Enhancing Negation Scope Detection using Multitask Learning

Harsh Patel
Indian Institute of Technology Gandhinagar
India
harsh.patel@iitgn.ac.in

Xulang Zhang
Nanyang Technological University
Singapore
XULANG001@e.ntu.edu.sg

Qian Liu
Nanyang Technological University
Singapore
liu.qian@ntu.edu.sg

***Abstract*—Negation is a linguistic phenomenon that usually occurs in a text for denial or refute of some occasion. Detection of such negative assertions is an essential sub-task in various applications of information extraction and data mining. In this paper, we present a deep multitask learning (MTL) framework to enhance the performance of Negation Scope detection using part-of-speech (POS) tagging as an auxiliary task. We show how the relationship between these two tasks, which do not seem to be easily linked from a linguistic point of view, is mutually beneficial.**

***Index Terms*—sentiment analysis, multitask learning, negation scope detection, pos-tagging, deep learning**

I. Introduction

In recent years, sentiment analysis has become increasingly popular for processing social media data on online communities [1], social networks [2], and microblogging platforms [3]. Sentiment analysis is a branch of affective computing research that aims to mine opinions from text (but sometimes also images [4] and videos [5]). Most of the literature is on English language but recently an increasing number of works are tackling the multilinguality issue [6], especially in booming online languages such as Arabic [7], Chinese [8], and Spanish [9]. Besides traditional domains like business intelligence [10] and recommendation systems [11], sentiment analysis applications also include many other areas like financial forecasting [12], healthcare [13], cyber-harassment prevention [14], political forecasting [15], and dialogue systems [16].

Negation handling is one of the most important sub-tasks of sentiment analysis as negation can flip the polarity of clauses and even entire sentences. The occurrence of negation cues like *not*, *don't* or *-less* etc, often affect the true meaning of the sentence. Although the contextual polarity of a sentence is affected by the presence of a negation cue, it is not necessary that the sentiments conveyed by all the tokens get reversed. This makes it quite essential to determine the scope of these negation cues [17]. While detecting the presence of a negation cue in domain-specific texts could be done with the help of simple keyword matching approaches [18], the problem of identifying the scope of a negation cue is significantly challenging as it may not be limited to a few subsequent words in the sentence. For example, in the sentence, ***John had*** *never* ***said as much before***, the negation cue: *never*, affects both the events, a) saying did not take place, b) John was not the one who said.

Regardless of the approach used, the syntactic structure of the sentence plays a significant role in negation scope detection. This is because the position of negation cue in the syntactical tree along with the projection of its parent are strong indicators of the patterns the scope is likely to span [19]. Similarly, in the task of POS tagging, the syntactical structure is again the key factor as the tokens with same tags display similar syntactic behavior and morphology.

This correlating nature between the two tasks motivates us to leverage the idea of multitask learning i.e., training models that simultaneously learn to solve both tasks by exploiting their commonalities. This can result in improved learning efficiency for the task-specific portions and indeed gain better accuracies.

We empirically show that this approach outperforms the results obtained for negation scope detection with a separate single-task learning model.

II. Related Work

Sentiment analysis techniques can be broadly categorized into symbolic and sub-symbolic approaches: the former include the use of lexicons [20], ontologies [21], and semantic networks [22] to encode the polarity associated with words and multiword expressions; the latter consist of supervised [23], semi-supervised [24] and unsupervised [25] machine learning techniques that perform sentiment classification based on word co-occurrence frequencies. Among these, the most popular are algorithms based on belief networks [26], randomized networks [27], generative adversarial networks [28], and capsule networks [29]. There are also some hybrid frameworks that leverage both symbolic and sub-symbolic approaches [30].

While most works approach it as a simple categorization problem, sentiment analysis is actually a complex research problem that requires tackling many NLP tasks [31] tasks such as microtext normalization [32], to decode informal text, subjectivity detection [33], to filter out neutral content, anaphora resolution [34], to link pronouns with the entities of a sentence, personality recognition [35], for distinguishing between different personality types of the users, and negation handling. Some early works in negation handling systems like [36], use a simple rule-based approach where the tokens between the negation cue and the first punctuation are marked as the scope. A more robust rule-based approach that uses information from parse trees is developed in [37].

2375-9259/21/$31.00 ©2021 IEEE
DOI 10.1109/ICDMW53433.2021.00029

With the advent of new annotated corpora like the Bioscope corpus [38], various advanced computational approaches have proved to achieve improvements in negation scope modelling. Based on the guidelines provided for the scope annotation in the bioscope corpus, [39] hand-crafted a set of heuristic grammatical rules to define the scope of each cue. Considering it as a classification problem, Morante et al. [40] use an ensemble-based machine learning method combining Support Vector Machines (SVMs) and Conditional Random Fields (CRFs). [41] uses an advanced CRF-based model with a set of syntactic and structural features to solve the scope detection task in *SEM shared task 2012 [42].

