IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 9, Issue 6, June 2020

Neural Metaphor Identification using Contextual Information

Kavya T S[1], Binu R[2]

Student, Department of Computer Science and Engineering, GEC Palakkad, Palakkad, India[1]

Assistant Professor, Department of Computer Science and Engineering, GEC Palakkad, Palakkad, India[2]

Abstract: Metaphoric expressions are regular in ordinary language. Metaphor identification is important in natural language processing since it comes in several common tasks. Conventional methodologies, like phrase-level metaphor identification, identify metaphors with word pairs, where an objective word whose metaphoricity is to be distinguished is given ahead of time. However, such objective words are not featured in genuine content information; a more up-to-date approach is sequential metaphor identification. Also, most of the conventional methodologies use restricted linguistic context to identify metaphors like by considering a single verbs argument or the sentence containing a phrase. Since context has an inevitable role in identifying metaphors, the wider context is critical in metaphor identification tasks. In this work, a novel neural sequential metaphor identification system, constrained to semantically correct input and considers a wider context area, has been proposed. The system is tested on two widely used metaphor datasets: VUA and MOH-X and outperforms the previous approaches.

Keywords: BERT, BiLSTM, Context- dependent, Metaphor.

I. INTRODUCTION

A metaphor is a figure of speech that describes an object or an action in a way that is not literally true but helps to explain an idea or make a comparison. Metaphor provides more clarity, distinction, and artistry to an expression. Also, it is likewise a significant linguistic tool that has become a part of every-day language since a long time ago. Commonly metaphor arises when one idea or domain is viewed in terms of characteristics of another. Recognizing them requires challenging contextual reasoning about whether specific circumstances can really occur. Thus, metaphor processing systems have seen a huge enthusiasm for the Natural Language Processing (NLP) people group. Natural language consists of metaphorical and literal dimensions. Separating metaphorical and non-metaphorical language-use might be profoundly significant for a variety of applications that depend on natural language understanding. Since there are vast texts, manually identifying each metaphorical usage is not at all practical there raises the need for automatic metaphor identification systems. Currently, many efforts have been taken for automatic metaphor identification. Most of them are phrase-level or word pair-based metaphor identification models that need a target word to be given in advance to check whether it is a metaphor or not. But in real-time texts, the target words are not highlighted. Sequential metaphor identification is a newer approach which detects metaphoricity of all the words in a text and thus no target word is to be given. Most of the past methodologies concentrated on constrained types of semantic context, for instance by just giving Subject-Verb-Object triples such as (car, drink, gasoline) to the model. While the verbal arguments give solid signs, giving the full sentential context supports progressively exact predictions. Indeed, even in the couple of situations when the full sentence is utilized existing models have utilized unigram-based features with constrained expressivity. Widening the context area helps to correctly interpret metaphors.

The experts started **examining** the Soviet Union with a microscope to study perceived changes.

He **defended** his position through his publications.

Her speech **attacked** his viewpoint.

My car **drinks** gasoline.

Fig. 1 Examples of sentences containing metaphor

Some examples of metaphoric sentences are shown in Fig. 2. Metaphors are indicated with bold letters in the table. In the first example, 'examining' is metaphor in light of the fact that it is impossible to literally use a 'microscope' to examine a whole nation. Similarly, 'defended', 'attacked', and 'drinks' are also metaphors w.r.t their contexts. Metaphor Identification Procedure (MIP) is a technique for distinguishing metaphorically used words in discourse. A metaphor is

distinguished if the literal meaning of a word differs from the meaning that word takes in this context. For this, context dependent and context independent word embeddings are used. For example, in 'he is drowning in debt', the contextual meaning of 'drown' is 'to have or experience too much', which contrasts with its literal meaning of 'to suffocate by submersion especially in water'. MIP analyses the relations among metaphors and their contexts to determine whether it is a metaphor. Information extraction, Machine translation, Dialog systems, Sentiment analysis, Report generation, Text analytics, Question Answering, Political discourse analysis are few applications of metaphor identification. Here a BiLSTM based approach is proposed which makes use of static (GloVe) and dynamic (BERT) embedding to capture contextual information. Next section discusses different research works done in the area of metaphor detection.

