Metaphor Detection for Low Resource Languages: From Zero-Shot to Few-Shot Learning in Middle High German

Anonymous ACL submission

Abstract

In this work, we present a novel unsupervised method for adjective-noun metaphor detection on low resource languages. We propose two new approaches: First, a way of artificially generating metaphor training examples and second, a novel way to find metaphors relying only on word embeddings. The latter enables application for low resource languages. Our method is based on a transformation of word embedding vectors into another vector space, in which the distance between the adjective word vector and the noun word vector represents the metaphoricity of the word pair. We train this method in a zero-shot pseudo-supervised manner by generating artificial metaphor examples and show that our approach can be used to generate a metaphor dataset with low annotation cost. It can then be used to finetune the system in a few-shot manner. In our experiments we show the capabilities of the method in its unsupervised and in its supervised version. Additionally, we test it against a comparable unsupervised baseline method and a supervised variation of it.

1 Introduction

The automatic detection of metaphors is a useful tool for literary studies. While many recent supervised approaches for common languages like English exist, those methods rely on large pretrained models like BERT (Devlin et al., 2019) transformers and on labeled metaphor datasets. Those pretrained models and labeled data can not be obtained for low resource languages like Middle High German. To enable metaphor detection in those low resource languages without annotated data we propose a novel unsupervised zero-shot approach based only on simple word embeddings. In our approach, adjective-noun metaphor word pairs are found by transforming their word embeddings into another vector space, where common word pairs are located near each other. At the same time, metaphoric word pairs have a large cosine distance between them. Their cosine distance then serves as a measurement of metaphoricity.

A metaphor, as a semantic figure of speech, is a way of referring to one concept by mentioning another. An example for this would be the phrase *the car drinks gasoline*, where the word *drinks* from the domain of food consumption is applied to word *car* from the domains of transportation and machines. It carries over its base meaning of consumption of liquids, so that the reader understands that the car consumes fuel. Another example would be the phrase *a sweet thought*. Here the word *sweet* from the domain of taste is applied to the word *thought*. While in its base meaning only physical objects can be sweet, the reader understands by their context knowledge and world knowledge that a sweet taste is considered pleasant and thus the aforementioned phrase means a pleasant thought.

In this work, we want to concentrate on adjective-noun pattern like *sweet thought*, *raw emotion*, or *clear answer*. While with the knowledge of syntactical dependencies also more complex forms can be analyzed, we want to limit our approach to methods also applicable to low resource languages like Middle High German, where no syntax parsing is available. Thus, we assume that part-of-speech tags, lemmas and token-based word embeddings like word2vec (Mikolov et al., 2013) or fasttext (Bojanowski et al., 2017) embeddings are obtainable. We do not rely on methods requiring large amounts of training data like transformer models or syntax parsers.

There are different ways to define adjective-noun metaphors to operationalize the search for them. One possibility is to define metaphors as a violation of selectional preference. The approach we focus on, defines the adjective that commonly occur together with a noun as their selection preference. When an adjective that does not typically appear together with the noun emerges, this anomaly is called a selection preference violation. This implies that an adjective from another source domain is used to describe something from the target domain of the noun. It fits our definition of a metaphor. Since our approach should also be applicable to new languages without an existing labeled metaphor dataset in that language, we need to develop an *unsupervised* approach. In Section 3. we explain how to derive such a method from a supervised method.

2 Related Work

The most common current approaches for metaphor detection like MelBERT (Choi et al., 2021) and DeepMet (Su et al., 2020) are based on supervised learning and transformer models. Those models require to be pretrained on a very large corpus with billions of tokens. However, if we want to search for metaphors in low resource languages like Middle High German, using such a large pretrained language model is not possible. Additionally, there may be no training dataset for supervised training available to finetune the model on.

Other approaches like (Reinig and Rehbein, 2019) use supersense taxonomies like GermaNet (Hamp and Feldweg, 1997; Henrich and Hinrichs, 2010), which is comparable to the English WordNet (Fellbaum, 1998), which deliver information about the domain that certain words belong to. However, those external sources of information are not present for low resource languages like Middle High German. In an earlier unsupervised approach, the authors of (Shutova and Sun, 2013) used grammatical relations between words as the basis for a clustering approach based on hierarchical graph factorization. For this approach syntax parsing is necessary, as well. The authors of (Navarro-Colorado, 2015) propose an unsupervised metaphor detection system based on topic modeling. In comparison, they do not search for adjective-noun pairs but instead for single words with metaphorical meaning inside a sentence.