More recently, neural network-based approaches have also been adopted for negation scope detection. Lazib et al. [43], considering it as a sequence labeling problem, proposed various Recurrent Neural Network (RNN) models to detect the scope of a negation cue automatically. Similarly, Fancellu et al. [19] use Bi-LSTM (Bidirectional Long Short-Term Memory) networks to solve the task and achieve outperforming results on the *SEM shared task 2012 [42]. Their approach includes the use of additional embedding vectors of POS tag information for improved modelling of the input sentences. The POS-related information leads to better performances in most of their model variants, helping us to assess the correlation between the two tasks. Qian et al. [44] proposed a Convolutional Neural network (CNN) with weighted average pooling to extract meaningful features from the syntactic paths between the cues and the potential tokens to address the problem of negation scope detection.

Various tasks in the NLP field are highly interrelated, making it natural to apply multitask learning. For example, [45] uses a cascading architecture to share the learnt contextual features of a POS tagging classifier to train the Semantic Role Labelling (SRL) model. Alternatively, [46] use a joint training approach and show improvements in results for POS tagging and noun-phrase chunking tasks. More such advanced joint training approaches have been used for other pre-processing tasks by [47]–[49]. [50] use a multitask, neural tensor fusion-based learning approach to model the synergy between sentiment and sarcasm classification tasks. Motivated by the success of neural networks models in solving numerous Natural Language Processing (NLP) tasks, especially Transformers [51], we propose in this paper a deep multitask learning framework that combines transformer encoders along with a Bi-LSTM network and CRF layers to solve the negation scope detection task with POS tagging as an auxiliary task.

III. Methodology

The syntactic nature of sentences plays a significant role while solving both negation scope detection and POS tagging tasks [19]. To leverage the commonalities, we use a multitask learning approach, where a single deep learning model is used to perform more than one task (in our case, negation scope detection and POS tagging). We specifically focus on improving the accuracy of negation scope detection by using POS tagging as an auxiliary task. Fig 1 shows our multitask learning architecture. The following sections describe the key components of our architecture:

A. Input Representation

We use trainable word embedding vectors (dimensions, $D_e = 256$) to represent each token in the input sequence. These variable-length input sequences are initially padded to a fixed length ($L = 85$, length of the longest sentence in the dataset) using null vectors. The padding does not contribute to the gradient during the training step in any component of the architecture.

B. Shared LSTM

The embedded input sequences are further passed through an LSTM network, i.e, a recurrent neural network capable of capturing long-term dependencies in the sequential inputs. We use Bi-LSTM network ($D_{lstm} = 128$) that constructs context-rich sentence-level token representations ($H_* = [h_{*1}, h_{*2}..., h_{*L}]$, with $h_{*i} \in R^{2D_{lstm}}$), taking into account the input sequence bi-directionally (* representing both backward and forward). The computation of this hidden layer at time t and the word-embedding vector x can be represented as following:

$$i_t = sigmoid(W_x^i x + W_h^i h_{t-1} + b^i) \tag{1}$$

$$f_t = sigmoid(W_x^f x + W_h^f h_{t-1} + b^f) \tag{2}$$

$$o_t = sigmoid(W_x^o x + W_h^o h_{t-1} + b^o) \tag{3}$$

$$c_t = c_{t-1} \cdot f_t + i_t \cdot tanh(W_x^c x + W_h^c h_{t-1} + b^c) \tag{4}$$

$$h_{back,forw} = o_t \cdot tanh(c_t) \tag{5}$$

where **W**s are the weight matrix and **b**s are the bias parameters, h_{t-1} is the hidden layer state at $time = t-1$, i_t, f_t, o_t are the input, forget, and output gates at $time = t$ of the network. c_t represents the cell activation vectors. The output sequence ($y_{lstm} = [y_1, y_2..., y_L]$, with $y_i \in R^{D_t}$), of size $D_t = 256$, is further calculated using the weight ($W_{lstm} \in R^{D_t \times 2*D_{lstm}}$) and bias parameters ($b_{lstm} \in R^{D_t}$). In our approach, the *Shared LSTM* is expected to learn the correlated features between the two tasks, i.e, negation scope detection and POS tagging.

