HYPERNYM DISCOVERY

Interim Report

COURSE CODE: INTRO TO NLP - CS7.401.S22

Advisor:
Prof. Manish Shrivastava

Team Name:
Natural Barrier

Project Mentor:
Nirmal Surange

Project Representatives:
Aditya Kumar Singh - 2021701010
Dhruv Srivastava - 2021701021
Nayan Anand - 2021701014

Academic year:
2021-2022

1. Project Description

1.1. Problem Statement

Hypernymy refers to the capability to relate generic terms or classes to their specific instances.[3]. In History there have been multiple studies done on Hypernymy which includes but is not limited to the tasks of identifying whether a hypernymic relation holds between a given candidate pair of terms.[1][7] However for most of the benchmarks established for hypernymy evaluation , a binary classification approach has been prevalent.[6] In this problem we look forward to explore available data (multiple sources) from different angles and attempt hypernym discovery for a given the search space. In this project we'll majorly the experiments done in the cited paper[2] and further improve the benchmark scores by adding a few of our own experiments.

1.2. Sections Overview

The following sections in the report outlines the progress made thus far. The project status is on-track. We have made attempts in multiple directions, not limited to unsupervised pattern-based approach, supervised projection learning approach, as well as the Hybrid approach (combining both supervised and unsupervised) for hypernym discovery which in turn has built up on top of discovering the Hypernyms using Hearst pattern[4]. Section **2** explains the dataset explored and Section **3** contains the detailed overview of the data-set being used in our project. Finally, we summarise the report by mentioning our timeline, challenges realised and future plans in Section **4** and **5** and **6** respectively.

2. Data Understanding

This section briefly outlines the individual data-sets we explored to assist our task- Hypernym discovery. All of them are publicly available and we plan to use a subset of them in this project.

2.1. List of data-sets explored

1. **WordNet dataset** (1985-2011 (last stable release): WordNet is a massive lexicon of English words. Nouns, verbs, adjectives, and adverbs are grouped into 'synsets,' which are collections of cognitive synonyms that individually express a distinct concept. Conceptual-semantic and lexical relations like hyponymy and antonymy are used to connect synsets. In WordNet terminology, each group of synonyms is a synset, and a synonym that forms part of a synset is a lexical variant of the same concept. It contains 155,327 words organized in 175,979 synsets for a total of 207,016 word-sense pairs; in compressed form, it is about 16 megabytes in size.
   The primary version of WordNet is publicly available under a BSD style license and it has multiple derivatives for other languages. The last stable version 3.1 (June 2011) is incorporated in the famous open-source Python library- Natural Language Toolkit (NLTK).
   Fig. 1 represents the hierarchy of words in WordNet data-set.

2. **BabelNet** (v5 released in Feb. 2021): BabelNet is a lexicalized semantic network which follows the same structure as WordNet. It was created automatically by connecting Wikipedia to WordNet and it is one of the largest computational lexicon for the English language. It is a multilingual dataset which covers over 500 languages and contains almost 20 million synets with over 1.4 billion word sense. The integration is accomplished through automatic mapping and statistical machine translation to fill in lexical gaps in resource-poor languages.
   Fig. 2 represents the sources used to build babel data set.

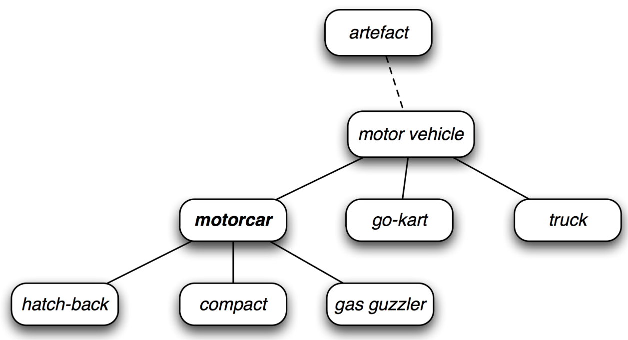


Figure 1: Hierarchical structure of WordNet data-set.

