

SSN_MLRG1 at SemEval-2022 Task 10: Structured Sentiment Analysis using 2-layer BiLSTM

Karun Anantharaman, Divyasri K, Jayanthan PT

Ms. S. Angel Deborah, Ms. S. Rajalakshmi,

Dr. R.S. Milton, Dr. T. T. Mirnalinee

Department of Computer Science and Engineering

SSN College of Engineering

Chennai 603 110, Tamil Nadu, India

karun19049@cse.ssn.edu.in,

{divyasri2011037, jayanthan2010606}@ssn.edu.in,

{rajalakshmis, angeldeborahs}@ssn.edu.in,

{miltonrs, mirnalineett}@ssn.edu.in

Abstract

Task 10 in SemEval 2022 is a composite task which entails analysis of opinion tuples, and recognition and demarcation of their nature. In this paper, we will elaborate on how such a methodology is implemented, how it is undertaken for a Structured Sentiment Analysis, and the results obtained thereof. To achieve this objective, we have adopted a bi-layered BiLSTM approach. In our research, a variation on the norm has been effected towards enhancement of accuracy, by basing the categorization meted out to an individual member as a by-product of its adjacent members, using specialized algorithms to ensure the veracity of the output, which has been modelled to be the holistically most accurate label for the entire sequence. Such a strategy is superior in terms of its parsing accuracy and requires less time. This manner of action has yielded an SF1 of 0.33 in the highest-performing configuration.

1 Introduction

Sentiment Analysis is a specialized area under Natural Language Processing which deals with the extraction of opinions and emotions from text which may include reviews, social media posts, forums or news. Sentiment Analysis has become a powerful tool for detecting the public opinion on any topic.

This research article evinces how the sentiments and opinions expressed by people in various environments could be understood by means of a structured sentiment analysis (SSA) approach. There are conventional means to

this end, in transition-based and graph-based techniques, but a potent alternative is the novel way to envision dependency parsing tasks as algorithmic pattern-recognition by way of sequence labelling. Sequence labeling is essentially an agglomeration of multiple discrete categorizations, extended on to each element of the entire sequence. The model is trained on 7 seven data sets with 5 different languages English, Spanish, Basque, Norwegian and Catalan.

In Subtask-A we try to separately identify the holder and target in the text, along with the sentiment expression used using a Sequence Labelling Approach.

In Subtask-B we predict whether the extracted tokens have a relation or the lack of one.

Subtask-C involves the concatenation of the results of Subtask-A and Subtask-B to provide the final predictions.

2 Background

2.1 Definitions

The following contains descriptions of the models made use of, and related terminology. Namely we will define LSTM, BiLSTM, Sigmoid, Linear, Max Pooling Layers and Word Embeddings and the BIO Sequence Labelling approach.

Long Short-Term Memory(LSTM)

Speech-recognition language models, among numerous other activities, have seen encouraging levels of success by availing Recurrent Neural Networks (RNN) (Mikolov et al., 2010, 2011) (Graves and Schmidhuber, 2005). The model is made to predict the current output by analysing the long-distance features, by virtue of how history information is incorporated into an RNN. Long-range dependencies are better detected and processed by LSTMs than RNNs (Graves and Schmidhuber, 2005) by virtue of their difference in terms of how the latter has purpose-built memory cells in lieu of the hidden layer updates in the former (Hochreiter and Schmidhuber, 1997). Aside from this striking feature, LSTMs and RNNs are essentially the same.

BiDirectional Long Short-Term Memory (BiLSTM) - It a sequence processing model that comprises of one LSTM receiving input in the forward direction, and another which drives input backwards. BiLSTMs are used to enhance the subject and context to be used by the network, by proliferating the quantity of data that could be accessed by the algorithm (e.g. cognizance of the words in front of, and following a certain word that is to be analyzed).

Sigmoid Layer The sigmoid layer is responsible for making sure that the output remains confined within the interval (0,1), through subjecting a sigmoid function on the input.

$$S(x) = \frac{1}{1 + e^{-x}}$$

Linear Layer Mathematically, this module is designed to calculate the linear equation $Ax = b$ where x is input, b is output, A is weight.

Max Pooling Layer Max pooling task results in a map output in which the significant features from the preceding feature map is reflected, considering how it extracts the most crucial constituents of the feature map under the purview of the filter.

Word Embeddings It refers to a specialized means of representation for input text, dependent on its connotation, thus attributing the same representation to words who depict a similar meaning. In this paper we use GloVe and FastText embeddings.

BIO Sequence Labelling - “B-I-O” is a tagging scheme, where either of “B” (Beginning), “I” (Inside), or “O” (Outside) labels denote the relative position of the part of speech. If none of these three assignments could be meted out to the text in question, a special label denoting that case could be assigned.