$$y_{lstm} = W_{lstm} Concat(H_{back}, H_{forw}) + b_{lstm} \tag{6}$$

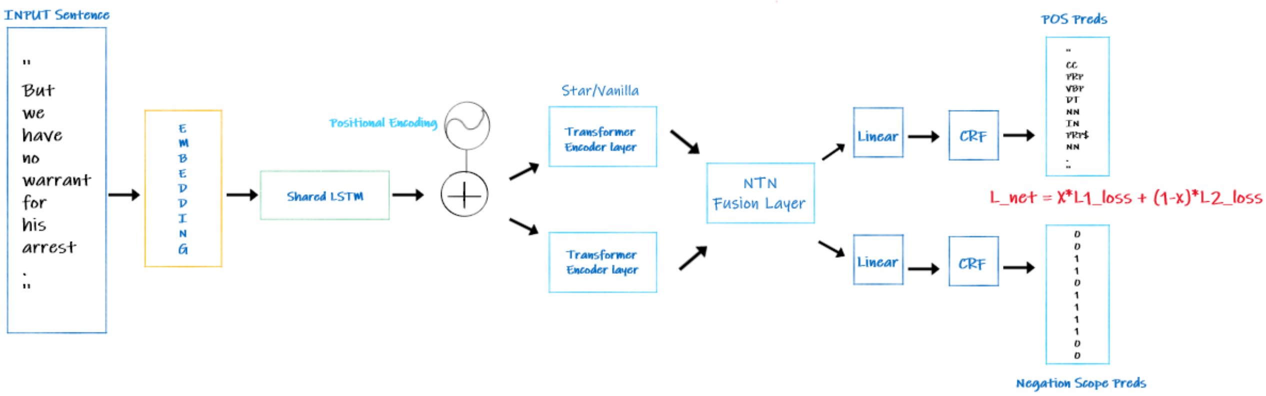

Fig. 1. Our multitask learning architecture along with input-output example.

C. Transformer Encoder

The context-rich sentence-level token representations are further fed to individual transformer encoder layers to map them to abstract representations that hold task-specific learned information for each token in the input sequence. Transformer encoder is a stack of multiple encoder blocks ($n_{layers} = 4$) that map the input sequence to a concrete and contextualized encoding sequence. Multi-headed attention networks ($n_h = 10$, equation 7) in the transformer weigh the relevance of each token in the sequence and prioritize them to produce more meaningful outputs. We use *positional-encodings* in order to inject the positional information in the sequence. Mathematical formulation:

$$Attention(Q, K, V) = softmax(\frac{QKT}{\sqrt{d_k}})V \tag{7}$$

$$MultiHeadAttn(Q, K, V) = Concat(A_1, ..., A_{n_h})W^O,$$
$$where A_i = Attention(QW_i^Q, KW_i^K, VW_i^V)$$

Each input vector in y_{lstm} is used with three different ways in the attention mechanism: the Query ($Q \in R^{d_k}$), the Key ($K \in R^{d_k}$) and the Value ($V \in R^{d_v}$). These vectors are further projected n_h times to allow the model to jointly use information from different representation by concatenating the results. The parameters include $W_i^Q \in R^{D_t \times d_k}, W_i^K \in R^{D_t \times d_k}, W_i^V \in R^{D_t \times d_v}$ and $W^O \in R^{n_h d_v \times D_t}$. The output of the attention network (X) is further passed through a fully connected feed-forward network (FFN) with two position-wise linear transformations and a ReLu activation.

$$FFN(X) = ReLu(XW_{f1} + b_{f1})W_{f2} + b_{f2}$$
$$where, W_{f1} \in R^{D_t \times d_f}, W_{f2} \in R^{\times D_t},$$
$$b_{f1} \in R^{d_f}, and b_{f2} \in R^{D_t}.$$

We use the above encoder setup for both negation scope detection ($H_{neg} = FFN_1(X)$) and POS tagging ($H_{pos} = FFN_2(X)$). Along with the standard transformers, we also experiment with a lightweight alternative, Star Transformer [52], for reducing the architecture complexity.

