University of West Bohemia in Pilsen
Department of Computer Science and Engineering
Univerzitni 8
30614 Pilsen
Czech Republic

Aspects of Sentiment Analysis

PhD Study Report

Tomáš Hercig

Technical Report No. DCSE/TR-2015-04
May, 2015

Distribution: public

Technical Report No. DCSE/TR-2015-04
May 2015

Aspects of Sentiment Analysis

Tomáš Hercig

Abstract

This report introduces the task of sentiment analysis, describes the core problems and presents the formal definition of sentiment analysis. The basic machine learning algorithms for text classification are described as well as the most commonly used features for sentiment analysis. Brief overview of distributional semantics is presented. Related work and the state-of-the-art approaches to sentiment analysis are thoroughly described and sorted by the granularity level of sentiment analysis. Great emphasis is on the sentiment analysis in the Czech environment.

Copies of this report are available on
http://www.kiv.zcu.cz/publications/
or by surface mail on request sent to the following address:

University of West Bohemia in Pilsen
Department of Computer Science and Engineering
Univerzitni 8
30614 Pilsen
Czech Republic

Contents

Chapter 1

Introduction

Sentiment analysis is a sub-field of natural language processing and employs machine learning, computational linguistics and data mining. Generally, it deals with the automatic extraction and analysis of sentiments, opinions, emotions and beliefs expressed in written text.

Sentiment analysis has become a mainstream research field since the early 2000s. Its impact can be seen in many practical applications, ranging from analysing product reviews [Stepanov and Riccardi, 2011] to predicting sales and stock markets using social media monitoring [Yu et al., 2013]. The users' opinions are mostly extracted either on a certain polarity scale, or binary (positive, negative); various levels of granularity are also taken into account, e.g., document-level, sentence-level, or aspect-based sentiment [Hajmohammadi et al., 2012].

Most of the research in automatic sentiment analysis of social media has been performed in English and Chinese, as shown by several recent surveys [Liu and Zhang, 2012, Tsytsarau and Palpanas, 2012].

The goal of sentiment analysis is to automatically detect the polarity of a text. The emphasis should be on the word automatically as the task has a particular focus on supervised and unsupervised machine learning.

If we understand the meaning (semantics) of a text, we will also uncover the sentiment hidden in the text. We believe that distributional semantics models are essential to understand the meaning and sentiment hidden in text.

1.1 Motivation

There are many researchers trying to surpass the latest best results and achieve the state-of-the-art in English sentiment analysis by using handcrafted features. This approach may result into overfitting the data. However, sentiment analysis in Czech has not yet been thoroughly targeted by the research community.

Czech as a representative of a inflective language is an ideal environment for the study of various aspects of sentiment analysis (overview or breadth study of sentiment analysis if you will) for inflectional languages. It is challenging because of its very flexible word order and many different word forms.

We conceive this study to deal with several aspects of sentiment analysis. The breadth of this study can lead to more general view and better understanding of sentiment analysis. We can reveal and overcome unexpected obstacles, create necessary evaluation datasets and even come up with new creative solutions to sentiment analysis tasks.

Thus the aim of the doctoral thesis is to study various aspects of sentiment analysis with the emphasis on the Czech language.

1.2 Outline

Chapter 2 describes the challenges in sentiment analysis and formulates the basic and aspect-based definitions.

It is necessary to define the state-of-the-art techniques and evaluation measures before some results are presented, thus Chapter 3 is devoted to machine learning techniques and evaluation measures. The most commonly used features for sentiment analysis are covered in Chapter 4. The features seems to have at least the same importance as the methods.

Distributional semantic models are introduced in Chapter 5. Semantics models can be used as additional sources of information for sentiment analysis classification.

The related work for sentiment analysis is presented in Chapter 6.

Chapter 7 summarizes the challenges of sentiment analysis and states the aims of the doctoral thesis.

Chapter 2

Sentiment Analysis

This chapter describes the core problems of the current state-of-the-art algorithms and present the formal definition of sentiment analysis.

Sentiment analysis in general is connected to not only to opinions but to emotions, feelings and attitudes as well. Sentiment polarity is only a part of this field which assigns a sentiment label (e.g. positive, negative and neutral) to texts. In this report we will mainly focus on the sentiment polarity task.

2.1 Challenges

A positive or negative sentiment word may have opposite orientations in different application domains. The word *"loud"* is generally negative (*"the fan is very loud"*) however in a certain situation it can be positive, e.g. *"wow the speakers are really loud"*.

A sentence containing sentiment words may not express any sentiment. This frequently happens in questions and conditional sentences, e.g. *"Could you tell me which printer is the best?"* and *"If I can find a good laptop in the shop, I will buy it."* Both these sentences contain a positive sentiment word, but neither expresses a positive or negative opinion on any specific product. However, not all questions and conditional sentences express no sentiments, e.g., *"Does anyone know how to get this terrible camera to work?"*.

Other aspects of subjective texts related to sentiment can be considered important as well. Various emotions such as anger, fear, disgust, happiness, sadness and surprise can be extracted from affected texts in order to determine the state of mind of the author. This affected state can be later used to switch to a different mode of sentiment interpretation or hateful posts filtering in forums.

Sarcastic sentences with or without sentiment words are hard to deal with, e.g., *"What a great car! It stopped working in two days."* Sarcasm will be discussed in more detail in Section 2.4.

Many sentences without sentiment words can also imply opinions. These sentences usually express some factual information in an objective manner. The sentence *"This printer uses a lot of ink"* implies a negative sentiment about the printer since it uses a lot of resource (ink). This sentence is objective as it states a fact.

Unlike factual information, opinions and sentiments have an important characteristic, namely, they are subjective. Single opinion from one person represents only the subjective view of that single person. It is thus important to examine a collection of opinions from many people rather than only a single person. Since product reviews are highly focused with little irrelevant information and opinion rich, they allow us to see different issues more clearly than from other forms of opinion text.

Twitter postings (tweets) are short (at most 140 characters) and informal, and use many Internet slangs and emoticons. Twitter postings are easier to analyse due to the length limit because the authors are usually straight to the point, but you have to deal with the Twitter specific slang.[Liu, 2012]

Forum discussions are perhaps the hardest to deal with because the users there can discuss anything and also interact with one another. Different application domains are also considered very difficult to deal with. Social and political discussions are much harder than opinions about products and services, due to complex topic and sentiment expressions.[Liu, 2012]

The task of aggregating and representing sentiment of a document or majority of documents is called sentiment summarization. Since the amount of information available on the Internet is huge a brief overview of market sentiment can be very helpful for both customers and producers. The automatic summarization should be unbiased, quicker and accurate, unlike humans. Moreover the average human reader could have considerable difficulty doing the same.

There are even individuals or organizations who give fake opinions in reviews and forum discussions to promote or to discredit target products, services, organizations, or individuals. Such individuals are called opinion spammers and the fake opinions are called opinion spam. Opinion spamming has become a major issue. There is no easy way to detect these fake opinions.

These issues all present major challenges. In fact, these are just some of the difficult problems.

2.2 Basic Sentiment Polarity Definition

An opinion is a quadruple (G, S, H, T) [Liu, 2012], where

- G is the sentiment target,
- S is the sentiment about the target,
- H is the opinion holder,
- T is the time when the opinion was expressed.

Sentiment analysis can be done on different levels of granularity.

- **Document level** is usually used on various reviews, where the task is to determine the overall sentiment towards the target (e.g. product or movie).
- **Sentence level** analyses the overall sentiment of a sentence.
- **Aspect-based** sentiment analysis focuses on the precise features (aspects) of the sentiment target. Both the document and sentence level of sentiment analysis fail to understand exactly which aspect of the target is branded by the opinion holder with the given sentiment. Aspect-based sentiment analysis will be discussed in Sections 2.3 and 6.3.
- **Word level** of sentiment analysis identifies the polarity of words. For more information see Section 6.2.

Let us use the term entity to denote the target object that has been evaluated.

An entity is a product, service, topic, issue, person, organization, or event. It is described with a pair, hierarchy of parts, sub-parts, and so

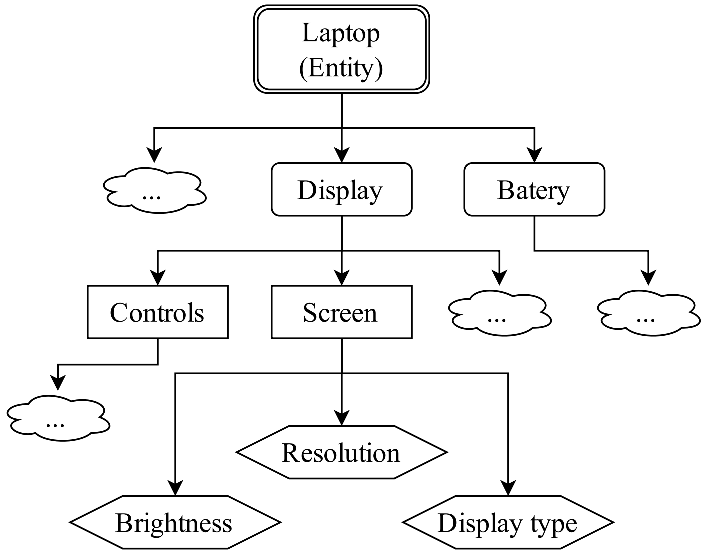


Figure 2.1: Example entity (laptop), its parts (rounded rectangle), sub-parts (rectangle) and attributes (hexagon). Clouds represent omitted hierarchical structures.

on, and a set of attributes. Each part or sub-part also has its own set of attributes [Liu, 2012]. Figure 2.1 shows an example of such hierarchy.

This entity as a hierarchy of any number of levels needs a nested relation to represent it. Recognizing parts and attributes of an entity at different levels of details is extremely hard, fortunately most applications do not need such complex analysis. Thus, we simplify the hierarchy to two levels and use the term aspects to denote both parts and attributes. In the simplified tree, the root node is still the entity itself, but the second level (also the leaf level) nodes are different aspects of the entity. This simplified framework (figure 2.2) is what is typically used in practical sentiment analysis systems. Note that in the research literature, entities are also called objects, and aspects are also called features (or product features).

2.3 Aspect-Based Sentiment Definition

An opinion is a quintuple $(E_i, A_{ij}, S_{ijkl}, H_k, T_l)$ [Liu, 2012], where

- E_i is the name of an entity,

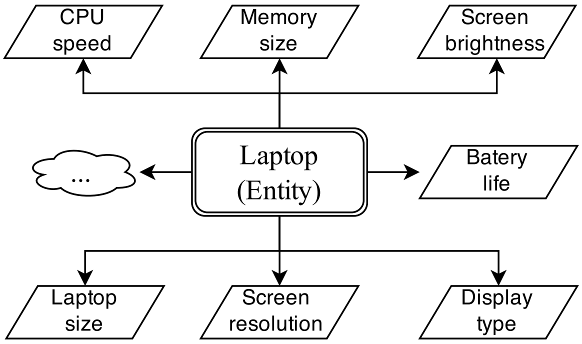


Figure 2.2: Example entity (laptop) and its aspects (rhomboids). Cloud represents omitted aspects.

