Back to the Roots:
Predicting the Source Domain of Metaphors using Contrastive Learning

Meghdut Sengupta and **Milad Alshomary** and **Henning Wachsmuth**
Leibniz University Hannover, Hannover, Germany
Institute of Artificial Intelligence
{m.sengupta, m.alshomary, h.wachsmuth}@ai.uni-hannover.de

Abstract

Metaphors frame a given target domain using concepts from another, usually more concrete, source domain. Previous research in NLP has focused on the identification of metaphors and the interpretation of their meaning. In contrast, this paper studies to what extent the source domain can be predicted computationally from a metaphorical text. Given a dataset with metaphorical texts from a finite set of source domains, we propose a contrastive learning approach that ranks source domains by their likelihood of being referred to in a metaphorical text. In experiments, it achieves reasonable performance even for rare source domains, clearly outperforming a classification baseline.

1 Introduction

Metaphors foster meaning in language by establishing a mapping between two conceptual domains, where concepts rooted in a usually rather concrete *source domain* are projected to a usually rather abstract *target domain* (Lakoff and Johnson, 2003). In other words, metaphors explain one concept in terms of another concept. For example, in the sentence "the sales tax would generate $12 billion in annual tax revenues", the target domain *taxation* is described through concepts from the source domain *machine*, as indicated by the verb "generate".

Recent research suggests that even state-of-the-art NLP models face problems with making inferences on figurative language such as metaphors (Chakrabarty et al., 2021). To better comprehend the meaning intended by metaphorical language, additional levels of understanding need to be incorporated. So far, past research in natural language processing has focused on the distinction of literal from metaphorical text (Shutova et al., 2010) as well as on the interpretation of metaphors in terms of understanding their literal meaning from their intended meaning and vice versa (Shutova et al., 2012; Stowe et al., 2021). For these tasks, the mapping between source and target domain has often been used as an effective cue. To the best of our knowledge, however, no work directly attempts the actual identification of the conceptual domains of metaphors from a given sentence. A reason behind may lie in the theoretical unboundedness of the number of concepts (and, as a result, the space of metaphors) associated with a single concept.

In this paper, we study to what extent source domains can be predicted computationally from given metaphorical sentences. We restrict our view to the slightly simplified setting in which a set of possible source domains is predefined (but possibly large). Conceptually, this makes the task a classification problem: Given the sentence, assign it to the correct source domain.

However, for larger numbers of source domains, it may be hard to learn a reliable classification model, particularly when annotated metaphor data is limited. Instead, we therefore propose a contrastive learning approach (Zhang et al., 2022) based on our hypothesis that the source domain and the metaphorical sentences are related linguistically. The approach ranks all source domains based on the similarity of their embeddings to the embedding of the given sentence. At inference time, it then chooses the top-ranked source domain.

We evaluate our approach on the corpus of Gordon et al. (2015), covering 1429 English metaphorical sentences and 138 source domains. With an accuracy of 0.619, our approach clearly outperforms transformer-based classification baselines, especially on rare source domains. Even though the unboundedness problem remains, we thereby contribute towards a better computational understanding of metaphorical language. To go beyond, we expect that modeling external knowledge about source domains will be needed.

2 Related Work

As stated above, past NLP research has tackled the study of metaphors mostly in the form of two

Proceedings of the 3rd Workshop on Figurative Language Processing (FLP), pages 137 - 142
December 8, 2022 ©2022 Association for Computational Linguistics

Sentence	Metaphor	Src. Domain
The sad news is with the exception of very few no firearm organisation is doing anything of the slightest value in fighting gun control.	fighting	Struggle, War
This is the historical context of Obama's election victory.	victory	Competition, Game, War
They attack ""rich people"" while enjoying all the spoils of their luck, I have zero problems with earned wealth, but these clowns literally lucked out in life.	attack	War

Table 1: Example sentences from the dataset having one or more than one concepts grouped as the source domain.

tasks: metaphor identification (Mao et al., 2018; Do Dinh and Gurevych, 2016) and metaphor interpretation (Beust et al., 2003; Shutova, 2010). Most works in these research fields build on the work of Lakoff and Johnson (2003) on the interpretation of intended meanings in metaphorical expressions. The author theorized different metaphors in terms of mapped concepts (source and target domains). Approaches to metaphor interpretation have particularly witnessed unsupervised extraction of source domains and target domains to interpret the intended meaning of metaphorical expressions (Li et al., 2013; Yu and Wan, 2019). In contrast, we seek to predict the source domain, even if it is not mentioned in the text.

