Target-Level Sentiment Analysis on Various Genres

Viktor Hangya

Supervisor: Dr. Richárd Farkas

Summary of the Ph.D. thesis

University of Szeged
Doctoral School of Computer Science

June 2019

Introduction

Due to the development of Web 2.0, a large amount of user generated data is produced daily. A significant portion of this data is in textual form which contains the opinions and sentiments of the authors. These contents are very useful to gather insights on the general opinion about various entities thus organizations started to exploit them in various applications, such as monitoring brands, predicting the results of elections and disaster management. Because of the large number of these textual contents, manual processing is not applicable, therefore an automatic analysis is needed.

More precisely, the task of sentiment analysis is to decide the sentiment polarity of a given sentence, which in most cases could be positive, negative or neutral. The general approach is to decide the overall sentiment value of a given sentence. However, when writing reviews about entities, people usually share their opinions about various properties of the given entity instead of writing just one global feeling about it. Furthermore, people often compare multiple entities in one sentence, which means that the same sentiment expression can be positive for one entity while it is negative for the other. Thus, *document-level* analysis is insufficient, in these cases it is important to be aware of the expressed sentiments for the different targets. The aim of *target-level* sentiment analysis, covered in the first part of the dissertation, is – given a pair of text and target entity – to classify the polarity of the expressed sentiments related to that target. We show the difficulties of the task compared to the traditional approach and contribute to the field by developing novel techniques for detecting relevant content and their sentiment more precisely, thus improving performance on this task.

An important aspect of sentiment analysis, and natural language processing in general, is the difference among various text domains. Texts from social media come from various genres and languages, which makes the creation of a broadly applicable sentiment analysis system nearly impossible because text styles can be very different. For example, considering texts coming only from social media platforms, blogs and news posts have the most standard language, they are the most correct in terms of spelling and usually there is no length limitation thus topics can be described in more detail. On the other hand, micro-blogs like Twitter have length limitations, which results in concise texts and they may contain a lot of slang words. In addition, they are used for quicker communication hence posts contain many misspellings. The second main topic of the dissertation focuses on the difficulties caused by domain, genre and language differences in the context of sentiment analysis. We show main differences of various domains and languages with a comparative analysis. We propose methods for domain adaptation of sentiment lexicons, which are important building blocks of systems, and for the adaptation of bilingual systems in low-resource setups.

Structure of the Dissertation

The goal of the dissertation is to push the boundaries of sentiment analysis methods. Various natural language processing techniques are introduced starting from specific preprocessing steps through feature engineering till different ways and levels of analysis. Most of the approaches are based on the more traditional manual feature engineering used with linear classifiers but techniques based on neural networks are also introduced.

As mentioned above, the dissertation comprises of two main topics involving 4 thesis points and an investigative chapter. The first part deals with target-level sentiment analysis while the second tackles the problem of handling genre, domain and language differences. In the following, the structure of the dissertation is briefly outlined along with the key publications of the author. A more detailed description of each chapter and the achieved results are discussed in the following sections.

Chapter 3 introduces the problems of target-level sentiment analysis. We propose novel techniques for the task based on both the surface form and the syntax of sentences, which help to detect target related information. It is shown that the introduced techniques reached state-of-the-art performance, moreover, the developed system performed among the best of the official participants of several shared tasks.

Most sentiment analysis systems predict one label on document- or target-level. Chapter 4 deals with fine-grained analysis, i.e. detecting the sentiments of each constituent of a sentence. Previous approaches relied on expensive resources, which made their application hard in many setups. To overcome this problem, a semi-supervised technique was developed and it is shown that by handling the sentiment of constituents as latent variables, performance can be improved. Results are presented on the traditional document-level sentiment analysis task and on target-level as well.

The second topic of this dissertation is introduced in Chapter 5. Genre, domain and language differences are serious issues of sentiment analysis. In this introductory chapter a detailed comparison of these differences is shown as the foundation of the last two chapters.

Chapter 6 focuses on the shortcomings of sentiment lexicons over various domains. The same expression can often have opposite sentiment polarities in different domains. To overcome this problem, methods for lexicon creation and adaptation are introduced using labeled and unlabeled lexical resources. Results are shown on Hungarian datasets.

Finally, chapter 7 addresses the problems of cross-lingual sentiment analysis related to domain differences. We introduce a semi-supervised method for domain adaptation which does not rely on additional labeled data. We show that both semantic knowledge from general domain data and domain specific information is needed to achieve good performance in bilingual transfer learning.

Table 1 summarizes the relationship among the thesis chapters and the key referred publications of the author.