II. RELATED WORK

The method proposed by Ekaterina Shutova et al. [1] captures metaphoric knowledge by using verb and noun clustering. It was the first method to employ unsupervised methods for metaphor identification. It points out lexico-syntactic similarity in the textual environment of metaphorical phrases derived from the same source concept. Clustering concepts using grammatical relations and lexical features allows capturing their relatedness by association and harvests a large number of metaphorical expressions. The acquired clusters then represent the possible source and target concepts between which metaphorical associations hold. The information on such associations is then used to annotate metaphoricity in a large corpus. Peter D. Turney et al. [2] introduced an algorithm based on the hypothesis that metaphorical word usage is correlated with the degree of abstractness of the word's context. It classifies a word sense in a given context as either literal or metaphorical. The introduced abstractness rating algorithm was used to generate feature vectors from a words context and training data was used to learn a logistic regression model that relates degrees of abstractness to the classes literal and metaphorical. The main advantage of this approach is that it readily generalizes to new words.

The approach proposed by Hyeju Jang et al. [3] classifies a target word by examining sentence-level topic transitions. They incorporated several indicators of sentence- level topic transitions as features, such as topic similarity between a sentence and its neighbouring sentences, measured by Sentence Latent Dirichlet Allocation (LDA). According to their perspective, metaphors are often used to express emotional experiences and occur more frequently around personal topics. Thus it gives importance to emotion and cognition words in sentences and their contexts. They tested the system on a breast cancer discussion forum dataset that features metaphors occurring in conversational text. Sunny Rai et al. [4] proposed classifies continuous text into Metaphor or otherwise using Conditional Random Fields (CRF) and a hybrid feature set. Unlike previous approaches, this system classifies an entire sentence not a target word. The feature set is a combination of syntactic, conceptual, affective and word embeddings based features. It uses MRC Psycholinguistic Database (MRCPD) and WordNet-Affect for extracting features.

The method proposed by Omnia Zayed et al. [5] was based on distributional semantics and identifies metaphors on the phrase-level. It is a semi-supervised approach that makes use of distributed representations of word meaning to capture metaphoricity. It extracts all verb-noun pairs from the data and then finds the semantic similarity between the candidate and a predefined seed set of metaphors using pre-trained word embeddings. Based on a threshold similarity value, it classifies the given candidate. The work proposed by Sunny Rai et al. [6] emphasis on local context combined with SVO relations. They use a graphical representation derived from a dependency parser to capture the syntactic structure of an utterance. The context for a word was extracted on the basis of its dependencies with surrounding content words. They assign weights to edges between a word from context and the root verb. The edge features basically includes the semantic relatedness between the connected vertices and their likelihood of co-occurrence. The edge features provide comparable performance as traditional features as per the results obtained. The system performance was evaluated on TroFi dataset.

In [7], proposed by Chuhan Wu et al., metaphor detection was done by a CNN-LSTM model. It was a sequential model thus it determines whether each word is a metaphor or literal. Through LSTM [8] and CNN [9] layers it utilizes both long-range and local information. A weighted softmax classifier was used to predict the label sequence of sentence. Word cluster features are also incorporated and are obtained by clustering the word embedding vectors via k-means [10] method. Their one-hot encoded values are combined with the word embeddings as the final word representations to input the neural network. CNN was used to extract the local contextual information and LSTM was used to extract the long-range information from the CNN outputs.

In the [11] introduced by Jesse Mu et al., they explored the benefits of using wider context area in metaphor detection. This work uses gradient boosting classifiers on representations of an utterance and its surrounding context. The contextual information is learned through a variety of document embedding methods. They examined the role of discourse level features in metaphor detection. They use concatenated pre-trained vector representations for lemmatized verb, its arguments and surrounding context as features to the classifier. The model was trained and tested on VU Amsterdam Metaphor Corpus.

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 9, Issue 6, June 2020

III. DATASET

Two publicly available metaphor datasets are used to train the model: VUA and MOH-X.

A. VUA Dataset

VU Amsterdam Metaphor Corpus (VUA) [12] is the largest publicly available hand-annotated metaphor dataset. It contains data distributed across four genres of the British National Corpus: Academic, Conversation, News, and Fiction. All words are annotated as metaphorical or literal in these texts based on MIP. It contains 10561 data samples and out of which 28\% are metaphorical samples.

B. *MOH-X Dataset*

The MOH-X dataset is created by Mohammad et. al. [13]. It is a hand-annotated dataset build by using WordNet [14]. In this corpus, only a single target verb in each sentence is annotated. During pre-processing task, for each sentence, label sequence for all words is created. This dataset contains 641 data samples and out of which 49\% metaphorical samples.

