

Targeted Aspect-Based Sentiment Analysis for Lithuanian Social Media Reviews

Mažvydas PETKEVIČIUS¹, Daiva VITKUTĖ-ADŽGAUSKIENĖ and Darius AMILEVIČIUS

Vytautas Magnus University, Lithuania

Abstract. The paper presents research results for solving the task of targeted aspect-based sentiment analysis in the specific domain of Lithuanian social media reviews. Methodology, system architecture, relevant NLP tools and resources are described, finalized by experimental results showing that our solution is suitable for solving targeted aspect-based sentiment analysis tasks for under-resourced, morphologically rich and flexible word order languages.

Keywords. Targeted aspect-based sentiment analysis, entity recognition, deep learning, social media

1. Introduction

While initially sentiment analysis (or opinion mining) was implemented by assigning sentiment polarities on the sentence level, it gradually evolved towards fine-grained aspect-based sentiment analysis (ABSA), allowing to identify sentiments of several aspects in a sentence.

In order to achieve even better accuracy of sentiment analysis, aspect-based sentiments are classified towards target entity mentions in given sentence, leading to so-called targeted ABSA (TABSA). Such a complex approach is related to the nature of user opinions or reviews in social media texts, where usually we have several entities mentioned in a text, together with different aspects and sentiment polarities of these entities being addressed.

Following this approach, several sequential tasks must be solved – identifying entity mentions, resolving aspect categories and classifying sentiment polarities with respect to aspect categories. Different combinations of information extraction, deep learning and natural language processing (NLP) techniques are used for these tasks.

Related work. While sentiment analysis on the sentence level shows really good results, especially when applying deep learning techniques, TABSA remains a difficult task. Related research addresses different issues, including building balanced datasets [1], choosing the right model architecture [2] and coping with language specifics [3].

¹Corresponding Author: Mažvydas Petkevičius; Vytautas Magnus university; E-mail: mazvydas.petkevicius@vdu.lt.

Different architectures are considered, especially neural sequential models, such as long short-term memory (LSTM) networks with different modifications - hierarchical attention mechanism [4,5], delayed memory update [6]. Another approach suggests convolutional neural network based (CNN-based) aspect-level sentiment classification models, consisting of two CNNs [7].

Problem statement. TABSA is normally a suitcase research problem [8] that requires tackling many natural language processing (NLP) tasks, including text preprocessing, named-entity recognition, dependency parsing, etc.

Also, dataset preparation involves the use of different NLP resources as annotated corpora, WordNet ontologies, or topical dictionaries. The quality of these resources directly influences the precision of the overall solution. While the necessary NLP tools are already mature for well-resourced languages, the situation is different for the Lithuanian language.

We propose a TABSA architecture for the Lithuanian language, combining CNN-based classification models and commonsense language knowledge embedded in NLP components. It is presented on the example of university study program sentiment analysis in social media reviews.

2. Methodology

Our TABSA approach considers sentiment analysis as aimed at different entities (targets) – organizations, products, services, persons, etc. We define entities as composite objects, having a set of components, and, also, a set of attributes. Each component may have its own sub-components and its set of attributes, and so on. Thus, an entity can be hierarchically decomposed based on the part-of relation [9]. The aspects of an entity, in general, are the components and attributes related to the entity or its constituent parts. In our example of study program sentiment analysis, entity (target) is hierarchically structured in three levels (university, study direction, and study program), and aspects, such as “teaching”, “studies”, “infrastructure”, “career”, can be attributed to any of these levels.

Overview of the approach. Joint recognition of targets and aspect-sentiment pairs in TABSA is a difficult task, and, usually, these two issues are separated in research works [6]. As we consider the case of targets having a composite and hierarchical structure, our approach handles target recognition and aspect-sentiment learning as separate tasks, sharing aspect classification results and corresponding datasets. Rule-based linking phase is included for aggregating the results of those two tasks at the end of the overall process.

Target recognition: task description. In our case, target recognition task is aimed at identifying composite structure of hierarchical named-entities and a corresponding aspect category. Rule-based matching is applied for named-entity recognition (NER), allowing either matching against a knowledge-base, or applying a partial rule-defined match [10]. Linking of all the Target components, including aspect category, obtained from aspect classifier, is performed in the context of a chunk by applying corresponding rules.

Aspect-sentiment classification: task description. Linguistic and statistic research of social media corpus texts revealed that aspect terms are expressed in nouns and noun-phrases (approx. 65 percent), and verbs and verb-phrases (around 35 percent). Both aspect and sentiment terms are explicit and implicit, however, in our case, we limit our-

selves to explicit terms as they constitute the majority. Supervised learning is applied with the aim of attributing classes to aspects and sentiments.