- A_{ij} is an aspect of E_i,
- S_{ijkl} is the sentiment about aspect A_{ij} of entity E_i expressed by H_k at the time T_l,
- H_k is the opinion holder,
- T_l is the time when the opinion is expressed by H_k.

The entity E_i and its aspects A_{ij} together represent the opinion target. The sentiment S_{ijkl} is positive, negative, or neutral, or expressed on a certain polarity scale, e.g., 1 to 5 stars as used by most review sites. Special aspect GENERAL is used to denote an opinion on the entity itself as a whole.

In this definition, subscripts are used to emphasize that the five pieces of information in the quintuple must correspond to one another. That is, the opinion S_{ijkl} must be given by opinion holder H_k about aspect A_{ij} of entity E_i at time T_l. Each of these five components is essential and any mismatch is problematic in general.

For example, in the sentence *"The English adore him but the Spanish hate him."*, it is clearly important to distinguish between the two opinion holders. The time component may seem not very important, but in practise an opinion expressed two years ago is not the same as an opinion expressed yesterday.

The definition does not cover all possible ways to express an opinion. The definition would be too complex if it did and thus make the problem extremely difficult to solve. However the definition is sufficient for most applications.

The limits of this simplification are evident e.g. in the case of a comparative opinion. Comparative opinion expresses a relation of similarities or differences between two or more entities and/or a preference of the opinion holder based on some shared aspects of the entities. [Liu, 2012]

There are other situations in which a more complex definition would be needed. For example, the situation in "This car is too small for a tall person," which does not say the car is too small for everyone. The context of the opinion is an important information, which is not covered in the simplified definition.

Furthermore, we simplified the hierarchical structure of entity. If we want to study different aspects of an aspect (e.g. phone battery and its price and capacity), then we need to treat an aspect (battery) of an entity (phone) as a separate entity.

Definition from Semantic Evaluation Workshop

The semantic evaluation workshop *SemEval* is an important series of workshops studying sentiment. There are several ways to define aspects and polarities. The SemEval2014's [Pontiki et al., 2014] definition distinguishes two types of aspect-based sentiment: aspect terms and aspect categories. The whole task is divided into four subtasks. Figure 2.3 gives examples for each subtask.

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.

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: Děti dostaly naprosto krvavé maso.
EN: The meat they brought to the kids was totally bloody.
→ {maso (meat)}
```

(a) Aspect term extraction

```
CZ: Děti dostaly naprosto krvavé maso.
EN: The meat they brought to the kids was totally bloody.
→ {maso (meat): negative}
```

(b) Aspect term polarity

```
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.
→ {služby (service), prostředí (ambience)}
```

(c) Aspect category detection

