Research Statement

Veselin Stoyanov

Overview

Opinion analysis (or *subjectivity analysis*) is concerned with extracting information about attitudes, beliefs, emotions, opinions, evaluations and sentiment expressed in texts. The field has received much recent attention, driven by the opportunity to create practical applications that automatically mine opinion information from the large amount of available electronic text on the Web (e.g. Popescu and Etzioni (2005) and Hu and Liu (2004)). Additionally, research has been motivated by the potential to effectively use opinion information in existing natural language processing (NLP) applications such as information retrieval and question answering (e.g. Riloff et al. (2005), Stoyanov et al. (2004) and Stoyanov et al. (2005)).

More precisely, my research falls in the area of *fine-grained opinion analysis*, which is concerned with extracting opinions at the level of sentences, clauses, or individual expressions of opinions. Each fine-grained opinion is characterised by a number of attributes: the words that signal an expression of opinion (the *opinion expression*), the *source* or *opinion holder*, the *topic* and the *polarity* of the opinion. Consider, for example the sentence:

(1)[$_{\text{OH}}$ John] likes [$_{\text{Topic}}$ New York] for its cultural diversity.

This example contains an opinion signaled by the OPINION EXPRESSION "likes". The SOURCE of the opinion is "John", the TOPIC is "New York" and the POLARITY is positive.

Using the above definitions, the next three sections discuss my research contributions as well as my future research plans.

Opinion Analysis

Research in the area of *opinion analysis* to date has concentrated on developing methods for the automatic extraction of opinions as well as their attributes. Previous efforts have focused on the extraction of opinion expressions in context (e.g. Bethard et al. (2004), Breck et al. (2007)), the assignment of polarity to these expressions (e.g. Wilson et al. (2005), Kim and Hovy (2006)), source extraction (e.g. Bethard et al. (2004), Choi et al. (2005)), and identification of the source-expresses-opinion relation (e.g. Choi et al. (2006)), i.e. linking sources to the opinions that they express. For different reasons, two important areas of fine-grained opinion analysis lack effective approaches: identifying opinion topics and opinion summarization. My research in opinion analysis has focused on addressing these two areas as discussed below.

Opinion Topics. Extracting topics of fine-grained opinions has proven to be a challenging task (Wiebe, 2005; Wilson, 2005) due to several issues described in Stoyanov and Cardie (2008a) and Stoyanov and Cardie (2008b). As a result, previous work has not been able to provide effective definitions of fine-grained opinion topics for general opinions or present successful methods for automatically extracting topics of such opinions.[1]

[1] An exception is the related area of opinion extraction from product reviews, where several approaches have been proposed (e.g. Kobayashi et al. (2004), Yi et al. (2003), Popescu and Etzioni (2005), Hu and Liu (2004)). However, topic extraction in the product review domain can be effectively substituted with a lexicon look-up, which is not the case for general fine-grained opinions.

[Source Zacarias Moussaoui] [– *complained*] at length today about [Topic his own lawyer], telling a federal court jury that [Topic he] was [– *more interested in achieving fame than saving Moussaoui's life*].

Mr. Moussaoui said he was appearing on the witness stand to tell the truth. And one part of the truth, [Source he] said, is that [Topic sending him to prison for life] would be "[– *a greater punishment*] than being sentenced to death."

"[– [Topic *You*] *have put your interest ahead of* [Source *my*] *life*]," [Source Mr. Moussaoui] told his court-appointed lawyer Gerald T. Zerkin.

...

But, [Source Mr. Zerkin] pressed [Topic Mr. Moussaoui], was it [– *not true*] that he told his lawyers earlier not to involve any Muslims in the defense, not to present any evidence that might persuade the jurors to spare his life?

...

[Source Zerkin] seemed to be trying to show the jurors that while [Topic the defendant] is generally [+ *an honest individual*], his conduct shows [Topic he] is [– *not stable mentally*], and thus [– *undeserving*] of [Topic the ultimate punishment].