2.2 Related Work

Sentiment analysis constitutes five operations in sequence as i) sentiment expression extraction, ii) sentiment target extraction, iii) sentiment-holder extraction, iv) definition of relationships between elements, v) assignment of polarity. (Yadav and Vishwakarma, 2020) compares various SOTA DL techniques which have been applied to this problem, including CNNs, Recursive Neural Nets, RNNs, LSTM, GRU and Deep Belief Networks, concluding that LSTMs give better results compared to other models. (Xu et al., 2019) explored the possibility of using BiLSTMs for sentiment analysis on comments, and found improved accuracy. (Phan et al., 2020) proposed an ensemble model of various feature vectors to form embeddings which were fed to a CNN, this method greatly improved accuracy on sentences with fuzzy sentiment. A Bidirectional RNN-CNN (Basiri et al., 2021) was also found to achieve SOTA results. Thus, emphasizing the fact that bidirectional models capture context better in textual data.

Sequence Labeling Typically, it would be advantageous to formulate an NLP operation akin to a general sequence-labeling task. Each element from a defined input sequence is considered, and a collection of labels is scrutinized to pick out one relevant to the text, and an as-

signment is made with the aforementioned. In this paper we propose a Part of Speech tagging approach. (Akhundov et al., 2018) have shown how BiLSTM-CRF can be used for this task. (Prasad and Kan, 2017) shows the extraction of keyphrases and relation prediction using CRFs for sequence labeling.

2.3 Data

The seven datasets in question encompass five different languages of which the expressions are composed of, to be subjected to structured sentiment analysis. They are made such as to typify the way in which the particular approach to this study performs, as necessitated by the task. The datasets are similar in terms of possessing holders, expressions and targets, while dissimilarities are by virtue of their frequency and distribution.

The maximum number of holders in an individual dataset would be the 2,054, featured in MPQA (Wiebe et al., 2005), while the lion’s share of targets (8,293) and expressions (11,115) both are allocated to NoReCFine (Øvrelid et al., 2020). The set DSUnis (Toprak et al., 2010) possesses the least amount of all among holders, targets, and expressions in 94, 1601, and 1082 respectively.

The dataset provided by openNER project (Agerri et al., 2013) are `opener_en` and `opener_es`. The number of targets and expressions in `opener_en`, `opener_es` are 1286, 1760, 1062 and 1625 respectively.

All the groups carry markers for polarity of the text, which shows the positive or negative connotation carried by each member. MPQA and DSUnis are distinctive in how they also include instances of “neutrality”, beside the extremities. In DSUnis, this feature is utilized while dealing with clauses that showcase varying degrees of both polarities, with contextual variance. The neutralities in MPQA are much less complicated, as they only vouch for words that are subjective, and may not necessarily have a polarizing effect.

MPQA is entirely in English, and carries text from news agencies. The two datasets involving critiques of hotels are MultiBEU

and MultiBCA (Barnes et al., 2018), which are Basque and Catalan, respectively. They include markers that further qualify each polarity, as “strong positive” or “strong negative”. DSUnis is essentially an agglomeration of user reviews in English from the internet towards e-commerce and educational institutions, with only the latter being considered as part of this research. This is a consequence of the e-commerce reviews mostly comprising only the relevant polar targets that account for polarity, without holding the expressions themselves. NoReCFine is a Norwegian dataset made up of professional reviews belonging to a multitude of domains, and is also the most voluminous set of the lot. It additionally shows the intensity of the polarity for each expression, as in slight, normal or strong, which is deemed beyond the scope of the study.

3 System Overview

Here we provide a model that first learns to extract the sub-elements (holders, targets, expressions) using sequence labelers, and then tries to classify whether or not they have a relationship.

Specifically, we first train three separate BiLSTM models to extract holders, targets, and expressions, respectively. We then train a relation prediction model, which uses a BiLSTM + max pooling to create contextualized representations of 1) the full text, 2) the first element (either a holder or target) and 3) the sentiment expression.

These three depictions are concatenated and sent to a linear layer, followed by a sigmoid function. The training consists of predicting whether two elements have a relationship or not, converting the problem in binary classification.

4 Experimental Setup

The sequence labelling model employed in our research essentially attempts to divaricate and demarcate various elements of the input text into tuples. To begin with, the starting file `Get_baseline.sh` is executed in order to call the subsequent files in `convert-to-bio.py`, and

Convert-to-rels.py. The former is responsible for converting the given statements into the workable format, as in the stratification of data into holder, target and expression parts, with “B-I-O” labels. The labels are also carriers of the polarity (positive, negative, or neutral) of each text. Next up, the Convert-to-rels.py file is called, upon which, it extracts the target/expression pairs and holder/expression pairs, and creates 2 new fields e1, e2. Furthermore, the BiLSTMs are trained to extract holder, target and expression from the data. Pretrained GloVe or FastText embeddings are also provided to the model for the operation of labeling. After passing the input through the aforementioned processing stages, the vocabulary obtained in the end is stored in a vocabulary dictionary. 3 BiLSTMs are trained separately for each labeling task, following which, is the relation prediction model trained. The full sentence, in addition to e1 and e2, are all scanned for relations. If relations are present, then predictions are made accordingly. Finally to get a consolidated prediction.json terminal output file, the inference.py file is called. The holders, targets, and expressions have been extracted using the trained BiLSTMs already and the polarity is available from the expression labelling. The data is formatted accordingly and lastly packed into a neat json format. The following model incorporates a learning rate of 0.01. We have run it for upto 10 epochs, to arrive at a considerable amount of accuracy. The number of hidden layers in this model is kept as 1 by default.