			Chapters				
			3	4	5	6	7
CLEF	2013	(Hangya and Farkas, 2013a)	•				
CogInfoCom	2013	(Hangya and Farkas, 2013b)	•				
ICCIA	2017	(Hangya et al., 2017)		•			
SEMEVAL	2013	(Hangya et al., 2013)			•		
AIRE	2017	(Hangya and Farkas, 2017)	•		•		
TSD	2015	(Hangya, 2015)				•	
ACL	2018	(Hangya et al., 2018)					•

Table 1. The relation between the thesis topics and the corresponding publications. In the results presented in Chapters 3, 5 and 6 the author's contribution was prominent while Chapters 4 and 7 are the results of joint works with other researchers.

Target-Level Sentiment Analysis

Chapter 3 introduces the task of target-level sentiment classification where instead of the detection of global sentiments in texts, the task is to classify only those sentiments which are related to a given target. This application is often employed by various entities such as product owners or public figures in order to monitor users' sentiments towards their products or actions. Various difficulties have to be overcome when developing such systems, e.g. selecting only the relevant parts of texts in case they contain sentiments related to different targets.

The author proposed a competitive system for target-level analysis and evaluated on multiple genres and languages. In order to reach good performance on tweets, a Twitter specific preprocessing was introduced. In order to make a document-level classifier target-aware, various target-level features were introduced. For detecting relevant parts of input texts both surfaceform- and syntax-based methods were proposed. The former includes re-weighting the importance of features describing the input sentences based on the position of the target. In addition, several features capturing a priori sentiment information about target entities and topics are proposed. In contrast, syntax-based methods rely on the syntactic structure of the sentences, in the form of constituency and dependency parses, in order to emphasize related information.

The competitiveness of the system was shown on both Twitter and more standard genres and also on Hungarian texts besides English ones. It was shown that surfaceform based methods perform better in the case of noisy texts, such as tweets, while syntax-based approaches help more for well formed inputs. Furthermore, the combination of the two methods gives the best performance. In addition to the experiments performed on various text sources, task-specific systems were also introduced, which ranked among the best systems in shared tasks, e.g. reaching 1^{st} place in the case of the RepLab 2013 shared task. Related to his publications (Hangya and Farkas, 2013a,b, 2017), the author regards the following results as his main contributions to the field:

- Surfaceform-based target-level feature engineering for emphasizing relevant content on less well formed texts;
- Syntax-based feature engineering, using both dependency and constituency parses, for target specific feature extraction on more standard texts;
- A hybrid system using both of the above techniques reaching best overall performance and its error analysis on various datasets.

Fine-Grained Sentiment Analysis

Chapter 4 performs sentiment analysis on a more fine-grained level. Most of the sentiment classification approaches predict one sentiment label for a sentence or document regardless of performing document- or target-level analysis. The next step in terms of granularity is to apply the analysis on phrase-base, i.e. to predict a sentiment score for each word, phrase and sub-sentence in a given sentence and for the whole sentence as well in a bottom-up fashion. The benefit of such analysis is twofold. First, by exploiting the sentiment scores of building blocks of a sentence and their composition, e.g. they can be contrastive or one can intensify the other, the system can understand the sentences better, thus improving performance on the document-level. Second, fine-grained analysis can be used for other tasks, e.g. target-level sentiment analysis, by creating features from the fine-grained output.

Previous approaches for this task showed that this approach can improve the sentence-level performance but they used either fine-grained annotated data, which is expensive to create and only accessible for a restricted set of domains and languages or language specific rules that are also hard to adapt to other use cases. The author together with his colleagues proposed a latent syntactic structure based system relying only on sentence-level annotations making the system easily applicable to any languages and domains. During training, only the sentiment value of the whole sentence is given while the rest of the values are handled as latent variables whose distribution is learned using an iterative procedure. The method employs general feature templates for capturing sentiment and compositionality of phrases making it easily applicable to many setups.

The performance of the system is tested on both document-level and target-level tasks and it is shown that the proposed system improves classification performance. Although other systems having access to fine-grained annotations performed better on the document-level, it was shown that the proposed method is more effective on domains where such data is not accessible, which is the case for most setups. The output of the proposed system was exploited for the target-level sentiment classification task, by extracting target-specific features from the generated sentiment trees, resulting performance improvement for this task as well. Related to his publication (Hangya et al., 2017), the author regards the following results as his main contributions to the field:

- Introduction of latent syntactic structure-based fine-grained sentiment analysis method using only sentence-level annotations as opposed to previous work;
- Improvement of document-level sentiment classification by exploiting sentence compositionality;
- Incorporating fine-grained analysis into target-level classification by proposing sentiment tree-based features.

Sentiment Analysis on Various Genres

Chapter 5 is an introduction to the second main topic of the dissertation. Domain differences represent a huge problem in general for machine learning. Systems developed and trained in one set of input might not perform well in the case of domain shift. The author comparatively analyzed sentiment classification on document-level English and Hungarian datasets coming from different domains and genres. As a basis for the next chapters, he pinpointed main issues when applying systems on various domains.

Chapter 5 is not a key thesis point in the dissertation. Related to his publications (Hangya et al., 2013; Hangya and Farkas, 2017), the author regards the following result as his main contribution to the field:

- In-depth quantitative and qualitative comparison of document-level sentiment analysis on various domains, genres and languages.

