

Sentilo: Semantic Web-based Sentiment Analysis

Valentina Presutti¹, Diego Reforgiato¹, Aldo Gangemi^{1,2}
Andrea Giovanni Nuzzolese^{1,3}, Sergio Consoli¹, and Daria Spampinato¹

¹ STLab-ISTC Consiglio Nazionale delle Ricerche, Rome, Italy

² LIPN Sorbonne-Cité, Université de Paris-13

³ Dipartimento di Scienze dell'Informazione, Università di Bologna, Italy.

Abstract. We present Sentilo⁴, an unsupervised, domain-independent system that performs sentiment analysis by hybridizing natural language processing techniques with semantic web technologies. Sentilo is able to recognize the holder of an opinion, to detect the topics and sub-topics in the scope of it, and to measure the sentiment expressed on each of them. All this information is represented formally by means of a RDF graph, and holders' and topics' identity is resolved on Linked Data. Sentilo is available as REST service, to be exploited by client applications, as well as as human-oriented demo that features an intuitive graphical user interface.

Acknowledgements. 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”.

1 Semantic Sentiment Analysis

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: [11] provides evidence that including semantic features to SA algorithms improves their performance. However, existing approaches at SA, even those that include semantic features, are mainly supervised and rely on the availability of manually annotated samples, hence they are usually domain-dependent. Differently, we have the ambitious goal of deeply hybridizing natural language processing and semantic technologies and building a domain-independent, unsupervised approach for computing SA of sentences.

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*, 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 [2,6,13]. However, in such approaches, the SA analysis task melts the topic detection task, which is never evaluated separately.

⁴ <http://wit.istc.cnr.it/stlab-tools/sentilo>

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 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.

In this paper we present a system named *Sentilo*, that, inspired by such an ideal behavior, analyses the sentiment of a sentence: it identifies the holder of an opinion, the topics and sub-topics of that opinion, and the sentiment expressed on each of them by the holder as well as the sentiment of the overall sentence. Topics, holder, and sentiments are represented as formal models (i.e., RDF/OWL graphs) and topics and holders are resolved on Linked Data in order to allow the aggregation of sentiments expressed on the same topic in different contexts or from different sources.

Sentilo⁵ can be used by humans through an intuitive graphical user interface (GUI) or can be accessed through its REST API. The GUI serves mainly as a demonstrator of its capability, while the REST service has a clear commercial potential: the granularity of its results allows a client application to aggregate opinion analysis coming from different sources, on a specific or a general domain, based on shared holders and/or topics. Potential clients could be Amazon, TripAdvisor, iTunes, magazines and news, etc.; all stakeholders dealing with opinions or reviews that have interest in performing data analytics on such opinions during time. Additionally, political parties as well as companies putting new products on the market would be potential users of Sentilo, in order to monitor the impact of a new brand/product on target users by analyzing their opinions.

2 Analyzing opinions

Sentilo implements a sophisticated approach for sentiment analysis. It makes the assumption that any sentence that it receives as input expresses an opinion. An opinionated sentence can or cannot carry an explicit sentiment. For example the sentence:

John said that the patient was suffering from tick bite fever.

does not express any explicit sentiment about the topic of the opinion, but it still expresses an opinion by John. Hence, in this case Sentilo only identifies the opinion holder, i.e. “John” - who can be sometimes implicit - the opinion trigger, i.e. “say” - that can be implicit as well - and the opinion topics. Notice that identifying the opinion trigger allows us to detect the opinion holder. Trigger verbs are recognized according to [12].

⁵ <http://wit.istc.cnr.it/stlab-tools/sentilo>

Sentilo implementation is inspired by Davidson’s view[4]: the world is described in terms of events and situations, and objects are almost always involved in a specific occurrence of one or the other. Based on this view of the world, sentences are represented as linked events or situations, with participating objects. We use DOLCE+DnS [5,7] as a vocabulary for events and situations, and VerbNet [10] as reference for thematic roles of events. Based on this rationale, we distinguish main topics from sub-topics of an opinion. For example, given the above example, the main topic of the expressed opinion would be the referred occurrence of the event “suffer”, and “the patient” would be a sub-topic.

