A Semantic Web Based Core Engine to Efficiently Perform Sentiment Analysis

Diego Reforgiato Recupero[1(✉)], Sergio Consoli[1], Aldo Gangemi[1,2], Andrea Giovanni Nuzzolese[1,3], and Daria Spampinato[1]

[1] STLab-ISTC Consiglio Nazionale delle Ricerche, Rome, Italy
{diego.reforgiato,sergio.consoli,aldo.gangemi,andrea.nuzzolese, daria.spampinato}@istc.cnr.it
[2] LIPN, University Paris 13, Sorbone Cit'e, UMR CNRS, Paris, France
[3] Department of Computer Science and Engineering, University of Bologna, Bologna, Italy

Abstract. In this paper we present a domain-independent framework that creates a sentiment analysis model by mixing Semantic Web technologies with natural language processing approaches (This work is supported by the project PRISMA SMART CITIES, funded by the Italian Ministry of Research and Education under the program PON.). Our system, called Sentilo, provides a core sentiment analysis engine which fully exploits semantics. It identifies the holder of an opinion, topics and sub-topics the opinion is referred to, and assesses the opinion trigger. Sentilo uses an OWL opinion ontology to represent all this information with an RDF graph where holders and topics are resolved on Linked Data. Anyone can plug its own opinion scoring algorithm to compute scores of opinion expressing words and come up with a combined scoring algorithm for each identified entities and the overall sentence.

Keywords: Sentic computing · Sentiment analysis · Semantic features

1 Introduction

Sentiment Analysis is a research area that involves the analysis of people's sentiments, opinions, emotions towards entities such as products, movies, services, etc. It is one of the hottest problems which belongs to the Natural Language Processing field which has been investigated only starting from the year 2000. So far, Sentiment Analysis approaches have used statistical classifiers, natural language processing techniques, data mining and lexical resources to identify the tone of a given sentence respect to a certain topic. For example, given the following opinion: *"Joy Ride is not an interesting film but the director John Dahl made a perfect work for his audience"*; an ideal system would be able to identify several topics referred to by such opinionated sentence. *"Joy Ride"* is certainly one, the *"work of John Dahl"* associated with this movie is another one, and finally *"John Dahl"*. Additionally, such ideal system would be able to analyze that the

V. Presutti et al. (Eds.): ESWC Satellite Events 2014, LNCS 8798, pp. 245–248, 2014.
DOI: 10.1007/978-3-319-11955-7_28

sentiment expressed on *"Joy Ride"* is negative, while the sentiment expressed on the work of *"John Dahl"*, and on *"John Dahl"* himself is slightly positive, and that the whole sentence carries both positive and negative sentiments.

The goal of Sentiment Analysis is to detect quintuples $(e_j, a_{jk}, so_{ijkl}, h_i, t_l)$ from unstructured text where e_j is the topic, a_{jk} is the aspect/feature of the topic e_j, so_{ijkl} is the sentiment value of the opinion from the opinion holder h_i on aspect a_jk at time t_l. Structure the unstructured data extracted from raw text is still a challenging task [1].

Semantics has been used only recently for Sentiment Analysis [10] where the authors provide evidence that the inclusion of semantics features in sentiment analysis algorithms improves the overall performance.

Semantic sentiment analysis can take advantage from linked data, ontologies, controlled vocabularies, and lexical resources (e.g. DBpedia, YAGO, ConceptNet [9], SenticNet [4], Nell[1], OIE[2], etc.), which help aggregating the conceptual and affective information associated with natural language opinions.

