# Social Media Monitoring in Real Life with Blogmeter Platform

Andrea Bolioli[1], Federica Salamino[2], and Veronica Porzionato[3]

[1] CELI srl, Torino, Italy,
abolioli@celi.it,
www.celi.it
[2] CELI srl, Torino, Italy,
salamino@celi.it
[3] Me-Source srl, Milano, Italy,
veronica.porzionato@blogmeter.it,
www.blogmeter.it

**Abstract.** A social media monitoring platform used by business clients has to face interesting and sometimes unexpected issues arising from real texts processing, in particular dealing with the task of sentiment analysis of word-of-mouth communication. In this paper we describe some of the solutions adopted by BlogMeter, a proprietary listening platform that helps agencies and brands to discover what is said online about brands, people, topics and companies. We present some real life case studies, some of the linguistic resources used in the semantic annotation pipeline, and we suggest some topics for future investigations.

**Keywords:** sentiment analysis, opinion mining, mood, social media monitoring

## 1 Introduction

A social media monitoring platform used by business clients has to face interesting and sometimes unexpected issues arising from real texts processing, in particular dealing with the task of sentiment analysis of word-of-mouth communication. As everybody knows, Sentiment Analysis (SA) is both a topic in natural language processing which has been investigated since several years and a tool for social media monitoring which is used in business services. Two classical and essential references on this topic are [2] and [12]; a recent survey that explores the latest trends is [5]. While the first attempts on english texts date back to the late 90's, SA on italian texts is a more recent task (probably the first scientific publication is [9]).

In this paper we will use the term "sentiment analysis" as a broad term, that includes the narrower terms "opinion mining" and "mood". When we use "opinion mining" we refer to the identification of a belief or estimation or judgement expressed upon an object or target (a comment upon something, in simple words). When we use "mood" we refer to the mood state or emotional state communicated in a portion of text.

We will describe some of the solutions adopted by BlogMeter, a proprietary listening platform, and we will present some real life case studies and some of the linguistic resources used in the semantic annotation pipeline.

The paper is organized as follows. Section 2 briefly describes the Blogmeter platform. Section 3 presents the annotation pipeline. In section 4 we touch upon two interesting topics in SA and in section 5 we presents case studies coming from real-life application of sentiment analysis.

## 2 Blogmeter Platform

BlogMeter[4] is a social media monitoring service operating since 2009 and used by private and public companies in order to collect consumer and market insights from social media and conversations taking places through them ([11]). The monitoring process includes three main phases:

- Listening: thanks to purpose-developed data acquisition systems, the platform detects and collects from the web the potentially interesting data.
- Understanding: a "semantic engine" is used to structure and classify the conversations in accordance to the defined drivers (topics and entities mentioned in the texts).
- Analysis: through the analysis platform the user can surf the conversations in a structured way, aggregate the drivers in one or more dashboards, discover unforeseen trends in the concept clouds and drill down the data to read the messages inside their original context.

In this paper we will focus on the Understanding phase, which includes automatic classification and SA. In detail it consists of:

- creation of a domain-based taxonomy (i.e. an ontology of brands, products, people, topics);
- identification and automatic classification of relevant documents (according to the taxonomy);
- sentiment evaluation and opinion mining (automatic or supervised).

The monitored sources are typically user-generated media, such as blogs, forums, social networks, news groups, content sharing sites, sites of questions and answers (Q&A), reviews of products / services, which are active in many countries and in different languages. The overall number of sources is more than 500,000 blogs (of which approximately 70,000 active, with a post in the last three months) and 700 gathering places (forums, newsgroups, Q&A sites, content sharing platforms, social networks). This computation considers Facebook as a single source, but in fact, it is the largest collector of conversations (the system monitors the public status updates and the production of over 4,000 Italian official pages). We also consider web services like Instagram, Google+, Tumblr, Twitter or sectoral services like Foursquare or TripAdvisor. On the average, every day the system analyzes the following number of "documents":

[4] www.blogmeter.eu

- 3.7 million post retrieved from web sources;
- over 2 million interactions from 1,000 Twitter business profiles and 4,000 Facebook business pages.