Overall model. Conceptual architecture of the proposed solution is presented in Figure-1.

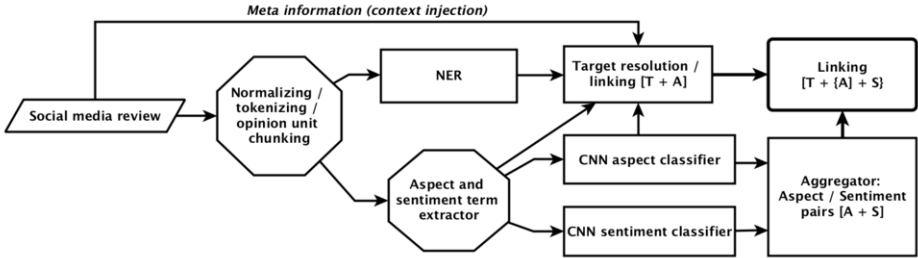


Figure 1. Conceptual TABSA architecture for the proposed solution

In our solution, we go in parallel with the target recognition and aspect-sentiment classification, linking the results afterwards. Target recognition is implemented as a sequence of NER and Target linking tasks. The NER component gets tokenized and normalized input, which is then segmented into meaningful chunks by a syntactic analyzer. Entity recognizer consists of two main components: the main logic, consisting of a rule-set and a knowledge base. Entities are recognized and linked throughout multiple phases: candidate entity generation, candidate entity ranking and unlinkable mention prediction. The recognized entities are then handed off to the target resolution and linking component, which combines target entities with the aspect. The recognized entities are then passed over to the target resolution and linking component, combining target entities with the aspect.

For the prediction of aspect and sentiment categories, we implement two Convolutional Neural Network (CNN) classifiers using the Keras² library with TensorFlow³ as the backend. Dependency-based noun and verb phrase extractor feeds candidate terms from an opinion unit into neural network-based classifiers. Predictions from classifiers go to an aggregator, where classified aspect and sentiment terms are aggregated on opinion unit syntax basis.

Finally, the results from both Target Recognizer and Aspect-Sentiment Classifier are passed over to the Linking component for aggregation, using rule-based techniques and aspect category as the key.

3. Experiment

Resources (Datasets). Lithuanian language is an under-resourced language, especially considering specific topical domains. Our base dataset consists of 13,000 sentences from various Lithuanian social media sources with reviews of a selected domain. 45 percent of the dataset are complex sentences, containing two or more meaningful opinion units

²<https://keras.io>

³<https://www.tensorflow.org>

(sub-sentences). Three human experts have annotated them with aspect category labels (“teaching”, “studies”, “infrastructure”, “career”, “common”, “NULL”), as well as with sentiment category labels (“very negative”, “negative”, “neutral”, “positive”, “very positive”). Generally, we assume that opinion cannot be “neutral”, as only facts can be considered as neutral. However, we include aspect category “NULL” and sentiment “neutral” for mark-up of some sub-sentences that are out of range of our domain evaluation. Experts also manually extracted aspect and sentiment terms. In order to resolve problems with insufficient dataset and to balance it for all categories, we used data augmentation techniques. For this purpose, we used synsets for aspect and sentiment terms from Lit-WordNet ⁴ [11] and Hunspell ⁵ engine with Lithuanian grammar and lexicon sets ⁶ for generating necessary morphological forms for synonym words. Our augmented corpus contains 250,000 words (19,500 unique word-types).

Table 1. Augmented corpus structure

	Infrastructure	Teaching	Studies	Carrier	Common	Neutral
Very positive	500	2,000	2,100	800	300	300
Positive	500	2,000	2,100	800	300	300
Neutral	500	2,000	2,100	800	300	300
Negative	500	2,000	2,100	800	300	300
Very negative	500	2,000	2,100	800	300	300
Total	2,500	10,000	10,500	4,000	1,500	1,500

Knowledge-base for entity recognition and linking was constructed on official data, including fields of studies, study programs and universities and their colloquialisms in social media. NER rule-based matcher was implemented using spaCy ⁷ library.