There are also unsupervised approaches that work without labeled data and do not use big pretrained transformer models. Our baseline (Pramanick and Mitra, 2018) uses an approach that clusters adjective-noun pairs using the kmeans algorithm. To cluster the data, six different features are used: (1) abstractness rating of the adjective; (2) abstractness rating of the noun; (3) difference between the abstractness ratings; (4) cosine similarity of the word embeddings of the noun; (5) edit distance from the adjective to the noun, normalized by the number of characters in the adjective; (6) edit distance from the noun to the adjective, normalized by the number of characters in the noun. Clusters are then interpreted as metaphors or non-metaphors. While this approach also uses information - *the abstractness rating* - that may not be present in low resource languages, we consider this a comparable baseline approach to our work. Due to its unsupervised nature, it can also be used on languages without existing metaphor dataset.

3 Method

Our contribution consists of two parts: First, we propose a feedforward neural network that maximizes the cosine distance between the word vectors of an adjective-noun word pair for metaphors and minimizes the distance otherwise. Second, a way to train this model in a zero-shot setting without any metaphor examples. It also covers a step to finetune the system on human annotated metaphors previously proposed by the unsupervised system.

3.1 Metaphor Ranking

The basic idea of our novel approach is to transform the word embeddings of the adjective and the noun into a vector space. The cosine distance between the transformed vectors is small if the adjective is meant literally and large if the adjective has a metaphorical function. The intuition behind this is that words which occur often next to each other should have a low distance by the nature of the word embeddings, while unusual combinations like metaphors should have a higher distance. We use a simple feedforward network with both the same weights for the embedding of the adjective and the embedding of the noun. As a result, we can transform the word vectors into a vector space where this distance property is ensured by the training. The cosine embedding function (Payer et al., 2018) is used as a training loss to maximize the cosine distance if the adjective has a metaphorical meaning and minimizes the distance if the adjective has a literal meaning. The cosine distance of the transformed vectors then represents the metaphoricity of a word pair an can be used to rank all possible metaphor candidates.

method	TSV	poems
supervised	0.90	0.82
SVM baseline features (+abst)	0.92	0.77
SVM baseline features	0.67	0.74
zero-shot	0.70	0.74
baseline (+abst)	0.85	0.81
baseline	0.52	0.83

Table 1: Results of two different experiments: numbers are the average precision, which is the area under the precision-recall-curve. Methods in italics are our approaches; methods marked with +abst use features that are not present in low resource languages.

iteration	GerDraCor	TSV	poems	MHG
base	0.26	0.70	0.74	0.22
iter 1	0.60	0.84	0.77	0.61
iter 2	0.71	0.67	0.74	0.25
iter 3	0.46	0.72	0.78	0.60
iter 4	0.73	0.70	0.77	0.40
iter 5	0.95	0.59	0.78	0.60
iter 6	0.60	0.70	0.82	0.66

Table 2: Results of the iteratively trained model on the GerDraCor corpus on the GerDraCor test set (precision at top 100) and on the TSV and poetry test sets (average precision): The MHG column shows the results on the Middle High German test set (precision at top 100).

3.2 Unsupervised Zero-Shot Training

As a goal, we also want to apply this method to low resource languages like Middle High German where we do not have a labeled metaphor dataset. This makes supervised training impossible. To mitigate this, we assume that the number of metaphorical adjectives in a text is low enough to make the majority of adjective-noun pairs in a text good examples for non-metaphors. Based on this assumption, we generate artificial metaphor examples by using the idea of selectional preference violation. As such, we shuffle the adjectives to generate random adjective-noun pairs and label those as metaphor examples. While this may not result in semantically useful metaphors, it still satisfies the idea of selectional preference violation. It also enables the classifier to distinguish between normal and anomalous pairs.

3.3 Few-Shot Finetuning

With the above mentioned idea, we get a classifier to rank the metaphoricity of adjective-noun pairs using no labeled training data. This approach can then be refined with a human-in-the-loop bootstrapping approach. Using the zero-shot classifier, we can rank all the adjective-noun pairs in the training corpus by their metaphoricity. A human annotator can then annotate the most promising metaphor candidates to generate a metaphor dataset without the need to annotate the whole text. This step can be repeated in an iterative manner, generating better metaphor examples with every annotation step.