Notable research combining metaphor identification and interpretation has been carried out by Shutova et al. (2013). The authors first identified metaphors by verb and noun clustering, followed by interpreting the intended meaning of the metaphors by addressing it as a paraphrasing task.

Li et al. (2013) modeled explicit conceptual metaphors (where the source domain and the target domain are situated as excerpts of text in the sentence) and implicit conceptual metaphors (where the two domains are not apparent), where they extracted source and target domains in an unsupervised approach. A limitation of their work is that no evaluation is provided regarding how authentic the source domains and the target domains are that are excavated.

Recently Stowe et al. (2021) have interpreted metaphors by extracting source and target domains from the semantic space of their associated concepts in FrameNet (Ruppenhofer et al., 2016), to generate metaphorical expressions. We complement this study in that we assess how well source domain prediction works when the set of domains is known in advance.

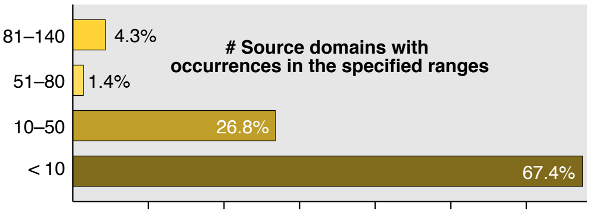

Figure 1: Insights into the distribution of the given data: 67.4% of the source domains are referred to in less than 10 metaphors, 4.3% occur between 81 to 140 times, etc.

Ahrens and Jiang (2020) developed an algorithm to identify source domains from text with the help of lexical resources like WordNet, which partially addressed the unboundedness problem of source domains. However, their work is essentially an annotation procedure for source domain verification.

The only datasets suitable for our purposes are the one of Shutova and Teufel (2010), where source and target domains have been annotated manually, and the one of Gordon et al. (2015) where both conceptual source and target domains and their linguistic triggers are given. We rely on the latter, since the former one has only 761 samples.

3 Data

To study the task of predicting the source domain of metaphors, we need data where source domains are annotated. We employ the dataset of Gordon et al. (2015), which was originally created to explore how the *meaning shift* (Shutova et al., 2013) happens between source and target domains. The dataset contains 1771 metaphorical sentences, spanning 70 source domains annotated for the linguistic metaphors (metaphorical text excerpts in the sentence corresponding to source and target domains). We use the "source linguistic metaphor" and henceforth refer to it simply as *metaphor*. For example, in the sentence "An invasion of wealth may not suit their interests", the metaphor is "invasion" and the annotated source domain is *War*.

Table 1 shows three example metaphors from the dataset. As can be seen, some metaphors pertain to more than one source domain. For example, in the sentence "This is the historical context of Obama's election victory", the metaphor "victory" has the source domains *Competition*, *Game*, and *War*. In

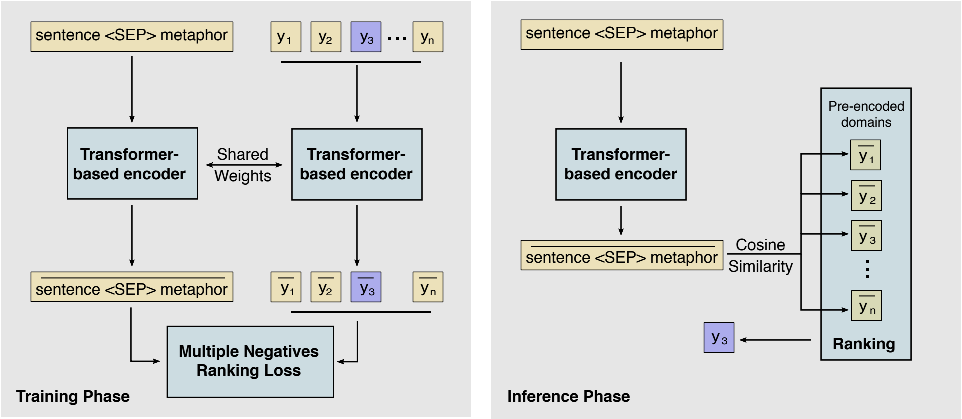

Figure 2: Our contrastive learning approach: During training, we optimize the transformer models based on Multiple Negatives Ranking Loss. At inference, we select the source domain most similar to a given metaphorical sentence.

this paper, we see such cases as composite source domains, that is, if a metaphor in a given sentence has multiple source domains, we treat them as one new source domain. As a result, the total number of source domains in our work is 138. Figure 1 shows the distribution of source domains in the whole dataset, underlining the complexity of the problem and the sparsity of the data.

