

Unsupervised Methods to Improve Aspect-Based Sentiment Analysis in Czech

Tomáš Hercig¹², Tomáš Brychcín², Lukáš Svoboda¹², Michal Konkol², Josef Steinberger¹

¹ University of West Bohemia, Department of Computer Science and Engineering,
Faculty of Applied Sciences, Plzeň,
Czech Republic

² University of West Bohemia, NTIS—New Technologies for the Information Society,
Faculty of Applied Sciences, Plzeň,
Czech Republic

tigi@kiv.zcu.cz, brychcin@kiv.zcu.cz, svobikl@kiv.zcu.cz,
konkol@kiv.zcu.cz, jstein@kiv.zcu.cz

Abstract. We examine the effectiveness of several unsupervised methods for latent semantics discovery as features for aspect-based sentiment analysis (ABSA). We use the shared task definition from SemEval 2014. In our experiments we use labeled and unlabeled corpora within the restaurants domain for two languages: Czech and English. We show that our models improve the ABSA performance and prove that our approach is worth exploring. Moreover, we achieve new state-of-the-art results for Czech. Another important contribution of our work is that we created two new Czech corpora within the restaurant domain for the ABSA task: one labeled for supervised training, and the other (considerably larger) unlabeled for unsupervised training. The corpora are available to the research community.

Keywords. Aspect-based sentiment analysis, latent semantics.

1 Introduction

Majority of recent approaches to sentiment analysis tries to detect the overall polarity of a sentence (or a document) regardless of the target entities (e.g. restaurants, laptops) and their aspects (e.g. food, price, battery, screen). In contrast, the current approach, aspect-based sentiment analysis (ABSA) identifies the aspects of a given target entity and estimates the sentiment polarity for each mentioned aspect.

In the context of the ABSA task, the bottleneck is the size of the annotated data, which should be considerably larger in order to simulate real world applications. Web content such as blogs, forums, reviews etc. present a large amount of easily accessible domain-relevant unlabeled data which we could use to add specific domain knowledge essential for improving the state of the art of sentiment analysis. Thus we try to demonstrate the usefulness of these data.

Most of the research in automatic sentiment analysis has been devoted to English. There have been several attempts in Czech as well [27, 8, 2], but all were focused on the global (sentence or document level) sentiment.

The first attempt at aspect-based sentiment analysis in Czech was presented in [25]. This work provides an annotated corpus of 1244 sentences from the restaurant reviews domain and a baseline model achieving 68.65% F-measure in aspect term extraction, 74.02% F-measure on aspect category extraction, 66.27% accuracy in aspect term polarity classification, and 66.61% accuracy in aspect category polarity classification. The work in [26] creates a dataset in the domain of IT product reviews. This dataset contains 200 annotated sentences and 2000 short segments, both annotated with sentiment and marked aspect

terms (targets) without any categorization and sentiment toward the marked targets.

The current state of the art of aspect-based sentiment analysis methods for English was presented at the latest SemEval ABSA tasks namely the SemEval 2014 - 2016 [23, 22, 21]. The detailed description of each system is beyond the scope of this paper.

Our main goal is twofold: to show how unsupervised methods can improve an ABSA system in different languages and the creation of sufficiently large corpora for the ABSA task in Czech.

2 The ABSA Task

Aspect-based sentiment analysis firstly identifies the aspects of the target entity and then assigns a polarity to each aspect. There are several ways to define aspects and polarities. We use the definition based on the SemEval 2014's ABSA task, which distinguishes two types of aspect-based sentiment: aspect terms and aspect categories. The whole task is divided into four subtasks. The later SemEval's ABSA tasks (2015 and 2016) further distinguish between more detailed aspect categories and associate aspect terms (targets) with aspect categories.

2.1 Subtask 1: Aspect Term Extraction

Given a set of sentences with pre-identified entities (e.g. restaurants), the task is to identify the aspect terms present in the sentence and return a list containing all the distinct aspect terms.

CZ: **Tlačenka** se rozpadla, **polévka** ušla.
 EN: The **porkpie** broke down, the **soup** was ok.
 → {Tlačenka (porkpie), polévka (soup)}

2.2 Subtask 2: Aspect Term Polarity

For a given set of aspect terms within a sentence, the task is to determine the polarity of each aspect term: positive, negative, neutral or bipolar (i.e. both positive and negative).

CZ: **Tlačenka** se rozpadla, **polévka** ušla.
 EN: The **porkpie** broke down, the **soup** was ok.
 → {Tlačenka (porkpie): negative,
 polévka (soup): positive}

2.3 Subtask 3: Aspect Category Extraction

Given a predefined set of aspect categories (e.g. price, food), the task is to identify the aspect categories discussed in a given sentence. Aspect categories are typically coarser than the aspect terms of Subtask 1, and they do not necessarily occur as terms in the given sentence. In the analysed domain of 'restaurants', the categories include food, service, price, and ambience.

