Sensorimotor Enhanced Neural Network for Metaphor Detection

Mingyu Wan[1,2], **Baixi Xing**[3], **Qi Su**[1], **Pengyuan Liu**[3] **and Chu-Ren Huang**[2]

[1]School of Foreign Languages, Peking University
[2]Department of Chinese Bilingual Studies, The Hong Kong Polytechnic University
[3]School of Information Science Language, Beijing Language and Culture University
{wanmy,Liupengyuan,sukia}@pku.edu.cn
{xingbaixi}@gmail.com,{churen.huang}@polyu.edu.hk

Abstract

Detecting metaphors is challenging due to the subtle ontological differences between metaphorical and non-metaphorical expressions. Neural networks have been widely adopted in metaphor detection and become the main stream technology. However, linguistic insights have been less utilized. This work proposes a linguistically enhanced model for metaphor detection extending one published work (WAN et al., 2020) by incorporating the modality norms into attention-based Bi-LSTM. Results show that the current model outperforms most recent works by 0.5%-11% F1, indicating the effectiveness of using modality norms for metaphor detection. This work provides a new perspective to detect token-level metaphoricity by leveraging the modality mismatch between words.

1 Introduction

Metaphors are prevalent in our everyday language even without our consciousness of its presence as we speak and write. It induces the unknown using the known, explains the complex using the simple, and helps us to emphasize the relevant aspects of meaning resulting in effective communication.

In general, metaphor involves certain concept transfer from one domain (Source) to another (Target), as in 'sweet voice' (using taste to describe sound). Lakoff (1980) describes metaphor as a cognitive mechanism (a property of language) reflected by our conceptual system for structuring our understanding of the world. It is a fundamental way to relate our physical and familiar social experiences to a multitude of other subjects and contexts (Lakoff and Johnson, 2008).

As a popular linguistic device, metaphors encode versatile ontological information, which usually involve e.g. domain transfer (Ahrens et al., 2003; Ahrens and Jiang, 2020), sentiment reverse (Steen et al., 2010) or modality shift (Winter, 2019) etc. Therefore, detecting the metaphors in texts is essential for capturing the authentic meaning of the texts, which can benefit many natural language processing applications, such as machine translation, dialogue systems and sentiment analysis (Tsvetkov et al., 2014).

To better understand the intrinsic properties of metaphors and to provide an in-depth analysis to this phenomenon, we propose a linguistically-enriched deep learning model extending one published work (WAN et al., 2020) at ACL Figlang 2020 workshop by incorporating the modality norms into attention-based BiLSTM. As a continuation of their work, we conduct the current research to further testify the effectiveness of leveraging conceptual norms for metaphor detection. For standard reference, we adopt the dataset of the first and second shared tasks of metaphor detection on verbs of the VUA corpus (Klebanov et al., 2018)[1]. Details about the experiment are given in Sections 3-5.

2 Related Work

Research on metaphors have been mainly explored in the context of political communication, mental health, teaching, discourse analysis, assessment

[1]http://www.vismet.org/metcor/documentation/home.html

of English proficiency, among others (Ahrens and Jiang, 2020; Thibodeau and Boroditsky, 2011; Kathpalia and Carmel, 2011; Klebanov et al., 2008; Semino, 2008; Billow et al., 1997; Bosman, 1987).

Over the last decade, automated detection of metaphor has gained increasing research interest among the Natural Language Processing community. Many approaches have been proposed with systems such as traditional machine learning classifiers, deep neural networks and sequential models etc., trained on features of word vectors, n-grams, lexical information, semantic classes, concreteness, word associations, constructions and frames etc. (Hong, 2016; Rai et al., 2016; Do Dinh and Gurevych, 2016; Klebanov et al., 2014; Wilks et al., 2013; Bizzoni and Ghanimifard, 2018; Klebanov et al., 2015).