```
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.
→ {služby (service): positive, prostředí (ambience): negative}
```

(d) Aspect category polarity

Figure 2.3: Subtasks examples of aspect-based sentiment analysis.

Subtask 3: Aspect Category Detection

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.

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.

2.4 Sarcasm Detection

Since the goal of sentiment analysis is to automatically detect the polarity of a text, misinterpreting irony and sarcasm represents a big challenge [Davidov et al., 2010].

As there is only a weak boundary in meaning between irony, sarcasm and satire [Reyes et al., 2012], we will use only the term. Sarcasm generally reverses the polarity of an utterance from positive or negative into its opposite, which deteriorates the results of a given NLP task. Therefore, correct identification of sarcasm can improve the performance. Bosco et al. [2013] claim that *"even if there is no agreement on a formal definition of irony, psychological experiments have delivered evidence that humans can reliably identify ironic text utterances from an early age in life"*.

2.5 Sentiment Analysis for Inflectional Languages

Highly inflectional languages such as Czech are hard to deal with because of the high number of different word forms. Czech is even more challenging because it has very flexible word order. Czech language permits and frequently uses double even a triple negative in one sentence, thus making it difficult for computers to understand the meaning of the sentence. Moreover the subject can be omitted if it is known from the context.

Text is often preprocessed by various techniques in order to reduce the dictionary size. The importance of this preprocessing phase depends on the language. For highly inflectional languages like Czech, *stemming* or *lemmatization* is almost mandatory because it is necessary to reduce the high number of different word forms.

Lemmatization identifies the base or dictionary form of a word which is known as the *lemma*.

Stemming finds the base form of each word, usually by removing all affixes. The result of *stemming* is called *stem*

Sometimes a list of stop words is used to filter out words which occur in most documents and have only a small impact on the results.

2.6 Evaluation Criteria

The performance of methods used for sentiment analysis is evaluated by calculating various metrics like accuracy, precision, recall and F-measure (also F-score or F_1 score).

We will define these measures on a binary classification of positive and negative labels, but in general any number of labels can be used. We can show the results in the form of a confusion matrix.

- Positive (P) - positive text classified as positive.
- Negative (N) - negative text classified as negative.
- False positive (FP)- negative text classified as positive.
- False negative (FN) - positive text classified as negative.

	Positive	**Negative**
classified as positive	positive (P)	false positive (FP)
classified as negative	false negative (FN)	negative (N)

Table 2.1: Confusion matrix.

Now we can easily define accuracy, precision, recall and F-measure as follows.

$$\text{Accuracy} = \frac{P + N}{P + N + FP + FN} \tag{2.1}$$

$$\text{Precision} = \frac{P}{P + FP} \tag{2.2}$$

$$\text{Recall} = \frac{P}{P + FN} \tag{2.3}$$

$$\text{F-measure} = \frac{2P}{2P + FP + FN} \tag{2.4}$$

Accuracy is a proportion of all correctly predicted labels compared to all sentences. Precision is a measure of trust, that the objects marked as positive are really positive. Recall is a measure of trust, that all the positive objects are marked. F-measure is a harmonic mean between precision and recall and it is considered to be an overall perspective.

Figure 2.4 shows the distribution of positive and negative objects. The dashed line represents the decision threshold of classifier. The areas marked as FN and FP contain incorrectly classified objects.

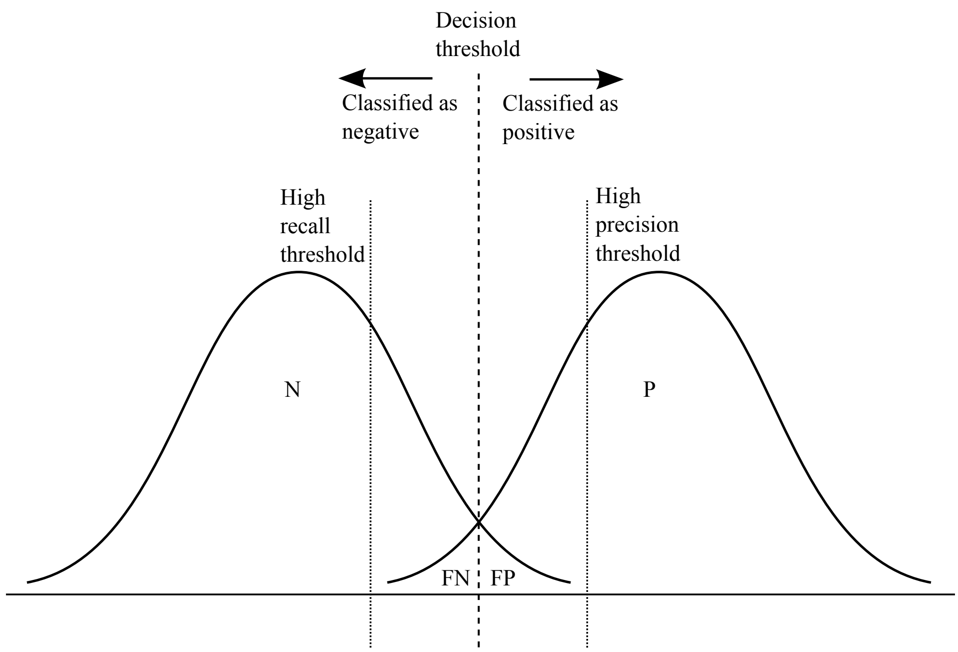


Figure 2.4: Precision and recall

Precision and recall compete against each other as shown on figure 2.4, the dotted lines represent the decision threshold for high recall or high precision. For example if the decision threshold is moved to the left, there will be fewer FN objects and more FP objects, resulting in high recall and lower precision. A high recall (or precision) classifier can be more suitable for various tasks. The commonly used evaluation metric is the harmonic mean between precision and recall usually called F-measure.

Chapter 3

Machine Learning

Sentiment analysis can be treated as a text classification problem. The standard approach is to classify a document as being positive or negative using a machine learning algorithm (classifier). The performance of sentiment analysis is strongly dependant on the applied classifier.

Machine learning algorithms essentially learn and store characteristics of a category from the data during a training phase. This is achieved by observing the properties of the annotated training data. The acquired knowledge is later applied to determine the best category for the unseen testing dataset. The training and testing datasets are both annotated by sentiment labels. Then depending on the data-size various model validation techniques can be used. *Cross-validation* is commonly used for sentiment analysis evaluations. The annotated dataset is split into k equal parts, then the first part is treated as the testing data and the rest as training data, this selection process is repeated for each of the parts. Each part is used exactly once as the testing data.

The de-facto standard for sentiment analysis is the Maximum Entropy classifier and Support Vector Machines (SVM) classifier, however a simple Naive Bayes classifier is often used as a baseline for evaluation.

3.1 Naive Bayes Classifier

The Naive Bayes (NB) classifier is a simple classifier commonly used as a baseline for many tasks. The model computes the posterior probability of a sentiment label based on predefined features in a given text as shown in equation 3.1, where s is the sentiment label and x is the given text.

$$P(s|x) = \frac{P(x|s)P(s)}{P(x)} \tag{3.1}$$

$$\hat{s} = \underset{s \in S}{\operatorname{argmax}}\ P(s)P(x|s) \tag{3.2}$$

The NB classifier is described by equation 3.2, where $\hat{s}$ is the assigned sentiment label. The NB classifier makes the decision based on the maximum a posteriori rule. In other words it picks the sentiment label that is the most probable. The NB classifier makes label conditional independence assumption.

3.2 Maximum Entropy Classifier

The Maximum Entropy (MaxEnt) classifier is based on the Maximum Entropy principle. The principle says that we are looking for a model which will satisfy all our constraints in the most general way (maximum entropy). To define a constraint we firstly need to define a feature. A feature is typically a binary function[1]. For example, consider the following dictionary feature designed to capture positive emoticons in the given text x.

$$f(x,s) = \begin{cases} 1 & \text{if } s \text{ is positive and } x \text{ contains a positive emoticon} \\ 0 & \text{otherwise} \end{cases} \tag{3.3}$$

The constraint is then defined as equality of mean values for a given feature.

$$E_p(f_i(x,s)) = E_{\tilde{p}}(f_i(x,s)) \tag{3.4}$$

$E_{\tilde{p}}(f_i(x,y))$ is the mean value of a feature computed over the training data and $E_p(f_i(x,y))$ is the mean value of the model. It is guaranteed that such a model exists, it is unique and follows the maximum-likelihood distribution (equation 3.5)[Berger et al., 1996].

$$p(s|x) = \frac{1}{Z(x)} \exp \sum_i \lambda_i f_i(x,s) \tag{3.5}$$

[1] In general any non-negative function can be used.

$f_i(x, s)$ is a feature and λ_i is a parameter to be estimated. $Z(x)$ is just a normalizing factor and ensures that $p(s|x)$ is a probability distribution.

$$Z(x) = \sum_s \exp \sum_i \lambda_i f_i(x, s) \tag{3.6}$$

Various training algorithms can be used for finding appropriate parameters. Limited memory BFGS (L-BFGS) method [Nocedal, 1980] proved very good performance.

3.3 SVM Classifier

Support Vector Machines (SVM) is a machine learning method based on vector spaces, where the goal is to find a decision boundary between two classes that represents the maximum margin of separation in the training data [Manning et al., 2008b].

SVM can construct a non-linear decision surface in the original feature space by mapping the data instances non-linearly to an inner product space where the classes can by separated linearly with a hyperplane.

Support Vector Machines

Following the original description [Cortes and Vapnik, 1995] we describe the principle in the simplest possible way. We will assume only binary classifier for classes $y = -1, 1$ and linearly separable training set $\{(x_i, y_i)\}$, so that the conditions 3.7 are met.

$$\begin{array}{rcrlrr} \mathbf{w} \cdot \mathbf{x_i} + b & \leq & -1 & \quad \text{if} & y_i = & -1 \\ \mathbf{w} \cdot \mathbf{x_i} + b & \geq & 1 & \quad \text{if} & y_i = & 1 \end{array} \tag{3.7}$$

Equation 3.8 combines the conditions 3.7 into one set of inequalities.

$$y_i \cdot (\mathbf{w}_0 \cdot \mathbf{x} + b_0) \geq 1 \quad \forall i \tag{3.8}$$

SVM search the optimal hyperplane (equation 3.9) that separates both classes with the maximal margin. The formula 3.10 measures the distance between the classes in the direction given by $\mathbf{w}$.

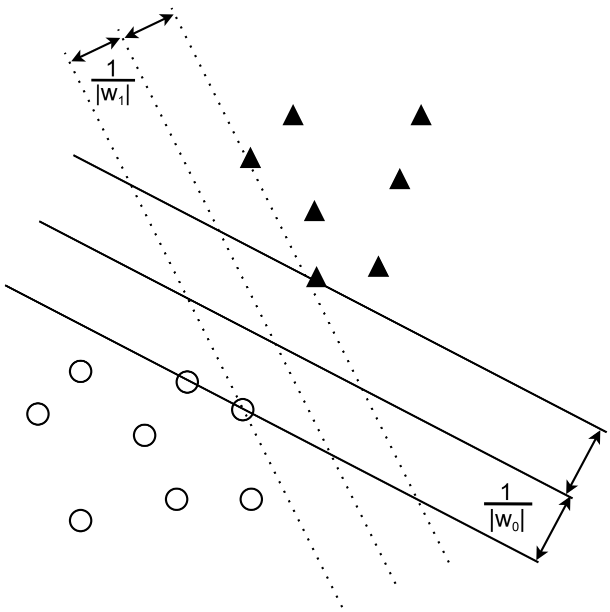


Figure 3.1: Optimal and suboptimal hyperplanes.

$$\mathbf{w}_0 \cdot \mathbf{x} + b_0 = 0 \tag{3.9}$$

$$d(\mathbf{w}, b) = \min_{x;y=1} \frac{\mathbf{x} \cdot \mathbf{w}}{|\mathbf{w}|} - \max_{x;y=-1} \frac{\mathbf{x} \cdot \mathbf{w}}{|\mathbf{w}|} \tag{3.10}$$

The optimal hyperplane, expressed in equation 3.11, maximizes the distance $d(\mathbf{w}, b)$. Therefore the parameters $\mathbf{w}_0$ and b_0 can be found by maximizing $|\mathbf{w}_0|$. For better understanding see the optimal and suboptimal hyperplanes on figure 3.1.

$$d(\mathbf{w}_0, b_0) = \frac{2}{|\mathbf{w}_0|} \tag{3.11}$$

The classification is then just a simple decision on which side of the hyperplane the object is.

Chapter 4

Features

Choosing the best feature set for sentiment analysis has the highest importance as it has a strong impact on the evaluation results. This chapter describes the most common features.

Features are often preprocessed by various techniques in order to reduce the feature space. The importance of this preprocessing phase depends on the language. For highly inflectional languages like Czech, *stemming* or *lemmatization* (see Section 2.5) is almost mandatory because it is necessary to reduce the high number of different word forms.

A *stem* or a *lemma* can be used directly as a feature similarly to a simple unigram feature. *Stemming* or *lemmatization* can also improve the performance of other features.

4.1 N-gram Features

N-grams and their frequency or presence is often used as a valid baseline. In some cases word positions and TF-IDF weighting scheme may be considered effective features.

N-gram features do not have to use only words, any item will do. For example POS patterns are simply POS n-grams.

N-gram Word n-grams are used to capture frequent word sequences. The presence of unigrams, bigrams and trigrams is often used as binary features. The feature space is pruned by the minimum n-gram occurrence (e.g. 5).

Note that this is the baseline feature in most of the related work.

Character n-gram Similarly to the word n-gram features, character n-gram features can be used, as proposed by, e.g., [Blamey et al., 2012]. Character trigrams are often used to capture frequent emoticons. The feature set usually contains 3-grams to 6-grams. The feature space is further pruned by the minimum occurrence of a particular character n-gram.

Skip-bigram Instead of using sequences of adjacent words (n-grams) skip-grams Guthrie et al. [2006], which skip over arbitrary gaps, can be used. Basic approach uses skip-bigrams with 2 or 3 word skips and removes skip-grams with a frequency ≤ 20.

Bag of words Set of words without any information on the word order is called bag of words.

4.2 POS-related Features

Direct usage of part-of-speech n-grams that cover sentiment patterns has not shown any significant improvement in the related work. Still, POS tags do provide certain characteristics of a text. Various POS-related features have been used in related work e.g., the number of nouns, verbs, and adjectives [Ahkter and Soria, 2010], the ratio of nouns to adjectives and verbs to adverbs [Kouloumpis et al., 2011], and the number of negative verbs obtained from POS tags.

