, with an approach used for comparing hierarchical clusters. The multi-faceted

¹¹We use the same Wikipedia dump to text extraction process from the SeedLing - Human Language Project (Emerson et al., 2014).

¹²i.e. words with minimum count of 1; other parameters set for the neural nets can be found on our GitHub repository.

	V	E	#c.c	cycles	#VC	%VC	#EC	%EC	:NE
Chemical	13785	30392	302	YES	13784	0.7838	2427	0.0977	1.1268
Equipment	337	548	28	YES	336	0.549	227	0.3691	0.5219
Food	1118	2692	23	YES	948	0.6092	428	0.2696	1.4265
Science	355	952	14	YES	354	0.7831	173	0.3720	1.6752
WN Chemical	1173	3107	31	YES	1172	0.8675	532	0.3835	1.8566
WN Equipment	354	547	43	YES	353	0.7431	149	0.3072	0.8206
WN Food	1200	3465	23	YES	1199	0.8068	549	0.3581	1.9021
WN Science	307	892	8	YES	306	0.7132	156	0.3537	1.6689

Table 1: Structural Measures and Comparison against Gold Standards for USAAR-WLV. The labels of the columns refer to no. of distinct vertices and edges in induced taxonomy ($|V|$ and $|E|$), no. of connected components ($\#c.c$), whether the taxonomy is a Directed Acyclic Graph (**cycles**), vertex and edge coverage, i.e. proportion of gold standard vertices and edges covered by system ($\%VC$ and $\%EC$), no. of vertices and edges in common with gold standard ($\#VC$ and $\#EC$) and ratio of novel edges ($:NE$).

	INRIASAC	LT3	NTNU	QASSIT	TALN-UPF	USAAR-WLV
Avg. F&M	0.3270	0.4130	0.0580	0.3880	0.2630	0.0770
Avg. Precision	0.1721	0.3612	0.1754	0.1563	0.0720	0.2014
Avg. Recall	0.4279	0.6307	0.2756	0.1588	0.1165	0.3139
Avg. F-Score	0.2427	0.3886	0.2075	0.1575	0.0798	0.2377
Avg. Precision of NE	0.4800	0.5960	0.3530	0.2470	0.1020	0.4200

Table 2: Averaged F&M Measure, Precision, Recall, F-score for All Systems Outputs when Compared to Gold Standard and Manually Evaluated Average Precision of Novel Edges.

evaluation scheme evaluates (i) the structural measures of the induced taxonomy (left columns of Table 1), (ii) the comparison against gold standard taxonomy (right columns of Table 1 and leftmost column of Table 2) and (iii) manual evaluation of novel edges precision (last row of Table 2).

Regarding the two types of automatic evaluation measures, the structural measures provides a gauge of the system’s coverage and the ontology structural integrity, i.e. “tree-likeness” of the ontology produced by the hypernym-hyponym pairs, and the comparison against the gold standards gives an objective measure of the “human-likeness” of the system in producing a taxonomy that is similar to the manually-crafted taxonomy.

5 Results

Table 1 presents the evaluation scores for our system in the TaxEval task, the $\%VC$ and $\%EC$ scores summarize the performance of the system in replicating the gold standard taxonomies.

In terms of vertex coverage, our system performs best in the chemical and WordNet chemical domain. Regarding edge coverage, our system achieves high-

est coverage for the science domain and WordNet chemical domain. Having high edge and vertex coverage significantly lowers false positive rate when evaluating hypernym-hyponyms pairs with precision, recall and F-score.

We also note that the wikipedia corpus extracted that we used to induce the vectors lacks coverage for the food domain. In the other domains, we discovered all terms in the wikipedia corpus plus the domains’ root hypernym (i.e. $|V| = \#VC + 1$).

Table 2 presents the comparative results between the participating teams in the TaxEval task averaged over all domains. We performed reasonable well as compared to the other systems in all measures. While our system’s F&M measure is low, it is only representative of the clusters we have induced as compared to the gold standard. To improve our F&M measure, we could reduce the number of redundant novel edges by pruning our system outputs and achieve comparable results to the other teams given our relatively precision of novel edges.

A detailed evaluation on the results for the individual domains is presented on Bordea et al. (2015).

6 Conclusion

In this paper, we have described our submissions to the Taxonomy Evaluation task for SemEval-2015. We have simplified a previously complex process of inducing a hypernym-hyponym ontology from a neural net by using the word vector for the non-content word text pattern, "is a".

Our system achieved modest results when compared against other participating teams. Given the simple approach to hypernym-hyponym relations, it is possible that future research can apply the method to other non-content words vectors to induce other relations between entities. The implementation of our system is released as open-source.

Acknowledgements

The research leading to these results has received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme FP7/2007-2013/ under REA grant agreement no 317471. We would like to thank the Daniel Cer and other anonymous reviewers for their helpful suggestions and comments.