Early studies of metaphor detection tend to adopt feature-engineering in a supervised machine learning paradigm, which construct feature vectors based on concreteness and imageability, semantic classification using WordNet, FrameNet, VerbNet, SUMO ontology, property norms and distributional semantic models, syntactic dependency patterns, sensorial and vision-based features (Alnafesah et al., 2020; Klebanov et al., 2016; Shutova et al., 2016; Gutierrez et al., 2016).

Recently, deep learning methods have been explored and become the main stream technology for metaphor detection (Mao et al., 2019; Dankers et al., 2019; Gao et al., 2018; Wu et al., 2018; Rei et al., 2017; Gutierrez et al., 2017). To name a few advances, Brooks and Youssef (2020) build up an ensemble of RNN models with Bi-LSTMs and bidirectional attention mechanisms. Chen et al. (2020) employs BERT to obtain the sentence embeddings, and then a linear layer is applied with softmax on each token to make predictions. Maudslay et al.(2020) combines the concreteness of a word with its static and contextual embeddings as inputs into a deep Multi-layer Perceptron network for predicting metaphoricity. Gong et al.(2020) used RoBERTa to obtain word embeddings and concatenate it with linguistic features (e.g. WordNet, VerbNet) as well as other features (e.g. POS, topicality, concreteness), and then feed them into a fully-connected Feedforward network to make predictions.

Despite many advances in the above studies, metaphor detection remains a challenging task. The semantic and ontological differences between metaphorical and non-metaphorical expressions are often subtle and their perception may vary from person to person. These methods show different strengths on detecting metaphors, yet each has its respective disadvantages, such as having generalization problems or lack association of their results with the intrinsic properties of metaphors. In Wan et al. (2020)'s work, they use conceptual features of modality and embodiment norms for metaphor detection based on traditional classifiers (Logistic Regression), which demonstrates the effectiveness of using modality exclusivity information for predicting metaphoricity. The current work aims to merge both strengths of linguistic wisdom and deep learning power into one architecture with the modality enriched neural networks, as illustrated in Section 4.

3 Data Description

3.1 The VUA Corpus

The VU Amsterdam Metaphor Corpus (VUA) (Tekiroğlu et al., 2015)[2] is used in the experiment for training and testing. The dataset consists of 117 fragments sampled across four genres from the British National Corpus: Academic, News, Conversation, and Fiction. The data is annotated using the MIPVU procedure (Steen, 2010) with a strong inter-annotator agreement (k>0.8). This dataset has been used as the competition corpus for two shared tasks on metaphor detection (Leong et al., 2018; Leong et al., 2020), which is publicly available for standard reference.

Information about the size of the sub-genres is given in Table 1. The training and testing texts, sentences, tokens and percentage of metaphors breakdown of the VUA verb track[3] is given in Table 2.

Text Genres	**No. of Tokens**	**No. of Fragments**
Academic texts	49,561 tokens	16 fragments
Conversation texts	48,001 tokens	24 fragments
Fiction texts	44,892 tokens	12 fragments
News texts	45,116 tokens	63 fragments
TOTAL	187,570 tokens	115 fragments

Table 1: Data components of the VUA corpus

[2] http://www.vismet.org/metcor/documentation/home.html

[3] The prediction and evaluation in this paper focuses on the verbs tokens only.

Dataset	Training	Testing
#texts	90	27
#sents	12,123	4,081
#tokens	17,240	5,873
%M	29%	-

Table 2: Number of texts, sentences, tokens, and percentage of metaphors for the VUA corpus

3.2 The Modality Norms

The Lancaster Sensorimotor norms (hereinafter modality norms) collected by Lynott (2019) is used for constructing the linguistic features in the deep learning model. The data include measures of sensorimotor strength (0-5 scale indicating different degrees of sense modalities/action effectors) for 39,707 English words across six perceptual modalities: touch, hearing, smell, taste, vision and interception, and five action effectors: mouth/throat, hand/arm, foot/leg, head (excluding mouth/throat), torso.[4]. Examples of five random words and their six main modality scores are demonstrated in Table 3.