4 Experiments

To evaluate our embedding approach as well as our unsupervised labeling approach, we conducted several experiments, which are explained below. We make our code publicly available [1]. Since we want to emulate the search for metaphors in low resource languages, we do not use all features that are possible in the German language. We exclude syntax trees, external knowledge bases like GermaNet and large pre-trained models like BERT (Devlin et al., 2019). We extracted PoS tags, lemmas and tokens using the spaCy (Honnibal et al., 2020) package. As annotated metaphor dataset we used the German version (Reinig and Rehbein, 2019) of the TSV metaphor dataset. Additionally, we used their annotated metaphor dataset from German poetry. However, their approach used features based on GermaNet, a supersese taxonomy which can not be assumed to exist for low resource languages. Hence, we did not compare our method to theirs.

As a corpus for the German case study to extract non-metaphors in an unsupervised manner, we used the GerDraCor (Fischer et al., 2019) corpus. For the case study on the low resource language Middle High German, we used the Referenzkorpus Mittelhochdeutsch (Klein et al., 2016) to train FastText (Bojanowski et al., 2017) word embeddings. We took 22 texts from the Mittelhochdeutsche Begriffsdatenbank (zep, 1992-2021 (laufend)) to analyze our approach on this language. The CLTK (Johnson et al., 2021) package was used to normalize the character representation of the Middle High German texts and to generate PoS tags.

4.1 Supervised Metaphor Retrieval

In the most simple case we have a dataset consisting of word pairs which are either labeled as a metaphor or as non-metaphor. Given these labels,

[1] link will be inserted in the camera ready version

our approach can be used without any modification. For our baseline, we trained an SVM with the features of the otherwise unsupervised baseline method. The baseline features contain an abstractness feature which may not be present in low resource languages. To enable a fair comparison, we used these features both with and without the abstractness feature present. Table 1 shows that our supervised approach achieves similar results to the supervised baseline features together with the abstractness. Without abstractness, our approach achieves a higher average precision by 0.18 percent points.

4.2 Unsupervised Metaphor Retrieval

In this experiment we again used the annotated TSV mataphor dataset and the poems dataset. However, we did not use any examples annotated as metaphors for our zero-shot approach. As explained in Section 3, we used randomly connected adjectives and nouns from the non-metaphor set as metaphor examples. Results in Table 1 (marked as *zero-shot*) show that we get slightly lower average precision than the baseline approach with the abstractness features. However, we get far better avarage precision numbers than the baseline approach without the abstractness features.

4.3 Baseline

As baseline experiments we used the methods explained in the related work section. Since the abstractness features are not present in low resource languages, we also conducted an experiment without these features. To compare this with the supervised approach, we also used the baseline features with a kernel SVM in a supervised manner.

4.4 Case Studies

Our main contribution is a method to generate a metaphor dataset and create a metaphor retrieval system for a low resource language with no previously annotated metaphor dataset. To analyze whether our approach is suitable for this, we conducted two case studies: One on German and one on Middle High German.

For the German texts we extracted adjective-noun pairs from one half of the GerDraCor corpus and used them to train the unsupervised zero-shot system. Two sets of random combinations of adjectives and nouns were used as pseudo metaphor examples. For the Middle High German Data we used eleven texts from the Mittelhochdeutsche Begriffsdatenbank to extract word pairs. In every iteration we then annotated the top 100 rated unannotated examples in the training corpus, the bottom 50 unannotated examples and another random 50 unannotated examples. This strategy allows to build a metaphor training dataset for both of these languages. We discarded multiple occurance of the same word pairs as well as ambiguous examples and detections based on errors like wrong PoS tagging. For German, the final dataset contained 390 metaphors and 449 non-metaphors, for Middle High German, it was 287 metaphors and 365 non-metaphors, respectively. For testing, we annotated the top 100 results on the other half of the GerDraCor corpus for German and the top 100 results on eleven other texts from the Mittelhochdeutsche Begriffsdatenbank for Middle High German.

The results in Table 2 show that the zero-shot classifier found 26 metaphors in the top-100 results for German and 22 metaphors in the top-100 results for Middle High German. After only one round of annotation, this already increased to 60 metaphors for German and 61 metaphors for Middle High German. However, it can also be seen that for further iterations this process is still not completely stable. While a tendency towards improvement can be seen, further investigation are necesssary.