IV. METHODOLOGY

The degree of dissimilarity between the vector representations of the source and target domains can be used to determine the dissimilarity between the mapped domains and thereby, detecting metaphorical usages [7]. For that static and dynamic embeddings are used. The issue of metaphor detection is posed either as a classification problem [11] that is, label a word as either metaphor or literal, or as a sequence labelling task [4] that is, assigning a sequence of labels to every word in a given sentence. Here, it is posed as a sequence labelling task. The basic architecture of the proposed system is shown in Fig. 2.

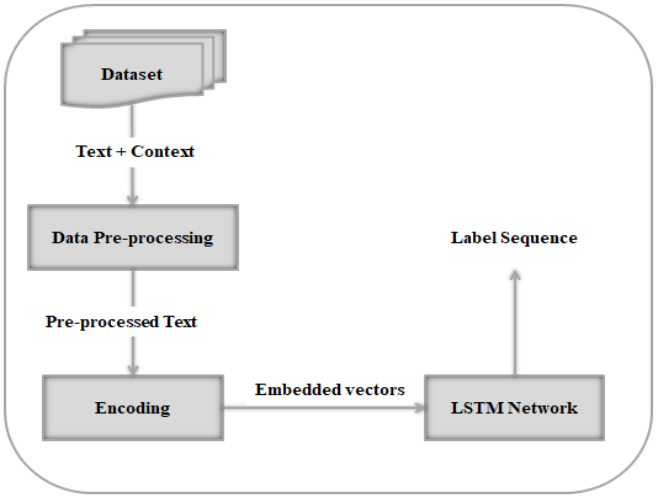

Fig. 2 Basic Architecture

A. Pre-processing

As a first step of pre-processing, all unwanted features are to be removed. Also the unnecessary characters in the dataset are to be removed. After that, the process comes up is tokenization. Since the model uses Bert embedding, Bert tokenization is to be used for tokenizing the data. Once this step is done, the next is converting all the characters present in the data to lowercase. Final step is generating POS tag sequence and label sequence. The model needs the POS tags and labels of words in the input text. The dataset contains only a single word's POS tag and label. Thus, the POS tags and labels for the remaining words are to be generated by using POS tagger, and string and array handling functions.

B. Encoding

Word embeddings are used to get the encoding of the input texts. A metaphor can be distinguished if the literal meaning of the word differs from the meaning that word takes in this context. Capturing the extent of dissimilarity between the mapped domains is really helpful to determine metaphoricity of an expression. For that, the extent of dissimilarity between the vector representation of the source and target domains can be used. Static (context-independent) and dynamic (context-dependent) embeddings are used for this purpose. For the same input, both the static and dynamic word embeddings are calculated and the combination of these vectors is fed to the next module. Glove embeddings are used to get the static vectors and BERT embeddings are used to get the dynamic vectors.

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 9, Issue 6, June 2020

As the literal meaning representation Pretrained GloVe is used since words have been embedded with their most common senses (trained on Wikipedia 2014 and Gigaword 5). GloVe, stands for Global Vectors, is a model for Word Representation. It is an unsupervised learning algorithm for getting vector representations for words. This is accomplished by mapping words into a meaningful space where the distance between words is based on semantic similarity. GloVe vectors of dimension 300 are obtained by training on the combination of Gigaword 5 and Wikipedia 2014 corpus, which contains 6 billion tokens, is used. The pretrained vector *glove.6B.300d.txt* is used as context-independent vectors.

BERT stands for Bidirectional Encoder Representations from Transformers [15]. Bert uses Transformer which is an attention mechanism that learns contextual relations between words (or sub-words) in a text. Normal directional models read the text input sequentially, like left-to-right or right-to-left. But the Transformer encoder reads the entire text (sequence of words) at once. This feature allows the model to learn the context of a word based on all of its surroundings (left and right of the word). BERT embedding is published by Google and they use this in their search engines. It is a new method to obtain pre-trained language model word representation. BERT offers an advantage over models like GloVe because while each word has a fixed representation under GloVe regardless of the context within which the word appears, BERT produces word representations of dimension 768 that are dynamically informed by the words around them. Besides catching clear contrasts, the context-informed word embeddings capture other forms of information that result in more accurate feature representations, which thus brings about better model performance.

The vectors, obtained through GloVe embedding and Bert embedding, are then concatenated using python *numpy* function. The so obtained combination vector is then fed into the model. Before feeding the vector to the model, it should be padded in order to make all sequences in a batch fit a common length.