System setup. To solve aspect-based sentiment multi-class classification problem, complex classifiers were used. A complex classifier consists of two classifiers with identical CNN architecture: one for aspect classification task, and the second - for sentiment classification task. CNN-based classifiers were used as CNNs provide a faster alternative to LSTM models at a comparable performance. They are faster to train and use fewer parameters. CNN models, when applied to text, are a reasonable alternative when there is no strong dependence on distant past of the input sequence, as is in our case, since opinion units are normally very short. In our experiments, complex classifier was used in two scenarios: 1) Baseline model (S1), and 2) Advanced model (S2). For both scenarios, datasets were divided into a training subset (80 percent) and testing subset (20 percent).

In scenario S1, simple CNN architecture was used for multi-class classification with mini-batch size of 256. Our model was starting to overfit at about epoch 30, so we trained it for 25 epochs. In this scenario, aspects and sentiment dictionaries, extracted and labeled by human experts, were used for lookup. A study review corpus, labeled by human experts, was also used. Embedding was done by Keras internally, using our custom studies’ review corpus.

⁴<http://mackus.vdu.lt/LitWordNet>

⁵<http://hunspell.github.io>

⁶<https://github.com/Semantika2>

⁷<https://spacy.io>

Scenario S2 was more complex. In this case, CNN consists of the following layers: embedding layer, two 1D convolutional layers, max pooling layer, global max pooling layer and two dense layers. Rectified linear unit activation function is used, and the final dense layer uses the Sigmoid activation function. The following CNN hyperparameters were used: dropout rate (p) of 0.5, kernel regularizer $l2=1e-4$, mini-batch size of 256 and 5 epochs. Since CNNs can learn patterns in word embeddings, we used FastText word embeddings to use sub-word information. For our purpose, we created our own FastText embedding model using our custom Lithuanian review corpus. CNNs were trained only on a targeted studies' review corpus, where reviews were labeled by human experts.

For Target recognition (NER) spaCy's rule-based matcher was utilized⁸, where a set of patterns comprises a ruleset. These rules were deduced from the knowledge base consisting of several types of entities:

- 57 universities and colleges;
- 107 fields of studies (with relations to universities and study programs);
- 1001 study programs (that are related to fields of studies).

These were then either lemmatized or converted into corresponding regular expressions and added as patterns to spaCy to match multi-word phrases.

Table 2. Experiment results: selected examples

Review	NER	Aspect	Sentiment
Complex sentence (two opinion units, two named entities):			
Opinion unit 1: Šiaip [university1] gali pasigirti geresne technika straipsnyje aiškiai parašyta tai (<i>In general [university1] can be proud of better equipment as stated in the article</i>)	[university1]	Infrastructure	Positive
Opinion unit 2: [university2] gal kai kurie dėstytojai aukštesnio lygio (<i>in [university2] maybe some teachers can be considered superior</i>)	[university1]	Teaching	Negative
Complex sentence (two opinion units, one named entity):			
Opinion unit 1: Šiaip [university1] dėstytojai puikūs (<i>In [university1] teachers are excellent</i>)	[university1]	Teaching	Positive
Opinion unit 2: bet technika pasenusi (<i>but the equipment is outdated</i>)	[injected university1]	Infrastructure	Negative
Non-complex sentence (one opinion unit, one named entity):			
Nieko nesupratau ką dėstė visą semestrą (<i>didn't understand a single thing they were teaching</i>)	[injected university1]	Teaching	Negative

⁸<https://spacy.io/usage/rule-based-matching>

Preprocessing. The preprocessing phase involves text tokenization and normalization. Since often one sentence contains two or more meaningful units with an [aspect, sentiment] pair, as a result we get a sequence of meaningful opinion units: e.g., “sentence 0: [{unit1}, {unit2}, ... {unitN}]. A hybrid approach, consisting of pattern-recognition and dependency-based parser (spaCy for noun phrase chunking and Textacy⁹ for verb-phrase chunking) was applied for splitting text into opinion units. spaCy parser is trained on Vytautas Magnus University “gold standard” treebank for Lithuanian language ALK-SNIS v2¹⁰.

Results. Our solution was tested on unseen reviews from Lithuanian social media sources. Examples of the obtained results for selected test sentences are presented in Table 2. Component-related quality measures are presented in Table 3.