CZ: Přivítala nás velmi příjemná servírka, ale také místnost s ošuntělým nábytkem.
 EN: We were welcomed by a very nice waitress and a room with time-worn furniture.
 → {service, ambience}

2.4 Subtask 4: Aspect Category Polarity

Given a set of pre-identified aspect categories (e.g. food, price), the task is to determine the polarity (positive, negative, neutral or bipolar) of each aspect category.

CZ: Přivítala nás velmi příjemná servírka, ale také místnost s ošuntělým nábytkem.
 EN: We were welcomed by a very nice waitress and a room with time-worn furniture.
 → {service: positive, ambience: negative}

3 Distributional Semantics in ABSA

The backbone principle of methods for discovering hidden meaning in a plain text is the formulation of the *Distributional Hypothesis* in [6]: "a word is characterized by the company it keeps."

The direct implication of this hypothesis is that the meaning of a word is related to the context where it usually occurs and thus it is possible to compare the meanings of two words by statistical comparisons of their contexts. In this paper we use six distributional semantics models (see the following list).

Hyperspace Analogue to Language (HAL) [16]

is a very simple method for building semantic space based on co-occurrence matrix.

Correlated Occurrence Analogue to Lexical Semantics (COALS) [24] is an extension of the HAL model which employs the Pearson's correlation and Singular Value Decomposition (SVD).

Continuous Bag-of-Words (CBOW) [18] is a model based on Neural Network Language Model (NNLM) that tries to predict the current word using a small context window around the word.

Skip-gram [19] is similar to CBOW but works in opposite direction.

It finds word patterns that are useful for predicting the surrounding words within a certain range in a sentence.

Global Vectors (GloVe) [20] use the ratios of the word-word co-occurrence probabilities to encode the meanings of the words.

Latent Dirichlet Allocation (LDA) [1] discovers hidden topics in the text.

Distributional semantics models typically represent the meaning of a word as a vector, that reflects the contextual information of the word throughout the training corpus. Each word is associated with a vector of real numbers. Represented geometrically, the word meaning is a point in a high-dimensional space. The words that are closely related in meaning tend to be closer in the space.

The ability to compare two words enables us to use a clustering method. Similar words are clustered into bigger groups of words (clusters). The clusters derived from semantics models are used as new sources of information for all four subtasks of ABSA (see Section 5).

4 The Data

The methods described in Section 3 require large unlabeled data in order to be trained. In this paper we used two types of corpora, labeled and unlabeled for both Czech and English. The properties of these corpora are shown in Table 1.

Labeled corpora for both languages are required to train the classifiers (see Section 5). For English, we use the corpora introduced in SemEval 2014 Competition Task 4 [23]. The main criterion in choosing the dataset was the dataset size (see Table 1).

For Czech, we extended the dataset from [25], nearly doubling its size. The annotation procedure was identical to that of the original dataset. The corpus was annotated by five native speakers. The majority voting scheme was applied to the gold label selection. Agreement between any two annotators was evaluated in the same way as we evaluate our system against the annotated data (taken as the gold standard). This means we take the output of the first annotator as the gold standard and the output of the second annotator as the output of the system. The same evaluation procedure as in [23] is used, i.e. the F -measure for the aspect term and aspect category extraction, and the accuracy for the aspect term and aspect category polarity. The resulting mean values of annotator agreement for the Czech labeled corpus are 82.91% (aspect term extraction), 88.02% (aspect category extraction), 85.71% (aspect term polarity) and 88.44% (aspect category polarity). We believe this testifies to the high quality of our corpus. The corpus is available for research purposes at <http://nlp.kiv.zcu.cz/research/sentiment>.

The labeled corpora for both languages use the same annotation scheme and are in the same domain. This allows us to compare the effectiveness of the used features on the ABSA task for these two very different languages.

The lack of publicly available data in the restaurant domain in Czech forced us to create a cross-domain unlabeled corpus for Czech. The Czech unlabeled corpus is thus composed of three related domains: recipes (8.8M tokens, 57.1%), restaurant reviews (2M tokens, 12.8%), and hotel reviews (4.7M tokens, 30.1%). We selected these three domains because of their close relations, which should be sufficient for the purposes of the ASBA task.

The English unlabeled corpus was downloaded from <http://opentable.com>.

5 The ABSA System

We use and extend the systems created by [3]. We implemented four separate systems – one for each subtask of ABSA. The required machine learning algorithms are implemented in the Brainy

Table 1. Properties of the SemEval ABSA tasks and corpora used in the experiments in terms of the number of *sentences*, aspect terms (*targets*), aspect categories (*categories*), *tokens* and unique *words*

Dataset	Sentences	Targets	Categories	Tokens	Words
English labeled 2016 train + test	2.7k	2.5k	3.4k	39.1k	4.4k
English labeled 2015 train + test	2k	1.9k	2.5k	29.1k	3.6k
English labeled 2014 train	3k	3.7k	3.7k	46.9k	4.9k
Czech labeled 2014 train	2.15k	3.3k	3k	34.9k	7.8k
English unlabeled	409k	—	—	27M	121k
Czech unlabeled	514k	—	—	15M	259k

machine learning library [12]. We further extended this system and competed in the SemEval 2016 ABSA task and we were ranked as one of the top performing systems [9].

