Filter and Annotate: Towards Automatic Identification of Genuine Metaphoricity

Erik-Lân Do Dinh, Iryna Gurevych
Ubiquitous Knowledge Processing (UKP) Lab
Technische Universität Darmstadt
http://www.ukp.tu-darmstadt.de

Petra Gehring
Institute of Philosophy
Technische Universität Darmstadt
gehring@phil.tu-darmstadt.de

***Abstract*—Natural Language Processing has largely addressed automatic metaphor detection on grounds of (cognitive) linguistic frameworks, especially the Conceptual Metaphor Theory of Lakoff and Johnson [1], which sees metaphor as ubiquitous. In contrast, in this work we view metaphor as an exceptional phenomenon [2]. This change in perspective affects applicability of machine learning approaches for metaphor detection, usage of corresponding features, as well as availability of datasets. We propose a combination of manual annotation and automatic filtering as an approach to conduct first steps into the direction of genuine metaphor detection.**

***Index Terms*—conceptual metaphor theory, genuine metaphors, automatic metaphor detection**

I. Motivation

Metaphors have been a long standing subject in Natural Language Processing (NLP). More than just mere literary devices, phrases like *green energy* or *collapsing economy* permeate our everyday language. Accordingly, NLP researchers often base their models for detecting such metaphorical phrases in text corpora on the Conceptual Metaphor Theory (CMT) [1]. In short, it states that a metaphor is constructed using a *source* and a *target* domain, and constitutes a mapping between these domains. For example, the metaphor *the economy collapsed* maps the concrete domain of BUILDING to the more abstract domain of SYSTEM, facilitating understanding of the latter. Further, such a metaphor may entail many more metaphorical linguistic expressions, e.g., an economy can be *rebuilt*, be supported by *pillars*, or even be found *in ruins*.

Due to the dominance of the CMT in NLP, other definitions of metaphor are usually not examined. While one of the prominent features of metaphors in the CMT is their ubiquity, Gehring stresses that rarity and interaction between an expression and its larger context, or frame, are instead defining attributes of metaphors [2], [3]. In difference to the CMT, this interaction does not need to be binary in its nature. For example, in the "genuine" metaphor

> "[we have to] possess ourselves and not let our machine possess us, we have to get the technological gallop under non-technological control" [4, p. 91]

This work has been supported by the German Federal Ministry of Education and Research (BMBF) under the promotional reference 01UG1816B (CEDIFOR).

"technology and the gallop and control (together with the question of possession and the machine) are comprehensible, but the metaphor unfolds in a complex manner" [3].

II. Related Work

The differentiation between the CMT and such genuine metaphors also has an impact on computational approaches. The CMT enables automatic methods to capitalize on the binary source/target domain interaction by using features such as concreteness difference between interacting terms (e.g., *soft [concrete] power [abstract]*; [5]), violations of selectional preferences (e.g., *swallow anger [emotion instead of expected fluid]*; [6]), or topic information (e.g., *capture [topic: hunting] a photo*; [7]). Usually, only certain grammatical constructions are investigated, e.g., adjective-noun or verb-object constructions [8]. However, problems arise due to the mentioned ubiquity of textual instantiations of conceptual metaphors—for many statistically collected features, the metaphorical sense of many words is already encoded. On the other hand, genuine metaphors are 1) inherently rare; and 2) do not necessarily conform to the conceptual binary source/target domain scheme; however, they are still constituted by a certain break in the interaction between (multiple) components. Certain features are still applicable (e.g., topic information), while others are more suited to the CMT (e.g., concreteness). Genuine metaphors are the exception to the norm, and thus are difficult to model. Consequently, we adapt and extend an existing approach to detect novel metaphors [9]. In it, Schulder and Hovy introduce the TF.IDF-based *term relevance* metric, which measures "how 'out of place' a word is in a given context."

III. Approach

As a first step, we propose to take this "negative" definition of genuine metaphors seriously, and introduce a filtering approach. As it is difficult to positively state what constitutes a metaphor, we instead filter out "normal" language use (including ubiquitous metaphors), using an approach similar to [9]. This may both mean normal with regard to general language use, e.g., in comparison to a large background corpus, and normal with regard to the text under investigation itself. Further, we employ the other criterion: the interaction between text components, or more specifically, the semantic break between an expression and its context, using topic and

978-1-5386-9156-4/18/$31.00 ©2018 IEEE
DOI 10.1109/eScience.2018.00067

sense information. For supervised approaches to detection of metaphors under the CMT, many, small annotated datasets exist, however often using a slightly different concrete realization of the definition. In contrast, there is no such dataset for genuine metaphoricity at all. Closest to our use case are the English datasets annotated for novel metaphor relations [10] and novel metaphor tokens [11].