Word	A	G	H	V	O	I
Adopt	1.222	0.056	1.056	**1.889**	0.111	1.222
Big	0.944	0.167	2.722	**3.889**	0.111	0.333
Daze	0.455	0.000	0.000	1.953	0.000	**3.253**
Eat	1.263	**4.526**	2.158	2.632	2.421	2.474
Learn	**3.941**	0.765	1.765	3.882	0.588	1.529

A: Auditory; G: Gustatory; H: Haptic;
V: Visual; O: Olfatory; I: Interoceptive

Table 3: Examples of the Modality Norms

The modality with the highest scores (highlighted) among the six senses of the words marks the dominant sense modality for each word, such as '**Visual**' for words '*Adopt*' and '*Big*'. As sensorimotor information plays a fundamental role in cognition, these norms provide a valuable knowledge representation to the conceptual categories of the tokens in the corpus which may serve as salient features for inferring metaphors. Motivated by the above idea, we propose a modality enriched neural network to further testify its effectiveness.

[4]https://osf.io/7emr6/

4 The Modality Enriched Model

In the modality enriched model, words are processed with the integration of linguistic features and word embedding. We map the modality scores of the words to the norms and obtain modality representations and then use them as inputs to neural networks. The architecture of the modality enriched model is demonstrated in Figure 1.

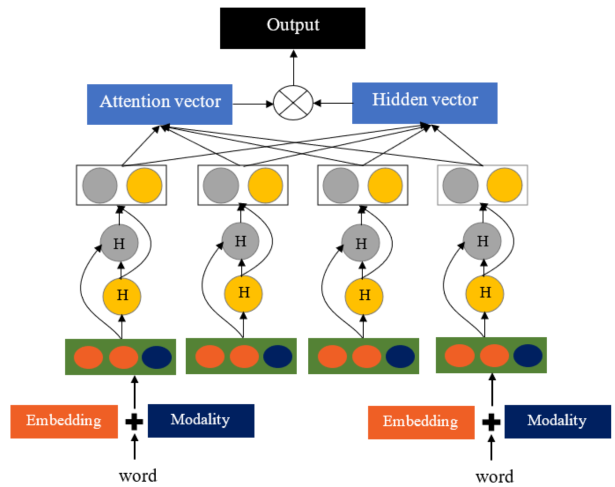

Figure 1: The Modality Enriched Model

Let H $\in$ R$^{d\times N}$ be a matrix consisting of hidden vectors $[h_1, h_2....h_N]$ that is produced by LSTM, where d is the size of hidden layers and N is the length of the given sentence. The attention mechanism will produce an attention weight α. The final sentence representation is given by:

$$h = H \times \alpha^T$$

We also add a additional Linear layer. The final probability distribution is:

$$y = softmax(W_s h + b_s)$$

Let y be the target distribution for sentence, $\hat{y}$ be the predicted sentiment distribution. Train to minimize the cross-entropy error between y and $\hat{y}$ for all sentences.

$$loss = -\sum_i \sum_j y_i^j log \hat{y}_i^j + \lambda \mid\mid \theta \mid\mid^2$$

We use glove embedding and modality vectors to represent the input data. The red circle denotes the usual embedding, the gray circle represents the linguistics feature. We concatenate both representation to generate a new representation as the input of the next layer. LSTM layer produces a hidden status of each word in a sentence. We use these status to calculate an attention weight which will be multiplied with output of LSTM layer. Finally, we get a probability distribution of 0-1 label to train the model and as the prediction result.