The systems share a simple preprocessing phase, in which we use a tokenizer based on regular expressions. The tokens are transformed to lower case. Punctuation marks and stop words are ignored for the polarity task. In the case of Czech, we also remove diacritics from all the words, because of their inconsistent use.

The feature sets created for individual tasks are based on features commonly used in similar natural language processing tasks, e.g. named entity recognition [14], document level sentiment analysis [7], and document classification [5]. The following baseline features were used:

Affixes (A) – Affix (length 2-4 characters) of a word at a given position.

Tf-idf (T) – Term frequency - inverse document frequency of a word.

Learned dictionary (LD) – Dictionary of aspect terms from training data.

Words (W) – The occurrence of word at a given position (e.g. previous word).

Bag of words (BoW) – The occurrence of a word in the context window.

Bigrams (B) – The occurrence of bigram at a given position.

Bag of bigrams (BoB) – The occurrence of a bigram in the context window.

The baseline feature set is then extended with semantic features. The features are based on the word clusters created using the semantic models described in Section 3. The following semantic features were used:

Clusters (C) – The occurrence of a cluster at a given position.

Bag of clusters (BoC) – The occurrence of a cluster in the context window.

Cluster bigrams (CB) – The occurrence of cluster bigram at a given position.

Bag of cluster bigrams (BoCB) – The occurrence of cluster bigram in the context window.

Each C (alternatively, CB, BoC, or BoCB) feature can be based on any of the models from Section 3. In the description of the systems for individual tasks, we use simply C to denote that we work with this type of feature. When we later describe the experiments, we use explicitly the name of the model (e.g. HAL).

5.1 Subtask 1: Aspect Term Extraction

The aspect term extraction is based on experiences in Named Entity Recognition (NER) [13, 14]. The NER task tries to find special expressions in a text and classify them into groups. The aspect term extraction task is very similar, because it also tries to identify special expressions. In contrast with NER, these expressions are not classified, and have different properties, e.g. they are not so often proper names.

We have decided to use Conditional Random Fields (CRF) [15], because they are regarded as the state-of-the-art method for NER. The baseline feature set consists of W , BoW , B , LD , and A . In our experiments, we extend this with the semantic features C and CB . The context for this task is defined as a five word window centred at the currently processed word.

5.2 Subtask 2: Aspect Term Polarity

Our aspect term polarity detection is based on the Maximum Entropy classifier, which works very well in many NLP tasks, including document-level sentiment analysis [7].

For each aspect term, we create a context window ten words to the left and right of the aspect term. The features for each word and bigram in this window are weighted based on their distance from the aspect term given by weighing function. This follows the general belief that close words are more important than distant words, which is used in several methods [16].

We have tested several weighing functions and selected the Gaussian function based on the results. The expected value μ and the variance σ^2 of the Gaussian function were found experimentally on the training data. We omit the description of these experiments, as they are outside the scope of this paper.

The feature set for our baseline system consists of BoW and BoB , and we further experiment with BoC and $BoCB$.

5.3 Subtask 3: Aspect category extraction

The aspect category extraction is based on research in multi-label document classification [5]. The multi-label document classification system tries to assign several labels to a document. We do exactly the same, although our documents are only single sentences and the labels are aspect term categories.

We use one binary Maximum Entropy classifier for each category. It decides whether the sentence belongs to the given category. The whole sentence is used as the context.

The baseline uses the features BoW , BoB , and T . We try to improve it with BoC and $BoCB$.

5.4 Subtask 4: Aspect Category Polarity

The aspect category task is very similar to document-level sentiment analysis [7] when the document is of similar length. We create one Maximum Entropy classifier for each category. For a given category, the classifier uses the same principle as in global sentiment analysis. Of course, the training data are different for each category. The context in this task is the whole sentence.

We use the following features as a baseline: BoW , BoB , and T . In our experiments, we extend this with BoC and $BoCB$.

6 Experiments

In the following presentation of the results of the experiments, we use the notation BL for a system with the baseline feature set (i.e. without cluster features). Cluster features based on HAL are denoted by HAL . For other semantic spaces, the notation is analogous.

Because Czech has rich morphology we use stemming to deal with this problem (stemming is denoted as S). Also we use the stemmed versions of semantic spaces (the corpora used for training semantics spaces are simply preprocessed by stemming). The system that uses this kind of cluster features is denoted by $S\text{-}HAL$ for the HAL model, and analogously for the other models.