Thus, our approach consists of two phases:

- First, using unsupervised methods (e.g., derived from [9]), we mark likely non-metaphoric sentences in given documents. The user can adjust thresholds and parameters to influence this pre-selection.
- Second, we use an interactive annotation approach to further narrow down metaphoric sentences. This phase alternates user annotation and supervised learning to further filter out non-metaphoric sentences.

For the first phase, we combine relatedness measures from three levels: global (external background corpus / document collection), regional (inner-document), and local (inner-sentence). For the second phase, we approach the problem of data sparseness using existing annotation tools. Further, while the actual metaphor can span arbitrarily many tokens, for easier modeling purposes we constrain ourselves to sentence-based classification and annotation, i.e., we annotate and classify sentences that contain genuine metaphoricity. After automatically filtering out non-metaphorical sentences with unsupervised approaches in the first phase, we present the remaining sentences in the annotation tool. Here, the user can label sentences as non-metaphoric and thus further refine the filter. While this initially presents the user with a large annotation task, our interactive learning approach aims to quickly reduce the amount of data a user has to review to find genuine metaphors. This process is inherently subjective; because metaphor is a highly nuanced phenomenon and can thus be subject to high judgment variability, our first goal is to guide the user in an exploratory search. As such, inter-annotator agreement between multiple users is not of primary concern. However, annotations of other users can be incorporated into the filtering and the supervised learning step.

At a later stage, our approach may be applied to a possibly more appropriate phrase instead of the sentence level. However, for such a refinement we would first need reliable dependency parsing to delineate the phrases. To the best of our knowledge, we are the first to attempt to exclusively identify genuine metaphoricity in document context, in contrast to ubiquitous conceptual metaphors.

References

[1] G. Lakoff and M. Johnson, *Metaphors we live by*. Chicago, IL, USA: University of Chicago Press, 1980.

[2] P. Gehring and I. Gurevych, “Suchen als Methode: Zu einigen Problemen digitaler Metapherndetektion,” *Journal Phänomenologie. Schwerpunkt: Metaphern als strenge Wissenschaft*, no. 41, pp. 99–109, 2014.

[3] P. Gehring, “Taking metaphor’s contextuality seriously: Sharpening theory – reinspecting methodology,” *mediaesthetics – Journal of Poetics of Audiovisual Images*, no. 2, 2017.

[4] H. Jonas, “Warum die Technik ein Gegenstand für die Ethik ist: fünf Gründe [Why technology is an object of ethics. Five reasons],” in *H. Lenk and G. Ropohl (eds.): Technik und Ethik*, pp. 81–91, 1987.

[5] Y. Tsvetkov, L. Boytsov, A. Gershman, E. Nyberg, and C. Dyer, “Metaphor detection with cross-lingual model transfer,” in *Proceedings of ACL 2014*, (Baltimore, MD, USA), pp. 248–258, Association for Computational Linguistics, 2014.

[6] Y. Wilks, “Making preferences more active,” *Artificial Intelligence*, vol. 11, no. 3, pp. 197–223, 1978.

[7] B. Beigman Klebanov, C. W. Leong, M. Heilman, and M. Flor, “Different texts, same metaphors: Unigrams and beyond,” in *Proceedings of the Second Workshop on Metaphor in NLP*, (Baltimore, MD, USA), pp. 11–17, Association for Computational Linguistics, 2014.

[8] M. Rei, L. Bulat, D. Kiela, and E. Shutova, “Grasping the finer point : A supervised similarity network for metaphor detection,” in *Proceedings of EMNLP 2017*, (Copenhagen, Denmark), pp. 1538–1547, Association for Computational Linguistics, 2017.

[9] M. Schulder and E. Hovy, “Metaphor detection through term relevance,” in *Proceedings of the Second Workshop on Metaphor in NLP*, (Baltimore, MD, USA), pp. 18–26, Association for Computational Linguistics, 2014.

[10] N. Parde and R. D. Nielsen, “A corpus of metaphor novelty scores for syntactically-related word pairs,” in *Proceedings of LREC 2018*, (Miyazaki, Japan), pp. 1535–1540, European Language Resources Association, 2018.

[11] E.-L. Do Dinh, H. Wieland, and I. Gurevych, “Weeding out conventionalized metaphors: A corpus of novel metaphor annotations,” in *Proceedings of EMNLP 2018*, (Brussels, Belgium), p. (in press), Association for Computational Linguistics, 2018.