Finding and Expanding Hypernymic Relations in the Music Domain

Luis ESPINOSA-ANKE [a,1], Sergio ORAMAS [b,2],
José CAMACHO-COLLADOS [c,3] and Horacio SAGGION [a,4]
[a] *Natural Language Processing Group, Pompeu Fabra University*
[b] *Music Technology Group, Pompeu Fabra University*
[c] *Linguistic Computing Laboratory, Sapienza University of Rome*

Abstract. Lexical taxonomies are tree or directed acyclic graph-like structures where each node represents a concept and each edge encodes a binary hypernymic (is-a) relation. These lexical resources are useful for AI tasks like Information Retrieval or Machine Translation. Two main trends exist in the construction and exploitation of these resources: On one hand, general purpose taxonomies like WordNet, and on the other, domain-specific databases such as the CheBi chemical ontology, or MusicBrainz in the music domain. In both cases these are based on finding correct hypernymic relations between pairs of concepts. In this paper, we propose a generic framework for hypernym discovery, based on exploiting linear relations between (term, hypernym) pairs in Wikidata, and apply it to the domain of music. Our promising results, based on several metrics used in Information Retrieval, show that in several cases we are able to discover the correct hypernym for a given novel term.

Keywords. Semantics, taxonomy learning, word sense disambiguation.

1. Introduction

Question Answering and Reasoning, as well as other applications in Artificial Intelligence, Natural Language Processing (NLP) and Music Information Retrieval, may benefit dramatically from semantic knowledge. Many approaches for creating and formalizing this knowledge are based on domain ontologies, whose backbone are *lexical taxonomies* [15]. The term taxonomy is used to refer to graph-like hierarchical structures where concepts are nodes organized over a predefined merging or splitting criterion [7]. For example, WordNet [13] groups words into sets of super and subordinate (is-a) relations. Taxonomies have proven beneficial for tasks like Question Answering [4].

In the music field, there have been some attempts to formalize knowledge in form of manually curated Knowledge Bases (richer taxonomies with a wider set of relations, and where there are often more than one graph layers), such as MUSICBRAINZ[5] and

[1]Email: luis.espinosa@upf.edu
[2]Email: sergio.oramas@upf.edu
[3]Email: collados@di.uniroma1.it
[4]Email: horacio.saggion@upf.edu
[5]http://musicbrainz.org/

DISCOGS[6]. Moreover, generic resources like WIKIPEDIA include a sizable amount of Music data. By extension, Knowledge Bases based on WIKIPEDIA, such as DBPEDIA[7] or FREEBASE[8], also include this information, with the benefit of it being presented in a structured way. However, their coverage is limited as they are not specifically targeting the music domain, and they may miss novel and independent artists, albums or songs, and also musical entities that are only locally relevant.

In this paper we propose to expand the musical subset of Wikidata by developing and evaluating a system for (term, hypernym) discovery, since it is widely agreed in the literature that a compulsory prior phase in taxonomy and ontology learning is the correct identification of is-a relations. We evaluate on Wikidata ground truth, obtaining encouraging results. In addition, a manual inspection of a sample of wrong hypernymic predictions also reveal that our method may be usable for incorporating additional relations to a lexical taxonomy, since in some cases it provides candidate hypernyms which are correct even if absent in the ground truth evaluation data[9].

2. Related Work

Building up on the pioneering work by [5] for hypernym discovery, later methods have leveraged linguistic regularities as a first step for taxonomy learning. Some of these works include contributions that exploit syntactic evidence together with a probabilistic framework [17], using WordNet hypernym relations to learn syntactic dependencies and introduce them as features into a logistic regression classifier. Taxonomies can also be constructed combining syntactic dependencies and structural information present in Wikipedia such as hyperlinks [3].

Taxonomy learning can also be cast as a clustering problem. For instance, [6] observe that multilingual distributional evidence can be effectively used for clustering terms hierarchically using the k-means algorithm. Furthermore, [9] propose a "knowledge + context" hierarchical clustering approach, where key domain terms are extracted from a general-purpose Knowledge Base (KB) and afterwards the web is used as source for contextual evidence. Contextual evidence is also used in [10], who assign a taxonomic relation to concept pairs according to predefined syntactic relations over dependency trees, e.g. if two terms appear in a *Subject-Verb-Object* pattern.

