TripleSent: a Triple Store of Events Associated with their Prototypical Sentiment

Veronique Hoste[1], Els Lefever[1],Stephan van der Waart van Gulik[2] and Bart Desmet[1]

[1]LT3 Language and Translation Technology Team
[2]Centre for Logic and Philosophy of Science
Ghent University, Belgium
email: firstname.lastname@ugent.be

***Abstract*—The current generation of sentiment analysis systems is limited in their real-world applicability because they cannot detect utterances that implicitly carry positive or negative sentiment. We present early stage research ideas to address this inability with the development of a dynamic triple store of events associated with their prototypical sentiment.**

Keywords—sentiment detection; triple store; implicit sentiment; natural language processing.

I. INTRODUCTION

In the last decades, state-of-the-art research in natural language processing (NLP) has made a shift from rule-based to statistical corpus-based approaches, which require high-quality electronic text corpora. Supervised and unsupervised statistical approaches to structure and interpret patterns in text and speech have been successfully developed on such corpora. Examples include part-of-speech taggers, parsers, named entity recognition, machine translation, speech recognition, text classification and summarization, sentiment analysis, etc. Some of these tasks can be performed with near-human accuracy (e.g., part-of-speech tagging), whereas for more complex tasks, such as sentiment analysis, performance is limited by the amount of available knowledge.

In sentiment analysis, the objective is to automatically determine the sentiment (positive, neutral or negative) expressed in an utterance, e.g., (a) *"I love to go shopping"*, (b) *"Coke tastes great",* (c) *"I bought the mattress a week ago, and a valley has formed"*. Most state-of-the-art sentiment analysis systems combine a statistical approach with lists of subjective words (*"love"*, *"great"*), such as the MPQA (Multi-Perspective Question Answering) [1] lexicon. As a result, they are capable of detecting expressions of sentiment only if they can learn them from annotated corpora or sentiment lexicons. While current sentiment analyzers can deal with expressions that address sentiments explicitly, as in examples (a) and (b), they **struggle with sentiments that are only implicitly present in so-called polar facts**, as is the case in example (c) [2]. Current systems fail to detect polar facts, which implicitly carry positive or negative sentiment. This is problematic, because implicit sentiment has been shown to account for more than half of the sentiment in certain domains (e.g., product reviews, *"Web surfing drains the battery"*, or financial reporting, *"Fed lowers interest rates"*) [3]. Progress in the automatic detection of ironic utterances such as *"Going to the dentist tomorrow yippee",* in which the expressed sentiment is not to be understood in its literal sense, also suffers from the lack of common sense knowledge [4][5].

As this severely limits the real-world applicability of the current generation of sentiment analyzers, we aim to investigate the feasibility of developing a dynamic triple store of events associated with their prototypical sentiment. Such common-sense knowledge could then complement other knowledge sources (e.g., sentiment lexicons) and other types of features derived from training data in a classification-based approach to sentiment analysis or irony detection.

Knowledge bases, such as WordNet, DBPedia, Freebase, OpenCyc, SUMO and Open Mind Common Sense, which store and structure lexical and factual knowledge in machine-readable formats, have been instrumental for the success of complex language understanding applications, such as the IBM Watson question answering system [6]. They are an essential resource for tasks that involve factual analysis, such as summarization, wikification, question answering and textual entailment. For sentiment analysis, however, there is an additional need for knowledge about the prototypical sentiments people hold towards entities and events. As "prototypical" sentiment, we consider sentiments that are commonly associated with a certain event, an event being the combination of a verb and a direct, indirect or prepositional object. Certain events may entail multiple prototypical sentiments, depending on perspective. As an example, the sentence *"Fed lowers interest rates"* will be considered prototypically positive for people who want to take out a loan, but it can also be considered negative in that it may cause inflation.

The remainder of this ideas paper is organized as follows. In Section 2, we propose the methodology we intend to use to build a knowledge base of events and their prototypical sentiment. In the last section, we present some prospects for future work beyond the construction of the knowledge base.

II. RESEARCH OBJECTIVES

We conceive TripleSent as consisting of two interacting layers: a **knowledge base** and a **reasoner**. The knowledge base contains events for which the prototypical sentiment is known with a high certainty. This information is stored in the form of sentiment triples. For example, the negative sentiment commonly associated with *"going to the dentist"* can be formally captured by the sentiment triple <visit-dentist, has-sentiment, negative> (note that there is some notational abuse here to facilitate the reader). The reasoner, on the other hand, is capable of inferring sentiment for events that are not stored in the database. When a user asks for the prototypical

 ISBN: 978-1-61208-472-5

sentiment for "visit the oncologist", the reasoner combines information from factual knowledge bases like WordNet [7] (which knows that oncologists, like dentists, are a kind of doctor) with the sentiment information from the triple store, to (conditionally) infer the expected sentiment triple <visit-oncologist, has-sentiment, negative>. Some of the inferences can be truly 'conditional' because whenever new, more reliable information contradicting the inferred triple is added or generated, the reasoner will need to revoke the inference (and all other inferences that rely on it). Like human reasoning, this requires a non-monotonic logic approach (see Objective 2).