5 Experimental Results

In order to evaluate the effectiveness of the proposed model for metaphor detection, we randomly select a development set (4,380 tokens) from the training set (17,240 tokens) in proportion to the Train/Test ratio of the task in Leong et al. (2020). The evaluation results are summarized in Table 4 below:

Category	Approach	P	R	F1
Baseline	uni-gram + LR	0.52	0.66	0.58
Linguistic	modality + linear	0.61	0.56	0.58
	modality + LSTM	0.70	0.68	0.69
Neural	Glove + LSTM	0.74	0.75	0.75
Enriched	modality + Glove + LSTM	**0.77**	**0.76**	**0.76**

Table 4: Evaluation Results of the System

In Table 4, the baseline of using unigram as features and logistic regression (LR) as the classifier is implemented for a basic comparison. It is a commonly adopted baseline in the tasks of metaphor detection. We also implement several sub-categories of approaches before trying the enriched model, including the linguistic and neural networks in separate and also in combination. The results show an 18% F1 improvement of the enriched model over the baseline, a 7% F1 improvement over pure linguistic model, a 1.5% F1 improvement over the pure neural network model, and this superiority is salient and consistent in terms of both P (Precision) and R (Recall).

To further demonstrate the effectiveness of our method, this following table presents the comparisons of our system to some highly related recent works on the same task. All the results are publicly available, as reported in Leong et al. (2020). The detailed results are displayed in Table 5 below:

Our method obtains very promising results: it outperforms 6/7 highly related works to a great extent (0.5%-11% F1 gain), also approaching a reachable performance (a 4% F1 discrepancy) to the Top 1 work in record (Su et al., 2020). Moreover, our results are consistently superior to the top baseline and other linguistically-based or deep learning approaches. This suggests the effectiveness of leveraging modality norms in neural networks for metaphor detection, echoing the hypothesis in Wan et al. (2020) that metaphor manifests a concept mismatch (modality shift in particular) between source and target.

6 Conclusions

We presented a linguistically enhanced method for metaphor detection of VUA verbs using modality features plus attention-based neural network in continuation of Wan et al. (2020)'s first implementation on using conceptual norms for metaphor detection. Inter- and cross-approach comparisons among state-of-the-arts all demonstrate the effectiveness of adding modality information into neural networks for enhancing the performance of metaphor detection. It reconfirms the hypothesis that metaphor manifests a concept mismatch (modality shift in particular) between source and target. Future work will expand the current experiment to predictions of all four lexical words across more datasets.

References

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

Kathleen Ahrens, Siaw Fong Chung, and Chu-Ren Huang. 2003. Conceptual metaphors: Ontology-based representation and corpora driven mapping principles. In *Proceedings of the ACL 2003 workshop on Lexicon and figurative language-Volume 14*, pages 36–42. Association for Computational Linguistics.

Ghadi Alnafesah, Harish Tayyar Madabushi, and Mark Lee. 2020. Augmenting neural metaphor detection with concreteness. In *Proceedings of the Second Workshop on Figurative Language Processing*, pages 204–210.

Richard M Billow, Jeffrey Rossman, Nona Lewis, Deberah Goldman, and Charles Raps. 1997. Observing expressive and deviant language in schizophrenia. *Metaphor and Symbol*, 12(3):205–216.

Yuri Bizzoni and Mehdi Ghanimifard. 2018. Bigrams and bilstms two neural networks for sequential

Work	Method	F1
Wan et al. (2020)	modality + other features + LR	0.652
Kuo and Carpuat (2020)	Bi LSTM+Embeddings+Unigram Lemmas+Spell Correction	0.686
Kumar & Sharma (2020)	Character embeddings+Similarity Networks+Bi-LSTM+Transformer	0.717
Liu et al. (2020)	BERT, XNET + POS tags + Bi-LSTM	0.730
Li et al. (2020)	ALBERT + BiLSTM	0.755
Top base: Devlin et al. (2018)	BERT: Pre-training of deep bidirectional transformers	0.756
The current study	**modality + Glove + LSTM**	**0.761**
Top 1: Su et al. (2020)	Global and local text information+Transformer stacks	0.804

Table 5: Comparison of Results of Our System to State-of-the-art Works

metaphor detection. In *Proceedings of the Workshop on Figurative Language Processing*, pages 91–101.

Jan Bosman. 1987. Persuasive effects of political metaphors. *Metaphor and Symbol*, 2(2):97–113.