We are unaware of methods specifically tackling the automatic learning of a lexical taxonomy in the music domain. However, attempts do exist in formalizing this specialized knowledge in several ways. Let us review some of them.

As mentioned in Section 1, MUSICBRAINZ and DISCOGS are two examples of curated musical Knowledge Bases. They are open Music encyclopedias of music metadata built collaboratively and available to the public. MUSICBRAINZ, in addition, is regularly published as Linked Data by the LINKEDBRAINZ project[10]. Generic Knowledge Bases based on WIKIPEDIA, like the ones described earlier, include a healthy amount of Music data, such as artist, album and song biographies, definitions of musical concepts and

[6] http://www.discogs.com
[7] http://dbpedia.org
[8] https://www.freebase.com/
[9] Available at: http://http://www.taln.upf.edu/resources/music_wikidata.
[10] http://linkedbrainz.org/

genres, or articles about music institutions and venues. However, their coverage is biased towards popular artists and works from Western culture. Finally, let us refer to GROVE MUSIC ONLINE[11], a Music encyclopedia containing over 60k articles written by Music scholars. However, this encyclopedia is not freely open and runs by subscription.

In conclusion, there seems to be a gap to be filled between the automatic formalization of knowledge and the music domain, which this paper aims at addressing at the (term, hypernym) level.

3. Method

Let us describe briefly the two main semantic resources on which we base our method:

BABELNET: We leverage BABELNET [14][12] for accessing the Wikidata musical subset, as in its current version it includes information on *domains*[2], one of them being Music. BABELNET currently constitutes the largest single multilingual repository of named entities and concepts, containing around 14M synsets enriched with a set of *definitions*, available thanks to the seamless integration of resources such as Wikipedia, OmegaWiki, Wiktionary, Wikidata and WordNet. Most relevant to this work is that Wikidata includes several thousands of (term, hypernym) concept pairs in the music domain.

SENSEMBED: We take advantage of a vector space representation of senses, namely SENSEMBED [8], which is exploited to associate each BABELNET sense (the *BabelNet* representation of a single disambiguated concept for a given word) with its corresponding vector in a vector space model. We opt for SENSEMBED because current representations like word embeddings [12] associate vectors to individual words only. In fact, words are potentially ambiguous and are thus not linkable to reference sense inventories, which we observe can provide significant support in term of is-a relations. SENSEMBED, on the other hand, constitutes a hybrid approach for obtaining latent continuous representations of individual word senses. It exploits the structured knowledge of a large sense inventory along with the distributional information gathered from text corpora. SENSEMBED vectors are trained on the English Wikipedia, with BABELNET as sense inventory.

As for our proposed method, it is based on the linguistic regularities of word embeddings, and the properties that so far have been extensively explored [12,11]. Briefly put, our algorithm trains a *transformation function* Ψ between vectors associated to terms and hypernyms from the music branch of Wikidata subsumed by BABELNET, and learns a linear transformation Ψ(term)$\rightarrow$hypernym between them. As it can be seen in Figure 1, given a pair (`catching_up_with_depeche_mode`$_{bn}$[13], `album`$_{bn}$), we obtain a set of vectors where each vector is associated to each of the lexicalizations (i.e. different ways to express the same concept) of the hypernym. In this example, these would be `album`, `album_charts` or `album_track`. This training data is used to train Ψ over SENSEMBED. This approach has been exploited for machine translation [11] or Twitter language normalization [18].

[11] http://www.oxfordmusiconline.com

[12] http://babelnet.org

[13] The $_{bn}$ subscript denotes concepts that are disambiguated and hence are directly associated with one *sense* in BABELNET. For example, for the term `apple`, two possible senses would be `apple`$_{bn:fruit}$ and `apple`$_{bn:company}$.

Formally, let $\mathcal{T}$ be a list of (term, hypernym) Wikidata pairs in the music domain. Each term is represented by a BabelNet sense from SENSEMBED vocabulary. For each training example $e \in \mathcal{T}$, t_i and h_i are term and hypernym, respectively. The term matrix $T = [t_1, t_2 \dots t_n]$ and the hypernym matrix $H = [h_1, h_2 \dots h_n]$ are given by their vector representations. Together, they constitute a set of training examples Φ, composed by vector pairs $\{t_i, h_i\}_{i=1}^{n}$, which are used to learn a linear transformation matrix Ψ.