Objective 1: Event extraction and enrichment

To kick-start the knowledge base, events will be collected for which the sentiment is known. These events will be obtained by extracting patterns for highly explicit sentiment expressions (e.g., *"I hate"* or *"I love"*) or from large web data crawls (e.g., commoncrawl.org), which will subsequently be syntactically and semantically parsed to extract events and sentiment triples. In the same vein, we will investigate leveraging existing large parsed datasets to extract high-confidence sentiment triples with minimal human intervention, using pattern-based and supervised sentiment analysis techniques [8]. Events for which both polarities are found frequently in the data will initially not be considered for further processing and will be investigated in more detail to understand the nature of this ambiguity. Given the linguistic diversity with which events can be expressed, the usefulness of the resulting triple store will also heavily depend on the ability to automatically handle orthographic variation (as for example in *"pediatrician"*, *"paediatrician"* or *"pediatrist"*), and syntactic and semantic synonymous structures (e.g., *"visit"*, *"going to"*, *"seeing"*, etc. *"a pediatrician"*).

In order to allow for the creation of new sentiment triples, explicit sentiment triples present in the knowledge base will be linked to ontological information provided by lexical resources and factual knowledge bases such as WordNet and DBPedia, respectively.

Objective 2: Opinion inferencing

The reasoner can infer all kinds of new sentiment triples from already known triples using (decidable) fragments of first-order predicate logic. However, in order to enable TripleSent to also deal with the expected sentiment for events that are not yet stored in the database, the reasoner should allow dynamic, conditional inferences of unseen triples. For example, starting from the explicit sentiment triple <visit-oncologist, has-sentiment, negative>, the reasoner relies on WordNet information like <oncologist, is-a, medical specialist> to (provisionally) derive <visit-medical-specialist, has-sentiment, negative>, and, again by relying on WordNet information, to (provisionally) derive <visit-dentist, has-sentiment, negative> and <visit-podologist, has-sentiment, negative>. Note that the last sentiment attribution is debatable, and can be revoked in the (future) presence of other, more reliable triples (stating explicitly, for example, that prototypical visits to podologists are not negative). For the implementation of this type of reasoning, we will evaluate different non-monotonic logic approaches, such as default logic [9], adaptive logics [10] or answer set programming [11].

In order to evaluate the event extraction, event enrichment and opinion inferencing, we will manually annotate test corpora by relying both on expert annotators and crowdsourcing. For the evaluation of the event extraction, we will assess precision both for the event extraction and the sentiment attached to these events. In order to also enable the measuring of recall, we will furthermore rely on an existing corpus for irony detection annotated with event-sentiment annotations [12]. As in previous annotation efforts, it was shown that crowdsourcing is a reliable and very cost-effective means of collecting human knowledge, we will also investigate the use of a **crowdsourcing methodology to validate and enrich the output of the platform**. Inferred sentiment triples will be presented to a crowd of human annotators who indicate what they consider to be the prototypical sentiment for the given event. This could provide additional high-confidence triples to be stored, contradicting evidence to inform non-monotonic decisions (e.g., exceptions such as <visit-podologist, has-sentiment, neutral>), and grounding that can be used in a feedback loop to improve the inference engine.

III. CONCLUSION AND FUTURE WORK

To date, there is a complete lack of reusable and dynamically growing knowledge bases linking events to implicit sentiment, which can be used for research and development in opinion inferencing. The TripleSent platform including the knowledge base and the automatic reasoner will open new perspectives in NLP research and can push the state-of-the-art in semantic text processing and inferencing, and more specifically in NLP applications such as sentiment analysis and irony detection.

REFERENCES

[1] J. Wiebe, T. Wilson, and C. Cardie, "Annotating expressions of opinions and emotions in language", Language Resources and Evaluation, vol. 39, issue 2-3, 2005, pp. 165-210.

[2] B. Liu, "Sentiment Analysis and Opinion Mining", Morgan & Claypool Publishers, May 2012.

[3] M. Van de Kauter, B. Desmet, and V. Hoste, "The good, the bad and the implicit: a comprehensive approach to annotating explicit and implicit sentiment", Language Resources and Evaluation, Springer Netherlands, vol. 49, 2015, pp. 685-720.

[4] E. Riloff, et al., "Sarcasm as Contrast between a Positive Sentiment and Negative Situation", Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (EMNLP 2013), 2013, pp. 704-714.

[5] C. Van Hee, E. Lefever, and V. Hoste, "LT3: Sentiment Analysis of Figurative Tweets: piece of cake #NotReally", Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), 2015, pp. 684-688.

[6] D. Ferrucci, et al., "Building Watson: An overview of the DeepQA project", AI Magazine, vol. 31, no. 3, 2010, pp. 59-79.

[7] C. Fellbaum, "WordNet: An Electronic Lexical Database", Cambridge, MA: MIT Press, 1998.

[8] C. Van Hee, M. Van de Kauter, O. De Clercq, E. Lefever and V. Hoste, "LT3: Sentiment Classification in User-Generated Content Using a Rich Feature Set", Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), 2014, pp. 406-410.

[9] R. Reiter, "A logic for default reasoning", Artificial Intelligence, vol. 13, no. 1, 1980, pp. 81-132.

[10] D. Batens, "A General Characterization of Adaptive Logics", Logique et Analyse, vol. 44, no. 173-175, 2003, pp. 45-68.

[11] M. Blondeel, S. Schockaert, D. Vermeir, and M. De Cock, "Fuzzy Answer Set Programming: An Introduction", Soft Computing: State of the Art Theory, vol. 291, 2013, pp. 209-222.

[12] C. Van Hee, E. Lefever, and V. Hoste, "Exploring the Realization of Irony in Twitter Data", Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16). European Language Resources Association (ELRA), accepted for publication.

[13] O. De Clercq, et al., "Using the Crowd for Readability Prediction". Natural Language Engineering, vol. 20, no. 3, 2014, pp. 293-235.