D. Neural Tensor Network Layer

We use a Neural Tensor Network (NTN) layer ($D_{ntn} = 256$) to fuse and model the relationships between the encoded representations of the tasks in our MTL framework. Neural Tensor layers are extended versions of the standard linear layer with a bi-linear tensor layer, relating the encoded vectors across multiple dimensions to extract information relevant to both the vectors.

$$y_{ntn} = tanh(H_{neg}W_n^{[1:D_{ntn}]}H_{pos}^T + Concat(H_{neg}, H_{pos})W_v + b_n) \tag{8}$$

where, $W_n \in R^{D_{ntn} \times D_t \times D_t}, W_v \in R^{2D_t \times D_{ntn}}$, $b_n, y_{ntn} \in R^{D_{ntn}}$. We use individual Linear layers (simple feed-forward networks with $D_l = 256$) to further map the fused representations to the desired number of features for both the individual tasks.

$$s_{neg} = W_1 y_{ntn} + b_1 \tag{9}$$

$$s_{pos} = W_2 y_{ntn} + b_2 \tag{10}$$

where, $W_1, W_2 \in R^{D_l \times D_{ntn}}$ and $b_1, b_2 \in R^{D_l}$ are trainable parameters.

E. Conditional Random Field (CRF)

Since both negation scope detection and POS tagging are sequence labelling tasks, we use task-specific CRF layers to model the dependency between each token and the entire sentence representation. Conditional Random Field is a probabilistic framework for labeling structured, sequential data. Unlike the Softmax layer, often used as an activation in the output of the neural network models for classification, CRFs do not assume an independent relationship between the token's outcome probability.

Thus, defining a conditional probability distribution over the label sequences provides a significant improvement in the sequence labelling task [53]. In our framework, two individual CRF layers, in the end, classify each token in the input sequence with the most relevant POS tag as well as indicate whether the token lies in the scope of the negation cue. The probability of final output y_* is calculated using:

$$P(y_*|s_*) = \frac{\sum_{t=1}^{L} e^{f(y_{*t-1}, y_{*t}, s_*)}}{\sum_{y'}^{Y(s_*)} \sum_{t=1}^{L} e^{f(y'_{t-1}, y'_t, s_*)}} \tag{11}$$

where * is for neg and pos. Here, $f(y_{*t-1}, y_{*t}, s_*)$ computes the transition score for y_t. Viterbi algorithm is used to optimize finding the maximum $P(y_*|s_*)$ probable sequence [54].

F. Training

We use a combination of the negative log-likehoods obtained from both the task-specific CRF layers as the loss function for training. We use Adam [55], a method for stochastic optimization of parameters along with an adaptive learning rate scheduler to train the architecture.

IV. Experiments

1) Datasets: We use the *SEM Shared Task 2012 dataset (Conan Doyle stories) [42] which consists of 885 training and 264 testing samples, each annotated with negation cues and their scope, as well as the POS tag information. The cues are the words that express negation, and the scope is the part of a sentence that is affected by the presence of such negation cues.

One variant of our model includes pre-training some layers of the model on a specific POS tagging dataset. We use the English Web Treebank [56], a corpus of $16,662$ sentences annotated using the Universal Dependencies annotation for this purpose.

2) Model Variants: We evaluate the following variants of our model.

A. Stardard Single-Task Learning

Here, we train our model by removing the shared layers between both the task, i.e., *Shared LSTM* and *NTN Fusion layer* to perform individual single-task learning over both the tasks.

B. Simple Multitask Learning with Shared LSTM

In this setting, both the tasks share the *Shared LSTM* layer (Section III-B) while minimizing the loss function.

C. Multitask Learning with NTN Fusion

Here, we train our model on the complete MTL architecture that we propose in Section III including both the important shared layers, *Shared LSTM* (Section III-B) and *NTN Fusion layer* (Section III-D).

D. Multitask Learning with separate pre-training of POS-related layers

In this variant, we first pre-train the POS tagging task-specific layers separately on the universal dataset (Section IV-1, $N_{epochs} = 60$, $lr = 1e-2$). We then train our complete MTL architecture by initializing the POS-specific layers with the pre-trained weights.

For all the experiments, we perform hyperparameter tuning to achieve the best loss convergence results.

V. Results and Discussion

Table I shows the results of negation scope detection on the Conan-Doyle dataset using different approaches used in the previous works. [57] uses a rule-based approach whereas, [58] uses a CRF-based sequence labelling approach to determine the scope of negation. However, to enhance the performance, more advanced computational approaches have been used by the community. [59] uses a support vector machine model along with a few heuristics to get improved results. [19] instead uses a recurrent neural network model to achieve outperforming results on the negation scope detection task.

Table II further shows the results on different variants of our MTL architecture. We perform an ablation study to understand and compare the essential components of our multitask learning framework. With the Standard Single-Task Leaning variant as a baseline, we see that the Simple MTL with Shared LSTM model performed slightly better in the negation scope detection task. This shows that the Shared LSTM plays a reasonable role in understanding the commonalities between the two tasks, which sets a good foundation for joint learning.