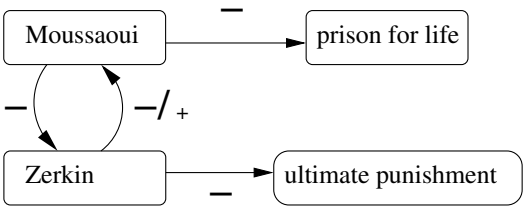

Figure 1: Example text containing opinions (above) and a summary of the opinions (below).

Part of my task, therefore, was to provide a new, operational definition of *opinion topic* in which the topic of an opinion depends on the context in which its associated opinion expression occurs (Stoyanov and Cardie, 2008a; Stoyanov and Cardie, 2008b). Using this definition, I extended an existing fine-grained opinion corpus (the MPQA corpus (Wiebe et al., 2005)) with manual annotations that encode topic information (Stoyanov and Cardie, 2008a). This work also suggested a novel method for general-purpose opinion topic identification that, following our new definition, treats the problem as an exercise in topic coreference resolution (Stoyanov and Cardie, 2008b). We evaluated our approach using the topic-annotated portion of the MPQA corpus.

Inter-annotator agreement results for the manual annotations are reasonably strong across a number of metrics and the results of our experiments are promising: using either automatically or manually identified topic spans, we achieve topic coreference scores that statistically significantly outperform two topic segmentation baselines across three coreference resolution evaluation measures (B^3, Krippendorff's α and CEAF). For the B^3 metric, for example, the best baseline achieves a topic coreference score of 0.55 while our topic coreference algorithm scores 0.57 and 0.71 using automatically, and manually, identified topic spans, respectively.

Opinion Summarization. While opinion information as extracted by existing systems (i.e., raw opinion information) is useful, the true potential of this information can be realized only after the raw information is consolidated in a meaningful way. I will use the term opinion summarization to describe the process of meaningfully aggregating opinions. An example of an opinion summary is shown in Figure 1. The example

shows a text segment containing fine-grained opinions (above) and a summary of those opinions (below). In the text, sources and topics of opinions are bracketed; opinion expressions are shown in italics and bracketed with their associated polarity, either positive (+) or negative (-). In the summary, entities involved in opinions are shown as nodes and aggregated opinions are shown as directed edges. Opinions from the same source on the same topic are combined, statistics are computed for each source/topic, and multiple opinions from the same source on the same topic are aggregated.

The task of opinion summarization for open domain opinions has received little research attention. Part of my dissertation research addresses this void. Specifically, we define two general representations for opinion summaries, identify the problems that need to be addressed by opinion summarization systems, develop methods to address these problems, introduce novel quantitative evaluation metrics for opinion summaries and construct and evaluate full opinion summaries for the documents in a standard opinion-oriented corpus (the aforementioned MPQA corpus). The problems that we identify and address are (1) *source coreference resolution* (Stoyanov and Cardie, 2006b; Stoyanov and Cardie, 2006a), the problem of linking together opinions that belong to the same source and (2) *topic coreference resolution* (Stoyanov and Cardie, 2008a; Stoyanov and Cardie, 2008b), the problem of linking together opinions that share the same topic. Results for the evaluation of complete summaries have been submitted for publication (Stoyanov and Cardie, 2009).

Future Work For the problem of opinion topic identification, I plan to evaluate several alternative approaches to defining and extracting opinion topics. In particular, we chose an approach to opinion topics in which it as assumed that each opinion has one "main" topic. An alternative approach is to allow each opinion to concern several relevant topics. Both approaches have their advantages and disadvantages and I plan to explore the multiple-topics-per-opinion alternative.

In the area of opinion summarization, I plan to perform experimental evaluation aimed at understanding the impact of different subsystems and incorporating different opinion topic definitions. Based on these experiments, I plan to attempt different approaches to the underlying problems of source and topic coreference resolution. Additionally, I plan to evaluate the usefulness of opinion summaries by extending existing NLP applications and by building user interfaces that will allow opinion summaries to be presented and manipulated by a user.