References

- Matthew Berland and Eugene Charniak. 1999. Finding Parts in Very Large Corpora. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*, pages 57–64.
- Norman Biggs, E. Keith Lloyd, and Robin J. Wilson. 1976. *Graph theory 1736-1936*. Clarendon Press.
- Georgeta Bordea, Paul Buitelaar, Stefano Faralli, and Roberto Navigli. 2015. Semeval-2015 task 17: Taxonomy Extraction Evaluation. In *Proceedings of the 9th International Workshop on Semantic Evaluation*.
- Sharon Ann Caraballo. 2001. *Automatic Construction of a Hypernym-labeled Noun Hierarchy from Text*. Ph.D. thesis, Providence, RI, USA. AAI3006696.
- Kirill Degtyarenko, Paula De Matos, Marcus Ennis, Janna Hastings, Martin Zbinden, Alan Mcnaught, Rafael Alcántara, Michael Darsow, Mickaël Guedj, and Michael Ashburner. 2008. ChEBI: A Database and Ontology for Chemical Entities of Biological Interest. *Nucleic acids research*, 36(suppl 1):D344–D350.
- Guy Emerson, Liling Tan, Susanne Fertmann, Alexis Palmer, and Michaela Regneri. 2014. SeedLing: Building and Using a Seed corpus for the Human Language Project. In *Proceedings of the 2014 Workshop on the Use of Computational Methods in the Study of Endangered Languages*, pages 77–85.
- Trevor Fountain and Mirella Lapata. 2012. Taxonomy Induction using Hierarchical Random Graphs. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 466–476.
- Ruiji Fu, Jiang Guo, Bing Qin, Wanxiang Che, Haifeng Wang, and Ting Liu. 2014. Learning Semantic Hierarchies via Word Embeddings. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1199–1209.
- Roxana Girju. 2003. Automatic Detection of Causal Relations for Question Answering. In *Proceedings of the ACL 2003 workshop on Multilingual summarization and question answering-Volume 12*, pages 76–83.
- Marti A Hearst. 1992. Automatic Acquisition of Hyponyms from Large Text Corpora. In *Proceedings of the 14th conference on Computational linguistics-Volume 2*, pages 539–545.
- Daniel Khashabi. 2013. On the Recursive Neural Networks for Relation Extraction and Entity Recognition. Technical report.
- 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.
- 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, June.
- Douglas B Lenat. 1995. CYC: A Large-Scale Investment in Knowledge Infrastructure. *Communications of the ACM*, 38(11):33–38.
- Dekang Lin and Patrick Pantel. 2001. Discovery of Inference Rules for Question-Answering. *Natural Language Engineering*, 7(04):343–360.
- Dekang Lin. 1998. Automatic Retrieval and Clustering of Similar Words. In *Proceedings of the 17th international conference on Computational linguistics-Volume 2*, pages 768–774.
- Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013a. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.
- 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.
- George A Miller. 1995. WordNet: A Lexical Database for English. *Communications of the ACM*, 38(11):39–41.
- Jeff Mitchell and Mirella Lapata. 2010. Composition in Distributional Models of Semantics. *Cognitive Science*, 34(8):1388–1439.
- Roberto Navigli, Paola Velardi, and Stefano Faralli. 2011. A Graph-Based Algorithm for Inducing Lexical Taxonomies from Scratch. In *IJCAI 2011, Proceedings of the 22nd International Joint Conference on Artificial Intelligence, Barcelona, Catalonia, Spain, July 16-22, 2011*, pages 1872–1877.
- Patrick Pantel and Deepak Ravichandran. 2004. Automatically Labeling Semantic Classes. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics: HLT-NAACL 2004*.
- Adam Pease, Ian Niles, and John Li. 2002. The Suggested Upper Merged Ontology: A Large Ontology for the Semantic Web and its Applications. In *Working Notes of the AAAI-2002 Workshop on Ontologies and the Semantic Web*, Edmonton, Canada.
- 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, May. ELRA.
- Luís Sarmento, Paula Carvalho, and Eugénio Oliveira. 2009. Exploring the Vector Space Model for Finding Verb Synonyms in Portuguese. In *Proceedings of the International Conference RANLP-2009*, pages 393–398.
- Andrew M. Saxe, James L. McClelland, and Surya Ganguli. 2013. Learning Hierarchical Category Structure in Deep Neural Networks. pages 1271–1276.
- Rion Snow, Daniel Jurafsky, and Andrew Y Ng. 2006. 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*, pages 801–808.
- Liling Tan. 2013. Examining crosslingual word sense disambiguation. Master’s thesis, Nanyang Technological University. pages 17-21.
- Stefan Thater, Hagen Fürstenau, and Manfred Pinkal. 2010. Contextualizing Semantic Representations using Syntactically Enriched Vector Models. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 948–957.
- Luu Anh Tuan, Jung-jae Kim, and Kiong See 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.
- Paola Velardi, Stefano Faralli, and Roberto Navigli. 2013. OntoLearn Reloaded: A Graph-based Algorithm for Taxonomy Induction. *Computational Linguistics*, 39(3):665–707.