The MTL with NTN fusion model outperforms both the standalone and shared LSTM models with a significant margin of improvement in the F1-Score for negation scope detection. This means that the additional NTN fusion network serves as a beneficial component for inheriting the inter-task resemblance into the model. All our multitask learning variants outperform the standard single-task learning model in our primal task of negation scope detection, exhibiting the potential of MTL in natural language processing tasks.

Comparing MTL with NTN Fusion and MTL with pre-trained POS layers, we can see that although the addition of POS information is beneficial for negation scope detection, there's a trade-off between the accuracies of the two tasks.

Even though the performance of our model is not quite on par with the state-of-the-art model [19], which utilizes embedded universal POS information, the proposed multitask learning framework has a lot of potentials. As we can infer from the results, POS tagging is related but not strongly tied to negation scope detection. More relevant syntactic processing tasks can be incorporated into the joint learning model to improve the performance, such as lemmatization and parsing.

VI. Conclusion

In this work, we leverage the idea of deep multitask learning to assess the potential of standard syntactic features in learning to solve natural language pre-processing tasks. We use shared

TABLE I
RESULTS OF DIFFERENT APPROACHES OF NEGATION SCOPE DETECTION ON THE CONAN-DOYLE DATASET

Approach	**Negation Scope Detection**		
	Presicion	**Recall**	**F1-Score**
Some Previous Models			
Rule-based (de Albornoz et al., 2012)	85.37	68.53	76.03
CRF-based (Lapponi et al., 2012)	82.25	82.16	82.20
SVM+heuristics (Read et al., 2012)	81.99	88.81	85.26
Bi-LSTM based (Fancellu et al., 2016)	92.62	85.13	**88.72**
MTL with NTN Fusion (Our model)	82.73	80.55	81.63

TABLE II
RESULTS FOR OUR DIFFERENT MODEL VARIANTS ON THE CONAN-DOYLE DATASET

Model Variant	**POS Tagging**	**Negation Scope Detection**		
	Accuracy	**Presicion**	**Recall**	**F1-Score**
Stardard Single-Task Learning	85.87%	79.44%	73.87%	76.56%
Simple MTL with Shared LSTM	82.54%	87.23%	69.57%	77.41%
MTL with NTN Fusion	87.09%	82.73%	80.55%	**81.63%**
MTL with separate pre-training of POS layer	**93.33%**	82.50%	78.25%	80.63%

neural network layers to exploit the linguistic relations between negation scope detection and POS tagging using various multitask learning architectures. Furthermore, we show that multitask learning-based methods for negation scope detection, using POS tagging as an auxiliary task, though not on par with the state-of-the-art, reasonably outperform the standard single-task learning model. We also concluded that while POS tagging is beneficial for negation scope detection, a trade-off can happen when trying to push the performance of both tasks.

For future work, it's worth further experimenting on the proposed MTL framework with other tasks related to negation scope detection, such as morphological tagging, lemmatization, and parsing.

ACKNOWLEDGMENT

We are grateful to Prof. Erik Cambria, Nanyang Technological University (NTU), Singapore, for his constant support and guidance. We would like to thank the School of Computer Science and Engineering (SCSE), NTU Singapore, for providing computational resources to perform various experiments.

REFERENCES

[1] S. Cavallari, E. Cambria, H. Cai, K. Chang, and V. Zheng, "Embedding both finite and infinite communities on graph," *IEEE Computational Intelligence Magazine*, vol. 14, no. 3, pp. 39–50, 2019.

[2] D. Camacho, A. Panizo-LLedot, G. Bello-Orgaz, A. Gonzalez-Pardo, and E. Cambria, "The four dimensions of social network analysis: An overview of research methods, applications, and software tools," *Information Fusion*, vol. 63, pp. 88–120, 2020.

[3] E. Cambria, H. Wang, and B. White, "Guest editorial: Big social data analysis," *Knowledge-Based Systems*, vol. 69, pp. 1–2, 2014.

[4] E. Ragusa, E. Cambria, R. Zunino, and P. Gastaldo, "A survey on deep learning in image polarity detection: Balancing generalization performances and computational costs," *Electronics 8 (7), 783*, vol. 8, no. 7, p. 783, 2019.

[5] L. Stappen, A. Baird, E. Cambria, and B. Schuller, "Sentiment analysis and topic recognition in video transcriptions," *IEEE Intelligent Systems*, vol. 36, no. 2, pp. 88–95, 2021.

[6] S. L. Lo, E. Cambria, R. Chiong, and D. Cornforth, "Multilingual sentiment analysis: From formal to informal and scarce resource languages," *Artificial Intelligence Review*, vol. 48, no. 4, pp. 499–527, 2017.