4.3 Lexical Features

Additional lexical resources such as sentiment lexicons or *SentiWordNet* [Baccianella et al., 2010] can be used as features. These resources use external knowledge to improve the results of sentiment analysis.

4.4 Semantic Features

Distributional semantics (see Section 5) represent the latest trend in sentiment analysis. This is because of their ability to represent the meaning of texts simply by using a statistical analysis. The direct application of a joint sentiment and topic model proved to be useful [Lin and He, 2009]. Alternatively, semantics models can be used as new sources of information for classification (e.g. n-gram features or as bag of clusters instead of bag of words).

4.5 Other Features

Syntactic Features Features trying to capture word dependencies and sentence structure usually by exploiting syntactic information generated from parse trees

Orthographic Features Features based on the appearance of the word (sometimes called word shape), e.g. the first letter is a capital letter, all letters are capital or the words consists of digits.

Emoticons Lists of positive and negative emoticons (e.q.Montejo-Ráez et al. [2012]) capture the number of occurrences of each class of emoticons within the text.

Punctuation-based Features Features consisting of special characters, number of words, exclamation marks, question marks, quotation marks. These features usually do not significantly improve the results.

4.6 Feature Selection

The basic reason for using feature selection (or reduction) methods for supervised sentiment analysis is twofold: first, the reduced feature set decreases the computing demands for the classifier, and, second, removing irrelevant features can lead to better classification accuracy. Furthermore, noise and redundancy in the feature space increase the likelihood of over-fitting [Abbasi et al., 2011].

A study by Sharma and Dey [2012] compares five methods for feature selection, namely Information Gain, Chi Square, Gain Ratio, Relief-F, and Document Frequency, with seven different classifiers. Results are reported on the widely-used movie review database from Pang et al. [2002]. The best performance was achieved by using the SVM classifier and the Gain Ratio selector with the number of features ranging from 2,000 to 8,000 and employing only unigrams as features sorted by their frequency.

Abbasi et al. [2008] proposed an entropy-weighted genetic algorithm that combines Information Gain with a genetic algorithm for selecting features in a bootstrapping manner, tuned on held-out data. They performed document-level binary sentiment of English and Arabic and used SVM as the main classifier. Their results were superior to other approaches, such as plain SVM or Information Gain selection. In their later work, Abbasi et al. [2011] proposed another feature selection method called the Feature Relation Network. This manually constructed network of feature dependencies (e.g., subsumption[1] or parallel relations of various n-grams) relies on *SentiWordNet* in order to assign the final feature weights.

Forman [2003] proposes a metric called Bi-Normal Separation and provides an extensive comparison with another twelve existing feature selection methods. Using SVM as the underlying classifier, the proposed method yields the best results and is suitable for skewed (imbalanced) classes. Other examples of feature selection methods for sentiment analysis or text classification can be found in, e.g., [Chen et al., 2009, Aghdam et al., 2009].

Since feature selection is also important outside the domain of text classification, Wasikowski and Chen [2010] conducted a systematic study, focusing on dealing with class imbalance on small samples. They compare seven selection methods on 11 small datasets with highly skewed classes and conclude by recommending two best-performing algorithms, especially for scenarios that require a small number of features. Another approach based on dynamic mutual information is presented in Liu et al. [2009]. Again, the experiments are conducted on 16 benchmark datasets with a rather small size (up to 8124 instances only) and a small number of features (from 18 to 279), which is a fundamentally different scenario from machine learning-based sentiment analysis.

Feature selection, however, does not have to lead to a better performance in all cases, as reported e.g. by Boiy and Moens [2009], who report Chi-square selection results in their preliminary tests without any success.

[1] *'is-a'* hierarchical relation

Chapter 5

Distributional Semantics

As mentioned in chapter 4, semantics models represent the latest trend in sentiment analysis. They can be applied directly to jointly model sentiment and topics or alternatively, the features derived from semantics models can be used as new sources of information for classification (e.g. n-gram features or as bag of clusters instead of bag of words).

The backbone principle of methods for discovering hidden meaning in a plain text is the formulation of the *Distributional Hypothesis* [Harris, 1954, Firth, 1957]. The famous quote of Firth [1957] says that *"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. This implication was confirmed by empirical tests carried out on human groups in [Rubenstein and Goodenough, 1965, Charles, 2000]. The models based on the Distributional Hypothesis are often referred to as *distributional semantics models.*

Some distributional semantic models use the *Bag-of-word* hypothesis (e.g. LDA). *Bag-of-word* hypothesis assumes that the word order has no meaning. The term bag means a set where the order of words has no role.

Distributional semantics models typically represent the meaning of a word as a vector: the vector reflects the contextual information of the word throughout the training corpus. Each word $w \in W$ (where W denotes the word vocabulary) is associated with a vector of real numbers $\boldsymbol{w} \in \mathbb{R}^k$. 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). Example of such a method is the k-means algorithm, which is often used because of its efficiency and acceptable computational requirements. As a measure of the similarity between two words, is commonly used the cosine similarity of word vectors, calculated as the cosine of the angle between the corresponding vectors.

5.1 HAL

HAL (Hyperspace Analogue to Language) [Lund and Burgess, 1996] is a very simple method for building semantic space. HAL records the co-occurring words into a matrix. The words are observed in a small context window[1] around the target word in the given corpus. The Co-occurring words are weighted inversely to their distance from the target word. This results in the co-occurrence matrix $\mathbb{M} = |W| \times |W|$, where $|W|$ is the size of the vocabulary. Finally, the row and column vectors of $\mathbb{M}$ represent the co-occurrence information of the words appearing before and after the target word.

5.2 COALS

COALS (Correlated Occurrence Analogue to Lexical Semantics) [Rohde et al., 2004] extends the HAL model. The difference is that after recording the co-occurrence information, the raw counts of $\mathbb{M}$ are converted into Pearson's correlations. Negative values are reset to zero and other values are replaced by their square roots. The optional final step, inspired by LSA [Deerwester et al., 1990], is to apply the SVD (singular value decomposition) to $\mathbb{M}$, resulting in a dimensionality reduction and also the discovery of latent semantic relationships between words.

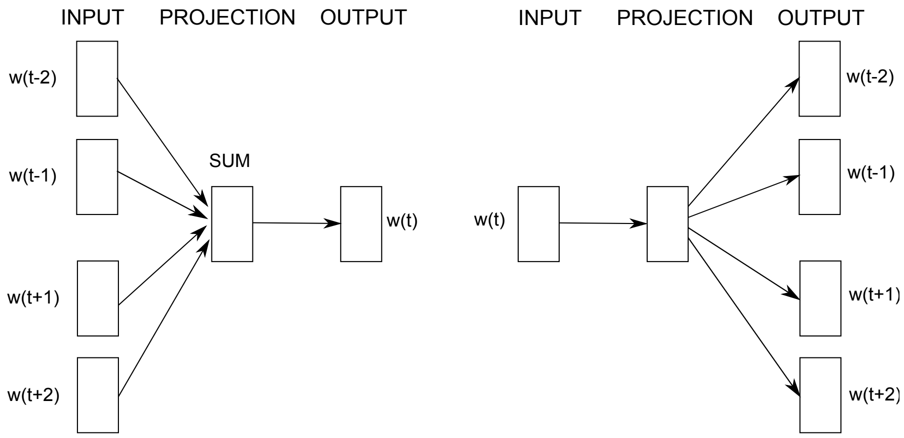
5.3 CBOW

CBOW (Continuous Bag-of-Words) [Mikolov et al., 2013a] tries to predict the current word using a small context window around the word. This model estimates word vector representation based on the context. The word vectors

[1]typically four words

can be compared using e.g. cosine similarity measure. Instead of using a co-occurrence matrix this model uses a neural network for the meaning extraction.

The architecture is similar to the feed-forward NNLM (Neural Network Language Model) proposed in Bengio et al. [2006]. The NNLM is computationally expensive between the projection and the hidden layer. Thus, in the proposed architecture CBOW, the (non-linear) hidden layer is removed and the projection layer is shared between all the words. The word order in the context does not influence the projection (see Figure 5.1a). This architecture has proved to be of low computational complexity.



(a) CBOW (b) Skip-gram

Figure 5.1: Neural network model architectures.

5.4 Skip-Gram

The Skip-gram architecture is similar to CBOW. However, instead of predicting the current word based on the context, it tries to predict a word's context based on the word itself [Mikolov et al., 2013b]. Thus, the intention of the Skip-gram model is to find word patterns that are useful for predicting the surrounding words within a certain range in a sentence (see Figure 5.1b). The Skip-gram model estimates the syntactic properties of words slightly worse than does the CBOW model, but it is much better at modeling their semantics [Mikolov et al., 2013a].

5.5 LDA

LDA (Latent Dirichlet Allocation) [Blei et al., 2003] is a well known topic model. LDA is based on the *Distributional Hypothesis* and the *Bag-of-words Hypothesis*, i.e., that the word order does not matter and there is some latent relation between the words within the same document (within the same context).

The underlying idea is that document is a mixture of topics and topic is a mixture of words. The vector representation of words can be used to model word meanings. The meaning of words can be represented by the associated topic distribution. The model can be extended to jointly model topic and sentiment [Lin and He, 2009].

Chapter 6

State of the art in Sentiment Analysis

There are many ways to categorize sentiment analysis approaches e.g. by their use of methods and resources (dictionary-based and machine learning-based). Whereas dictionary-based methods usually depend on a sentiment dictionary (or a polarity lexicon) and a set of handcrafted rules [Taboada et al., 2011], machine learning-based methods require labeled training data that are later represented as features (see Section 4) and fed into a classifier (see Section 3). Recent attempts have also investigated semi-supervised methods that incorporate unlabeled data [Zhang et al., 2012].

However the granularity level of sentiment analysis seems to be the most natural way to categorize the related work.

The most of the research in automatic sentiment analysis has been devoted to English. There were several attempts in other languages (e.g. Banea et al. [2010], Ghorbel and Jacot [2011], VILARES et al. [2015], Basile and Nissim [2013]), but we will focus only on Czech and English.

Although we have devoted substantial effort to clearly describe all methods in the following Sections in detail, we would like to direct curious readers to in-depth surveys Pang and Lee [2008], Liu and Zhang [2012], Tsytsarau and Palpanas [2012] and Martínez-Cámara et al. [2014][1] for additional information.

[1]This survey is focused on sentiment analysis in Twitter

6.1 Document-Level and Sentence-Level

The key point of using machine learning for sentiment analysis lies in engineering a representative set of features (see Section 4). Pang et al. [2002] experimented with unigrams (presence of a certain word, frequencies of words), bigrams, part-of-speech (POS) tags, and adjectives on a movie review dataset. Martineau and Finin [2009] tested various weighting schemes for unigrams based on the TFIDF model [Manning et al., 2008a] and proposed delta weighting for a binary scenario (positive, negative). Their approach was later extended by Paltoglou and Thelwall [2010] who proposed further improvements in delta TFIDF weighting achieving the accuracy of 96.9% on the movie review dataset and 85.04% on the BLOG06 dataset.

The focus of current sentiment analysis research is shifting towards social media, mainly targeting Twitter [Kouloumpis et al., 2011, Pak and Paroubek, 2010] and Facebook [Go et al., 2009, Ahkter and Soria, 2010, Zhang et al., 2011, López et al., 2012]. Analyzing media with a very informal language benefits from involving novel features, such as emoticons [Pak and Paroubek, 2010, Montejo-Ráez et al., 2012], character n-grams [Blamey et al., 2012], POS and POS ratio [Ahkter and Soria, 2010, Kouloumpis et al., 2011], or word shape [Go et al., 2009, Agarwal et al., 2011].