4 Approach

For a predefined set of domains, we here model source domain prediction as a ranking task. Given a metaphorical sentence as input, we rank all candidate domains by their likelihood of being the source domain based on their semantic similarity to the sentence. Then, we choose the top-ranked domain as the predicted source domain.

To that end, we develop a contrastive learning approach which compares the semantic representations of the input sentence and the candidate domains. Figure 2 gives an overview.

4.1 Training Phase

On a training set, our approach learns to minimize the semantic distance of the correct source domain from the given metaphorical sentence. For representing the data at hand, we build on the recent success of sentence transformers (Reimers and Gurevych, 2019), which leverage efficient representations for different downstream tasks. We fine-tune a sentence transformer as follows:

1. We pass the sentence (concatenated with its metaphor by a separator token) and each source domain through two transformer-based encoders with shared weights, in order to obtain an embedding for each. Our central idea revolved around exploring how our approach works. To test the approach to it's full potential we refrain from using large transformer based encoders like T5 (Raffel et al., 2020) - which we think may affect the model performance to the extent, where understanding what is responsible for a good model performance - the approach or the encoder - would be difficult. Hence, we simply use BERT (Devlin et al., 2019) and DistilBERT (Sanh et al., 2019) [1] as encoders for creating the sentence representations.

2. For a vector of sentences $\mathbf{x}$ and corresponding correct source domains $\mathbf{y}$, that is, with only positive instance pairs (x_i, y_i) with $x_i \in \mathbf{x}$ and $y_i \in \mathbf{y}$ like Reimers and Gurevych (2020) we rely on *Multiple Negatives Ranking Loss* (Henderson et al., 2017), where x_i along with each domain y_j, $j \neq i$, is used as a negative pair. Let $k = |X| = |Y|$ be the number of pairs, then we compute the loss as:

$$\begin{aligned} & \mathcal{L}(\mathbf{x}, \mathbf{y}, \theta) \\ = \; & -\frac{1}{k} \cdot \sum_{i=1}^{k} \log P_{\text{approx}}(y_i | x_i) \\ = \; & -\frac{1}{k} \cdot \sum_{i=1}^{k} \Big(S(x_i, y_i) - \log \sum_{j=1}^{k} e^{S(x_i, y_j)} \Big) \end{aligned}$$

In line with Henderson et al. (2017), $S(x, y)$ is the score of an instance computed from the sentence embeddings. The ranking function is defined

[1] Specifically, we use 'bert-base-uncased' and 'distilbert-base-uncased' as the pre-trained checkpoints. These are the variants with the lowest number of parameters of BERT and DistilBERT respectively.

Approach	Encoder	Accuracy
Majority baseline	–	0.063
Classification baseline	BERT	0.421
	DistilBERT	0.473
Contrastive learning	BERT	**0.619**
	DistilBERT	0.612

Table 2: Main results: Accuracy of our approach and the baselines. Using BERT, our approach performs best.

by θ which is a vector storing the current parameters of the transformer-based encoders. Following the idea of contrastive learning, the loss will be minimized, if positive instances get high scores and negative instances low scores.

4.2 Inference Phase

At inference time, the input is just a sentence concatenated with its metaphor. We pass this input through the encoder to obtain its embedding. Using a ranking evaluator, we next compute the cosine similarity in terms of the paired cosine distance between the sentence embedding and the pre-encoded embeddings of each of the candidate source domains. Then, we take the most similar source domain as our predicted output, that is, the one whose embedding has the minimum distance to the sentence embedding.

5 Experiments

This section reports on first experiments that we carried out to evaluate our approach to source domain prediction against different baselines. The goal was to study whether and when contrastive learning provides advantages over standard classification in the given task. [2]

5.1 Experimental Setup

We relied on the following experimental setup:

Data From the dataset described in Section 3, we omitted two instances that were corrupt. We also removed a few duplicates: These instances had the same sentence and source domain, but a different value for some attribute that we did not use (e.g., "schema slot"). Afterwards, we split the remaining 1429 texts randomly into 70% for training (1000 texts), 10% for validation (128 texts), and 20% for testing (301 texts). The split is preserved for reproducibility. We evaluate our model with top-1 accuracy score with our ranking evaluator as mentioned previously.

Majority Baseline To assess how much can be learned from the data, we employ a majority baseline that always predicts the majority source domain found in the training set.