[7] O. Oueslati, E. Cambria, M. B. HajHmida, and H. Ounelli, "A review of sentiment analysis research in arabic language," *Future Generation Computer Systems*, vol. 112, pp. 408–430, 2020.

[8] H. Peng, E. Cambria, and A. Hussain, "A review of sentiment analysis research in chinese language," *Cognitive Computation*, vol. 9, no. 4, pp. 423–435, 2017.

[9] I. Chaturvedi, E. Cambria, and D. Vilares, "Lyapunov filtering of objectivity for Spanish sentiment model," in *IJCNN*, 2016, pp. 4474–4481.

[10] Y. Li, S. Wang, Y. Ma, Q. Pan, and E. Cambria, "Popularity prediction on vacation rental websites," *Neurocomputing*, vol. 412, pp. 372–380, 2020.

[11] Y. Li, Q. Pan, T. Yang, S. Wang, J. Tang, and E. Cambria, "Learning word representations for sentiment analysis," *Cognitive Computation*, vol. 9, no. 6, pp. 843–851, 2017.

[12] A. Picasso, S. Merello, Y. Ma, L. Oneto, and E. Cambria, "Technical analysis and sentiment embeddings for market trend prediction," *Expert Systems with Applications*, vol. 135, pp. 60–70, 2019.

[13] E. Cambria, A. Hussain, T. Durrani, C. Havasi, C. Eckl, and J. Munro, "Sentic computing for patient centered applications," in *IEEE ICSP*, 2010, pp. 1279–1282.

[14] J. A. Kumar, S. Abirami, T. E. Trueman, and E. Cambria, "Comment toxicity detection via a multichannel convolutional bidirectional gated recurrent unit," *Neurocomputing*, vol. 441, pp. 272–278, 2021.

[15] A. Khatua, A. Khatua, and E. Cambria, "Predicting political sentiments of voters from twitter in multi-party contexts," *Applied Soft Computing*, vol. 97, no. 106743, 2020.

[16] Y. Ma, K. L. Nguyen, F. Xing, and E. Cambria, "A survey on empathetic dialogue systems," *Information Fusion*, vol. 64, pp. 50–70, 2020.

[17] M. Dadvar, C. Hauff, and F. de Jong, "Scope of negation detection in sentiment analysis," in *Proceedings of the Dutch-Belgian Information Retrieval Workshop, DIR 2011*. University of Amsterdam, Feb. 2011, pp. 16–20, null ; Conference date: 04-02-2011 Through 04-02-2011.

[18] W. W. Chapman, W. Bridewell, P. Hanbury, G. F. Cooper, and B. G. Buchanan, "Evaluation of negation phrases in narrative clinical reports." *Proceedings of the AMIA Symposium*, pp. 105–109, 2001.

[19] F. Fancellu, A. Lopez, and B. Webber, "Neural networks for negation scope detection," in *Proceedings of the 54th Annual Meeting of the*

