Semantic Web-based Sentiment Analysis

Diego Reforgiato Recupero[1], Sergio Consoli[1], Aldo Gangemi[12], Andrea Giovanni Nuzzolese[13], Valentina Presutti[1], and Daria Spampinato[1]

[1] National Research Council (CNR), Institute of Cognitive Sciences and Technologies, Semantic Technology Laboratory, Via Gaifami 18, 95028 Catania (CT)
[2] LIPN, University Paris 13, Sorbone Cit'e, UMR CNRS, France
[3] Department of Computer Science and Engineering, University of Bologna, Mura Anteo Zamboni 7, 40127 Bologna (BO), Italy
{diego.reforgiato,sergio.consoli,aldo.gangemi, andrea.nuzzolese,valentina.presutti,daria.spampinato}@istc.cnr.it

Abstract. [1] The introduction of semantics in Sentiment Analysis research has proved to bring several benefits for what performances are concerned and has allowed to identify new challenging tasks to be accomplished. Semantics helps structuring the plain natural language text with formal representation. The current system we are developing performs sentiment analysis by hybridizing natural language processing techniques with Semantic Web technologies. Our system, called Sentilo, is able to recognize the holder of an opinion, to detect the topics and sub-topics in its scope, and to measure the sentiment expressed by them. This information is formally represented by means of RDF graphs according to an OWL opinion ontology, while holders and topics identity is resolved on the Linked Open Data cloud.

Keywords: Sentiment Analysis, Sentic Computing, Semantic features

1 Introduction

Sentiment Analysis (SA) is one of the hottest problems currently studied in Natural Language Processing (NLP), and recently it has entered the Semantic Web world: [16] provides evidence that including semantic features to SA algorithms improves their performance. However, existing approaches at SA, even those that include semantic features, are basically supervised and rely on the availability of manually annotated samples, hence they are usually domain-dependent. Semantic sentiment analysis can take advantage from linked data, ontologies, controlled vocabularies, and lexical resources (e.g. DBpedia, YAGO, ConceptNet [13], SenticNet [4], Nell [11], OIE [7], etc.), which help aggregating the conceptual and affective information associated with natural language opinions.

Combining NLP and Semantic Web approaches could provide us with the flexibility of language processing techniques, as well as with the depth of semantic

[1] This work is supported by the project PRISMA SMART CITIES, funded by the Italian Ministry of Research and Education under the program Programma Operativo Nazionale

knowledge bases, through which also sentiments that are expressed in a subtle manner can be detected, as in the case of concepts that do not explicitly convey any emotion, but which are implicitly linked to other concepts that do so. What is challenging is the way those techniques can be used and combined to yield highly performant systems. With semantics, we can expand 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.

Another common aspect of most existing SA methods is that they neglect the identification of holders and topics of an opinion as a task per se. In fact, they mainly focus on interpreting the tone of a sentence by identifying terms that carry a particular sentiment polarity; it has been demonstrated that including topic detection in models used by algorithms for SA improves their results [2, 12, 17]. However, in such approaches, the SA task melts with the topic detection task, which is never evaluated separately.

Sentilo[2] is a semantic SA system introduced in [10] that analyses the sentiment of a sentence: it identifies the holder of an opinion, the topics and sub-topics of that opinion, the sentiment expressed in each of them by the holder, as well as the sentiment of the overall sentence. Topics, holders, and sentiments are represented as RDF graphs compliant to an OWL ontology [15] described in [10], while topics and holders are resolved on the Linked Open Data cloud in order to aggregate sentiments expressed on the same topic in different contexts or from different sources.

2 Analyzing Opinions

Sentilo implementation is inspired by Davidsons view [6]: 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 [8, 9] as a vocabulary for events and situations, and VerbNet [14] as reference for thematic roles of events. Based on this rationale, we distinguish main topics from sub-topics of an opinion. The distinction between topics and subtopics, as well as the event- and situation-based representation of opinions, impacts on the strategy used for computing the sentiment scores of individual topics and of the whole sentences. To compute sentiment scores we rely on two resources: Sentic.net [5, 3], a publicly available resource that provides polarized scores of concepts, and SentiWordNet [1], a lexical resource for opinion mining. Given an entity, identified as a topic of an opinion (either a main or sub-topic), we compute its sentiment score by combining the scores of its associated opinion features, which are extracted from the RDF graph representing the opinionated sentence. If the topic participates in an event or a situation occurrence, we say that such occurrence provides a context to it, and affects its sentiment score.

We also want to tackle issues contained in sentences like the following: *"John is happy because President Alvarez was arrested"*. For such a sentence, a common

[2] http://wit.istc.cnr.it/stlab-tools/sentilo/service

reader would understand a positive emotion for John as he is happy and a negative event (not opinion as that would depend on the context) for the President Alvarez as he was arrested. A careful reader however would also consider John as the holder of a negative opinion for the President Alvarez as John is having a positive reaction to a bad event happened to President Alvarez. To this aim we introduce the concepts of *Role sensitivity* and *Factual impact*. These concepts have been the basis for the design of a novel resource of annotated verbs, named *SentiloNet*. A role is sensitive with respect to an event (referred to by a verb) if it is indirectly affected by an opinion directly expressed on the event. As far as the annotation of a verb (frame) is concerned, the *sensitivity* is an attribute of its thematic roles. The value of the sensitivity attribute of a role with respect to a verb can be either *true* or *false*, meaning that the role is sensitive or is not, respectively. Factual impact indicates that an event (referred to by a verb) has either a positive or negative impact on a specific role. As far as the annotation of a verb is concerned, the *factual impact* is an attribute of its sensitive roles. The value of this attribute for a role can be *positive*, *negative*, meaning that the inherited opinion will keep its polarity or change it, respectively. The current version of SentiloNet includes 1,100 annotated verbs. Given the high number of different thematic roles of verbs, we have devised a heuristics that allowed us to manually annotate a good amount of verbs in a rather limited amount of time. SentiloNet indicates, for 1,100 verbs, the value of *sensitivity* and *factual impact* attributes for each class of roles.