The union of feature sets is denoted by the operator $+$. E.g. $BL+S\text{-}BL+S\text{-}GloVe$ denotes the baseline feature set extended by stemmed baseline features and by a stemmed version of GloVe clusters.

The number of clusters for a particular semantic space is always explicitly mentioned in the following tables.

6.1 Unsupervised Model Settings

All unsupervised models were trained on the unlabeled corpora described in Section 4.

The implementations of the HAL and COALS algorithms are available in an open source package S-Space [10]¹. The settings of the GloVe, CBOW, and Skip-gram models reflect the results of these methods in their original publications [20, 18] and were set according to a reasonable proportion of the complexity and the quality of the resulting word vector outputs. We used the GloVe implementation provided on the official website², CBOW and Skip-gram models use the Word2Vec³ implementation and the LDA implementation comes from the MALLET [17] software package.

The detailed settings of all these methods are shown in Table 2.

Table 2. Model settings

	dimension	window	special settings
HAL	50,000	4	
COALS	14,000	4	without SVD
GloVe	300	10	100 iterations
CBOW	300	10	100 iterations
SKIP	300	10	100 iterations
LDA	100	sentence	1000 iterations

CLUTO software package [11] is used for words clustering with the k -means algorithm and cosine similarity metric. All vector space models in this paper cluster the word vectors into four different numbers of clusters: 100, 500, 1000, and 5000. For stemming, we use the implementation of HPS [4]⁴ that is the state-of-the-art unsupervised stemmer.

6.2 Results

We experimented with two morphologically very different languages, English and Czech. English,

¹Available at <https://code.google.com/p/airhead-research/>.

²Available at <http://www-nlp.stanford.edu/projects/glove/>.

³Available at <https://code.google.com/p/word2vec/>.

⁴Available at <http://liks.fav.zcu.cz/HPS>.

as a representative of the Germanic languages, is characterized by almost no inflection. Czech is a representative of the Slavic languages, and has a high level of inflection and relatively free word order.

We provide the same evaluation as in the SemEval 2014 [23]. For the aspect term extraction (TE) and the aspect category extraction (CE) we use F -measure as an evaluation metric. For the sentiment polarity detection of aspect terms (TP) and aspect categories (CP), we use accuracy.

We use 10-fold cross-validation in all our experiments. In all the tables in this section, the results are expressed in percentages, and the numbers in brackets represents the absolute improvements against the baseline.

We started our experiments by testing all the unsupervised models separately. In the case of Czech, we also tested stemmed versions of all the models. For English, we did not use stemming, because it does not play a key role [7]. The results are shown in Tables 3 and 4.

Each model brings some improvement in all the cases. Also, the stemmed versions of the models are almost always better than the unstemmed models. Thus, we continued the experiments only with the stemmed models for Czech. The stems are used as a separate features and are seen to be very useful for Czech (see Table 6).

In the subsequent experiments, we tried to combine all the clusters from one model. We assumed that different clustering depths could bring useful information into the classifier. These combinations are shown in Table 5 for English and Table 6 for Czech. We can see that the performance was considerably improved. Taking these results into account, the best models for ABSA seem to be GloVe and CBOW.

To prevent overfitting, we cannot combine all the models and all the clustering depths together. Thus, we only combined the two best models (GloVe, CBOW). The results are shown again in Tables 5 and 6 in the last row. In all the subtasks, the performance stagnates or slightly improves.

Our English baseline extracts aspect terms with 75.6% F -measure and aspect categories with 77.6% F -measure. The Czech baseline is considerably worse, and achieves the results

Table 3. Aspect term and category extraction (TE, CE) and polarity (TP, CP) results on English dataset