In many cases, the gold data for training and testing the classifiers are created semi-automatically [Kouloumpis et al., 2011, Go et al., 2009, Pak and Paroubek, 2010]. In the first step, random samples from a large dataset are drawn according to the presence of emoticons (usually positive and negative) and are then filtered manually. Although large high-quality collections can be created very quickly with this approach, it makes a strong assumption that every positive or negative post must contain an emoticon.

Balahur and Tanev [2012] performed experiments with Twitter posts as part of the CLEF 2012 RepLab.[2] They classified English and Spanish tweets with a small but precise lexicon, which also contained slang, combined with a set of rules that captured the manner in which sentiment is expressed in social media.

Balahur and Turchi [2012] studied the manner in which sentiment analysis can be done for French, German and Spanish, using Machine Translation. They employed three different MT systems (Google Translate, Bing Translator, and Moses [Koehn et al., 2007]) in order to obtain training and test data for the three target languages. Subsequently, they extracted features for a machine learning model. They additionally employed meta-

[2] `http://www.limosine-project.eu/events/replab2012`

classifiers to test the possibility to minimize the impact of noise (incorrect translations) in the obtained data. Their experiments showed that training data obtained using machine translation do not significantly decrease performance of sentiment analysis and thus it can be a solution in the case of unavailability of the target language annotated data.

Socher et al. [2013] present their *Recursive Neural Tensor Network* and *Sentiment Treebank*. The *Sentiment Treebank* contains five sentiment labels (very positive to very negative) for 215,154 phrases in the parse trees of 11,855 sentences. They train the *Recursive Neural Tensor Network* on the new treebank and evaluate against the state of the art methods. This model outperforms all previous methods on several metrics and pushes the state of the art in binary sentence-level classification on the Rotten Tomatoes dataset from 80% up to 85.4%. The accuracy of predicting the five sentiment labels for all phrases reaches 80.7%, an improvement of 9.7% over bag of features baselines. This is due to the fact that the model accurately captures sentence composition and the effects of negation at various tree levels for both positive and negative phrases.

Kiritchenko et al. [2014b], Zhu et al. [2014] describe a state-of-the-art sentiment analysis system that detects the sentiment of short informal textual messages (tweets and SMS messages) and the sentiment of terms. Their supervised system is based on a machine learning approach leveraging a variety of features. They employ automatically generated lexicons using tweets with sentiment-word hashtags and tweets with emoticons. Separate sentiment lexicon captures negated words. The system ranked first in the SemEval-2013 shared task “Sentiment Analysis in Twitter” (Task 2), obtaining an F-measure of 69.02% in the message-level task and 88.93% in the term-level task. Additional improvements boost the F-measure to 70.45% (message-level task) and 89.50% (term-level task).

6.2 Word-Level

Identifying the semantic orientation of subjective terms[3] (words or phrases) is a fundamental task for sentiment lexicon generation. These sentiment or opinion lexicons are compiled in an automatic manner with an optional final human check. The task of identifying semantic word orientation is also called words polarity detection. There are publicly available resources containing sentiment polarity of words e.g. *General Inquirer*[4], *Dictionary of*

[3] Also called sentiment words, opinion words and polar words
[4] http://www.wjh.harvard.edu/ inquirer/

Affect of Language[5], *WordNet-Affect*[6] or *SentiWordNet* [Baccianella et al., 2010] These resources are mainly used for computing the sentence or document sentiment by dictionary methods or as features for machine learning methods. Another use is the generation of a domain specific lexicon.

Turney [2002], Turney and Littman [2003] estimate the semantic orientation of words by computing the *Pointwise Mutual Information* (PMI) between the given word and paradigm words (e.g. good, bad, nice, nasty). Another approach [Kamps et al., 2004] measures the synonym relation of words based on *WordNet*[7].

Another popular way of using *WordNet* obtains a list of sentiment words by an iterative process of expanding the initial set with synonyms and antonyms Kim and Hovy [2004], Hu and Liu [2004]. Kim and Hovy [2004] determine the sentiment polarity of unknown words according to the relative count of their positive and negative synonyms.

Wiebe et al. [2005], Wilson et al. [2005] create the *Multi-Perspective Question Answering* (MPQA) corpus containing 535 news articles from a wide variety of news sources and describe the overall annotation scheme. They also compile a subjectivity lexicon with tagged prior[8] polarity values of words.

Rao and Ravichandran [2009] treat the sentiment polarity detection as a semi-supervised label propagation problem in a graph, where nodes represent words and edges are the relations between words. They use *WordNet* and OpenOffice thesaurus and positive and negative seed sets.

As demonstrated by Fahrni and Klenner [2008] the polarity of words is domain specific and lexicon-based approaches have difficulty with some domains. Machine learning algorithms naturally adapt to the corpus domain by training. Statistical approach to lexicon generation adapts the lexicon to the target domain. Fahrni and Klenner [2008] propose to derive posterior polarities using the co-occurrence of adjectives to create a corpus-specific dictionary.

He et al. [2008] use *Information Retrieval* methods to build a dictionary by extracting frequent terms from the dataset. The sentiment polarity of

[5]http://www.hdcus.com/

[6]http://wndomains.fbk.eu/wnaffect.html

[7]WordNet Miller and Fellbaum [1998] is a hierarchical lexical database containing nouns, verbs, adjectives and adverbs grouped into synonym sets (synsets). The synsets are related by different types of relationships to other synsets.

[8] *"Prior polarity refers to the sentiment a term evokes in isolation, as opposed to the sentiment the term evokes within a particular surrounding context."*[Pang and Lee, 2008]

each document is computed as a relevance score to a query composed of the top terms from this dictionary. Finally, the opinion relevance score is combined with the topic relevance score, providing a ranking of documents on that topic.

Choi and Cardie [2008]determine the polarity of terms using a structural inference motivated by compositional semantics. Their experiments show that lexicon–based classification with compositional semantics can perform better than supervised learning methods that do not incorporate compositional semantics (accuracy of 89.7% vs. 89.1%), but a method that integrates compositional semantics into the learning process performs better than the previous approaches (90.7%). The results were achieved on the MPQA dataset. Later they study the adaptability of lexicons to other domains using an integer linear programming approach [Choi and Cardie, 2009].

Xu et al. [2012] have developed an approach based on HAL (see Section 5.1) called *Sentiment Hyperspace Analogue to Language* (S-HAL). The semantic orientation of words is characterized by a specific vector space. This feature vectors were used to train a classifier to identify the sentiment polarity of terms.

Saif et al. [2014] adapt the social-media sentiment lexicon from Thelwall et al. [2012] by extracting contextual semantics of words to update prior sentiment strength in lexicon and apply it to three different Twitter datasets achieving an average improvement of 2.46% and 4.51% in terms of accuracy and F-measure respectively.

6.3 Aspect-Based Sentiment Analysis

Recently a lot of attention has been targeted on sentiment analysis at finer levels of granularity, namely, aspect-based sentiment analysis (ABSA). The goal of ABSA is to extract aspects and to estimate the sentiment associated with the given aspect [Liu, 2012]. For the task definition see Section 2.3.

6.3.1 Aspect Term Extraction

The basic approach to aspect extraction is finding frequent nouns and noun phrases [Liu et al., 2005, Blair-Goldensohn et al., 2008, Moghaddam and Ester, 2010, Long et al., 2010].

Sequential learning methods (e.g. *Hidden Markov Models* [Rabiner, 2010] or *Conditional Random Fields* [Lafferty et al., 2001]) can be applied to aspect extraction. This approach treats aspect extraction as a special case of the general information extraction problem.

Hu and Liu [2004] extract the most frequent features (noun or noun phrases) and then remove meaningless feature phrases and redundant single-word features. Wei et al. [2010] further prune the feature space using a list of subjective (positive and negative) adjectives. Pavlopoulos and Androutsopoulos [2014] propose adding a pruning mechanism that uses continuous space vector representations of words and phrases to further improve the results.

Another widely used approach to this problem is the use of topic models. Brody and Elhadad [2010] present a system that uses local (sentence-level) LDA (see Section 5.5) to discover aspect terms (nouns). Observing that every opinion has a target, a joint model can be designed to model the sentiment of words and topics at the same time [Xianghua et al., 2013, Mei et al., 2007, Titov and McDonald, 2008a]. A topic-based model for jointly identifying aspect and sentiment words was proposed by Zheng et al. [2014].

6.3.2 Aspect Sentiment Classification

Aspect sentiment classification can be divided into lexicon-based approaches and machine learning approaches. Machine learning performs better in a particular application domain but it is difficult to be scaled up to a large number of domains. Lexicon-based techniques are more suitable for open-domain applications [Liu, 2012].

Lexicon-based approaches (e.g. [Xianghua et al., 2013, Ding et al., 2008, Hu and Liu, 2004]) use a list of aspect-related sentiment phrases as the core resource for aspect sentiment classification.

Jiang et al. [2011] use a dependency parser to generate a set of aspect dependent features for classification. Boiy and Moens [2009] weights each feature based on the position of the feature relative to the target aspect in the parse tree.

Brody and Elhadad [2010] extract sentiment polarity from a constructed conjunction polarity graph.

Jo and Oh [2011] propose probabilistic generative models that outperform other generative models and are competitive in terms of accuracy with supervised aspect sentiment classification methods.

Semantic Evaluation Workshop SemEval 2014

The current state of the art of aspect-based sentiment analysis methods was presented at the SemEval 2014 Task 4 [Pontiki et al., 2014]. The detailed description of each system is beyond the scope of this paper, however we try to briefly describe the highest ranking systems. The task description was introduced in Section 2.3. The comparison of the systems are shown in table 6.1.

Kiritchenko et al. [2014a] (NRC-Canada) proposed a hybrid system that incorporates both machine learning ngram features and automatically constructed sentiment lexicons for affirmative and negated contexts.

Brun et al. [2014] (XRCE) train one classifier to detect the categories and for each category they train a separate classifier for category detection of the corresponding polarities They extend their previous system built on a robust deep syntactic parser, which calculates semantic relations of words. The adaptation includes additional hand-written rules (regular expressions), extending dependency grammar and lexicons.

Castellucci et al. [2014] (UNITOR) exploit kernel methods within the SVM framework. They model the aspect term extraction task as a sequential tagging task by using SVM^{hmm}. The aspect term polarity, aspect category detection and aspect category polarity detection are tackled as a classification problem where multiple kernels are linearly combined to generalize several linguistic information. Tree kernels proposed in Collins and Duffy [2001] are adopted to model syntactic similarity through convolutions among syntactic tree substructures.

Chernyshevich [2014](IHS_RD) relies on a rich set of lexical, syntactic and statistical features and the CRF model to correctly extract the aspect terms. She also runs a preprocessing step that performs e.g. slang and misspelling corrections, POS tagging, parsing, noun phrase extraction, semantic role labeling, entity boundary detection.