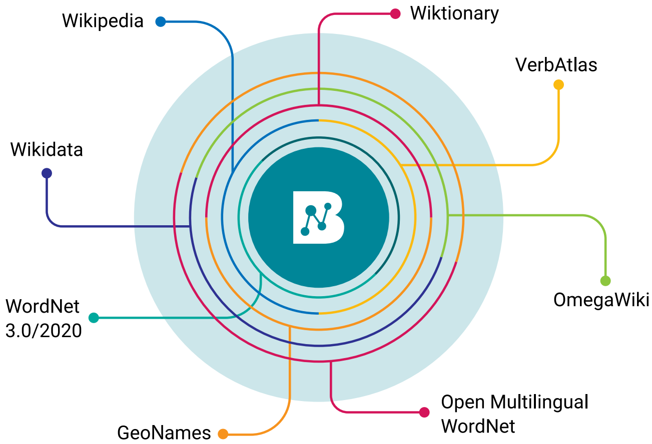


Figure 2: Different sources used to build BabelNet data-set

3. **SemEval 2018 Task 9: Hypernym Discovery data:** We are given five different corpora, three of which are language vocabulary with words in English, Italian, and Spanish, and the remaining two corpora are medical and music vocabulary.
   All the corpus are divided into train, validation and test splits along with additional vocabulary (upto tri-grams) which contains the subset of primary hypernyms that occured less than 5-3 times and are over-generic in nature. These filtered vocabulary can be used to reduce search space for potential candidates. The hyponyms are divided into two categories- Entity (names, location, etc) and Concepts (phenomenon, activity, etc). Some of the samples from this data-set are-

Hyponym	**Hypernym**	Corpus
sodium nitrite	chemical series bond, inorganic compound, chemical bond, chemical group	English
hymn	religious song, vocal, song, track, human voice church music, religious music musical composition, piece of music, opus, composition, christian music, musical style, music genre musical work, work of art	Music
childhood obesity	malady, illness, sickness, disorder, disease	Medical
cantinero	camarero, barman, trabajador, empleado, persona	Spanish
gettone	dischetto, moneta, compenso, disco, cerchio	Italian

Table 1: Samples of Hyponyms and corresponding Hypernyms from SemEval 2018 Task 9 data-set.

3. Ongoing Approach

3.1. Introduction

As suggested above hypernym discovery in layman terms is dicovering the genreral category ("h") of that query ("q"). Or in other case what genralised picture we develop when we ask a quesry "q". For example, "what a *cat* is?" "what is an egg?" and so on. We, more or less, capture the global meaning and most importantly what all similar queries can be categorised into this domain and also what all domains can further be linked to this query is the *paramount* question that lies behind this simple "hierarchial" organising task. At the end our hypernymy detection system should be capable of learning "taxonomy" and "lexical" semantics, including pattern-based methods and graph based methods.

3.2. Simple Base-Line Model

We took the word embeddings of query (trained through `word2vec`) model with default parameters) and match them (using *Cosine* similarity) with their corresponding hypernym's embeddings (also denoted as *Term Embed-*

ding Averaging) due to averaging going for multiple hypernyms (or hyponyms) if any, to discover the hypernymic-relationship.

3.3. Supervised Approach

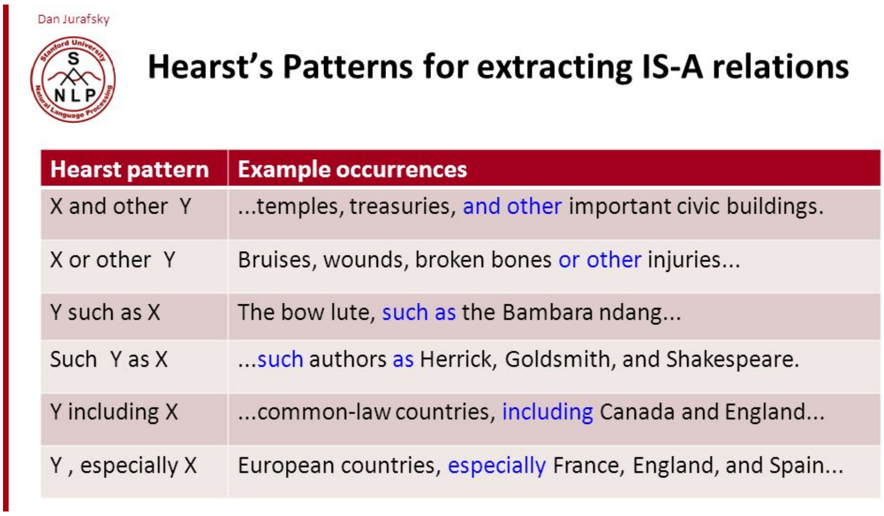
In this approach we divide our task into two major subtask:

1. "Embedding learning" (through `word2vec`).
2. Hypernym-hyponym relation learning.

Now within the supervised approach we try to extract the latent representation of embeddings of (q, h) using a neural network model (e.g., GRU and LSTM) and compute the similarity between e_q (query's latent representation) and e_h (corresponding hypernym's representation) through cosine similarity. As per literature, above methodologies have already been explored and would serve as supervised baselines in our project.