Noun Phrase Coreference Resolution

As part of my work on opinion summarization, I address the problems of source and topic coreference resolution. Both of these problems are similar to the task of *noun phrase coreference resolution* or simply *coreference resolution*, which is defined as the problem of identifying all noun phrases (NPs) in a document that refer to the same real-world entity. As I became familiar with research in the area of coreference resolution, two serious issues became obvious:

1. In spite of the availability of several benchmark data sets (e.g. MUC-6 (1995), ACE NIST (2006)) and their use in many formal evaluations, there are surprisingly few conclusive statements about the state-of-the-art in NP coreference resolution. In particular, it remains difficult to compare the effectiveness of different approaches to coreference resolution even in relative terms, results vary widely across data sets and there is little understanding of which aspects of the coreference resolution problem are handled well or poorly by state-of-the-art systems.

2. There are few publicly available research software tools for coreference resolution.

To address the issues above I initiated an effort to create a coreference resolution research platform. The goal of the effort is twofold: to study the state-of-the-art in coreference resolution and to provide the research community with much needed software tools. We aim for a coreference resolution research platform that satisfies the following characteristics: (i) implements the basic underlying architecture of contemporary

coreference resolution systems, (ii) includes several state-of-the-art coreference algorithms, (iii) can run on the majority of the standard coreference resolution data sets, (iv) implements most standard coreference resolution scoring algorithms and, (v) is relatively fast, easy to configure and run and can be extended with new methods and features with little effort. With the help of other researchers at the University of Utah and the Lawrence Livermore National Laboratory we created Reconcile, a coreference resolution research platform that meets these five criteria.

With Reconcile in hand, we were then able to study the state of the art along three axes. First, we examined three subproblems that play a role in coreference resolution: named entity recognition, anaphoricity determination, and coreference element detection. We measured the impact of each subproblem on coreference resolution, and confirmed that certain assumptions regarding these subproblems in the evaluation methodology can dramatically simplify the overall task. Second, we measured the performance of a state-of-the-art coreference resolver on several classes of anaphora and used these results to develop a quantitative measure for estimating coreference resolution performance on new data sets. Third, we are now performing a comparative study of the relative performance of different state-of-the-art coreference resolution methods and approaches. Results of the first two studies have been published (Stoyanov et al., 2009) and we are in the process of concluding and publishing the third study.

Additionally, we are currently preparing Reconcile for release to the research community. Even at this early stage, there has been significant interest in using Reconcile both as a research tool as well as a black-box coreference resolution system.

Future Work First, I plan to extend and complete the comparative study of state-of-the-art approaches to coreference resolution. Subsequently, I plan to use the information from our study and address the aspects of coreference resolution that are not handled well by existing systems. Finally, I plan to perform theoretical and empirical comparison of existing coreference resolution evaluation metrics.

Information Extraction: Future Work

Fine-grained opinion extraction bears many similarities to tasks in information extraction, which is the problem of automatically populating scenario templates for a given event or entity type. For example, an IE application targeting disasters may have to detect mentions of disasters and populate a template including fields such as the date, time, location and the damage caused by the disaster. This is analogous to extracting fine-grained opinions with their attributes such as the source, topic and polarity of the opinion. In fact, fine-grained sentiment analysis can be considered an information extraction task.

Similar to fine-grained opinion information, automatically extracted information in other domains needs to be aggregated to be completely useful. IE algorithms often extract partially filled templates, which need to be effectively consolidated, subject to solving problems similar, in many respects, to the tasks in opinion summarization that I have addressed.

My concrete plans in the area of information extraction include applying techniques that I have used for opinion summarization to other IE tasks. For example, I believe that our approach to source coreference resolution as *partially supervised clustering* can be used effectively for other domains such as extracting information about events.