Jennifer Brooks and Abdou Youssef. 2020. Metaphor detection using ensembles of bidirectional recurrent neural networks. In *Proceedings of the Second Workshop on Figurative Language Processing*, pages 244–249.

Xianyang Chen, Chee Wee Leong, Michael Flor, and Beata Beigman Klebanov. 2020. Go figure! multi-task transformer-based architecture for metaphor detection using idioms: Ets team in 2020 metaphor shared task. In *Proceedings of the Second Workshop on Figurative Language Processing*, pages 235–243.

Verna Dankers, Marek Rei, Martha Lewis, and Ekaterina Shutova. 2019. Modelling the interplay of metaphor and emotion through multitask learning. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 2218–2229.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

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.

Ge Gao, Eunsol Choi, Yejin Choi, and Luke Zettlemoyer. 2018. Neural metaphor detection in context. *arXiv preprint arXiv:1808.09653*.

Hongyu Gong, Kshitij Gupta, Akriti Jain, and Suma Bhat. 2020. Illinimet: Illinois system for metaphor detection with contextual and linguistic information. In *Proceedings of the Second Workshop on Figurative Language Processing*, pages 146–153.

E Dario Gutierrez, Ekaterina Shutova, Tyler Marghetis, and Benjamin Bergen. 2016. Literal and metaphorical senses in compositional distributional semantic models. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 183–193.

E Dario Gutierrez, Guillermo A Cecchi, Cheryl Corcoran, and Philip Corlett. 2017. Using automated metaphor identification to aid in detection and prediction of first-episode schizophrenia. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2923–2930.

Jisup Hong. 2016. Automatic metaphor detection using constructions and frames. *Constructions and frames*, 8(2):295–322.

Sujata S Kathpalia and Heah Lee Hah Carmel. 2011. Metaphorical competence in esl student writing. *Relc Journal*, 42(3):273–290.

Beata Beigman Klebanov, Daniel Diermeier, and Eyal Beigman. 2008. Lexical cohesion analysis of political speech. *Political Analysis*, pages 447–463.

Beata Beigman Klebanov, Ben Leong, Michael Heilman, and Michael Flor. 2014. Different texts, same metaphors: Unigrams and beyond. In *Proceedings of the Second Workshop on Metaphor in NLP*, pages 11–17.

Beata Beigman Klebanov, Chee Wee Leong, and Michael Flor. 2015. Supervised word-level metaphor detection: Experiments with concreteness and reweighting of examples. In *Proceedings of the Third Workshop on Metaphor in NLP*, pages 11–20.

Beata Beigman Klebanov, Chee Wee Leong, E Dario Gutierrez, Ekaterina Shutova, and Michael Flor. 2016. Semantic classifications for detection of verb metaphors. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 101–106.

Beata Beigman Klebanov, Chee Wee Leong, and Michael Flor. 2018. A corpus of non-native written english annotated for metaphor. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 86–91.

Tarun Kumar and Yashvardhan Sharma. 2020. Character aware models with similarity learning for metaphor detection. In *Proceedings of the Second Workshop on Figurative Language Processing*, pages 116–125.

Kevin Kuo and Marine Carpuat. 2020. Evaluating a bi-lstm model for metaphor detection in toefl essays. In *Proceedings of the Second Workshop on Figurative Language Processing*, pages 192–196.

George Lakoff and Mark Johnson. 1980. Metaphors we live by. *Chicago, IL: University of Chicago*.

George Lakoff and Mark Johnson. 2008. *Metaphors we live by*. University of Chicago press.

Chee Wee Leong, Beata Beigman Klebanov, and Ekaterina Shutova. 2018. A report on the 2018 vua metaphor detection shared task. In *Proceedings of the Workshop on Figurative Language Processing*, pages 56–66.

Chee Wee Leong, Beata Beigman Klebanov, Chris Hamill, Egon Stemle, Rutuja Ubale, and Xianyang Chen. 2020. A report on the 2020 vua and toefl metaphor detection shared task. In *Proceedings of the Second Workshop on Figurative Language Processing*, pages 18–29.