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 [3,9], 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.

Sentilo sentiment score of a topic t can be defined as a function taking the following arguments:

$$sentilo-sc(t) = f(\sum_{i=0}^n sc(q_i(t)), \sum_{i=0}^m sc(type_i(t)), truth(t), sc(trig), sc(ctx(t)), mod(t)) \quad (1)$$

where

- $sc(x)$ is the score of x as provided by a lexical resource for opinion mining (e.g., Senti.net);
- q_i is the object value of a triple `t dul:hasQuality qi`. Such triples represent in the RDF graph adjective and adverbs associated with entities;
- $type_i(t)$ is a 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;
- $trig$ is the opinion trigger verb;
- $ctx(t)$ is the context of t , if any.

In summary, given an opinionated sentence, Sentilo performs the following analyses:

- holder identification: if the holder is explicitly mentioned in a sentence, it is identified and represented in a RDF graph;

- topic detection: it identifies all topics that are in the scope of the expressed opinion and represents them as a RDF graph;
- holders and topics resolution: it resolves holders and topics identities on the Semantic Web;
- sentiment score computation: it calculates the sentiment score for each topic, considering also their inter-relationships and dependencies - if any - according to a Davidsonian representation of the world.

3 Sentilo at work

In this section, we use an example sentence for showing Sentilo at work. Sentilo⁶ is meant to be used as a REST service for sentence-based sentiment analysis. However, it is available also as a user-oriented application. Specifically, after a splash page (depicted in Figure 1) where a user finds a brief explanation of the service, Sentilo is accessible through two different user interfaces:

- a friendly interface targeted at generic web users, which measures the sentiment of a sentence and of its topics;
- a friendly interface targeted a developers, which visualizes the graph representation of a sentence (as from its RDF representation) enriched with opinion-related information, e.g. opinion holder, topics, sentiment scores, etc.

Let us consider the following sample sentence: *Tim Burton thinks that Johnny Depp is a great actor*. Figure 2 shows the first interface (targeted at generic web users). On the top-left of the page there is a text area where a user can type an opinionated sentence. On the right side of the text area there are two “sentimeters”, which initially are set to zero. The first one is green and measures the positive sentiment of the whole sentence, the second one is red and measures the negative sentiment of the whole sentence. Both can have values between 0 and 1.

As explained in Section 2, Sentilo identifies all topics that are in the scope of an opinion and computes a score for each of them. Such topics and associated scores are visualized, under the text area. For each topic, the sentimeters are three: the first (green) that measures (with values from 0 to 1) the positive charge of the sentiment associated with the topic, the second (red) that measures (with values from 0 to 1) the negative charge of the sentiment associated with the topic, and the third (red to green) that measures the average sentiment score of the topic. Furthermore, if a topic identity can be resolved on Linked Data, a descriptive image for the topic is shown, as well as additional information of possible interest for users.

An alternative interface is targeted at developers of applications that exploit Sentilo. This interface allows these users to enter a text (similarly to the other interface), and returns a graphical visualization of the RDF graph that represents

⁶ <http://wit.istc.cnr.it/stlab-tools/sentilo/>



Fig. 1: Sentilo splash page.

the sentence enriched with opinion-related data. Figure 3 shows a relevant part of such a visualization for the sentence *Tim Burton thinks that Johnny Depp is a great actor.*⁷

The entity `fred:Tim_burton`, which is resolved as `dbpedia:Tim_Burton`, is recognized as the holder of the opinion expressed by the sentence. The event “think” is recognized as the trigger, i.e. an event that announces the presence of an opinion in the sentence. The main topic of the opinionated sentence (identified by the property `sentilo:hasTopic`) is the fact (situation) that “Johnny Depp is a great actor”, which involves the entity `fred:Johnny_depp` (resolved as `dbpedia:Johnny_Depp`) as sub-topic. The opinion feature “great” is represented in the graph through the property value of a `dul:hasQuality` triple having `fred:Johnny_depp` as subject.