Toh and Wang [2014] (DLIREC) ranked the first in the aspect term extraction task in the restaurant domain and second in the laptop domain. They use a *Conditional Random Field* based classifier for aspect term extraction and linear classifier for aspect term polarity classification with lexicon, syntactic and semantic features. They created semantic clusters using word2vec tool Mikolov et al. [2013c][9]

[9] https://code.google.com/p/word2vec/

		Aspect detection					Aspect polarity		
		Const.	Team	P[%]	R[%]	F_1[%]	Const.	Team	ACC[%]
Restaurants	Term	U	DLIREC	85.35	82.71	84.01	C	DCU	80.95
		C	XRCE	86.25	81.83	83.98	SC	NRC-Can.	80.16
		C	NRC-Can.	84.41	76.37	80.19	U	UWB	77.69
		C	UNITOR	82.45	77.87	80.09	C	XRCE	77.69
	Category	C	NRC-Can.	91.04	86.24	88.58	C	NRC-Can.	82.92
		U	UNITOR	84.98	85.56	85.27	C	XRCE	78.15
		C	XRCE	83.23	81.37	82.29	U	UNITOR	76.29
		U	UWB	84.36	78.93	81.55	C	SAP_RI	75.61
Laptops	Term	SC	IHS_RD	84.80	66.51	74.55	C	DCU	70.49
		U	DLIREC	81.90	67.13	73.78	C	NRC-Can.	70.49
		C	DLIREC	79.31	63.30	70.41	C	SZTE-NLP	66.97
		C	NRC-Can.	78.77	60.70	68.57	C	UBham	66.66

Table 6.1: Comparison of the four best participating systems in each subtask. (SC) indicates a strongly constrained system that was not trained on the in-domain training data, (C) constrained system that was trained on the in-domain training data and (U) unconstrained system. ACC, P, R, and F_1 denote accuracy, precision, recall and F-measure, respectively.

Wagner et al. [2014] (DCU) combine various lexicons such as MPQA, *SentiWordNet* and *General Inquirer* and use both rule-based and machine learning approach. They focus on fine tuning of parameters and the systems efficiency.

[Brychcín et al., 2014] (UWB) present a system based on supervised machine learning extended by unsupervised methods for latent semantics discovery (LDA and semantic spaces - HAL and COALS see Section 5) and sentiment vocabularies. Their approach to aspect term extraction is based on *Conditional Random Fields*.

6.4 Summarization in Sentiment Analysis

One opinion from a single source is usually not sufficient for sentiment analysis applications thus some form of a summary of opinions is necessary.

A common form of summary is aspect-based opinion summarization. Hu and Liu [2004] simply count positive and negative sentences regarding aspects of target entity and then ranked based on their frequency. Finally, top-ranking sentences are selected to form the summaries.

Titov and McDonald [2008b] use statistical model based on LDA (see Section 5.5) to discover corresponding topics in text and extract supporting opinions as evidence for given aspect sentiment rating.

Paul et al. [2010] summarize multiple contrastive viewpoints of opinionated text using probabilistic topic model and a random walk formulation to score sentences and pairs of sentences from opposite viewpoints based on both their representativeness of the collection as well as their contrastiveness with each other.

Fang et al. [2012] propose cross-perspective topic model, that simulates the generative process of how sentiment words occur in different collections. They conduct a set of experiments on political domain to demonstrate the qualitative and quantitative properties of their approach.

Wang et al. [2014] develop a submodular function-based framework for query-focused opinion summarization. They also studied different metrics on text similarity estimation and their effect on summarization and tested the proposed framework on community question answering summarization and blog summarization. A human evaluation task was conducted to show the quality and information diversity of generated summaries.

6.5 Sentiment Analysis in the Czech Environment

Veselovská et al. [2012] presented an initial research on Czech sentiment analysis. They created a corpus which contains polarity categories of 410 news sentences. They used the Naive Bayes classifier and a classifier based on a lexicon generated from annotated data. The corpus is not publicly available, and because of its small size no strong conclusions can be drawn. Error analysis of lexicon-based classifier on this dataset was done by Veselovská and Hajič jr. [2013].

Subjectivity Lexicon for Czech [Veselovská, 2013, Veselovská et al., 2014] consists of 4947 evaluative items annotated with part of speech and tagged with positive or negative sentiment polarity. Although the lexicon did not significantly help to improve the polarity classification it is still a lexical resource worth mentioning.

Steinberger et al. [2012] proposed a semi-automatic "triangulation" approach to creating sentiment dictionaries in many languages, including Czech. They first produced high-level gold-standard sentiment dictionaries for two languages and then translated them automatically into a third language by means of a state-of-the-art machine translation service. Finally, the resulting sentiment dictionaries were merged using the overlap of the two automatic translations.

A multilingual parallel news corpus annotated with opinions on entities was presented in Steinberger et al. [2011]. Sentiment annotations were projected from one language to several others, which saved annotation time and guaranteed comparability of opinion mining evaluation results across languages. The corpus contains 1,274 news sentences where an entity (the target of the sentiment analysis) occurs. It contains seven languages including Czech. The research targets fundamentally different objectives from our research as it focuses on news media and aspect-based sentiment analysis.

The first extensive evaluation of Czech sentiment analysis was done by Habernal et al. [2013]. Three different classifiers, namely Naive Bayes, SVM (Support Vector Machines) and Maximum Entropy classifiers were tested on large-scale labeled corpora (10k Facebook posts, 90k movie reviews, and 130k product reviews). Habernal et al. [2014] further experimented with feature selection methods.

Habernal and Brychcín [2013] used semantic spaces (see Brychcín and Konopík [2014]) created from unlabeled data as an additional source of information to improve results. Brychcín and Habernal [2013] explored the benefits of the global target context and outperformed the previous unsupervised approach.

Steinberger et al. [2014] present the first attempt at aspect-level sentiment analysis in Czech and provide an annotated corpus of 1244 sentences from the restaurant reviews domain.

6.6 Sarcasm Detection

The issue of automatic sarcasm detection has been addressed mostly in English, although there has been some research in other languages, such as Dutch [Liebrecht et al., 2013], Italian [Bosco et al., 2013], or Brazilian Portuguese [Vanin et al., 2013]. The first attempt at sarcasm detection in Czech was done in Ptáček et al. [2014].

Chapter 7

Future Work and Preliminary Results

This chapter describes preliminary results and new ideas for future work that imply the aims of PhD thesis.

Automatic sentiment analysis in the Czech environment had not yet been thoroughly targeted by the research community. Therefore it was necessary to create a publicly available labeled dataset as well as to evaluate the current state of the art for two reasons. First, many NLP methods must deal with high flection and rich syntax when processing the Czech language. Dealing with these issues may lead to novel approaches to sentiment analysis as well. Second, freely accessible and well-documented datasets, as known from many shared NLP tasks, may stimulate competition, which usually leads to the production of cutting-edge solutions.[1]

We have done an in-depth research on machine learning methods for sentiment analysis of Czech social media in Habernal et al. [2013, 2014]. Three different classifiers, namely Naive Bayes, SVM (Support Vector Machines) and Maximum Entropy classifiers were tested on large-scale labeled corpora (10k Facebook posts, 90k movie reviews, and 130k product reviews). We explored different pre-processing techniques and employed various features and classifiers. We also experiment with five different feature selection algorithms and investigate the influence of named entity recognition and preprocessing on sentiment classification performance. We significantly outperformed the baseline (unigram feature without preprocessing) in three-class classification and achieved an F-measure of 0.69 using a combination of

[1]E.g., named entity recognition based on Conditional Random Fields emerged from CoNLL-2003 named entity recognition shared task.

features (unigrams, bigrams, POS features, emoticons, character n-grams) and preprocessing techniques (unsupervised stemming and phonetic transcription).

We aim to investigate the effectiveness of several unsupervised methods for latent semantics discovery as new features for sentiment analysis. We believe that semantics contains hidden information that can improve sentiment analysis.

Czech as a representative of a inflective language is an ideal environment for the study of various aspects of sentiment analysis for inflectional languages. It is challenging because of its very flexible word order and many different word forms.

We conceive this study to deal with several aspects of sentiment analysis – the document and sentence level of sentiment analysis, aspect-based sentiment analysis and the influence of figurative language on sentiment analysis. The breadth of this study can lead to more general view and better understanding of sentiment analysis. We can reveal and overcome unexpected obstacles, create necessary evaluation datasets and even come up with new creative solutions to sentiment analysis tasks.

The detection of sarcasm in Czech was done in Ptáček et al. [2014]. The next step is to incorporate sarcasm detection into the sentiment analysis process and compare the results with and without the additional information about sarcasm.

7.1 Aims of the PhD Thesis

The goal of the doctoral thesis is to propose novel methods for improving performance of sentiment analysis with special emphasis on inflectional languages (e.g. Czech). The work will be focused on the following research tasks:

- Deal with specific properties of Czech language in the sentiment analysis environment.
- Use additional semantic and/or syntactic information to improve sentiment analysis.
- Explore the influence of figurative language (e.g. sarcasm) on sentiment analysis.

Bibliography

Ahmed Abbasi, Hsinchun Chen, and Arab Salem. Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums. *ACM Trans. Inf. Syst.*, 26(3):12:1–12:34, 2008. ISSN 1046-8188. doi: 10.1145/1361684.1361685.

Ahmed Abbasi, Stephen France, Zhu Zhang, and Hsinchun Chen. Selecting attributes for sentiment classification using feature relation networks. *IEEE Transactions on Knowledge and Data Engineering*, 23(3):447–462, 2011. ISSN 1041-4347.

Apoorv Agarwal, Boyi Xie, Ilia Vovsha, Owen Rambow, and Rebecca Passonneau. Sentiment analysis of twitter data. In *Proceedings of the Workshop on Languages in Social Media*, LSM '11, pages 30–38, Stroudsburg, PA, USA, 2011. Association for Computational Linguistics.

Mehdi Hosseinzadeh Aghdam, Nasser Ghasem-Aghaee, and Mohammad Ehsan Basiri. Text feature selection using ant colony optimization. *Expert Syst. Appl.*, 36(3):6843–6853, April 2009. ISSN 0957-4174.

Julie Kane Ahkter and Steven Soria. Sentiment analysis: Facebook status messages. Technical report, Stanford University, 2010. Final Project CS224N.

Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner, and Daniel Tapias, editors, *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta, may 2010. European Language Resources Association (ELRA). ISBN 2-9517408-6-7.

Alexandra Balahur and Hristo Tanev. Detecting entity-related events and sentiments from tweets using multilingual resources. In *Proceedings of* *the 2012 Conference and Labs of the Evaluation Forum Information Access Evaluation meets Multilinguality, Multimodality, and Visual Analytics*, 2012.

Alexandra Balahur and Marco Turchi. Multilingual sentiment analysis using machine translation? In *Proceedings of the Association for Computational Linguistics: 3rd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*, ACL-WASSA'12, pages 52–60, 2012.

C. Banea, R. Mihalcea, and J. Wiebe. Multilingual subjectivity: Are more languages better? In *Proceedings of COLING*, 2010.

Valerio Basile and Malvina Nissim. Sentiment analysis on italian tweets. In *Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 100–107, 2013.