Additionally, I have general interest in the area of information extraction. I believe that the advent of the Semantic Web presents an unprecedented opportunity to develop IE approaches with immediate practical impact.

May 14, 2009

References

ACE. 2006. ACE 2005 evaluation, November.

S. Bethard, H. Yu, A. Thornton, V. Hativassiloglou, and D. Jurafsky. 2004. Automatic extraction of opinion propositions and their holders. In *2004 AAAI Spring Symposium on Exploring Attitude and Affect in Text*.

Eric Breck, Yejin Choi, and Claire Cardie. 2007. Identifying expressions of opinion in context. In *Proceedings of IJCAI*.

Y. Choi, C. Cardie, E. Riloff, and S. Patwardhan. 2005. Identifying sources of opinions with conditional random fields and extraction patterns. In *Proceedings of EMNLP*.

Yejin Choi, Eric Breck, and Claire Cardie. 2006. Joint extraction of entities and relations for opinion recognition. In *Proceedings of EMNLP*.

Minqing Hu and Bing Liu. 2004. Mining opinion features in customer reviews. In *AAAI*, pages 755–760.

Soo-Min Kim and Eduard Hovy. 2006. Extracting opinions, opinion holders, and topics expressed in online news media text. In *Proceedings of ACL/COLING Workshop on Sentiment and Subjectivity in Text*.

Nozomi Kobayashi, Kentaro Inui, Yuji Matsumoto, Kenji Tateishi, and Toshikazu Fukushima. 2004. Collecting evaluative expressions for opinion extraction. In *IJCNLP*.

Ana-Maria Popescu and Oren Etzioni. 2005. Extracting product features and opinions from reviews. In *HLT/EMNLP*.

E. Riloff, J. Wiebe, and W. Phillips. 2005. Exploiting subjectivity classification to improve information extraction. In *Proceedings of AAAI*.

V. Stoyanov and C. Cardie. 2006a. Partially supervised coreference resolution for opinion summarization through structured rule learning. In *Proceedings of EMNLP*.

V. Stoyanov and C. Cardie. 2006b. Toward opinion summarization: Linking the sources. In *Proceedings of the ACL Workshop on Sentiment and Subjectivity in Text*.

V. Stoyanov and C. Cardie. 2008a. Annotating topics of opinions. In *Proceedings of LREC*.

V. Stoyanov and C. Cardie. 2008b. Topic identification for fine-grained opinion analysis. In *Proceedings of COLING*.

V. Stoyanov and C. Cardie. 2009. Generating and evaluating summaries of fine-grained opinions. In *submission*.

V. Stoyanov, C. Cardie, J. Wiebe, and D. Litman. 2004. Evaluating an opinion annotation scheme using a new Multi-Perspective Question and Answer corpus. In *2004 AAAI Spring Symposium on Exploring Attitude and Affect in Text*.

V. Stoyanov, C. Cardie, and J. Wiebe. 2005. Multi-Perspective question answering using the OpQA corpus. In *Proceedings of EMNLP*.

Veselin Stoyanov, Claire Cardie, Nathan Gilbert, and Ellen Riloff. 2009. Conundrums in coreference resolution: Making sense of the state of the art. In *Proceedings of ACL (to appear)*.

Marc Vilain, John Burger, John Aberdeen, Dennis Connolly, and Lynette Hirschman. 1995. A model-theoretic coreference scoring scheme. In *Proceedings of the 6th conference on Message understanding.*

Janyce Wiebe, Theresa Wilson, and Claire Cardie. 2005. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, 1(2).

J. Wiebe. 2005. Personal communication.

Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of HLT/EMNLP.*

T. Wilson. 2005. Personal communication.

Jeonghee Yi, Tetsuya Nasukawa, Razvan C. Bunescu, and Wayne Niblack. 2003. Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques. In *Proceedings of ICDM.*