Clusters Task	100		500		1000		5000	
	TE	TP	TE	TP	TE	TP	TE	TP
BL+HAL	79.3 (+3.7)	69.0 (+1.6)	78.7 (+3.1)	68.8 (+1.4)	78.3 (+2.7)	69.2 (+1.8)	78.6 (+3.0)	69.6 (+2.3)
BL+COALS	77.6 (+1.9)	67.5 (+0.1)	77.3 (+1.7)	67.7 (+0.3)	77.3 (+1.7)	67.5 (+0.1)	76.6 (+0.9)	68.8 (+1.5)
BL+CBOW	78.4 (+2.8)	69.4 (+2.0)	78.6 (+3.0)	69.4 (+2.0)	79.1 (+3.5)	69.3 (+2.0)	77.8 (+2.2)	69.8 (+2.4)
BL+SKIP	77.8 (+2.2)	69.9 (+2.5)	77.6 (+2.0)	68.2 (+0.9)	77.9 (+2.3)	68.7 (+1.4)	77.6 (+2.0)	68.3 (+0.9)
BL+GLOVE	77.5 (+1.9)	69.2 (+1.8)	77.6 (+2.0)	69.8 (+2.4)	77.7 (+2.0)	69.3 (+2.0)	77.1 (+1.5)	68.8 (+1.4)
BL+LDA	77.4 (+1.8)	68.5 (+1.1)	77.2 (+1.5)	68.6 (+1.2)	77.8 (+2.2)	68.7 (+1.3)	77.1 (+1.4)	68.9 (+1.5)
Clusters Task	100		500		1000		5000	
	CE	CP	CE	CP	CE	CP	CE	CP
BL+HAL	78.6 (+1.2)	68.4 (+0.2)	79.3 (+1.9)	69.0 (+0.7)	78.5 (+1.0)	69.7 (+1.5)	78.8 (+1.3)	69.5 (+1.2)
BL+COALS	78.2 (+0.7)	68.1 (-0.2)	78.8 (+1.3)	68.7 (+0.4)	78.1 (+0.6)	68.9 (+0.6)	77.9 (+0.4)	69.5 (+1.2)
BL+CBOW	78.8 (+1.3)	70.5 (+2.2)	79.3 (+1.8)	70.7 (+2.4)	79.0 (+1.6)	69.9 (+1.6)	78.7 (+1.3)	70.9 (+2.6)
BL+SKIP	78.4 (+0.9)	69.4 (+1.1)	78.1 (+0.7)	70.0 (+1.7)	78.7 (+1.3)	70.6 (+2.3)	79.3 (+1.8)	70.4 (+2.1)
BL+GLOVE	79.3 (+1.8)	69.8 (+1.5)	79.1 (+1.6)	70.1 (+1.8)	79.4 (+1.9)	70.4 (+2.1)	78.8 (+1.3)	69.9 (+1.6)
BL+LDA	78.4 (+0.9)	69.6 (+1.3)	78.5 (+1.0)	69.5 (+1.2)	78.6 (+1.1)	68.8 (+0.5)	77.7 (+0.3)	69.1 (+0.8)

Table 4. Aspect term and category extraction (TE, CE) and polarity (TP, CP) results on Czech dataset