5 Conclusion

In this work, we presented a novel unsupervised method to enable metaphor detection. We demonstrated that our approach improves over comparable baseline approaches. The design of our method allows us to apply it to low resource languages without further changes. Our method produces excellent results when used in a supervised manner. While the results are worse when the method is used without labeled data, the method can still be used to enable a bootstrapping approach. There, metaphor candidates are extracted from a text in an unsupervised manner, labeled, and then used to train the supervised version method. Thus, our approach on the one hand enables metaphor detection in uninvestigated low resource languages, and on the other hand serves as a powerful supervised tool once the first metaphors have been discovered. An interesting next step would be to combine our approach with other unsupervised approaches mentioned in the related work section, that are applicable for low resource languages.

References

1992-2021 (laufend). Mittelhochdeutsche begriffsdatenbank (mhdbdb).

Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5:135–146.

Minjin Choi, Sunkyung Lee, Eunseong Choi, Heesoo Park, Junhyuk Lee, Dongwon Lee, and Jongwuk Lee. 2021. MelBERT: Metaphor detection via contextualized late interaction using metaphorical identification theories. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1763–1773, Online. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Christiane Fellbaum. 1998. *WordNet: An Electronic Lexical Database*. Bradford Books.

Frank Fischer, Ingo Börner, Mathias Göbel, Angelika Hechtl, Christopher Kittel, Carsten Milling, and Peer Trilcke. 2019. Programmable corpora: Introducing dracor, an infrastructure for the research on european drama.

Birgit Hamp and Helmut Feldweg. 1997. GermaNet - a lexical-semantic net for German. In *Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications*.

Verena Henrich and Erhard Hinrichs. 2010. GernEdiT - the GermaNet editing tool. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta. European Language Resources Association (ELRA).

Matthew Honnibal, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. 2020. spaCy: Industrial-strength Natural Language Processing in Python.

Kyle P. Johnson, Patrick J. Burns, John Stewart, Todd Cook, Clément Besnier, and William J. B. Mattingly. 2021. The Classical Language Toolkit: An NLP framework for pre-modern languages. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: System Demonstrations*, pages 20–29, Online. Association for Computational Linguistics.

Thomas Klein, Klaus-Peter Wegera, Stefanie Dipper, and Claudia Wich-Reif. 2016. Referenzkorpus mittelhochdeutsch (1050-1350), version 1.0).

Tomás Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. In *1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings*.

Borja Navarro-Colorado. 2015. A fully unsupervised topic modeling approach to metaphor identification - una aproximacion no supervisada a la deteccion de metaforas basada en topic modeling.

Christian Payer, Darko Štern, Thomas Neff, Horst Bischof, and Martin Urschler. 2018. Instance segmentation and tracking with cosine embeddings and recurrent hourglass networks. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018 - 21st International Conference, 2018, Proceedings*, Lecture Notes in Computer Science, pages 3–11. Springer Verlag Heidelberg. 21st International Conference on Medical Image Computing and Computer Assisted Intervention, MICCAI 2018 ; Conference date: 16-09-2018 Through 20-09-2018.

Malay Pramanick and Pabitra Mitra. 2018. Unsupervised detection of metaphorical adjective-noun pairs. In *Proceedings of the Workshop on Figurative Language Processing*, pages 76–80, New Orleans, Louisiana. Association for Computational Linguistics.

Ines Reinig and Ines Rehbein. 2019. Metaphor detection for german poetry. In *Proceedings of the 15th Conference on Natural Language Processing (KONVENS 2019): Long Papers*, pages 149–160, Erlangen, Germany. German Society for Computational Linguistics & Language Technology.

Ekaterina Shutova and Lin Sun. 2013. Unsupervised metaphor identification using hierarchical graph factorization clustering. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 978–988, Atlanta, Georgia. Association for Computational Linguistics.

Chuandong Su, Fumiyo Fukumoto, Xiaoxi Huang, Jiyi Li, Rongbo Wang, and Zhiqun Chen. 2020. DeepMet: A reading comprehension paradigm for token-level metaphor detection. In *Proceedings of the Second Workshop on Figurative Language Processing*, pages 30–39, Online. Association for Computational Linguistics.