C. Model Generation

Deep learning is a branch of machine learning based on artificial neural networks with representation learning. Recurrent Neural Network is a very popular Deep Learning model which uses recursion technique. Here, a supervised BiLSTM based RNN model is used to implement automatic metaphor detection system. As per the problem definition, the aim is to build a model that can generate a sequence of labels for each word in the sentence which tells whether it is a metaphor or literal. LSTM networks are usually preferred when the system needs to produce a sequential output. Here, a Bidirectional LSTM (BiLSTM) is used since it can process data in both directional and considers both past and future information. This feature is really helpful to learn the contextual information.

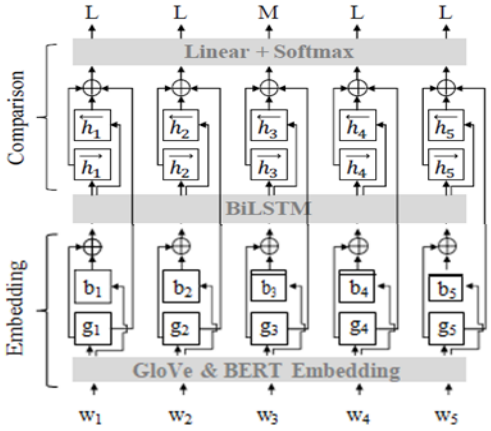

Fig. 3 Model Architecture

The proposed model is a supervised model since both the VUA and MOH-X datasets used to train the model are labelled. The model is based on concept of MIP. That is, a metaphor is classified by the difference between a word's contextual and literal meanings. The contextual meaning representation is concatenated with the literal meaning representation to facilitate the model to make the comparison. Fig. 2 shows the proposed model architecture.

The BiLSTM hidden states are used as the contextual meaning representations, where the hidden state of a word is encoded by its forward and backward contexts and itself. Pretrained GloVe is considered as the literal meaning representation. The comparison of literal and contextual can be seen at the top of Figure 2. It is the comparison stage where the GloVe embedding (literal) from below joins the hidden state from the BiLSTM (contextual). As an input feature of BiLSTM, the combination of Bert and GloVe embedding is used. It can be obtained by concatenating Bert

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 9, Issue 6, June 2020

vector (b) with Glove vector (g). For example, $[g_t ; b_t]$ is the combined vector of word at position t. Thus, the BiLSTM hidden state h_t can be given by $h_t = f_{BiLSTM}([g_t ; b_t] , \overrightarrow{h_{t-1}} , \overleftarrow{h_{t+1}})$

The proposed model has two layers: a BiLSTM layer and a liner layer. A dropout layer is also there in between them. The BiLSTM layer has 300 neurons, and the linear layer has two neurons as there are two categories (metaphor and literal).

V. RESULT AND DISCUSSION

For testing how well the system generates the metaphorical label sequence and how much precise is this sequence with the context of the input, model evaluation metrics are used. Even though it is a sequence generating model, each element in the sequence is used to classify words as metaphor or literal. The task can be considered as a binary classification task with two classes: Metaphor (1) and Literal (0). Thus, the model can be evaluated with the same evaluation measures used for any classification task. Some of the major evaluation metrics used for classification are discussed in the next paragraph. In order to obtain a good model, several experiments have been tried. From pre-processing phase to model generation phase, several alternatives were considered. In deep learning, tuning model parameters and hyper parameters has a great role in model performance. The choice of these parameters, activation function and optimization algorithm affects the model. The selection of input features to the model is also a main factor which can affect the model performance.

Based on training dataset: The choice of training dataset has a major impact on defining how well the model is. Without an establishment of high-quality training data, even the most performing algorithms can be rendered useless. Here the model performance is evaluated by training on two different datasets. The VUA dataset is the largest metaphor corpus that is publicly available. But the metaphoric content in that is only 28\%. MOH-X is a small dataset when compared to VUA. But it is a balanced dataset since its metaphoric content is 49\%. The performance metrics of the model trained on these datasets are provided in TABLE I.