Clusters Task	100		500		1000		5000	
	TE	TP	TE	TP	TE	TP	TE	TP
BL+HAL	75.6 (+4.2)	67.4 (+0.0)	75.4 (+4.0)	68.3 (+0.9)	75.5 (+4.0)	68.5 (+1.1)	75.0 (+3.5)	69.9 (+2.6)
BL+S-HAL	74.0 (+2.5)	66.8 (-0.5)	75.4 (+4.0)	68.5 (+1.1)	75.8 (+4.3)	69.4 (+2.0)	76.9 (+5.5)	70.3 (+2.9)
BL+COALS	75.1 (+3.6)	67.0 (-0.4)	74.5 (+3.0)	67.9 (+0.5)	74.6 (+3.2)	68.0 (+0.6)	74.3 (+2.8)	68.7 (+1.3)
BL+S-COALS	75.2 (+3.7)	69.5 (+2.1)	75.4 (+4.0)	69.2 (+1.9)	75.4 (+4.0)	68.8 (+1.4)	75.5 (+4.1)	69.5 (+2.1)
BL+CBOW	75.4 (+3.9)	68.2 (+0.9)	75.4 (+3.9)	69.7 (+2.3)	75.7 (+4.3)	70.5 (+3.1)	75.3 (+3.9)	70.1 (+2.7)
BL+S-CBOW	75.8 (+4.3)	69.6 (+2.2)	74.9 (+3.5)	70.4 (+3.0)	75.6 (+4.2)	70.1 (+2.7)	73.2 (+1.8)	71.1 (+3.7)
BL+SKIP	74.9 (+3.4)	69.4 (+2.0)	74.8 (+3.3)	70.2 (+2.8)	75.9 (+4.5)	70.8 (+3.4)	74.8 (+3.3)	69.4 (+2.0)
BL+S-SKIP	75.4 (+4.0)	69.6 (+2.2)	75.3 (+3.8)	70.6 (+3.2)	75.9 (+4.5)	69.7 (+2.3)	75.5 (+4.0)	69.9 (+2.5)
BL+GLOVE	74.3 (+2.8)	68.9 (+1.5)	75.4 (+4.0)	69.1 (+1.7)	75.6 (+4.2)	68.7 (+1.3)	74.4 (+3.0)	69.1 (+1.7)
BL+S-GLOVE	75.1 (+3.7)	69.0 (+1.6)	76.3 (+4.9)	70.0 (+2.6)	76.0 (+4.6)	69.5 (+2.1)	75.5 (+4.0)	69.6 (+2.2)
BL+LDA	74.4 (+2.9)	68.9 (+1.5)	74.5 (+3.0)	69.6 (+2.3)	73.7 (+2.3)	69.2 (+1.8)	73.1 (+1.7)	69.2 (+1.9)
BL+S-LDA	74.7 (+3.3)	69.2 (+1.8)	74.8 (+3.4)	70.2 (+2.8)	74.6 (+3.1)	69.5 (+2.1)	74.7 (+3.2)	69.8 (+2.4)
Clusters Task	100		500		1000		5000	
	CE	CP	CE	CP	CE	CP	CE	CP
BL+HAL	76.1 (+4.3)	68.9 (-0.8)	74.7 (+3.0)	71.1 (+1.4)	74.6 (+2.9)	70.8 (+1.1)	74.1 (+2.4)	71.7 (+2.0)
BL+S-HAL	75.4 (+3.7)	70.1 (+0.4)	75.6 (+3.8)	71.1 (+1.4)	75.2 (+3.5)	70.7 (+1.0)	76.6 (+4.9)	73.2 (+3.5)
BL+COALS	75.1 (+3.4)	69.9 (+0.2)	74.5 (+2.8)	72.5 (+2.8)	74.1 (+2.3)	70.9 (+1.2)	73.5 (+1.7)	70.9 (+1.2)
BL+S-COALS	75.0 (+3.2)	71.8 (+2.1)	75.7 (+3.9)	72.8 (+3.1)	74.8 (+3.1)	71.4 (+1.7)	74.8 (+3.1)	72.6 (+2.9)
BL+CBOW	74.8 (+3.0)	71.3 (+1.6)	74.9 (+3.2)	72.3 (+2.6)	74.4 (+2.7)	72.5 (+2.8)	74.6 (+2.8)	73.0 (+3.3)
BL+S-CBOW	76.2 (+4.5)	72.1 (+2.4)	74.0 (+2.3)	73.1 (+3.4)	75.3 (+3.6)	73.8 (+4.1)	75.3 (+3.5)	73.9 (+4.2)
BL+SKIP	75.6 (+3.9)	73.6 (+3.9)	74.7 (+3.0)	73.2 (+3.5)	75.9 (+4.1)	74.1 (+4.4)	74.1 (+2.3)	72.3 (+2.6)
BL+S-SKIP	75.9 (+4.1)	73.3 (+3.6)	74.5 (+2.8)	74.1 (+4.4)	76.1 (+4.4)	73.0 (+3.3)	75.6 (+3.8)	73.3 (+3.6)
BL+GLOVE	76.5 (+4.7)	70.9 (+1.2)	75.5 (+3.7)	71.7 (+2.0)	75.8 (+4.1)	72.1 (+2.4)	73.3 (+1.6)	71.9 (+2.2)
BL+S-GLOVE	77.2 (+5.5)	70.9 (+1.2)	77.1 (+5.4)	73.5 (+3.8)	77.4 (+5.7)	73.4 (+3.7)	75.9 (+4.2)	73.2 (+3.5)
BL+LDA	73.3 (+1.5)	72.3 (+2.6)	73.9 (+2.2)	72.9 (+3.2)	73.9 (+2.1)	72.8 (+3.1)	73.1 (+1.3)	72.0 (+2.3)
BL+S-LDA	73.4 (+1.7)	72.0 (+2.3)	74.4 (+2.7)	73.4 (+3.7)	74.1 (+2.3)	73.0 (+3.3)	74.2 (+2.4)	72.6 (+2.9)

Table 5. Models combinations on English dataset

Task	TE	TP	CE	CP
BL	75.6	67.4	77.5	68.3
BL+HAL	80.3 (+4.6)	70.6 (+3.2)	79.5 (+2.0)	69.5 (+1.3)
BL+COALS	78.7 (+3.0)	69.0 (+1.6)	78.6 (+1.1)	69.2 (+0.9)
BL+CBOW	80.6 (+5.0)	71.1 (+3.7)	79.3 (+1.8)	71.4 (+3.2)
BL+SKIP	78.9 (+3.2)	69.9 (+2.5)	79.6 (+2.1)	70.8 (+2.6)
BL+GLOVE	78.7 (+3.0)	70.2 (+2.8)	79.5 (+2.1)	70.8 (+2.5)
BL+LDA	78.5 (+2.9)	69.8 (+2.4)	78.4 (+0.9)	70.0 (+1.8)
BL+CBOW+GLOVE	80.4 (+4.8)	70.9 (+3.5)	80.6 (+3.1)	72.1 (+3.8)

Table 6. Model combinations on Czech dataset

Task	TE	TP	CE	CP
BL	71.4	67.4	71.7	69.7
BL+S-BL	74.9 (+3.4)	69.0 (+1.6)	73.6 (+1.9)	71.3 (+1.6)
BL+S-BL+S-HAL	78.5 (+7.0)	70.5 (+3.1)	78.5 (+6.8)	72.3 (+2.6)
BL+S-BL+S-COALS	77.8 (+6.3)	70.9 (+3.6)	77.5 (+5.7)	73.1 (+3.4)
BL+S-BL+S-CBOW	77.9 (+6.4)	72.1 (+4.7)	78.1 (+6.4)	73.6 (+3.9)
BL+S-BL+S-SKIP	77.8 (+6.3)	71.6 (+4.3)	78.0 (+6.3)	75.2 (+5.5)
BL+S-BL+S-GLOVE	78.5 (+7.1)	71.3 (+3.9)	79.5 (+7.8)	74.1 (+4.4)
BL+S-BL+S-LDA	77.4 (+6.0)	70.2 (+2.9)	75.6 (+3.8)	73.4 (+3.7)
BL+S-BL+S-CBOW+S-GLOVE	78.7 (+7.3)	72.5 (+5.1)	80.0 (+8.3)	74.0 (+4.3)