In this paper we describe Sentilo, a sentic computing system introduced in [8] that can be used as a sentiment analysis core engine to structure text and detect sentiment quintuples according to an ontology defined ad-hoc for the sentiment analysis tasks. Sentilo produces a RDF representation of an opinion sentence that allows the identification of holders, topics (resolved on Linked data to allow aggregation of sentiments on the same topic in different contexts/sources) and opinion triggers with high accuracy. With the use of semantics, we can extend the current state of the art in sentiment analysis to track, correlate, and compare sentiment of specific entities or group of related entities over time and across different contexts. Sentilo core engine prototype can be accessed through its REST API[3] and extended with sentiment scoring modules focusing on the features/domain that researchers want to target.

2 Sentilo Semantic Model

Sentilo consists of a set of components connected in a pipeline [8]. Given a sentence, the syntactic constructs are provided by C&C [6], a highly efficient linguistically parser using a tightly-integrated supertagger, which assigns combinatory categorial grammar lexical categories to words in a sentence. On top of that, the data are processed by Boxer [2], an open-domain software component for semantic analysis of text. It is compatible with first-order logic and builds upon the combinatory categorial grammar and discourse representation theory (DRT). DRT uses an explicit semantic structured language called Discourse Representation Structure (DRS). In Boxer, DRS are enriched with the VerbNet[4] inventory of thematic roles. Output of Boxer is then processed by FRED[5], a tool that uses

[1] http://rtw.ml.cmu.edu/rtw/
[2] http://ai.cs.washington.edu/projects/open-information-extraction
[3] http://wit.istc.cnr.it/stlab-tools/sentilo/service
[4] http://verbs.colorado.edu/~mpalmer/projects/verbnet.html
[5] http://wit.istc.cnr.it/stlab-tools/fred

frame-based design and a set of heuristics in order to produce correct terminology and structure according to Semantic Web design practices. FRED transforms the logical output of Boxer with frames into RDF/OWL in compliance with linked data principles as existing vocabularies are re-used whenever possible, named entities are resolved over resources existing in RDF datasets of the linked data cloud and, terms are disambiguated against WordNet and foundational ontologies. FRED is inspired by Davidson's view [5]: events and situations are primary objects for the representation of a domain. Based on this view of the world, sentences are represented as linked events or situations, with participating objects. We use DOLCE+DnS [7][6] as a vocabulary for events and situations, and VerbNet as reference for thematic roles of events. On top of FRED, we have developed an *opinion model annotator* (see [8] for the new and re-used components employed in Sentilo), a component that implements a set of heuristics that extract, from the FRED's graph of a given sentence, information about holders of an opinion sentence, its topics, and its opinion expressing words (i.e., opinion features). To the best of our knowledge, only a few of semantic models have been provided for sentiment analysis. One of the most relevant includes the MARL model which has been adopted in [3] to represent languages resources for sentiment analysis in a Linked Data conform way enabling leveraging of existing Semantic Web technologies. Sentilo enriches the RDF/OWL semantic representation of an opinion sentence with annotation triples based on *OntoSentilo*, an ontology for opinion sentences that we have defined in [8]. OntoSentilo represents concepts and relations existing between entities composing an opinion sentence. Figure 1 shows a fragment of the RDF graph that represents the sentence *You may think that the summer weather provides the perfect backdrop to a big day.* The use of the prefix `sentilo:` is intended for the local namespaces of concepts and relations added by Sentilo. Sentilo formally represents the holder of the opinion, i.e. person, the main topic of the sentence, i.e. the event occurrence `fred:think_1`, and its subtopics, i.e. the event occurrence `fred:backdrop_1`. Opinion features are identified as values of the relation `dul:hasQuality`, in this case `fred:perfectl`

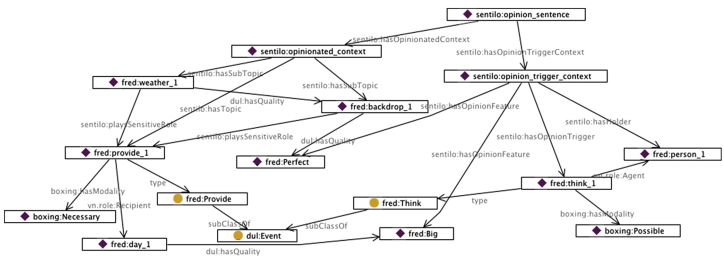

Fig. 1. An extract of the semantic representation for *"You may think that the summer weather provides the perfect backdrop to a big day"*. Note all the semantic relations provided by the framework that can be used for different purposes.