Yoshua Bengio, Holger Schwenk, Jean-Sébastien Senécal, Fréderic Morin, and Jean-Luc Gauvain. Neural probabilistic language models. In *Innovations in Machine Learning*, pages 137–186. Springer, 2006.

Adam L Berger, Vincent J Della Pietra, and Stephen A Della Pietra. A maximum entropy approach to natural language processing. *Computational linguistics*, 22(1):39–71, 1996.

Sasha Blair-Goldensohn, Kerry Hannan, Ryan McDonald, Tyler Neylon, George Reis, and Jeff Reynar. Building a sentiment summarizer for local service reviews. In *Proceedings of WWW-2008 workshop on NLP in the Information Explosion Era*, 2008.

Ben Blamey, Tom Crick, and Giles Oatley. R U : -) or : -(? character- vs. word-gram feature selection for sentiment classification of OSN corpora. In *Proceedings of AI-2012, The Thirty-second SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence*, pages 207–212. Springer, 2012.

D. M. Blei, A. Y. Ng, M. I. Jordan, and J. Lafferty. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:2003, 2003.

Erik Boiy and Marie-Francine Moens. A machine learning approach to sentiment analysis in multilingual web texts. *Information retrieval*, 12(5): 526–558, 2009.

Cristina Bosco, Viviana Patti, and Andrea Bolioli. Developing corpora for sentiment analysis: The case of irony and senti-tut. *IEEE Intelligent Systems*, 28(2):55–63, 2013. ISSN 1541-1672. doi: http://doi.ieeecomputersociety.org/10.1109/MIS.2013.28.

Samuel Brody and Noemie Elhadad. An unsupervised aspect-sentiment model for online reviews. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, HLT '10, pages 804–812, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics. ISBN 1-932432-65-5. URL `http://dl.acm.org/citation.cfm?id=1857999.1858121`.

Caroline Brun, Diana Nicoleta Popa, and Claude Roux. Xrce: Hybrid classification for aspect-based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 838–842, Dublin, Ireland, August 2014. Association for Computational Linguistics. URL `http://www.aclweb.org/anthology/S14-2149`.

Tomáš Brychcín, Michal Konkol, and Josef Steinberger. Uwb: Machine learning approach to aspect-based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 817–822. Association for Computational Linguistics, 2014. URL `http://aclweb.org/anthology/S14-2145`.

Tomáš Brychcín and Ivan Habernal. Unsupervised improving of sentiment analysis using global target context. In *Proceedings of the International Conference Recent Advances in Natural Language Processing RANLP 2013*, pages 122–128, Hissar, Bulgaria, September 2013. INCOMA Ltd. Shoumen, BULGARIA. URL `http://www.aclweb.org/anthology/R13-1016`.

Tomáš Brychcín and Miloslav Konopík. Semantic spaces for improving language modeling. *Computer Speech & Language*, 28(1):192 – 209, 2014. ISSN 0885-2308. doi: http://dx.doi.org/10.1016/j.csl.2013.05.001.

Giuseppe Castellucci, Simone Filice, Danilo Croce, and Roberto Basili. Unitor: Aspect based sentiment analysis with structured learning. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 761–767, Dublin, Ireland, August 2014. Association for Computational Linguistics. URL `http://www.aclweb.org/anthology/S14-2135`.

W. G. Charles. Contextual correlates of meaning. *Applied Psycholinguistics*, 21(04):505–524, 2000.

Jingnian Chen, Houkuan Huang, Shengfeng Tian, and Youli Qu. Feature selection for text classification with naïve bayes. *Expert Syst. Appl.*, 36 (3):5432–5435, April 2009. ISSN 0957-4174.

Maryna Chernyshevich. Ihs r&d belarus: Cross-domain extraction of product features using crf. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 309–313, Dublin, Ireland, August 2014. Association for Computational Linguistics. URL http://www.aclweb.org/anthology/S14-2051.

Yejin Choi and Claire Cardie. Learning with compositional semantics as structural inference for subsentential sentiment analysis. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 793–801. Association for Computational Linguistics, 2008.

Yejin Choi and Claire Cardie. Adapting a polarity lexicon using integer linear programming for domain-specific sentiment classification. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 2-Volume 2*, pages 590–598. Association for Computational Linguistics, 2009.

Michael Collins and Nigel Duffy. Convolution kernels for natural language. In *Advances in neural information processing systems*, pages 625–632, 2001.

Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Mach. Learn.*, 20(3):273–297, September 1995. ISSN 0885-6125. doi: 10.1023/A:1022627411411. URL http://dx.doi.org/10.1023/A:1022627411411.

Dmitry Davidov, Oren Tsur, and Ari Rappoport. Semi-supervised recognition of sarcastic sentences in twitter and amazon. In *Proceedings of the Fourteenth Conference on Computational Natural Language Learning*, CoNLL '10, pages 107–116, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics. ISBN 978-1-932432-83-1. URL http://dl.acm.org/citation.cfm?id=1870568.1870582.

S. Deerwester, S.T. Dumais, G.W. Furnas, T.K. Landauer, and R.A. Harshman. Indexing by latent semantic analysis. *Journal of the American Society for Information Science 41*, pages 391–407, 1990.

Xiaowen Ding, Bing Liu, and Philip S. Yu. A holistic lexicon-based approach to opinion mining. In *Proceedings of the Conference on Web Search and Web Data Mining*, 2008.

Angela Fahrni and Manfred Klenner. Old wine or warm beer: Target-specific sentiment analysis of adjectives. In *Proc. of the Symposium on Affective Language in Human and Machine, AISB*, pages 60–63, 2008.

Yi Fang, Luo Si, Naveen Somasundaram, and Zhengtao Yu. Mining contrastive opinions on political texts using cross-perspective topic model. In *Proceedings of the fifth ACM international conference on Web search and data mining*, pages 63–72. ACM, 2012.

John R. Firth. A Synopsis of Linguistic Theory, 1930-1955. *Studies in Linguistic Analysis*, pages 1–32, 1957.

George Forman. An extensive empirical study of feature selection metrics for text classification. *J. Mach. Learn. Res.*, 3:1289–1305, March 2003. ISSN 1532-4435.

Hatem Ghorbel and David Jacot. Sentiment analysis of french movie reviews. In Vincenzo Pallotta, Alessandro Soro, and Eloisa Vargiu, editors, *Advances in Distributed Agent-Based Retrieval Tools*, volume 361 of *Studies in Computational Intelligence*, pages 97–108. Springer Berlin Heidelberg, 2011. ISBN 978-3-642-21383-0. doi: 10.1007/978-3-642-21384-7_7. URL `http://dx.doi.org/10.1007/978-3-642-21384-7_7`.

Alec Go, Richa Bhayani, and Lei Huang. Twitter sentiment classification using distant supervision, 2009. CS224N Project Report, Stanford.

David Guthrie, Ben Allison, Wei Liu, Louise Guthrie, and Yorick Wilks. A closer look at skip-gram modelling. In *Proceedings of the 5th international Conference on Language Resources and Evaluation (LREC-2006)*, pages 1–4, 2006.

Ivan Habernal and Tomáš Brychcín. Semantic spaces for sentiment analysis. In Ivan Habernal and Václav Matoušek, editors, *Proceedings of the 16th international conference on Text, Speech and Dialogue (TSD'13)*, volume 8082 of *Lecture Notes in Computer Science*, pages 484–491. Springer, Berlin/Heidelberg, 2013.

Ivan Habernal, Tomáš Ptáček, and Josef Steinberger. Sentiment analysis in czech social media using supervised machine learning. In *Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 65–74, Atlanta, Georgia, June 2013. Association for Computational Linguistics. URL `http://www.aclweb.org/anthology/W13-1609`.

Ivan Habernal, Tomáš Ptáček, and Josef Steinberger. Supervised sentiment analysis in czech social media. *Information Processing & Management*, 50(5):693–707, 2014.

Mohammad Sadegh Hajmohammadi, Roliana Ibrahim, and Zulaiha Ali Othman. Opinion mining and sentiment analysis: A survey. *International Journal of Computers & Technology*, 2(3), 2012.

Zellig Harris. Distributional structure. *Word*, 10(23):146–162, 1954.

Ben He, Craig Macdonald, Jiyin He, and Iadh Ounis. An effective statistical approach to blog post opinion retrieval. In *Proceedings of the 17th ACM conference on Information and knowledge management*, pages 1063–1072. ACM, 2008.

Minqing Hu and Bing Liu. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '04, pages 168–177, New York, NY, USA, 2004. ACM.

Long Jiang, Mo Yu, Ming Zhou, Xiaohua Liu, and Tiejun Zhao. Target-dependent twitter sentiment classification. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, 2011.

Yohan Jo and Alice H. Oh. Aspect and sentiment unification model for online review analysis. In *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, WSDM '11, pages 815–824, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0493-1. doi: 10.1145/1935826.1935932. URL `http://doi.acm.org/10.1145/1935826.1935932`.

Jaap Kamps, MJ Marx, Robert J Mokken, and Maarten De Rijke. Using wordnet to measure semantic orientations of adjectives. 2004.

Soo-Min Kim and Eduard Hovy. Determining the sentiment of opinions. In *Proceedings of the 20th international conference on Computational Linguistics*, page 1367. Association for Computational Linguistics, 2004.

Svetlana Kiritchenko, Xiaodan Zhu, Colin Cherry, and Saif Mohammad. Nrc-canada-2014: Detecting aspects and sentiment in customer reviews. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 437–442, Dublin, Ireland, August 2014a. Association for Computational Linguistics. URL `http://www.aclweb.org/anthology/S14-2076`.

Svetlana Kiritchenko, Xiaodan Zhu, and Saif M Mohammad. Sentiment analysis of short informal texts. *Journal of Artificial Intelligence Research*, 50:723–762, 2014b.

Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondřej Bojar, Alexandra Constantin, and Evan Herbst. Moses: open source toolkit for statistical machine translation. In *Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions*, ACL '07, pages 177–180, Stroudsburg, PA, USA, 2007. Association for Computational Linguistics.

Efthymios Kouloumpis, Theresa Wilson, and Johanna Moore. Twitter sentiment analysis: The good the bad and the OMG! In *Proceedings of the Fifth International Conference on Weblogs and Social Media, Barcelona, Catalonia, Spain, July 17-21, 2011*. The AAAI Press, 2011.

John D. Lafferty, Andrew McCallum, and Fernando C. N. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the Eighteenth International Conference on Machine Learning*, pages 282–289, San Francisco, CA, USA, 2001. Morgan Kaufmann Publishers. URL `http://dl.acm.org/citation.cfm?id=645530.655813`.

Christine Liebrecht, Florian Kunneman, and Antal Van den Bosch. The perfect solution for detecting sarcasm in tweets #not. In *Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 29–37, Atlanta, Georgia, June 2013. Association for Computational Linguistics. URL `http://www.aclweb.org/anthology/W13-1605`.

Chenghua Lin and Yulan He. Joint sentiment/topic model for sentiment analysis. In *Proceedings of the 18th ACM conference on Information and knowledge management*, pages 375–384. ACM, 2009.

Bing Liu. Sentiment analysis and opinion mining. *Morgan & Claypool Publishers*, 2012.

Bing Liu and Lei Zhang. A survey of opinion mining and sentiment analysis. In *Mining Text Data*, pages 415–463. Springer, 2012.

Bing Liu, Minqing Hu, and Junsheng Cheng. Opinion observer: Analyzing and comparing opinions on the web. In *Proceedings of International Conference on World Wide Web*, 2005.

Huawen Liu, Jigui Sun, Lei Liu, and Huijie Zhang. Feature selection with dynamic mutual information. *Pattern Recogn.*, 42(7):1330–1339, July 2009. ISSN 0031-3203.

Chong Long, Jie Zhang, and Xiaoyan Zhu. A review selection approach for accurate feature rating estimation. In *Proceedings of Coling 2010: Poster Volume*, 2010.

Roque López, Javier Tejada, and Mike Thelwall. Spanish sentistrength as a tool for opinion mining peruvian facebook and twitter. In *Artificial Intelligence Driven Solutions to Business and Engineering Problems*, pages 82–85. ITHEA, Sofia, Bulgaria, 2012.

K. Lund and C. Burgess. Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods Instruments and Computers*, 28(2):203–208, 1996.

Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. *Introduction to Information Retrieval.* Cambridge University Press, New York, NY, USA, 2008a.

Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. *Introduction to Information Retrieval*. Cambridge University Press, 2008b. ISBN 0521865719. URL http://nlp.stanford.edu/IR-book/.

Justin Martineau and Tim Finin. Delta TFIDF: An improved feature space for sentiment analysis. In *Proceedings of the Third International Conference on Weblogs and Social Media, ICWSM 2009, San Jose, California, USA*. The AAAI Press, 2009. ISBN 978-1-57735-421-5.