71.4% and 71.7% *F*-measures in the same subtasks. The behaviour of our baselines for sentiment polarity tasks is different. The baselines for aspect term polarity and aspect category polarity in both languages perform almost the same: the accuracy ranges between 67.4% and 69.7% for both languages.

In our experiments, the word clusters from semantic spaces (especially CBOW and GloVe models) and stemming by HPS proved to be very useful. Large improvements were achieved for all four subtasks and both languages.

The aspect term extraction and aspect category extraction *F*-measures of our systems improved to approximately 80% for both languages. Similarly, the polarity detection subtasks surpassed 70% accuracy, again for both languages.

7 Summary

In this paper, we explored several unsupervised methods for word meaning representation. We created word clusters and used them as features for the ABSA task. We achieved considerable improvements for both the English and Czech languages. We also used the unsupervised stemming algorithm called HPS, which helped us to deal with the rich morphology of Czech.

Out of all the tested models, GloVe and CBOW seem to perform the best, and their combination together with stemming for Czech was able to improve all four ABSA subtasks. To the best of our knowledge, these results are now the state of the art for Czech.

We created two new Czech corpora within the restaurant domain for the ABSA task: one labeled for supervised training, and the other (considerably larger) unlabeled for unsupervised training. The corpora are available to the research community.

Since none of the methods used to improve ABSA in this paper require any external information about the language, we assume that similar improvements can be achieved for other languages. Thus, the main direction for future research is to experiment with more languages from different language families.

Acknowledgements

This work was supported by the project LO1506 of the Czech Ministry of Education, Youth and Sports, by Grant No. SGS-2016-018 Data and Software Engineering for Advanced Applications and by project MediaGist, EUs FP7 People Programme (Marie Curie Actions), no. 630786. Computational resources were provided by the CESNET LM2015042 and the CERIT Scientific Cloud LM2015085, provided under the programme "Projects of Large Research, Development, and Innovations Infrastructures".

References

1. Blei, D. M., Ng, A. Y., Jordan, M. I., & Lafferty, J. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, Vol. 3.
2. Brychcín, T. & Habernal, I. (2013). Unsupervised improving of sentiment analysis using global target context. *Proceedings of the International Conference Recent Advances in Natural Language Processing RANLP 2013*, INCOMA Ltd., Shoumen, Bulgaria, pp. 122–128.
3. Brychcín, T., Konkol, M., & Steinberger, J. (2014). UWB: Machine learning approach to aspect-based sentiment analysis. *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, Association for Computational Linguistics, pp. 817–822.
4. Brychcín, T. & Konopík, M. (2015). HPS: High precision stemmer. *Information Processing & Management*, Vol. 51, No. 1, pp. 68–91.
5. Brychcín, T. & Král, P. (2014). Novel unsupervised features for Czech multi-label document classification. *13th Mexican International Conference on Artificial Intelligence (MICAI 2014)*, Springer-Verlag, Berlin. Accepted.
6. Firth, J. R. (1957). A Synopsis of Linguistic Theory, 1930–1955. *Studies in Linguistic Analysis*, pp. 1–32.
7. Habernal, I., Ptáček, T., & Steinberger, J. (2014). Supervised sentiment analysis in Czech social media. *Information Processing & Management*, Vol. 50, No. 5, pp. 693–707.
8. Habernal, I., Ptáček, T., & Steinberger, J. (2013). Sentiment analysis in czech social media using supervised machine learning. *Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, Association for Computational Linguistics, Atlanta, GA, USA, pp. 65–74.
9. Hercig, T., Brychcín, T., Svoboda, L., & Konkol, M. (2016). Uwb at semeval-2016 task 5: Aspect based sentiment analysis. *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, Association for Computational Linguistics, San Diego, California, pp. 354–361.
10. Jurgens, D. & Stevens, K. (2010). The S-Space package: An open source package for word space models. *System Papers of the Association of Computational Linguistics*.
11. Karypis, G. (2003). CLUTO—A clustering toolkit.
12. Konkol, M. (2014). Brainy: A machine learning library. In Rutkowski, L., Korytkowski, M., Scherer, R., Tadeusiewicz, R., Zadeh, L. A., & Zurada, J. M., editors, *Artificial Intelligence and Soft Computing*, volume 8468 of *Lecture Notes in Computer Science*. Springer-Verlag, Berlin.
13. Konkol, M., Brychcín, T., & Konopík, M. (2015). Latent semantics in named entity recognition. *Expert Systems with Applications*, Vol. 42, No. 7, pp. 3470–3479.
14. Konkol, M. & Konopík, M. (2013). CRF-based Czech named entity recognizer and consolidation of Czech NER research. In Habernal, I. & Matoušek, V., editors, *Text, Speech and Dialogue*, volume 8082 of *Lecture Notes in Computer Science*. Springer-Verlag, Berlin, pp. 153–160.
15. Lafferty, J. D., McCallum, A., & Pereira, F. C. N. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. *Proceedings of the Eighteenth International Conference on Machine Learning*, Morgan Kaufmann Publishers, San Francisco, CA, USA, pp. 282–289.
16. Lund, K. & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods Instruments and Computers*, Vol. 28, No. 2, pp. 203–208.