[6] http://ontologydesignpatterns.org/ont/dul/DUL.owl

is a quality of the subtopic `fred:backdrop_1`. As an example of scoring, let us assume that `perfect` is assigned a score of 0.8. Then we can easily associate that score to the entity `backdrop` whose holder is already provided by the framework. The scores to assign to words in the model depends on the domain to focus and on the kind of feelings that want to be extracted. For example, one may want to extract feelings related to fear/bravery and provide scoring for words in that domain. Sentilo performances have been computed in [8] for time and accuracy of topic and sub-topic detection. A deep evaluation on the use of semantics to improve the sentiment analysis tasks (and comparisons) has to be done yet.

3 Conclusions

In this paper we have shown Sentilo, a semantic sentiment analysis core engine able to identify holders, topics, subtopics, opinion triggers, semantic sentiment relationships between terms. Anyone can use the information structured by Sentilo according to a sentiment analysis ontology and design his own sentiment analysis scoring algorithms to build on top of our framework in order to provide entity and sentence level sentiment scores.

References

1. Bing, L.: Sentiment analysis and subjectivity. In: Invited Chapter for the Handbook of Natural Language Processing (2010)
2. Bos, J.: Wide-coverage semantic analysis with boxer. In: Bos, J., Delmonte, R. (eds.) STEP 2008 Conference Proceedings on Semantics in Text Processing. vol. 1. Research in Computational Semantics, pp. 277–286. College Publications (2008)
3. Buitelaar, P., Arcan, M., Iglesias, C.A., Sanchez, J.F., Strapparava, C.: D3.1 - language resource model. Specification for semantic interoperability. In: EUROSENTIMENT Project FP7, pp. 1–21 (2013)
4. Cambria, E., Havasi, C., Hussain, A.: SenticNet 2: a semantic and affective resource for opinion mining and sentiment analysis. In: Youngblood, G.M., McCarthy, P.M. (eds.) FLAIRS Conference, pp. 202–207. AAAI Press (2012)
5. Davidson, D.: The logical form of action sentences. In: Rescher, N. (ed.) The Logic of Decision and Action, pp. 81–120. University of Pittsburgh Press, Pittsburgh (1967)
6. Djordjevic, B., Curran, J.R., Clark, S.: Improving the efficiency of a wide-coverage CCG parser. In: Proceedings of the 10th International Conference on Parsing Technologies, pp. 39–47. Association for Computational Linguistics (2007)
7. Gangemi, A.: What's in a schema? In: Huang, C., Calzolari, N., Gangemi, A., Lenci, A., Oltramari, A., Prevot, L. (eds.) Ontology and the Lexicon, pp. 144–182. Cambridge University Press, Cambridge (2010)
8. Gangemi, A., Presutti, V., Reforgiato, D.: Frame-based detection of opinion holders and topics: a model and a tool. IEEE Comput. Intell. Mag. **9**(1), 31–43 (2014)
9. Liu, H., Singh, P.: ConceptNet: a practical commonsense reasoning toolkit. BT Technol. J. **22**(4), 211–226 (2004)
10. Saif, H., He, Y., Alani, H.: Semantic sentiment analysis of twitter. In: Cudré-Mauroux, P., Heflin, J., Sirin, E., Tudorache, T., Euzenat, J., Hauswirth, M., Parreira, J.X., Hendler, J., Schreiber, G., Bernstein, A., Blomqvist, E. (eds.) ISWC 2012, Part I. LNCS, vol. 7649, pp. 508–524. Springer, Heidelberg (2012)