Following the notation in [18], this transformation can be depicted as Eq. 1:

$$T\Psi = H \tag{1}$$

We follow Mikolov et al.'s original approach and compute Ψ_τ as follows (Eq. 2):

$$\min_{\Psi} \sum_{i=1}^{|\Phi|} \|\Psi t_i - h_i\|^2 \tag{2}$$

Then, for each input term of the test set the trained function is applied to the term and we obtain the closest concepts to the resulting vector by using cosine similarity. Specifically, we keep the 10 closest concepts and these are the ones that are matched against the gold hypernym.

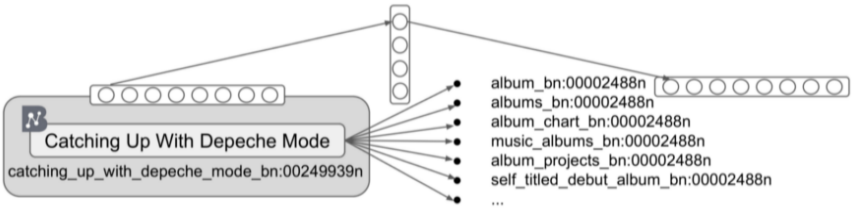

Figure 1. Learning a transformation matrix for sense-based hypernym detection.

As for training and testing statistics, we experimented with several training sizes, as it has been shown in previous work that increasing the size of semantically homogeneous training data (e.g. everything related to music, or the relation object of study being clearly defined) may boost the performance of the system [11]. We report numbers for training with 1k, 5k, 10k, 15k, 20k and 25k (term, hypernym) pairs, and evaluate each run on a held-out test set of 250 (term, hypernym) pairs. Since our algorithm provides ranked lists of candidates, it can be cast as an Information Retrieval-like problem, and therefore metrics such as P@K, MRR or R-Precision are relevant. See [1] for details on each of these metrics.

4. Results and Discussion

From the results provided in Table 1, we may conclude that our algorithm effectively improves as training data increases. This is consistent with the conclusions reached in [11],

Table 1. Results for the hypernym discovery algorithm at different training sizes.

Training Examples	P@1	P@2	P@3	P@4	P@5	R Prec.	MAP	MRR
1000	0.14	114	0.09	0.07	0.06	0.12	0.14	0.16
5000	0.19	0.15	0.12	0.1	0.08	0.18	0.21	0.23
10000	0.192	0.17	0.15	0.138	0.133	0.19	0.226	0.266
15000	0.264	0.232	0.19	0.167	0.158	0.21	0.27	0.34
20000	0.264	0.246	0.222	0.196	0.186	0.25	0.309	0.38
25000	**0.288**	**0.286**	**0.252**	**0.225**	**0.208**	**0.269**	**0.33**	**0.409**

with the difference that in their case the increase is in the hundreds of thousands, while in ours is in the thousands. In fact, the results reported in [16] suggest that, depending of the semantic relation which is being captured, massive amounts of training data may not be necessary.

In terms of MRR, which is probably the most relevant metric as it rewards those cases where one single good match appears high in the ranked list of results, the results are encouraging considering that this average is strict, as it is computed against all possible cases, even for those where the first ranking candidate had a very low cosine score with respect to $t_i\Psi$, which could have been discarded heuristically.

Finally, we sampled 50 *false positives* (FPs), i.e. pairs where, for a given Wikidata concept, our algorithm returned a wrong hypernym. We observed that 18% of our sampled FPs were correct, which suggests that our approach could potentially be used for extending existing ground truth repositories of is-a relations.

5. Conclusions and Future Work

We have described a method for discovering hypernymic relations in the music domain, which is based on linear relations of word embeddings and is designed on the back of BABELNET, as a reference repository of concepts and senses, as well as the backbone of SENSEMBED, a vector space representation of senses. We have provided experimental results at several degrees of training data, showing that the more training data, the better, although our experiments are several orders of magnitude below Mikolov et al.'s original work. We assessed the results yielded by several metrics, and found that the best scoring system is the most informed one.

As for future work, we would like to include a taxonomy induction module, i.e. inducing a full-fledged DAG taxonomy. Another potential avenue for future work may consist on capturing semantic relations beyond hypernymy, perhaps using WordNet relations, exploiting WordNet's mapping to BABELNET.