Association for Computational Linguistics (Volume 1: Long Papers). Berlin, Germany: Association for Computational Linguistics, Aug. 2016, pp. 495–504. [Online]. Available: https://aclanthology.org/P16-1047
[20] F. Xing, F. Pallucchini, and E. Cambria, "Cognitive-inspired domain adaptation of sentiment lexicons," *Information Processing and Management*, vol. 56, no. 3, pp. 554–564, 2019.
[21] I. Donadello, "OntoSenticNet 2: Enhancing reasoning within sentiment analysis," *IEEE Intelligent Systems*, vol. 36, no. 5, 2021.
[22] E. Cambria, Y. Li, F. Xing, S. Poria, and K. Kwok, "SenticNet 6: Ensemble application of symbolic and subsymbolic AI for sentiment analysis," in *CIKM*, 2020, pp. 105–114.
[23] I. Chaturvedi, E. Ragusa, P. Gastaldo, R. Zunino, and E. Cambria, "Bayesian network based extreme learning machine for subjectivity detection," *Journal of The Franklin Institute*, vol. 355, no. 4, pp. 1780–1797, 2018.
[24] A. Hussain and E. Cambria, "Semi-supervised learning for big social data analysis," *Neurocomputing*, vol. 275, pp. 1662–1673, 2018.
[25] E. Cambria, T. Mazzocco, A. Hussain, and C. Eckl, "Sentic medoids: Organizing affective common sense knowledge in a multi-dimensional vector space," in *Advances in Neural Networks*, ser. Lecture Notes in Computer Science, D. Liu, H. Zhang, M. Polycarpou, C. Alippi, and H. He, Eds., vol. 6677. Springer-Verlag, 2011, pp. 601–610.
[26] I. Chaturvedi, Y.-S. Ong, I. Tsang, R. Welsch, and E. Cambria, "Learning word dependencies in text by means of a deep recurrent belief network," *Knowledge-Based Systems*, vol. 108, pp. 144–154, 2016.
[27] G.-B. Huang, E. Cambria, K.-A. Toh, B. Widrow, and Z. Xu, "New trends of learning in computational intelligence," *IEEE Computational Intelligence Magazine*, vol. 10, no. 2, pp. 16–17, 2015.
[28] Y. Li, Q. Pan, S. Wang, T. Yang, and E. Cambria, "A generative model for category text generation," *Information Sciences*, vol. 450, pp. 301–315, 2018.
[29] W. Zhao, H. Peng, S. Eger, E. Cambria, and M. Yang, "Towards scalable and reliable capsule networks for challenging NLP applications," in *ACL*, 2019, pp. 1549–1559.
[30] Y. Susanto, E. Cambria, B. C. Ng, and A. Hussain, "Ten years of sentic computing," *Cognitive Computation*, vol. 13, 2021.
[31] E. Cambria, S. Poria, F. Bisio, R. Bajpai, and I. Chaturvedi, "The CLSA model: A novel framework for concept-level sentiment analysis," in *LNCS*. Springer, 2015, vol. 9042, pp. 3–22.
[32] R. Satapathy, E. Cambria, A. Nanetti, and A. Hussain, "A review of shorthand systems: From brachygraphy to microtext and beyond," *Cognitive Computation*, vol. 12, no. 4, pp. 778–792, 2020.
[33] I. Chaturvedi, E. Cambria, R. Welsch, and F. Herrera, "Distinguishing between facts and opinions for sentiment analysis: Survey and challenges," *Information Fusion*, vol. 44, pp. 65–77, 2018.
[34] R. Sukthanker, S. Poria, E. Cambria, and R. Thirunavukarasu, "Anaphora and coreference resolution: A review," *Information Fusion*, vol. 59, pp. 139–162, 2020.
[35] Y. Mehta, N. Majumder, A. Gelbukh, and E. Cambria, "Recent trends in deep learning based personality detection," *Artificial Intelligence Review*, vol. 53, pp. 2313–2339, 2020.
[36] B. Pang and L. Lee, "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts," in *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, ser. ACL '04. USA: Association for Computational Linguistics, 2004, pp. 271–es. [Online]. Available: https://doi.org/10.3115/1218955.1218990
[37] L. Jia, C. Yu, and W. Meng, "The effect of negation on sentiment analysis and retrieval effectiveness," in *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, ser. CIKM '09. New York, NY, USA: Association for Computing Machinery, 2009, pp. 1827–1830. [Online]. Available: https://doi.org/10.1145/1645953.1646241
[38] G. Szarvas, V. Vincze, R. Farkas, and J. Csirik, "The bioscope corpus: annotation for negation, uncertainty and their scope in biomedical texts," in *BioNLP*, 2008.
[39] L. Øvrelid, E. Velldal, and S. Oepen, "Syntactic scope resolution in uncertainty analysis," in *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*. Coling 2010 Organizing Committee, Aug 2010, pp. 1379–1387. [Online]. Available: https://aclanthology.org/C10-1155
[40] R. Morante and W. Daelemans, "A metalearning approach to processing the scope of negation," in *CoNLL*, 2009.
[41] J. P. White, "UWashington: Negation resolution using machine learning methods," in **SEM 2012: The First Joint Conference on Lexical and Computational Semantics – Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012)*. Montréal, Canada: Association for Computational Linguistics, 7-8 Jun. 2012, pp. 335–339. [Online]. Available: https://aclanthology.org/S12-1044
[42] R. Morante and E. Blanco, "*sem 2012 shared task: Resolving the scope and focus of negation," in **SEMEVAL*, 2012.
[43] L. Lazib, Y. Zhao, B. Qin, and T. Liu, "Negation scope detection with recurrent neural networks models in review texts," *International Journal of High Performance Computing and Networking*, vol. 13, no. 2, p. 211, 2019.
[44] Z. Qian, P. Li, Q. Zhu, G. Zhou, Z. Luo, and W. Luo, "Speculation and negation scope detection via convolutional neural networks," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Nov 2016, pp. 815–825. [Online]. Available: https://aclanthology.org/D16-1078
[45] S. S. Pradhan, W. H. Ward, K. Hacioglu, J. H. Martin, and D. Jurafsky, "Shallow semantic parsing using support vector machines," in *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics: HLT-NAACL 2004*. Association for Computational Linguistics, May 2004, pp. 233–240. [Online]. Available: https://aclanthology.org/N04-1030
[46] C. Sutton, A. McCallum, and K. Rohanimanesh, "Dynamic conditional random fields: Factorized probabilistic models for labeling and segmenting sequence data," *Journal of Machine Learning Research*, vol. 8, no. 25, pp. 693–723, 2007. [Online]. Available: http://jmlr.org/papers/v8/sutton07a.html
[47] S. Miller, H. Fox, L. Ramshaw, and R. Weischedel, "A novel use of statistical parsing to extract information from text," in *Proceedings of the 1st North American chapter of the Association for Computational Linguistics conference*, ser. NAACL 2000. Association for Computational Linguistics, Apr 2000, pp. 226–233.
[48] C. Sutton and A. McCallum, "Composition of conditional random fields for transfer learning," in *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, ser. HLT '05. Association for Computational Linguistics, Oct 2005, pp. 748–754. [Online]. Available: https://doi.org/10.3115/1220575.1220669
[49] ——, "Joint parsing and semantic role labeling," in *Proceedings of the Ninth Conference on Computational Natural Language Learning (CoNLL-2005)*. Association for Computational Linguistics, Jun 2005, pp. 225–228. [Online]. Available: https://aclanthology.org/W05-0636
[50] N. Majumder, S. Poria, H. Peng, N. Chhaya, E. Cambria, and A. Gelbukh, "Sentiment and sarcasm classification with multitask learning," *IEEE Intelligent Systems*, vol. 34, no. 3, pp. 38–43, 2019.
[51] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, u. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, ser. NIPS'17. Red Hook, NY, USA: Curran Associates Inc., 2017, pp. 6000–6010.
[52] Q. Guo, X. Qiu, P. Liu, Y. Shao, X. Xue, and Z. Zhang, "Star-transformer," in *Proceedings of the 2019 Conference of the North*. Association for Computational Linguistics, 2019, pp. 1315–1325. [Online]. Available: http://aclweb.org/anthology/N19-1133
[53] J. D. Lafferty, A. McCallum, and F. C. N. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *Proceedings of the Eighteenth International Conference on Machine Learning*, ser. ICML '01. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2001, pp. 282–289.
[54] H. Yan, B. Deng, X. Li, and X. Qiu, "TENER: adapting transformer encoder for named entity recognition," *CoRR*, vol. abs/1911.04474, 2019. [Online]. Available: http://arxiv.org/abs/1911.04474
[55] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *CoRR*, vol. abs/1412.6980, 2015.
[56] N. Silveira, T. Dozat, M. De Marneffe, S. Bowman, M. Connor, J. Bauer, and C. Manning, "A gold standard dependency corpus for english," in *Proceedings of the 9th International Conference on Language Resources and Evaluation, LREC 2014*, ser. Proceedings of the 9th International Conference on Language Resources and Evaluation,