## 3 Semantic Annotation Pipeline

Documents extracted from the web in the form of unstructured information are made available to the semantic annotation pipeline which analyze and classify them according to the domain-based taxonomies defined for the client. The annotation pipeline uses the UIMA framework (the Unstructured Information Management Architecture originally developed by IBM and now by the Apache Software Foundation [14]). UIMA annotators enrich the documents in terms of linguistic information, recognition of entities and concepts, identification of relations between concepts, entities and attitudes expressed in the text (opinions, mood states and emotions). Some linguistic resources and annotators are common to different application domains, while others are domain dependent. We will not describe here the pipeline modules in details, and we will focus on the main linguistic resource used in the SA module, i.e. a concept-level sentiment lexicon for italian. The sentiment lexicon is used by the semantic annotator, which recognizes opinions and expressions of mood and emotions, and it associates them with the opinion targets (when performing opinion mining). This component operates both on the sentence level (in order to treat linguistic phenomena such as negation and quantification) and on the document level, in order to identify relations between elements that are in different sentences.

### 3.1 A Sentiment Lexicon for Italian

In this section we describe the ”sentiment lexicon” used by the semantic annotator, i.e. the repository containing terms, concepts and patterns used in the SA annotation. Researchers have been building sentiment lexica for many years, in particular for the english language, and a review on recent results can be found for example in [6].

Our sentiment lexicon for Italian contains about 10.000 entries (6.200 single words and 3.400 multi-word expressions); each entry has information about sentiment, i.e. polarity, emotions, and domain application. It has been created and updated during the past three years, performing social media monitoring and SA in different application domains. Recently, an italian lexicon for sentiment analysis (Sentix) has been developed by [1], as the result of the alignment of several resources. One aspect it is worth mentioning is that the valence of many words can change in different context and domains. The single word “accuratezza” (”accuracy”), for example, has a default positive valence (express a positive attitude), just as it is for ”affare d’oro” (”to do a roaring trade”). On the contrary, “andare a casa” (“going home”) has no polarity in a neutral context, as long as it is not used in an area such as sentiment on Sanremo Festival, where it instead means being eliminated from the singing competition. Similarly, "truccato" ("to have make up on" or “to be rigged"), would not have negative polarity if the domain was a fashion show in Milan. Instead, in the field of online games or betting, the perspective changes.

The semantic annotator is a pattern matching component, which uses the sentiment lexicon, operates on the previous linguistic annotations and creates the corresponding sentiment concepts. The annotator can therefore recognize multi-word expressions that don't explicitly convey polarity and emotions but are related to concepts that do.

## 4 Hot Topics in Social Media Monitoring

Social media and users' opinions and mood states are increasingly linked. Social networks were born as a means of interaction and places for sharing contents; now it is widespread the desire to share emotions and opinions quickly and with as many people as possible. An example of a highly *chatted* domain on the web is Social TV, as people love expressing their opinions about TV hosts and participants.

### 4.1 Irony Detection

We had the opportunity to work on the TV show "The Voice", which has put us face to face with one of the hottest topics for those involved in Sentiment Analysis, namely irony recognition. Conveying a meaning that is the opposite of its literal meaning may cause troubles to linguists struggling with Sentiment Analysis. When the aim is to establish the polarity of a document, the problem that a machine will meet with is its lack of awareness about irony mechanism: only context and common background can help the disambiguation. In order to deal with this issue we collaborated on the creation of a corpus of ironic tweets, namely SentiTUT ([4]). We then proceeded by identifying recurring patterns in ironic tweets, trying to find a common motivation behind their use. Here we present two cases, among those observed, which contain food for thought and possible clues that point out the recognition of irony.

a) Comparisons

"Carolina e Troiano simpatici come le emorroidi a grappolo."

("Carolina and Troiano nice as cluster hemorrhoids.")

"Troiano ha la stessa grinta di una mummia."

("Troiano has the same grit of a mummy.")

"Troiano mi emoziona, lo vedo bene a passeggio con Benedetto XVI."

("Troiano moves me, I can imagine him walking together with
Benedict XVI.")

The examples just reported would create positive opinions in a keyword spotting approach. The context instead suggests that the opposite is true. So, only in the domain in question we can state that the same expressions show reversed polarity.

b) Questions

"Tre tweet a tuo favore su diecimila? Troiano sei un ottimista!"
("Three tweets out of ten thousand in your favor?
Troiano you are an optimist!")
"E come ogni giovedì il solito interrogativo: Troiano, perché?"
("As every Thursday the same question: Troiano, why?")
"Troiano migliorato? Non ho più parole."
("Troiano improved? I have no more words.")