Eugenio Martínez-Cámara, M. Teresa Martín-Valdivia, L. Alfonso Ureña-López, and Arturo Montejo-Raéz. Sentiment analysis in twitter. *Natural Language Engineering*, 20(01):1–28, 2014.

Qiaozhu Mei, Xu Ling, Matthew Wondra, Hang Su, and ChengXiang Zhai. Topic sentiment mixture: modeling facets and opinions in weblogs. In *Proceedings of International Conference on World Wide Web*, 2007.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013a.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems*, pages 3111–3119, 2013b.

Tomas Mikolov, Wen-tau Yih, and Geoffrey Zweig. Linguistic regularities in continuous space word representations. In *HLT-NAACL*, pages 746–751, 2013c.

George Miller and Christiane Fellbaum. Wordnet: An electronic lexical database, 1998.

Samaneh Moghaddam and Martin Ester. Opinion digger: an unsupervised opinion miner from unstructured product reviews. In *Proceeding of the ACM conference on Information and knowledge management*, 2010.

A. Montejo-Ráez, E. Martínez-Cámara, M. T. Martín-Valdivia, and L. A. Ureña López. Random walk weighting over sentiwordnet for sentiment polarity detection on twitter. In *Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis*, WASSA '12, pages 3–10, Stroudsburg, PA, USA, 2012. Association for Computational Linguistics.

Jorge Nocedal. Updating Quasi-Newton Matrices with Limited Storage. *Mathematics of Computation*, 35(151):773–782, 1980. URL http://www.jstor.org/stable/2006193.

Alexander Pak and Patrick Paroubek. Twitter as a corpus for sentiment analysis and opinion mining. In Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner, and Daniel Tapias, editors, *Proceedings of the International Conference on Language Resources and Evaluation, LREC 2010*. European Language Resources Association, 2010.

Georgios Paltoglou and Mike Thelwall. A study of information retrieval weighting schemes for sentiment analysis. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, ACL '10, pages 1386–1395, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics.

Bo Pang and Lillian Lee. Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, 2(1-2):1–135, 2008.

Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing - Volume 10*, EMNLP '02, pages 79–86, Stroudsburg, PA, USA, 2002. Association for Computational Linguistics.

Michael J Paul, ChengXiang Zhai, and Roxana Girju. Summarizing contrastive viewpoints in opinionated text. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 66–76. Association for Computational Linguistics, 2010.

John Pavlopoulos and Ion Androutsopoulos. *Proceedings of the 5th Workshop on Language Analysis for Social Media (LASM)*, chapter Aspect Term Extraction for Sentiment Analysis: New Datasets, New Evaluation Measures and an Improved Unsupervised Method, pages 44–52. Association for Computational Linguistics, 2014. URL `http://aclweb.org/anthology/W14-1306`.

Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. Semeval-2014 task 4: Aspect based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 27–35, Dublin, Ireland, August 2014. Association for Computational Linguistics and Dublin City University. URL `http://www.aclweb.org/anthology/S14-2004`.

Tomáš Ptáček, Ivan Habernal, and Jun Hong. Sarcasm detection on czech and english twitter. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages

213–223, Dublin, Ireland, August 2014. Dublin City University and Association for Computational Linguistics. URL `http://www.aclweb.org/anthology/C14-1022`.

Lawrence Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. In *Proceedings of the IEEE*, pages 257–286, 2010.

Delip Rao and Deepak Ravichandran. Semi-supervised polarity lexicon induction. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, pages 675–682. Association for Computational Linguistics, 2009.

Antonio Reyes, Paolo Rosso, and Davide Buscaldi. From humor recognition to irony detection: The figurative language of social media. *Data Knowl. Eng.*, 74:1–12, April 2012. ISSN 0169-023X. doi: 10.1016/j.datak.2012.02.005. URL `http://dx.doi.org/10.1016/j.datak.2012.02.005`.

D. L. T. Rohde, L. M. Gonnerman, and D. C. Plaut. An improved method for deriving word meaning from lexical co-occurrence. *Cognitive Psychology*, 7:573–605, 2004.

Herbert Rubenstein and John B. Goodenough. Contextual correlates of synonymy. *Communications of the ACM*, 8(10):627–633, October 1965.

Hassan Saif, Yulan He, Miriam Fernandez, and Harith Alani. Adapting sentiment lexicons using contextual semantics for sentiment analysis of twitter. In *The Semantic Web: ESWC 2014 Satellite Events*, pages 54–63. Springer, 2014.

Anuj Sharma and Shubhamoy Dey. A comparative study of feature selection and machine learning techniques for sentiment analysis. In *Proceedings of the 2012 ACM Research in Applied Computation Symposium*, RACS '12, pages 1–7, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1492-3.

Richard Socher, Alex Perelygin, Jean Y Wu, Jason Chuang, Christopher D Manning, Andrew Y Ng, and Christopher Potts. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the conference on empirical methods in natural language processing (EMNLP)*, volume 1631, page 1642, 2013.

J. Steinberger, P. Lenkova, M. Kabadjov, R. Steinberger, and E. van der Goot. Multilingual entity-centered sentiment analysis evaluated by parallel corpora. In *Proceedings of the 8th International Conference Recent Advances in Natural Language Processing*, RANLP'11, pages 770–775, 2011.

J. Steinberger, M. Ebrahim, Ehrmann M., A. Hurriyetoglu, M. Kabadjov, P. Lenkova, R. Steinberger, H. Tanev, S. Vázquez, and V. Zavarella. Creating sentiment dictionaries via triangulation. *Decision Support Systems*, 53:689—694, 2012.

Josef Steinberger, Tomáš Brychcín, and Michal Konkol. Aspect-level sentiment analysis in czech. In *Proceedings of the 5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, Baltimore, USA, June 2014. Association for Computational Linguistics.

E.A. Stepanov and G. Riccardi. Detecting general opinions from customer surveys. In *Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on*, pages 115–122, 2011.

Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. Lexicon-based methods for sentiment analysis. *Computational Linguistics*, 37(2):267–307, 2011.

Mike Thelwall, Kevan Buckley, and Georgios Paltoglou. Sentiment strength detection for the social web. *Journal of the American Society for Information Science and Technology*, 63(1):163–173, 2012.

Ivan Titov and Ryan McDonald. Modeling online reviews with multi-grain topic models. In *Proceedings of International Conference on World Wide Web*, 2008a.

Ivan Titov and Ryan T McDonald. A joint model of text and aspect ratings for sentiment summarization. In *ACL*, volume 8, pages 308–316. Citeseer, 2008b.

Zhiqiang Toh and Wenting Wang. Dlirec: Aspect term extraction and term polarity classification system. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 235–240, Dublin, Ireland, August 2014. Association for Computational Linguistics. URL `http://www.aclweb.org/anthology/S14-2038`.

Mikalai Tsytsarau and Themis Palpanas. Survey on mining subjective data on the web. *Data Mining and Knowledge Discovery*, 24(3):478–514, May 2012.

P.D. Turney and M.L. Littman. Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems*, 21(4):315–346, 2003. doi: 10.1145/944012.944013. URL `http://www.scopus.com/inward/record.url?eid=2-s2.0-2442507763&partnerID=40&md5=7bda5fbb202c15eea323c6c7c8d4a36f`. cited By 559.

Peter D Turney. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th annual meeting on association for computational linguistics*, pages 417–424. Association for Computational Linguistics, 2002.

Aline A Vanin, Larissa A Freitas, Renata Vieira, and Marco Bochernitsan. Some clues on irony detection in tweets. In *Proceedings of the 22nd international conference on World Wide Web companion*, pages 635–636. International World Wide Web Conferences Steering Committee, 2013.

Kateřina Veselovská. Czech subjectivity lexicon: A lexical resource for czech polarity classification. In Katarína Gajdošová and Adriána Žáková, editors, *Proceedings of the Seventh International Conference Slovko 2013; Natural Language Processing, Corpus Linguistics, E-learning*, pages 279–284, Lüdenscheid, Germany, 2013. Slovak National Corpus, L'. Štúr Institute of Linguistics, Slovak Academy of Sciences, RAM-Verlag. ISBN 978-3-942303-18-7.

Kateřina Veselovská and Jan Hajič jr. Why words alone are not enough: Error analysis of lexicon-based polarity classifier for czech. In *Proceedings of the 6th International Joint Conference on Natural Language Processing*, pages 1–5, Nagoya, Japan, 2013. Asian Federation of Natural Language Processing, Asian Federation of Natural Language Processing. ISBN 978-4-9907348-0-0.

Kateřina Veselovská, jr. Jan Hajič, and Jana Šindlerová. Subjectivity lexicon for czech: Implementation and improvements. *Journal for Language Technology and Computational Linguistics*, 29(1):47–61, 2014. ISSN 2190-6858.

Kateřina Veselovská, Jan Hajič jr., and Jana Šindlerová. Creating annotated resources for polarity classification in Czech. In Jeremy Jancsary, editor, *Proceedings of KONVENS 2012*, pages 296–304. ÖGAI, September 2012. PATHOS 2012 workshop.

DAVID VILARES, MIGUEL A. ALONSO, and CARLOS GÓMEZ-RODRÍGUEZ. A syntactic approach for opinion mining on spanish reviews. *Natural Language Engineering*, 21:139–163, 1 2015. ISSN 1469-8110. doi: 10.1017/S1351324913000181. URL `http://journals.cambridge.org/article_S1351324913000181`.

Joachim Wagner, Piyush Arora, Santiago Cortes, Utsab Barman, Dasha Bogdanova, Jennifer Foster, and Lamia Tounsi. Dcu: Aspect-based polarity classification for semeval task 4. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 223–229, Dublin, Ireland, August 2014. Association for Computational Linguistics. URL `http://www.aclweb.org/anthology/S14-2036`.

Lu Wang, Hema Raghavan, Claire Cardie, and Vittorio Castelli. Query-focused opinion summarization for user-generated content. In *Proceedings of COLING*, pages 1660–1669, 2014.

Mike Wasikowski and Xue-wen Chen. Combating the small sample class imbalance problem using feature selection. *IEEE Trans. on Knowl. and Data Eng.*, 22(10):1388–1400, October 2010. ISSN 1041-4347.

Chih-Ping Wei, Yen-Ming Chen, Chin-Sheng Yang, and ChristopherC. Yang. Understanding what concerns consumers: a semantic approach to product feature extraction from consumer reviews. *Information Systems and e-Business Management*, 8(2):149–167, 2010. ISSN 1617-9846. doi: 10.1007/s10257-009-0113-9. URL `http://dx.doi.org/10.1007/s10257-009-0113-9`.

Janyce Wiebe, Theresa Wilson, and Claire Cardie. Annotating expressions of opinions and emotions in language. *Language resources and evaluation*, 39(2-3):165–210, 2005.

Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the conference on human language technology and empirical methods in natural language processing*, pages 347–354. Association for Computational Linguistics, 2005.

Fu Xianghua, Liu Guo, Guo Yanyan, and Wang Zhiqiang. Multi-aspect sentiment analysis for chinese online social reviews based on topic modeling and hownet lexicon. *Knowledge-Based Systems*, 37(0):186 – 195, 2013. ISSN 0950-7051. doi: http://dx.doi.org/10.1016/j.knosys.2012.08.003. URL `http://www.sciencedirect.com/science/article/pii/S0950705112002158`.

Tao Xu, Qinke Peng, and Yinzhao Cheng. Identifying the semantic orientation of terms using s-hal for sentiment analysis. *Knowledge-Based Systems*, 35:279–289, 2012.

Liang-Chih Yu, Jheng-Long Wu, Pei-Chann Chang, and Hsuan-Shou Chu. Using a contextual entropy model to expand emotion words and their intensity for the sentiment classification of stock market news. *Knowledge Based Syst*, 41:89–97, March 2013.

Dan Zhang, Luo Si, and Vernon J. Rego. Sentiment detection with auxiliary data. *Information Retrieval*, 15(3-4):373–390, 2012.

Kunpeng Zhang, Yu Cheng, Yusheng Xie, Daniel Honbo, Ankit Agrawal, Diana Palsetia, Kathy Lee, Wei keng Liao, and Alok N. Choudhary.

SES: Sentiment elicitation system for social media data. In *Data Mining Workshops (ICDMW), 2011 IEEE 11th Conference on, Vancouver, BC, Canada, December 11, 2011*, pages 129–136. IEEE, 2011.

Xiaolin Zheng, Zhen Lin, Xiaowei Wang, Kwei-Jay Lin, and Meina Song. Incorporating appraisal expression patterns into topic modeling for aspect and sentiment word identification. *Knowledge-Based Systems*, 61(0):29 – 47, 2014. ISSN 0950-7051. doi: http://dx.doi.org/10.1016/j.knosys.2014.02.003. URL `http://www.sciencedirect.com/science/article/pii/S0950705114000434`.

Xiaodan Zhu, Svetlana Kiritchenko, and Saif M Mohammad. Nrc-canada-2014: Recent improvements in the sentiment analysis of tweets. *SemEval 2014*, page 443, 2014.