3.4. Unsupervised Approach

The most popular approach was introduced by Hearst (1992) [5] i.e., the pattern based approach who defined special textual patterns (e.g. *Y such as X* which also called as **Hearst patterns** or Lexical-Syntactic patterns) to mine hyponym/hypernym pairs from corpora. This approach is known to *suffer from low recall* because it assumes that hyponym/hypernym pairs will occur together in one of these patterns, which is often not the case. For instance, using the training data of *sub-task 1A*, we found that the majority of training pairs never co-occur within the same paragraph in *corpus 1A*, let alone within a pattern that suggests hypernymy. Later in our future work we'll discuss a more robust method on pattern-based hypernym discovery and how we ovecome this *lower recall* problem.



Hearst pattern	Example occurrences
X and other Y	...temples, treasuries, and other important civic buildings.
X or other Y	Bruises, wounds, broken bones or other injuries...
Y such as X	The bow lute, such as the Bambara ndang...
Such Y as X	...such authors as Herrick, Goldsmith, and Shakespeare.
Y including X	...common-law countries, including Canada and England...
Y , especially X	European countries, especially France, England, and Spain...

Figure 3: Source: *Jurafsky's* Lecture on Hearst Pattern

3.5. Hybrid Approach: CRIM Model

The team at **C**omputer **R**esearch **I**nstitute of **M**ontreal (CRIM) exploited combination of both "unsupervised pattern-based approach" and a "supervised projection learning approach" [2] to model the *hypernymy* relations. Currently this is the SoTA according to the scores evaluated in terms of *MAP*, *MRR*, and *P@1* on dataset provided in SemEval 2018 challenge.

On unsupervised part, to solve the problem of recall, they employed *three* techniques:

1. First, identify co-hyponyms for each query and add the hypernyms discovered for these terms to those found for the query. These co-hyponyms are identified using patterns, and filtered based on distributional similarity using the embeddings(e.g., co-hyponym patterns like *X, Y, and Z*).
2. Discover additional hypernyms using a method based on most multi-word expressions as they are compositional, and the prevailing head-modifier relation is hypernymy. (e.g., for multi-word expressions such as *cold ice cream*, *ice cream* would become hypernym (since it is the headword), whereas, *cold ice cream* becomes it's hyponym.)
3. Extending the set of Hearst-like patterns which we selected empirically (e.g. *Y such as X, Y other than*

X, not all Y are X, Y including X, Y especially X, Y like X, Y for example X, Y which includes X, X are also Y, X are all Y, not Y so much as X).

And they defined their algorithm as follows:

1. Create the empty set Q, which will contain an extended set of queries.
2. Search for the co-hyponym patterns in the corpus to discover co-hyponyms of q. Add these to Q and store their frequency (number of times a given co-hyponym was found using these patterns).
3. Score each co-hyponym $q_0 \in Q$ by multiplying the frequency of q_0 by the cosine similarity of the embeddings of q and q_0 . Rank the co-hyponyms in Q according to this score, keep the top n (setting $n = 5$ emperically), and discard the rest.
4. Add the original query q to Q.
5. Create the empty set of hypernyms H_q.
6. For each query $q_0 \in Q$, search for the hypernym patterns in the corpus to discover hypernyms of q_0. Add these to H_q.
7. Add the head of each term in H_q to this set, as well as the head of the original query q.
8. Score each candidate $cinH_q$ by multiplying its normalized frequency by the cosine similarity between the embeddings of c and q, and rank the candidates according to this score.

Note: Since this is embedding based, patern-based co-hyponyms and hypernyms can find terms not included in the vocabulary. But those "query" which are out-of-vocabulary we simply discard them as we don't have their embeddings.

On Supervised part, with **projection learning** we learn a function that takes as input the word embeddings of a query q and a candidate hypernym h and outputs the likelihood that there is a hypernymy relationship between q and h. To discover hypernyms for a given query q (rather than classify a given pair of words), we apply this decision function to all candidate hypernyms, and select the most likely candidates (or all those classified as hypernyms). This decision function can be learned in a supervised fashion using examples of pairs of words that are related by hypernymy and pairs that are not.

$$
\begin{aligned}
P_i &= (\phi_i.e_q)^T \quad \phi_i \in R^{d\times d} \text{ for } i \in 1, \ldots, k \\
s &= P.e_h \quad s \in R^{k\times 1} \\
u &= \sigma(W.s + b) \quad P \in R^{k\times d} \\
H(q,h,t) &= t \times \log(y) + (1-t) \times \log(1-y)
\end{aligned}
$$

Above set of equations are heart of supervised algorithm.

How they are being fused?

1. Select top 100 candidates according to each hyponym, normalize their scores and sum them.
2. Rerank the candidates according to this new score. This reranking function favours candidates found by both systems, but also gives a chance to strong candidates found by a single system.

With that being said, our main goal is to design model that mostly motivated from **CRIM** and on top of that w.r.t projection matrix what all alternatives we can look upon is what remaining to get answered in our future work.