Classification Baselines As discussed initially, the given task conceptually defines a classification problem. Accordingly for baselines, we fine-tune attention-based sequence-to-sequence language transformers in symmetry with the encoders of our contrastive learning approach, namely BERT and DistilBERT, to directly classify the source domains.[3] We report the final score in terms of the average accuracy over 20 iterations of each model. We optimized both models with AdamW (Loshchilov and Hutter, 2017) in six epochs, batches of size 32, a learning rate of 5^{-5}.

Contrastive Learning (Approach) The two configurations of our approach follows the concept discussed in Section 4. Also here, we report the average accuracy over 20 iterations for each model. We optimized both variants in 6 epochs, batches of size 32, and a learning rate of 5^{-5}.

5.2 Main Results

Table 2 presents the results of all evaluated models on the test set. The majority baseline achieves an accuracy of 0.063. While the classifier based on DistilBERT predicts a little less than half of all source domains correctly (0.473), our contrastive learning approaches clearly outperform all baselines, supporting our hypothesis. Still, the highest accuracy (0.619 based on BERT) reveals room for improvement, possibly suggesting a need for more knowledge about source domains and their connections to the concepts being mentioned.

5.3 Results across Source Domains

One major challenge regarding the task is the number of source domains involved and their distribution. As shown in figure 1, 67.4% of the source domains occur in less than 10 metaphors - indicating there are less than 10 instances of these source

[2] The experiment code can be found at `https://github.com/webis-de/FIGLANG-22`.

[3] Due to the high number of source domains (i.e., classes here) in the data, we considered grouping similar source domains and performing the classification in a two step process. We decided against, though, since many of the source domains occur rarely only (see Figure 1), so we would lose a substantial amount of information during grouping.

		# Src. Domain Occurrences			
Approach	**Encoder**	**< 10**	**10–50**	**51–80**	**81–140**
Classification baseline	BERT	0.000	0.214	0.504	0.823
	DistilBERT	0.000	0.376	**0.522**	**0.856**
Contrastive learning	BERT	0.480	**0.694**	0.511	0.632
	DistilBERT	**0.512**	0.664	0.500	0.615

Table 3: Result analysis: Accuracy on different subsets of the test set, partitioned based on the occurrences of the source domains in accordance with Figure 1.

domains in the dataset. This is particularly important because this represents the real-life scenario about how source domains occur in metaphors. Ideally, an approach for identifying source domains should be able to perform well in this scenario.

To see how our approach compares to the classification baseline across the distribution of the source domains in the dataset, we partitioned the test instances into four subsets depending on the occurrences of source domains (using the ranges from Figure 1).

Table 3 reports the average accuracy over 5 iterations on each subset, keeping all other hyperparameters same as discussed previously. As can be seen, our approach consistently outperforms the classification baselines in the case of rarer source domains (< *10* and *10–50*), which denotes the vast majority of the dataset. In contrast, the classification baselines perform better on the subsets with frequent source domains (*51–80* and *81-140*) While this suggest that more data may make classification suitable, the unboundedness of metaphors renders sufficient data unlikely in general. We thus conclude that our approach generalizes better to real-world scenarios with multiple source domains likely to be present in scanty data distributions.

6 Conclusion

Understanding a metaphor includes the recognition of the source domain from which concepts are projected to the target domain being discussed. In this paper, we have proposed a contrastive learning approach to recognize the source domain from a given metaphorical text computationally, when the set of domains is predefined. Experiments suggest that the approach works reasonably well, particularly for source domains that are represented scarcely, which we expect to likely happen often in real-world situations. However, the obtained results also reveal notable room for improvement. In future work, we plan to investigate the impact of modeling external knowledge about the domains as well as the recognition of source domains in unbounded settings.

Acknowledgments

This work has been supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), under project number TRR 318/1 2021 – 438445824.

Limitations

In our work, we have formulated our approach on the assumption that a given set of metaphors have a finite predefined set of source domains. In a real-world scenario, however, the possible candidates for a source domain of a metaphor are theoretically unbounded. Hence, while our assumption is a start towards modeling source domain prediction, it definitely leaves questions to be answered in this context. Moreover, we restricted our view to classification and contrastive learning approaches in this paper as an initial investigation of the task. Other NLP techniques may be worth considering, such as few-shot learning and active learning. We plan to investigate these in the future to get a better idea of the capabilities of our approach. Finally, we point that the observations we make in this paper about metaphor may not all generalize to other languages than English. Metaphor use has language-specific peculiarities that we left untouched here.