Table 3. Evaluation

	NER	Aspect classifier (S1)	Sentiment classifier (S1)	Aspect classifier (S2)	Sentiment classifier (S2)
Precision	0.66	0.88	0.89	0.89	0.92
Recall	0.76	0.85	0.83	0.86	0.86
Accuracy	83 %	93 %	91 %	94 %	93 %
F1-score	0.71	0.86	0.85	0.87	0.88

4. Discussion and Conclusions

Comparing the results of both experiments, the second setup (scenario S2 with a more complex CNN architecture) has shown slightly better results, but some considerations should be taken into account. In real-life applications, not only benchmark results, but also economic aspects are important. Though scenario S1 with a simple CNN architecture has shown only slightly worse results, this scenario requires significantly less computational resources. On the other hand, scenario S1 requires much more human expert efforts (manual aspect and sentiment term extraction and annotation). The scenario S2 is the opposite. It requires much more computational resources, but, at the same time, much less human expert efforts. In addition, learning transfer can be used in scenario S2. It must also be noted, that for under-resourced language or/and under-resourced domain cases, efficient and high-quality data augmentation technique is needed in both scenarios. That, in its turn, requires additional human expert efforts for augmented data set evaluation as well as additional computational expenses and availability of certain NLP tools.

Since our final scope is an end-to-end production-ready solution, we found that our proposed model performed very well in both scenarios with topical social media reviews. Unfortunately, we have no possibility to compare the results with other similar systems,

⁹<https://github.com/chartbeat-labs/textacy>

¹⁰https://github.com/UniversalDependencies/UD_Lithuanian-ALKSNIS

because most of them are for the English language. This is the first attempt to construct TABSA for Lithuanian.

A TensorFlow friendly combo: tagger and dependency-based parser is of crucial importance for TABSA. Nowadays, the best solution is spaCy, but all the implementations for spaCy's Lithuanian support must be retrained for the Lithuanian language.

The language normalizer plays a very important role in our solution since social media texts are full of misspellings.

Our data augmentation methodology demonstrated its suitability in solving insufficient dataset, domain-oriented and imbalanced dataset problems for the Lithuanian language.

Our future work is planned towards trying and comparing different neural net architectures and different word-embedding and language representation models. For example, BERT language representation model (might provide more advanced embeddings than FastText, but is more computationally expensive). On the other hand, FastText has already proved its ability to deal well with morphologically rich languages, especially solving Out-Of-Vocabulary (OOV) problem. For this reason, we used FastText embedding for our initial TABSA trials.

Lower performance of entity recognizer can be attributed to several factors such as word ambiguity, fuzzy partial matches, incorrect entity linking due to the limited context etc., although it performs reasonably well with exact entity matches.

References

- [1] Saeidi M, Bouchard G, Liakata M, Riedel S. SentiHood: Targeted Aspect Based Sentiment Analysis Dataset for Urban Neighbourhoods. Proceedings of COLING 2016, Osaka, Japan, 2016.
- [2] Augustyniak L, Kajdanowicz T, Kazienko P. Comprehensive Analysis of Aspect arXiv preprint arXiv:1909.04917, 2019. <https://arxiv.org/pdf/1909.04917.pdf>
- [3] Loukachevitch N, Kotelnikov E, Blinov P. Types of Aspect Terms in Aspect-Oriented Sentiment Labeling. Proceedings of the 5th Workshop on Balto-Slavic Natural Language Processing, Hissar, Bulgaria, 2015.
- [4] Ma Y, Peng H, Cambria E. Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive lstm. In Proceedings of AAAI 2018, New Orleans, USA, 2018.
- [5] Chen P, Sun Z, Bing L, Yang W. Recurrent attention network on memory for aspect sentiment analysis. In Proceedings of EMNLP 2017, Copenhagen, Denmark, 2017.
- [6] Liu F, Cohn T, Baldwin T. Recurrent Entity Networks with Delayed Memory Update for Targeted Aspect-Based Sentiment Analysis. Proceedings of NAACL-HLT (2018), New Orleans, USA, 2018.
- [7] Noh Y, Park S, Park SB. Aspect-Based Sentiment Analysis Using Aspect Map. Appl. Sci. 2019, 9 (16), 3239.
- [8] Cambria E, Poria S, Gelbukh A, Thelwall M. Sentiment Analysis is a Big Suitcase. IEEE Intelligent Systems, Vol. 32, Issue 6, 2017.
- [9] Zhang L, Liu B. Aspect and Entity Extraction for Opinion Mining. In: Chu W. (eds) Data Mining and Knowledge Discovery for Big Data. Studies in Big Data, vol 1. Springer, Berlin, Heidelberg, 2014.
- [10] Shen W, Wang J, Han J. Entity Linking with a Knowledge Base: Issues, Techniques, and Solutions, IEEE Transactions on Knowledge and Data Engineering, Vol. 27, Issue 2, 2014.
- [11] Vitkutė-Adžgauskienė D, Dainauskas J, Amilevičius D, Utkā A. Lithuanian Word Network – LitWord-Net. Darbai ir Dienos. 2016. Vol. 64 (In Lithuanian).