In the same application domain we detected a high percentage of correspondence between ironic tweets and questions: actually, since this TV program is not a cultural show in which questions and answers are the fundamental part, the ironic nature of questions co-occurring with a NE could be taken for granted.

Before proceeding with the identification of algorithms for the automatic recognition of irony, we chose to focus on the in-depth knowledge of specific domains through the research of recurring elements in order to understand how those domains work. In future developments we will test the validity and representativeness of examples like those reported above.

### 4.2 Emotions

The interest for emotion detection in social media monitoring grew in 2011 after the publication of the paper [3], where the authors argued that the analysis of mood in twitter posts could be used to predict stock market movements up to 6 days in advance. In details, they identified "calmness" as the predictive mood dimension, within a set of 6 different mood dimensions (happiness, kindness, alertness, sureness, vitality and calmness).

The definition of a set of basic (or primary) emotions is a debated topic, and the study and analysis of emotions and their expression in texts obviously has a long tradition in philosophy and psychology (see for example [10]). In NLP tasks, Ekman's six basic emotions (anger, disgust, fear, joy, sadness, surprise) has been often used (e.g. in [13]). In the Blogmeter platform we adopt Ekman list of emotions and "love", which is a primary emotion in Parrot's classification.

An interesting task we are investigating is trying to understand which kind of relationship does exist between emotions and irony ([4]).

The manual annotation of emotions in a reference italian corpus would be a useful advance for testing the accuracy of the automatic system.

## 5 Case studies and Examples

In this section we present some case studies and charts that visualize mood and opinion trends generated in different contexts.

### 5.1 Mood Analysis

As seen before, one dimension of the mood analysis is the main polarity expressed in a text. Blogmeter mood analysis has been used:

- as a gauge for the common feelings expressed through a peculiar social network and/or in a given span of time. The figure 1 for example shows the mood expressed in italian Twitter in the period January-April 2013.

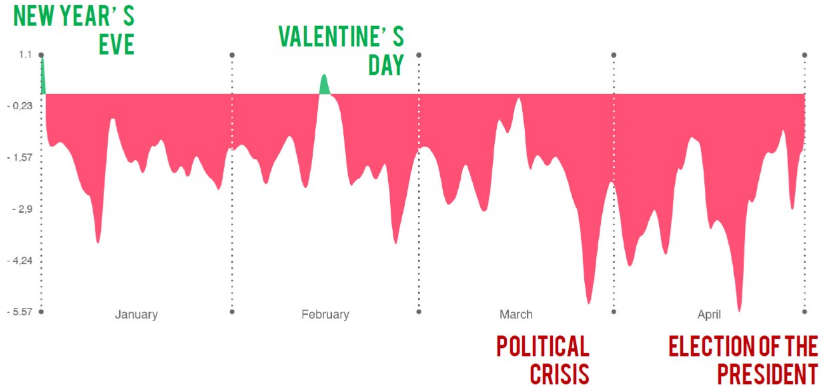

**Fig. 1.** Italians mood expressed by tweets: daily average relation between positive and negative moods during the period January-April 2013

- as a marker of the general mood during specific events. This kind of indicator was very useful during the tracking of live TV shows, due to its capacity to highlight positive and negative peaks on the social network in relation with a show's progression.

### 5.2 Opinion Mining

When the goal becomes more specific and the need is linking a specific subject (i.e. a target that could be a brand, role model or personality) with its related opinions throughout the post, sentiment analysis allows you to automatically examine thousands of messages in depth. Blogmeter's opinion mining has been applied for different industries, such as politics, banking and telecom, where the buzz has quite high volumes and at the same time very polarized opinions.

Another advanced application is near real-time semantic alerting. When sentiment analysis uncovers critical messages they are automatically labelled as negative and sent by email to those who can promptly intervene. This is important for instance in the transportation industry, where users need frequently updated information and feedback.

### 5.3 Emotions and Attitudes

Recently there has been a growing interest for emotion analysis. This kind of investigation can be very useful when the two poles, negative and positive moods, have results that are too streamlined to explain more complex feelings. For instance, this analysis has been used to explore the emotions behind social, political and natural events like the Italian earthquake in 2012 (Fig. 3).