6. Acknowledgements

This work is partially funded by the Spanish Ministry of Economy and Competitiveness under the Maria de Maeztu Units of Excellence Programme (MDM-2015-0502), and under the TUNER project (TIN2015-65308-C5-5-R, MINECO/FEDER, UE). We also acknowledge support from Dr. Inventor (FP7-ICT-2013.8.1611383).

References

[1] Jiang Bian, Yandong Liu, Eugene Agichtein, and Hongyuan Zha, 'Finding the right facts in the crowd: factoid question answering over social media', in *Proceedings of the 17th international conference on World Wide Web*, pp. 467–476. ACM, (2008).

[2] José Camacho-Collados, Mohammad Taher Pilehvar, and Roberto Navigli, 'Nasari: Integrating explicit knowledge and corpus statistics for a multilingual representation of concepts and entities', *Artificial Intelligence*, **240**, 36–64, (2016).

[3] Tiziano Flati, Daniele Vannella, Tommaso Pasini, and Roberto Navigli, 'Two is bigger (and better) than one: the wikipedia bitaxonomy project', in *ACL*, (2014).

[4] Sanda M Harabagiu, Steven J Maiorano, and Marius A Pasca, 'Open-domain textual question answering techniques', *Natural Language Engineering*, **9**(03), 231–267, (2003).

[5] Marti A Hearst, 'Automatic acquisition of hyponyms from large text corpora', in *Proceedings of the 14th conference on Computational linguistics-Volume 2*, pp. 539–545. Association for Computational Linguistics, (1992).

[6] Hans Hjelm and Paul Buitelaar, 'Multilingual evidence improves clustering-based taxonomy extraction.', in *ECAI*, pp. 288–292, (2008).

[7] Sung Ju Hwang, Kristen Grauman, and Fei Sha, 'Semantic kernel forests from multiple taxonomies', in *Advances in Neural Information Processing Systems*, pp. 1718–1726, (2012).

[8] Ignacio Iacobacci, Mohammad Taher Pilehvar, and Roberto Navigli, 'Sensembed: Enhancing word embeddings for semantic similarity and relatedness', in *Proceedings of ACL*, Beijing, China, (July 2015). Association for Computational Linguistics.

[9] Xueqing Liu, Yangqiu Song, Shixia Liu, and Haixun Wang, 'Automatic taxonomy construction from keywords', in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1433–1441. ACM, (2012).

[10] Tuan Luu Anh, Jung-jae Kim, and See Kiong Ng, 'Taxonomy Construction Using Syntactic Contextual Evidence', in *EMNLP*, pp. 810–819, (October 2014).

[11] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean, 'Efficient Estimation of Word Representations in Vector Space', *arXiv preprint arXiv:1301.3781*, (2013).

[12] Tomas Mikolov, Wen-tau Yih, and Geoffrey Zweig, 'Linguistic Regularities in Continuous Space Word Representations.', in *HLT-NAACL*, pp. 746–751, (2013).

[13] George A. Miller, R.T. Beckwith, Christiane D. Fellbaum, D. Gross, and K. Miller, 'WordNet: an online lexical database', *International Journal of Lexicography*, **3**(4), 235–244, (1990).

[14] Roberto Navigli and Simone Paolo Ponzetto, 'BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network', *Artificial Intelligence*, **193**, 217–250, (2012).

[15] Roberto Navigli, Paola Velardi, and Stefano Faralli, 'A graph-based algorithm for inducing lexical taxonomies from scratch', in *IJCAI*, pp. 1872–1877, (2011).

[16] Sara Rodríguez-Fernández, Luis Espinosa-Anke, Roberto Carlini, and Leo Wanner, 'Semantics-driven recognition of collocations using word embeddings', in *Proceedings of the 2016 Annual Meeting of the Association for Computational Linguistics (ACL)*, Berlin, Germany, (2016).

[17] Rion Snow, Daniel Jurafsky, and Andrew Y Ng, 'Semantic taxonomy induction from heterogenous evidence', in *Proceedings of COLING/ACL 2006*, pp. 801–808. Association for Computational Linguistics, (2006).

[18] L. Tan, H. Zhang, C. Clarke, and M. Smucker, 'Lexical comparison between wikipedia and twitter corpora by using word embeddings', in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pp. 657–661, Beijing, China, (July 2015). Association for Computational Linguistics.