A Addressing Semantic Web Challenge Requirements

Table 1 describes how Sentilo addresses minimal requirement specified by the Semantic Web Challenge call. Sentilo addresses also a number of desirable features which are described in Table 2.

⁷ For space problems we cut part of the graph and only show a relevant portion. The reader can test the sentence on the online interface in order to visualize the entire graph.



Fig. 2: Sentilo interface with *sentimeters*, targeted at generic web users: results for the sentence *Tim Burton thinks that Johnny Depp is a great actor*.

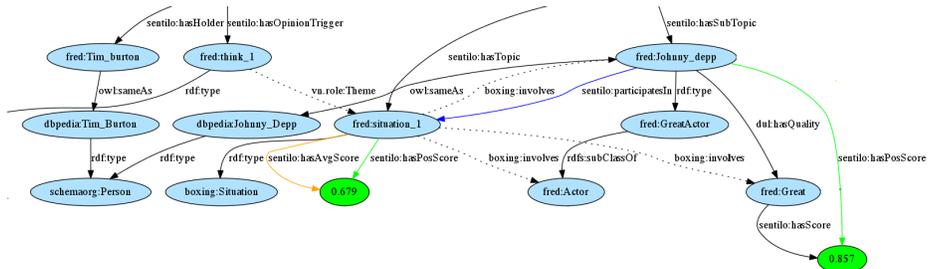


Fig. 3: Sentilo: graphical visualization of the RDF graph returned by the REST service for the sentence *Tim Burton thinks that Johnny Depp is a great actor*.

References

1. A. E. S. Baccianella and F. Sebastiani. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, may 2010.
2. K. Cai, S. Spangler, Y. Chen, and L. Zhang. Leveraging sentiment analysis for topic detection. *Web Intelligence and Intelligent Agent Technology, IEEE/WIC/ACM International Conference on*, 1:265–271, 2008.
3. E. Cambria and A. Hussain. *Sentic computing: Techniques, tools, and applications*. Dordrecht, Netherlands: Springer, ISBN: 978-94-007-5069-2, 2012.
4. D. Davidson. The logical form of action sentences. In N. Rescher, editor, *The Logic of Decision and Action*. University of Pittsburgh Press, Pittsburgh, 1967.

The application has to be an end-user application, i.e. an application that provides a practical value to general Web users or, if this is not the case, at least to domain experts.
Sentilo has a practical value to both general Web users and application developers. General Web users are typically supported by the “sentimeters” interface, while opinion mining experts can appreciate the graph representation under the hood, and application developers can use the REST API for empowering applications with sentimentrics.
The information sources used should be under diverse ownership or control, should be heterogeneous (syntactically, structurally, and semantically), and should contain substantial quantities of real world data (i.e. not toy examples).
We use an extension of VerbNet ⁸ in RDF (currently under our control), and WordNet-RDF [14] as lexical linked data, DBpedia as world linked data, SentiWordNet [1] and SenticNet ⁹ as sentiment score information, as well as an English movie reviews dataset: http://www.cs.cornell.edu/people/pabo/movie-review-data/ for testing. However, in principle Sentilo can be used for processing any (limited to English in this release) natural language resource on the web.
The meaning of data has to play a central role. Meaning must be represented using Semantic Web technologies. Data must be manipulated/processed in interesting ways to derive useful information and this semantic information processing has to play a central role in achieving things that alternative technologies cannot do as well, or at all;
Sentilo semantics catches the meaning of diverse resources: factual data from either structured or unstructured sources, lexical data, and sentiment data, by reusing RDF as a common representation language and OWL for the ontologies that homogeneously describe the data and the opinion model (triggers, holders, topics, etc.). Data are manipulated in non-trivial ways: factual and lexical knowledge are tightly coupled based on a semiotic meta-model (cf. FRED paper [8]), SPARQL querying is extensively used to morph FRED graphs into new “opinion graphs” appropriate to the opinion model. Without meta-modeling and Semantic Web standards it would be impossible or at least much harder to obtain the same results as Sentilo, especially considering the fine granularity of opinion-related information that it produces i.e., holder, topics, sub-topics, context, opinion features, etc.