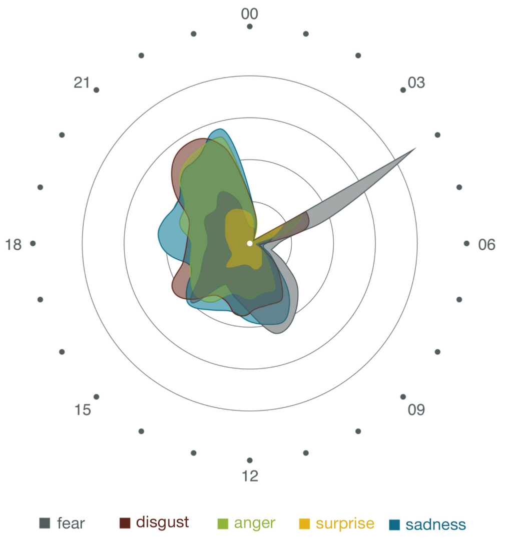

**Fig. 2.** Emotions revealed by tweets during Italian earthquake (05/20/2012, 00 a.m - 12 p.m) and the peak of fear at 4 a.m.

Semantic analysis can also be powerful in order to detect additional meanings, which are not covered in the mood/opinion/emotion dimension, expressing other kinds of people attitudes. In one of these applications, Blogmeter worked on clear voting intentions expressed on the social web, searching for declarations like "Ill vote X" or "I'll choose Y" (and not just "I like Z"). During the final days of the last Italian political campaign, the analysis revealed the striking rise of Beppe Grillo's party.

## 6 Conclusions

We presented Blogmeter, a social media listening service that provides interesting insights about common feelings expressed in social media, opinions about specific subjects and declared attitudes towards real actions or events.

We showed how, in order to achieve those results, it is important to exploit the potential of a well structured linguistic annotation pipeline, but also a domain-specific concept-level sentiment lexicon (also called "contextualized sentiment lexicon" in the literature).

We also presented some case studies, as examples of mood, opinion and emotion recognition in real life use cases. We leave the issue of automatic recognition of irony for further investigation. we hope to have the opportunity to compare the accuracy of Blogmeter system with other ones using an official italian corpus for sentiment analysis (such as SentiTUT).

## Acknowledgments

We would like to thank Sacha Monotti Graziadei, Vittorio Di Tomaso and Vincenzo Cosenza for always stimulating and leading new researches. Eugenia Burchi and Meghan White for their fundamental supervision of the paper; Matteo Casu for the essential help in its preparation. Last but not least, all Blogmeter colleagues for always giving their daily contributions.

## References

1. Basile, V., Nissim, M.: Sentiment Analysis on Italian tweets. Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, pages 100107, Atlanta, Georgia (2013)
2. Bing Liu: Sentiment Analysis and Opinion Mining. Morgan & Claypool Publishers (2012)
3. Bollen, J., Mao, H., Zeng, X.: Twitter mood predicts the stock market. Journal of Computational Science, 2(1) (2011)
4. Bosco, C., Patti, V., Bolioli, A.: Developing corpora for sentiment analysis and opinion mining: the case of irony and Senti-TUT. IEEE Intelligent Systems, vol. 28, no. 2, pp. 55-63 (2013)
5. Cambria, E. New Avenues in Opinion Mining and Sentiment Analysis IEEE Intelligent Systems, vol. 28, no. 2, pp. 15-21 (2013)
6. Cambria, E. et al. Knowledge-Based Approaches to Concept-Level Sentiment Analysis IEEE Intelligent Systems, vol. 28, no. 2 (2013)
7. Chihli Hung, Hao-Kai Lin: Using Objective Words in SentiWordNet to Improve Word-of-Mouth Sentiment Classification. IEEE Intelligent Systems, vol. 28, no. 2, pp. 47-54 (2013)
8. Cosenza, V.: Social Media ROI. Apogeo (2012).
9. Dini, L. and Mazzini, G.: Opinion classication Through information extraction. Proceedings of the Conference on Data Mining Methods and Databases for Engineering, Finance and Other Fields, pp. 299-310 (2002)
10. Galati, D.: Prospettive sulle emozioni e teorie del soggetto. Bollati Boringhieri (2002)
11. Pancaldi, V.: L'azienda centrata sull'ascolto del cliente. FrancoAngeli (2013)
12. Pang, B. and Lee, L.: Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval 2(1-2), pp. 1135 (2008)
13. Strapparava, C. and Valitutti, A.: "WordNet-Affect: an Affective Extension of WordNet", in Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC), pp. 1083-1086, Lisbon (2004).
14. UIMA Specifications http://uima.apache.org/uima-specification.html The Apache Software Foundation