Shuqun Li, Jingjie Zeng, Jinhui Zhang, Tao Peng, Liang Yang, and Hongfei Lin. 2020. Albert-bilstm for sequential metaphor detection. In *Proceedings of the Second Workshop on Figurative Language Processing*, pages 110–115.

Jerry Liu, Nathan O'Hara, Alexander Rubin, Rachel Draelos, and Cynthia Rudin. 2020. Metaphor detection using contextual word embeddings from transformers. In *Proceedings of the Second Workshop on Figurative Language Processing*, pages 250–255.

Dermot Lynott, Louise Connell, Marc Brysbaert, James Brand, and James Carney. 2019. The lancaster sensorimotor norms: multidimensional measures of perceptual and action strength for 40,000 english words. *Behavior Research Methods*, pages 1–21.

Rui Mao, Chenghua Lin, and Frank Guerin. 2019. End-to-end sequential metaphor identification inspired by linguistic theories. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3888–3898.

Rowan Hall Maudslay, Tiago Pimentel, Ryan Cotterell, and Simone Teufel. 2020. Metaphor detection using context and concreteness. In *Proceedings of the Second Workshop on Figurative Language Processing*, pages 221–226.

Sunny Rai, Shampa Chakraverty, and Devendra K Tayal. 2016. Supervised metaphor detection using conditional random fields. In *Proceedings of the Fourth Workshop on Metaphor in NLP*, pages 18–27.

Marek Rei, Luana Bulat, Douwe Kiela, and Ekaterina Shutova. 2017. Grasping the finer point: A supervised similarity network for metaphor detection. *arXiv preprint arXiv:1709.00575*.

Elena Semino. 2008. *Metaphor in discourse*. Cambridge University Press Cambridge.

Ekaterina Shutova, Douwe Kiela, and Jean Maillard. 2016. Black holes and white rabbits: Metaphor identification with visual features. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 160–170.

Gerard J Steen, Aletta G Dorst, J Berenike Herrmann, Anna A Kaal, and Tina Krennmayr. 2010. Metaphor in usage. *Cognitive Linguistics*, 21(4):765–796.

Gerard Steen. 2010. *A method for linguistic metaphor identification: From MIP to MIPVU*, volume 14. John Benjamins Publishing.

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.

Serra Sinem Tekiroğlu, Gözde Özbal, and Carlo Strapparava. 2015. Exploring sensorial features for metaphor identification. In *Proceedings of the Third Workshop on Metaphor in NLP*, pages 31–39.

Paul H Thibodeau and Lera Boroditsky. 2011. Metaphors we think with: The role of metaphor in reasoning. *PloS one*, 6(2):e16782.

Yulia Tsvetkov, Leonid Boytsov, Anatole Gershman, Eric Nyberg, and Chris Dyer. 2014. Metaphor detection with cross-lingual model transfer. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 248–258.

Mingyu WAN, Kathleen Ahrens, Emmanuele Chersoni, Menghan Jiang, Qi Su, Rong Xiang, and Chu-Ren Huang. 2020. Using conceptual norms for metaphor detection. In *Proceedings of the Second Workshop on Figurative Language Processing*, pages 104–109, Online, July. Association for Computational Linguistics.

Yorick Wilks, Adam Dalton, James Allen, and Lucian Galescu. 2013. Automatic metaphor detection using large-scale lexical resources and conventional metaphor extraction. In *Proceedings of the First Workshop on Metaphor in NLP*, pages 36–44.

Bodo Winter. 2019. Synaesthetic metaphors are neither synaesthetic nor metaphorical. *Perception metaphors*, pages 105–126.

Chuhan Wu, Fangzhao Wu, Yubo Chen, Sixing Wu, Zhigang Yuan, and Yongfeng Huang. 2018. Thu ngn at naacl-2018 metaphor shared task: Neural metaphor detecting with cnn-lstm model. In *Proceedings of the Workshop on Figurative Language Processing, New Orleans, LA*.