4. Project Status

Sl. No.	Milestone	Deliverable	Timeline	Status
1	Project Outline Submission		16^{th} March 2022	Done
2	Data Pre-processing	Constructing *word2vec* Embeddings	14^{th} March 2022	Done
3	Mid-Evaluation		15^{th} March 2022	Done
4	Model Building	Baseline-Model Supervised Models **CRIM** Model	20^{th} March 2022 1^{st} April 2022 15^{th} April 2022	In-progress In-progress To be done
5	Result Compilation		TBA	To be done
6	Final Report		21^{st} April, 2022	To be done
7	Final Presentation		TBA	To be done

5. Challenges

1. Not all Hypernyms are identified for a given Hyponym. For this we explore the Hypernyms derived from other co-hyponym.
2. However, our concerned task, hypernym discovery, is rather more challenging since it requires the systems to explore the semantic connection with all the exhausted candidates in the latent space and rank a candidate set instead of a binary classification in previous work.
3. The other challenge is representation for terms, including words and phrases, where the phrase embedding could not be obtained by word embeddings directly
4. Hypernym discovery in case of entity name cannot be determined by using co-hypernyms. (For eg : For original Hyponym "Musk", the Hypernyms generated by co-hyponyms can be "MuskMelon" ,"Musk Perfume" etc whereas we would be looking for Hypernyms like "Entrepreneur" , "CEO" , "Billionare" etc in case the original Hyponym represents "Elon Musk")
5. Additional contextual information shall be required for word having multiple meanings to identify the correct Hypernyms. (For eg : If our original hyponym is "Bank " in the context of Bank the financial institution, there is always a possibility of getting hypernym for bank from the context of "River Bank" which isn't suitable in our desired context.) Hence in such cases multiple embeddings becomes a necessity . Furthermore it shall be challenging to determine how many such embeddings are required to cover the entire vocabulary.
6. Selecting a specific embedding type has to be chosen very carefully as different embedding styles shall have different embedding space creation thereby possibly returning different Hypernyms for the same search query. (For eg : If we choose between word embedding and sense embedding the Hypernyms discovered in case of sense embedding for domain specific tasks would be more volatile and hence unreliable.) [8])

6. Future Tasks

As per the proposed plan, we are currently implementing the unsupervised pattern based approach after which the next immediate task is implementation of supervised projection based approach as well as Hybrid approach. Furthermore, we would be performing a few additional experiments as an attempt to improve the results obtained by the previous approaches. We shall also try to explore sequential pattern mining technique to automatically extract frequent sequential pattern between hyponym terms and their given hypernyms from the corpus. Final evaluation report shall contain a thorough comparison of results obtained by all the approaches as well as the improvement brought out by additional experiments performed.

References

[1] Marco Baroni and Alessandro Lenci. How we blessed distributional semantic evaluation. In *Proceedings of the GEMS 2011 Workshop on GEometrical Models of Natural Language Semantics*, pages 1–10, 2011.

[2] Gabriel Bernier-Colborne and Caroline Barrière. CRIM at SemEval-2018 task 9: A hybrid approach to hypernym discovery. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 725–731, New Orleans, Louisiana, June 2018. Association for Computational Linguistics.

[3] Jose Camacho-Collados, Claudio Delli Bovi, Luis Espinosa-Anke, Sergio Oramas, Tommaso Pasini, Enrico Santus, Vered Shwartz, Roberto Navigli, and Horacio Saggion. Semeval-2018 task 9: Hypernym discovery. In *Proceedings of the 12th International Workshop on Semantic Evaluation (SemEval-2018); 2018 Jun 5-6; New Orleans, LA. Stroudsburg (PA): ACL; 2018. p. 712–24.* ACL (Association for Computational Linguistics), 2018.

[4] Marti A Hearst. Automatic acquisition of hyponyms from large text corpora. In *COLING 1992 Volume 2: The 14th International Conference on Computational Linguistics*, 1992.

[5] Marti A. Hearst. Automatic acquisition of hyponyms from large text corpora. In *COLING 1992 Volume 2: The 14th International Conference on Computational Linguistics*, 1992.

[6] Rion Snow, Daniel Jurafsky, and Andrew Ng. Learning syntactic patterns for automatic hypernym discovery. *Advances in neural information processing systems*, 17, 2004.

[7] Yogarshi Vyas and Marine Carpuat. Detecting asymmetric semantic relations in context: A case-study on hypernymy detection. In *Proceedings of the 6th Joint Conference on Lexical and Computational Semantics (* SEM 2017)*, pages 33–43, 2017.

[8] Zhuosheng Zhang, Jiangtong Li, Hai Zhao, and Bingjie Tang. Sjtu-nlp at semeval-2018 task 9: Neural hypernym discovery with term embeddings. *arXiv preprint arXiv:1805.10465*, 2018.