LREC 2014, N. Calzolari, K. Choukri, S. Goggi, T. Declerck, J. Mariani, B. Maegaard, A. Moreno, J. Odijk, H. Mazo, S. Piperidis, and H. Loftsson, Eds. European Language Resources Association (ELRA), 2014, pp. 2897–2904, 9th International Conference on Language Resources and Evaluation, LREC 2014 ; Conference date: 26-05-2014 Through 31-05-2014.

[57] J. Carrillo de Albornoz, L. Plaza, A. Díaz, and M. Ballesteros, "UCM-I: A rule-based syntactic approach for resolving the scope of negation," in **SEM 2012: The First Joint Conference on Lexical and Computational Semantics – Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012)*. Montréal, Canada: Association for Computational Linguistics, 7-8 Jun. 2012, pp. 282–287. [Online]. Available: https://aclanthology.org/S12-1037

[58] E. Lapponi, E. Velldal, L. Øvrelid, and J. Read, "UiO 2: Sequence-labeling negation using dependency features," in **SEM 2012: The First Joint Conference on Lexical and Computational Semantics – Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012)*. Montréal, Canada: Association for Computational Linguistics, 7-8 Jun. 2012, pp. 319–327. [Online]. Available: https://aclanthology.org/S12-1042

[59] J. Read, E. Velldal, L. Øvrelid, and S. Oepen, "UiO1: Constituent-based discriminative ranking for negation resolution," in **SEM 2012: The First Joint Conference on Lexical and Computational Semantics – Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012)*. Montréal, Canada: Association for Computational Linguistics, 7-8 Jun. 2012, pp. 310–318. [Online]. Available: https://aclanthology.org/S12-1041