5 Results

The efficacy of the sentiment analysis model towards encapsulating the full sentiment graph can be depicted in terms of the criteria as enumerated by two benchmarks, in Sentiment Graph F1 (SF1) and Non-polar Sentiment Graph F1 (NSF1). The sentiment graphs are considered in terms of holder, target, expression, and polarity, for evaluation by SF1, whereas NSF1 takes into account all elements of the tuple, except polarity for scrutinization. In this case, a perfect match on the graph, with

respect to the mean of all three spans, and including weights pertaining to the gold and expected spans for each member is considered as a true positive. The precision value is a ratio where the numerator is the sum of predicted tokens that are found to be right, with the denominator being the total sum of all predicted tokens (amount of gold tokens is the denominator in case of recall). Empty targets and tokens are also taken into consideration.

Dataset	SF1-Score
norec	0.191
multibooked_ca	0.323
multibooked_eu	0.331
opener_en	0.306
opener_es	0.257
mpqa	0.015
darmstadt_unis	0.104
average	0.218

Table 1: Results

6 Conclusion

In summation, our research work presents the three-tier model which we constructed, and is found to have the best accuracy of 0.33. The model has been proven to consume significantly less time as opposed to graph parsing. An inconsistency however has been documented, with regard to the processing of complex sentences carrying multiple expressions, in how it considers only the polarity attributed to the terminal element. Going forward, the component BiLSTM could be supplanted with BiLSTM CRF and run, owing to the nature of the latter to be robust and independent of word embedding, and capability to provide superior accuracy levels on Parts of Speech tagging, Name Entity Recognition of data sets and chunking.

References

Rodrigo Agerri, Montse Cuadros, Sean Gaines, and German Rigau. 2013. OpeNER: Open polarity enhanced named entity recognition. In *Sociedad Española para el Procesamiento del Lenguaje Natural*, volume 51, pages 215–218.

- Adnan Akhundov, Dietrich Trautmann, and Georg Groh. 2018. [Sequence labeling: A practical approach](#). *CoRR*, abs/1808.03926.
- Jeremy Barnes, Toni Badia, and Patrik Lambert. 2018. [MultiBooked: A corpus of Basque and Catalan hotel reviews annotated for aspect-level sentiment classification](#). In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).
- Mohammad Ehsan Basiri, Shahla Nemati, Moloud Abdar, Erik Cambria, and U. Rajendra Acharya. 2021. [Abcdm: An attention-based bidirectional cnn-rnn deep model for sentiment analysis](#). *Future Generation Computer Systems*, 115:279–294.
- Alex Graves and Jürgen Schmidhuber. 2005. [Framewise phoneme classification with bidirectional lstm and other neural network architectures](#). *Neural Networks*, 18(5):602–610. IJCNN 2005.
- Sepp Hochreiter and Jürgen Schmidhuber. 1997. [Long Short-Term Memory](#). *Neural Computation*, 9(8):1735–1780.
- Tomas Mikolov, Martin Karafiát, Lukas Burget, Jan Cernocký, and Sanjeev Khudanpur. 2010. Recurrent neural network based language model. volume 2, pages 1045–1048.
- Tomas Mikolov, Stefan Kombrink, Lukas Burget, J.H. Cernocky, and Sanjeev Khudanpur. 2011. [Extensions of recurrent neural network language model](#). pages 5528 – 5531.
- Lilja Øvrelid, Petter Mæhlum, Jeremy Barnes, and Erik Veldal. 2020. [A fine-grained sentiment dataset for Norwegian](#). In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 5025–5033, Marseille, France. European Language Resources Association.
- Huyen Trang Phan, Van Cuong Tran, Ngoc Thanh Nguyen, and Dosam Hwang. 2020. [Improving the performance of sentiment analysis of tweets containing fuzzy sentiment using the feature ensemble model](#). *IEEE Access*, 8:14630–14641.
- Animesh Prasad and Min-Yen Kan. 2017. [WING-NUS at SemEval-2017 task 10: Keyphrase extraction and classification as joint sequence labeling](#). In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 973–977, Vancouver, Canada. Association for Computational Linguistics.
- Cigdem Toprak, Niklas Jakob, and Iryna Gurevych. 2010. [Sentence and expression level annotation of opinions in user-generated discourse](#). In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 575–584, Uppsala, Sweden. Association for Computational Linguistics.
- Janyce Wiebe, Theresa Wilson, and Claire Cardie. 2005. [Annotating expressions of opinions and emotions in language](#). *Language Resources and Evaluation*, 39(2):165–210.
- Guixian Xu, Yueting Meng, Xiaoyu Qiu, Ziheng Yu, and Xu Wu. 2019. [Sentiment analysis of comment texts based on bilstm](#). *IEEE Access*, 7:51522–51532.
- Ashima Yadav and Dinesh Kumar Vishwakarma. 2020. [Sentiment analysis using deep learning architectures: a review](#). *Artificial Intelligence Review*, 53(6):4335–4385.