17. **McCallum, A. K. (2002).** Mallet: A machine learning for language toolkit.
18. **Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013).** Efficient estimation of word representations in vector space. *CoRR*, Vol. abs/1301.3781.
19. **Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013).** Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, pp. 3111–3119.
20. **Pennington, J., Socher, R., & Manning, C. D. (2014).** Glove: Global vectors for word representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1532–1543.
21. **Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., AL-Smadi, M., Al-Ayyoub, M., Zhao, Y., Qin, B., Clercq, O. D., Hoste, V., Apidianaki, M., Tannier, X., Loukachevitch, N., Kotelnikov, E., Bel, N., Jiménez-Zafra, S. M., & Eryiğit, G. (2016).** SemEval-2016 task 5: Aspect based sentiment analysis. *Proceedings of the 10th International Workshop on Semantic Evaluation*, SemEval '16, Association for Computational Linguistics, San Diego, California.
22. **Pontiki, M., Galanis, D., Papageorgiou, H., Manandhar, S., & Androutsopoulos, I. (2015).** Semeval-2015 task 12: Aspect based sentiment analysis. *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, ACL, Denver, Colorado, pp. 486–495.
23. **Pontiki, M., Galanis, D., Pavlopoulos, J., Papageorgiou, H., Androutsopoulos, I., & Manandhar, S. (2014).** SemEval-2014 Task 4: Aspect based sentiment analysis. *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, ACL and Dublin City University, Dublin, Ireland, pp. 27–35.
24. **Rohde, D. L. T., Gonnerman, L. M., & Plaut, D. C. (2004).** An improved method for deriving word meaning from lexical co-occurrence. *Cognitive Psychology*, Vol. 7, pp. 573–605.
25. **Steinberger, J., Brychcín, T., & Konkol, M. (2014).** Aspect-level sentiment analysis in czech. *Proceedings of the 5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, Association for Computational Linguistics, Baltimore, MD, USA.
26. **Tamchyna, A., Fiala, O., & Veselovská, K. (2015).** Czech aspect-based sentiment analysis: A new dataset and preliminary results. *Proceedings of the 15th conference ITAT 2015*, pp. 95–99.
27. **Veselovská, K., Hajíč Jr., J., & Šindlerová, J. (2012).** Creating annotated resources for polarity classification in Czech. *Proceedings of the 11th Conference on Natural Language Processing*, ÖGAI.

Tomáš Hercig received his master degree in Computer Science from the University of West Bohemia, Czech Republic, in 2012. He is currently finishing his Ph.D. at the University of West Bohemia and works as researcher at the New Technologies for the Information Society (NTIS) Research Centre at the same university. His research is oriented on natural language processing, especially sentiment analysis.

Tomáš Brychcín finished his Ph.D. in computer science at the University of West Bohemia in 2015. Now he works as researcher at the New Technologies for the Information Society (NTIS) Research Centre at the same university. His research is oriented in natural language processing with special focus on distributional semantics, language modeling and sentiment analysis. He is also interested in NPC problems, cryptography and automatic speech recognition. He is co-founder of startup SentiSquare analyzing user generated data.

Lukáš Svoboda is currently Ph.D. student at University of West Bohemia. He received his master degree in computer science from the same university in 2013. His research is oriented on natural language processing, especially usage of recursive neural network in NLP. He is also interested in machine learning algorithms, artificial intelligence and embedded systems.

Michal Konkol is a researcher at University of West Bohemia. He received his master degree in computer science in 2010 and his Ph.D. in 2016, both at UWB. His research focuses on machine learning and natural language processing. He is

also a co-founder of SentiSquare, a startup which analyses customer-written texts.

Josef Steinberger works as an associate professor at the University of West Bohemia, where he defended the habilitation thesis in 2013. Before, he spent three years in the Europe Media Monitor team at the Joint Research Centre (JRC) of the

European Commission. His research focuses mainly on summarisation and sentiment analysis. He is also a co-founder of SentiSquare, a startup which analyses customer-written texts.

Article received on 22/12/2015; accepted on 02/03/2016.
Corresponding author is Tomáš Hercig.