Sentilo sentiment score $sc_{Sentilo}$ of a topic t can be defined as a function f taking the following arguments:

$$sc_{Sentilo}(t) = f(\sum_{i=0}^{n} sc(q_i(t)), \sum_{i=0}^{m} sc(type_i(t)), truth(t), sc(trig), sc(ctx(t)), mod(t))$$

- *sc(x)* is the score of x as provided by Sentic.net or SentiWordNet;
- q_i is the object value of a triple `t dul:hasQuality q`$_i$. Such triples represent the opinion features, i.e. adjectives and adverbs, associated with entities composing the opinion sentence;
- *type*$_i$*(t)* is the type of t expressed in the RDF graph by means of `rdf:type` triples;
- *truth(t)* is a truth value associated with an entity in the graph, typically an event or situation occurrence, or a quality. If its value is *false* it means that the entity is negated. E.g. in a sentence such as *"John is not a good guy"*, a RDF triple `situation_1 boxing:hasTruthValue fred:False` would be included in the graph, and its effect would be to change the sign of the sentiment score assigned to the feature *good*;
- *trig* is the opinion trigger verb;
- *ctx(t)* is the context of t, if any. It can be either a situation or an event to which t participates in;
- *mod(t)* is the modality of the verb t, if any. E.g. in a sentence such as *I would like a dog*, an RDF relationship `fred:like_1 boxing:hasModality boxing:Necessary` would be included.

3 Conclusions

In this paper we have given our view on SA and shown an example with Sentilo, a semantic SA system that we are currently developing. Sentilo is able to analyses the sentiment of a sentence, identify holders, topics and subtopics. As future direction we are designing a sentiment scoring algorithm that takes into account all the semantics information provided by Sentilo in order to correctly propagate the scores from topics/sub-topics to situations/events and viceversa.

References

1. S. Baccianella, A. Esuli, and F. Sebastiani. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *LREC*, 2010.
2. K. Cai, W. S. Spangler, Y. Chen, and L. Z. 0007. Leveraging Sentiment Analysis for Topic Detection. *Web Intelligence and Agent Systems*, 8(3):291–302, 2010.
3. E. Cambria. Senticnet. `http://sentic.net`, Dec. 2013.
4. E. Cambria, C. Havasi, and A. Hussain. SenticNet 2: A Semantic and Affective Resource for Opinion Mining and Sentiment Analysis. In G. M. Youngblood and P. M. McCarthy, editors, *FLAIRS Conference*, pages 202–207. AAAI Press, 2012.
5. E. Cambria and A. Hussain. *Sentic computing: Techniques, tools, and applications*, volume 2. Springer, 2012.
6. D. Davidson. The Logical Form of Action Sentences. In N. Rescher, editor, *The Logic of Decision and Action*, pages 81–120. University of Pittsburgh Press, 1967.
7. O. I. Extraction. Artificial intelligence - university of washington. `http://ai.cs.washington.edu/projects/open-information-extraction`.
8. A. Gangemi. What's in a Schema? In C. Huang, N. Calzolari, A. Gangemi, A. Lenci, A. Oltramari, and L. Prevot, editors, *Ontology and the Lexicon*, pages 144–182. Cambridge University Press, 2010.
9. A. Gangemi. Dolce Ultra Lite Ontology. `http://ontologydesignpatterns.org/ont/dul/DUL.owl`, Dec. 2013.
10. A. Gangemi, V. Presutti, and D. Reforgiato. Frame-based detection of opinion holders and topics: A model and a tool. *IEEE Computational Intelligence Magazine*, 2014.
11. N.-E. L. Learning. Research project at carnegie mellon university. `http://rtw.ml.cmu.edu/rtw/`, 2010.
12. C. Lin, Y. He, R. Everson, and S. M. Rüger. Weakly Supervised Joint Sentiment-Topic Detection from Text. *IEEE Trans. Knowl. Data Eng.*, 24(6):1134–1145, 2012.
13. H. Liu and P. Singh. ConceptNet: A Practical Commonsense Reasoning Toolkit. *BT Technology Journal*, 22(4):211–226, 2004.
14. M. Palmer. The VerbNet Project. `http://verbs.colorado.edu/~mpalmer/projects/verbnet.html`, Dec. 2013.
15. D. Reforgiato, A. G. Nuzzolese, S. Consoli, A. Gangemi, and V. Presutti. Sentilo. `http://wit.istc.cnr.it/stlab-tools/sentilo/service`, Dec. 2013.
16. H. Saif, Y. He, and H. Alani. Semantic Sentiment Analysis of Twitter. pages 508–524, Boston, UA, 2012. Springer.
17. I. Titov and R. McDonald. Modeling Online Reviews with Multi-grain Topic Models. In *WWW '08: Proceeding of the 17th international conference on World Wide Web*, pages 111–120, New York, NY, USA, 2008. ACM.