Table 1: Addressing Minimal Requirements

5. A. Gangemi. *What’s in a Schema?*, pages 144–182. Cambridge University Press, Cambridge, UK, 2010.
6. C. Lin, Y. He, R. Everson, and S. Ruger. Weakly supervised joint sentiment-topic detection from text. *IEEE Transactions on Knowledge and Data Engineering*, 24(6):1134–1145, 2012.
7. D. U. L. Ontology. <http://www.ontologydesignpatterns.org/ont/dul/dul.owl>.
8. V. Presutti, F. Draicchio, and A. Gangemi. Knowledge extraction based on discourse representation theory and linguistic frames. In *Knowledge Engineering and Knowledge Management*, volume 7603 of *Lecture Notes in Computer Science*, pages 114–129. Springer Berlin Heidelberg, 2012.
9. T. S. project. <http://sentic.net/>.
10. T. V. project. <http://verbs.colorado.edu/mpalmer/projects/verbnet.html>.
11. H. Saif, Y. He, and H. Alani. Semantic sentiment analysis of twitter. In *Proceedings of the 11th international conference on The Semantic Web - Volume Part I*, ISWC’12, pages 508–524, Berlin, Heidelberg, 2012. Springer-Verlag.
12. V. Subrahmanian and D. Reforgiato. Ava: Adjective-verb-adverb combinations for sentiment analysis. *IEEE Intelligent Systems*, 23:43–50, 2008.
13. I. Titov and R. McDonald. Modeling online reviews with multi-grain topic models. In *Proceedings of the 17th international conference on World Wide Web*, WWW ’08, pages 111–120, New York, NY, USA, 2008. ACM.
14. M. van Assem, A. Gangemi, and G. Schreiber. Conversion of WordNet to a standard RDF/OWL representation. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC’06)*, Genoa, Italy, May 2006.

The application provides an attractive and functional Web interface (for human users)
Sentilo provides two interfaces for humans, and a REST service. They address generic users, advanced users that want to look “under the hood”, and application developers.
The application should be scalable (in terms of the amount of data used and in terms of distributed components working together). Ideally, the application should use all data that is currently published on the Semantic Web.
Sentilo is built on top of a complex pipeline, which takes data from NLP machine reading components, interprets them as RDF graphs, links them to existing Linked Data, reinterprets them as opinionating/opinionated content, and finally associates analytic and synthetic scoring to them. It integrates Stanbol EntityHub ¹⁰ , which ensures scalability of entity linking and identity resolution on Linked Data.
Rigorous evaluations have taken place that demonstrate the benefits of semantic technologies, or validate the results obtained.
We have performed a rigorous evaluation on the precision and recall of the topic/holder detection tasks, which are excellent on holder detection (F1=.95) and good for topic detection (F1=.72). The evaluation procedure will be published elsewhere.
Novelty, in applying semantic technology to a domain or task that have not been considered before
In the introduction we explain that Semantic Web approaches to sentiment analysis so far are focused on including semantic features in the model used by learning algorithms. Our approach is strongly oriented to Semantic Web methods, including representation and query languages for the data, ontologies for models and meta-models, linked data practices to generate results directly reusable within Semantic Web applications. To our knowledge, Sentilo is the first Semantic Web approach to SA.
Functionality is different from or goes beyond pure information retrieval
Obviously Sentilo is not an information retrieval system: it applies machine reading and Semantic Web techniques to make opinion semantics explicit.
The application has clear commercial potential and/or large existing user base
In the introduction, we briefly mention that Sentilo has a clear commercial potential. Potential clients/stakeholders are iTunes, Amazon, TripAdvisor, journals and magazines, as well as any information provider or executive/consultant that wants to monitor opinions running in textual sources.

Table 2: Addressing Additional Desirable Feature