Table I Evaluation based on datasets

Evaluation Metric	Dataset	
	MOH-X	VUA
Precision	87.50	86.56
Recall	84.00	89.98
F1-Score	85.71	88.24
Accuracy	89.06	92.44

Based on input features: The choice of input features is another factor that has high influence on model performance. Choosing the right set of input features is critical for a model. Here, the model is trained with three different feature sets so as to compare the performance. Most of the existing models were used GloVe embeddings as the input feature. The proposed approach uses a combination of GloVe and Bert embeddings as the input feature. GloVe embeddings can represent the lexical features and Bert embeddings can represent the contextual features. In order to make a comparison, the model is also trained with Bert embedding as the input feature. As the TABLE II shows, the model with GloVe and Bert has higher scores when compared with other two models.

Table III Evaluation based on features

Evaluation Metric	Feature		
	GloVe	Bert	GloVe + Bert
Precision	69.23	76.92	87.50
Recall	72.00	80.00	84.00
F1-Score	70.59	78.43	85.71
Accuracy	76.56	82.81	89.06

VI. CONCLUSION

The proposed system is a neural metaphor detection system which considers contextual aspects. The model combines GloVe and Bert embeddings to capture both lexical and contextual information to represent the input. And also considers representations of wider discourse area around the input sentence and is constrained to semantically correct input texts. The model is trained on both VUA and MOH-X metaphor datasets and obtained accuracy of 92.44\% and 89.06\%, respectively. The performance gains of the model show that incorporating broader discourse information is a powerful

feature for metaphor identification systems, aligning with the qualitative analysis and the theoretical evidence suggesting metaphor comprehension is heavily influenced by wider context. Also, the contextualized word embedding essentially improves metaphor detection. The sequence labeling approach enhances performance when used to mutually predict the metaphoricity of all words in a sentence. In future, different semantic similarity measures can be combined for the task of metaphor identification. Also discourse can be used in more sophisticated ways, for example by modeling discourse relations or dialog state tracking.

REFERENCES

[1]. E. Shutova, L. Sun, and A. Korhonen, "Metaphor identification using verb and noun clustering," in Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010), pp. 1002–1010, 2010.
[2]. P. Turney, Y. Neuman, D. Assaf, and Y. Cohen, "Literal and metaphorical sense identification through concrete and abstract context in Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, pp. 680–690, 2011.
[3]. H. Jang, Y. Jo, Q. Shen, M. Miller, S. Moon, and C. Rose, "Metaphor detection with topic transition, emotion and cognition in context," in Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 216–225, 2016.
[4]. S. Rai, S. Chakraverty, and D. K. Tayal, "Supervised metaphor detection using conditional random fields," in Proceedings of the Fourth Workshop on Metaphor in NLP, pp. 18–27, 2016.
[5]. O. Zayed, J. P. McCrae, and P. Buitelaar, "Phrase-level metaphor identification using distributed representations of word meaning," in Proceedings of the Workshop on Figurative Language Processing, pp. 81–90, 2018.
[6]. S. Rai, S. Chakraverty, D. K. Tayal, and Y. Kukreti, "A study on impact of context on metaphor detection," The Computer Journal, vol. 61, no. 11, pp. 1667–1682, 2018.
[7]. C. Wu, F. Wu, Y. Chen, S. Wu, Z. Yuan, and Y. Huang, "Neural metaphor detecting with cnn-lstm model," in Proceedings of the Workshop on Figurative Language Processing, pp. 110–114, 2018.
[8]. J. Schmidhuber and S. Hochreiter, "Long short-term memory," Neural Comput, vol. 9, no. 8, pp. 1735–1780, 1997.
[9]. K. O'Shea and R. Nash, "An introduction to convolutional neural networks," arXiv preprint arXiv:1511.08458, 2015.
[10]. J. M. Keller, M. R. Gray, and J. A. Givens, "A fuzzy k-nearest neighbor algorithm," IEEE transactions on systems, man, and cybernetics, no. 4, pp. 580–585, 1985.
[11]. J. Mu, H. Yannakoudakis, and E. Shutova, "Learning outside the box: Discourse-level features improve metaphor identification," arXiv preprint arXiv:1904.02246, 2019.
[12]. G. Steen, A method for linguistic metaphor identification: From MIP to MIPVU, vol. 14. John Benjamins Publishing, 2010.
[13]. S. Mohammad, E. Shutova, and P. Turney, "Metaphor as a medium for emotion: An empirical study," in Proceedings of the Fifth Joint Conference on Lexical and Computational Semantics, pp. 23–33, 2016.
[14]. C. Fellbaum, "Wordnet: An electronic lexical database and some of its applications," 1998.
[15]. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," arXiv preprint arXiv:1810.04805, 2018.