Domain Specific Sentiment Lexicons

Chapter 6 focuses on the creation and domain adaptation of sentiment lexicons which are key external resources for sentiment analyzers. These lexicons contain words with their sentiment polarities and can be used for adding external knowledge to the feature extraction phase of systems. As mentioned before, text domains can differ to a large extent, which means that a given expression can convey different sentiments in different domains. For example, the word *loud* has a positive meaning in the speaker reviews domain while it is negative in the kitchen appliances domain. Since creating sentiment lexicons is time consuming and expensive, most of them are general purpose resources lacking domain specific knowledge.

The author shows that even a good quality lexicon can have negative effects on the performance if used in the wrong domain. To overcome this limitation, automatic methods were proposed to create and adapt lexicons to improve performance in a given domain. A method was proposed for extending an initial small lexicon by propagating sentiment polarity values based on various word relation information. In addition, a method creating sentiment lexicons from scratch using statistical information in annotated data was developed as well.

Since most lexicons are available in English, the author experimented with Hungarian sentiment analysis. The quality of sentiment lexicons were compared using the sentiment analysis task by relying on lexicon based features. It was shown that by extending a cheap domain specific seed lexicon using different lexical resources or by building one from scratch relying on annotated data, better performance can be reached compared to when using out of domain manually created lexicons. Related to his publication (Hangya, 2015), the author regards the following results as his main contributions to the field:

- Language independent automatic methods for creating domain-specific sentiment lexicons: extending a small seed lexicon and building from scratch using annotated data;

- Showing that domain specific lexicons are crucial in order to achieve good performance on Hungarian datasets.

Cross-Lingual Domain Adaptation for Sentiment Analysis

Chapter 7 introduces methods developed for domain adaptation for cross-lingual sentiment classification. For many languages there is an insufficient amount of labeled data to build good quality sentiment classifiers. Word embeddings gained a lot of attention in the neural network era because semantic information in unlabeled corpora can be exploited and the performance of systems relying on them can be improved. By representing words in a shared vector space, bilingual word embeddings can to some extent bridge the gap between languages. By relying on these resources, cross-lingual transfer learning made it possible to build models using annotations from a resource rich language and apply it to resource poor ones.

Although these methods work well in many scenarios, they often have low quality when the training and test data come from different domains. The author together with his colleagues showed that bilingual word embeddings built from general domain resources are lacking in-domain specific words and represent the incorrect meaning of some others. To overcome this issue, a two step domain adaptation method was proposed relying only on monolingual unlabeled data in contrast to other works. The method first adapts bilingual word embeddings to the target domain by enriching it with domain specific information. In the second step a semi-supervised approach incorporates further domain specific unlabeled data to improve cross-lingual knowledge transfer.

The performed experiments showed that by adapting bilingual word embeddings to the target domain, performance of cross-lingual sentiment classification can be improved. In addition to relying only on monolingual data, it was shown that the method is also easy to use for other tasks as well, such as bilingual lexicon induction. Related to his publication (Hangya et al., 2018), the author regards the following results as his main contributions to the field:

- Simple, yet effective method for domain adaptation of bilingual word embeddings;
- Semi-supervised method for incorporating unlabeled task specific target language/domain data to improve cross-language knowledge transfer;
- Domain adaptation of cross-lingual tasks using unlabeled data only.

Bibliography

Viktor Hangya. 2015. Automatic Construction of Domain Specific Sentiment Lexicons for Hungarian. In *Proceedings of the 18th International Conference on Text, Speech and Dialogue*, pages 201–208.

Viktor Hangya, Gábor Berend and Richárd Farkas. 2013. SZTE-NLP: Sentiment Detection on Twitter Messages. In *Proceedings of the 7th International Workshop on Semantic Evaluation*, pages 549–553.

Viktor Hangya, Fabienne Braune, Alexander Fraser and Hinrich Schütze. 2018. Two Methods for Domain Adaptation of Bilingual Tasks: Delightfully Simple and Broadly Applicable. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, pages 810–820.

Viktor Hangya and Richárd Farkas. 2013a. Filtering and Polarity Detection for Reputation Management on Tweets. In *Working Notes of the 2nd Conference and Labs of the Evaluation Forum.*

Viktor Hangya and Richárd Farkas. 2013b. Target-oriented opinion mining from tweets. In *Proceedings of the 4th IEEE International Conference on Cognitive Infocommunications*, pages 251–254.

Viktor Hangya and Richárd Farkas. 2017. A comparative empirical study on social media sentiment analysis over various genres and languages. *Artificial Intelligence Review*, 47(4):485–505.

Viktor Hangya, Zsolt Szántó and Richárd Farkas. 2017. Latent Syntactic Structure-Based Sentiment Analysis. In *Proceeding of the 2nd IEEE International Conference on Computational Intelligence and Applications.*