Ethical Statement

We do not see any immediate ethical concerns with the study presented in this paper. The data we used is freely available, and a potential misuse of the approach we develop for ethically doubtful use cases seems not apparent to us.

References

Kathleen Ahrens and Menghan Jiang. 2020. Source domain verification using corpus-based tools. *Metaphor and Symbol*, 35(1):43–55.

Pierre Beust, Stéphane Ferrari, Vincent Perlerin, et al. 2003. Nlp model and tools for detecting and interpreting metaphors in domain-specific corpora. In *Proceedings of the Corpus Linguistics 2003 conference*, pages 114–123. Citeseer.

Tuhin Chakrabarty, Debanjan Ghosh, Adam Poliak, and Smaranda Muresan. 2021. Figurative language in recognizing textual entailment. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 3354–3361, 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.

Erik-Lân Do Dinh and Iryna Gurevych. 2016. Token-level metaphor detection using neural networks. In *Proceedings of the Fourth Workshop on Metaphor in NLP*, pages 28–33.

Jonathan Gordon, Jerry Hobbs, Jonathan May, Michael Mohler, Fabrizio Morbini, Bryan Rink, Marc Tomlinson, and Suzanne Wertheim. 2015. A corpus of rich metaphor annotation. In *Proceedings of the Third Workshop on Metaphor in NLP*, pages 56–66, Denver, Colorado. Association for Computational Linguistics.

Matthew Henderson, Rami Al-Rfou, Brian Strope, Yun hsuan Sung, László Lukács, Ruiqi Guo, Sanjiv Kumar, Balint Miklos, and Ray Kurzweil. 2017. Efficient natural language response suggestion for smart reply. *ArXiv e-prints*.

George Lakoff and Mark Johnson. 2003. *Metaphors We Live By*. University of Chicago Press.

Hongsong Li, Kenny Q. Zhu, and Haixun Wang. 2013. Data-driven metaphor recognition and explanation. *Transactions of the Association for Computational Linguistics*, 1:379–390.

Ilya Loshchilov and Frank Hutter. 2017. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*.

Rui Mao, Chenghua Lin, and Frank Guerin. 2018. Word embedding and wordnet based metaphor identification and interpretation. In *Proceedings of the 56th annual meeting of the association for computational linguistics*. Association for Computational Linguistics (ACL).

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, Peter J Liu, et al. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *J. Mach. Learn. Res.*, 21(140):1–67.

Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*.

Nils Reimers and Iryna Gurevych. 2020. Making monolingual sentence embeddings multilingual using knowledge distillation.

Josef Ruppenhofer, Michael Ellsworth, Myriam Schwarzer-Petruck, Christopher R Johnson, and Jan Scheffczyk. 2016. Framenet ii: Extended theory and practice. Technical report, International Computer Science Institute.

Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *CoRR*, abs/1910.01108.

Ekaterina Shutova. 2010. Automatic metaphor interpretation as a paraphrasing task. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 1029–1037, Los Angeles, California. Association for Computational Linguistics.

Ekaterina Shutova, Lin Sun, and Anna Korhonen. 2010. Metaphor identification using verb and noun clustering. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pages 1002–1010, Beijing, China. Coling 2010 Organizing Committee.

Ekaterina Shutova and Simone Teufel. 2010. Metaphor corpus annotated for source - target domain mappings. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta. European Language Resources Association (ELRA).

Ekaterina Shutova, Simone Teufel, and Anna Korhonen. 2013. Statistical metaphor processing. *Computational Linguistics*, 39(2):301–353.

Ekaterina Shutova, Tim Van de Cruys, and Anna Korhonen. 2012. Unsupervised metaphor paraphrasing using a vector space model. In *Proceedings of COLING 2012: Posters*, pages 1121–1130, Mumbai, India. The COLING 2012 Organizing Committee.

Kevin Stowe, Nils Beck, and Iryna Gurevych. 2021. Exploring metaphoric paraphrase generation. In *Proceedings of the 25th Conference on Computational Natural Language Learning*, pages 323–336, Online. Association for Computational Linguistics.

Zhiwei Yu and Xiaojun Wan. 2019. How to avoid sentences spelling boring? towards a neural approach to unsupervised metaphor generation. 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 861–871, Minneapolis, Minnesota. Association for Computational Linguistics.

Rui Zhang, Yangfeng Ji, Yue Zhang, and Rebecca J. Passonneau. 2022. Contrastive data and learning for natural language processing. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Tutorial Abstracts*, pages 39–47, Seattle, United States. Association for Computational Linguistics.