

Sentiment Analysis for Reputation Management: Mining the Greek Web

Georgios Petasis^{1,2}, Dimitrios Spiliotopoulos^{1,3},
Nikos Tsirakis⁴, and Panayiotis Tsantilas⁴

¹ Intellitech Digital Technologies PC

`petasisg@yahoo.gr`, `dimitris@intellitech.gr`

² Software and Knowledge Engineering Laboratory,

Institute of Informatics and Telecommunications,

National Centre for Scientific Research (N.C.S.R.) “Demokritos”,

GR-153 10, P.O. BOX 60228, Aghia Paraskevi, Athens, Greece

`petasis@iit.demokritos.gr`

³ Athens Technology Centre

Rizariou 10, Chalandri, GR-15233, Athens, Greece

`d.spiliotopoulos@atc.gr`

⁴ Palo Ltd.

`{nt,pt}@palo.gr`

Abstract. Harvesting the web and social web data is a meticulous and complex task. Applying the results to a successful business case such as brand monitoring requires high precision and recall for the opinion mining and entity recognition tasks. This work reports on the integrated platform of a state of the art Named-entity Recognition and Classification (NERC) system and opinion mining methods for a Software-as-a-Service (SaaS) approach on a fully automatic service for brand monitoring for the Greek language. The service has been successfully deployed to the biggest search engine in Greece powering the large-scale linguistic and sentiment analysis of about 80.000 resources per hour.

1 Introduction

Sentiment analysis and opinion mining are relatively new areas of natural language processing that seek to capture an aspect of text beyond the purely factual. Contrary to facts, which are objective expressions about entities, events and their attributes, opinions are subjective expressions of emotions, feelings, attitudes or sentiments towards entities, events and their properties. One important aspect of opinions is the fact that they have targets: opinions are expressed for objects (i.e. entities or events) and their attributes. There are several levels of granularity regarding the detailing of the target identification in sentiment analysis. The vast majority of approaches that have been presented in the literature can be classified in the following three categories: *a)* Document level: determine whether a document expresses opinion and identify the sentiment of the document as a whole. *b)* Sentence level: identify if the sentence contains opinions and determine

the sentiment of the whole sentence. *c)* Attribute level: identify object attributes and determine the sentiment towards these attributes.

Reputation management on the other hand, relates to monitoring the reputation or the public opinion of an individual, a brand or a product. Social Web is of course a valuable resource for detecting and monitoring customer feedback, in order to detect early warning signals to reputation problems and content which damages the reputation of an entity. However, the detection of this information is not an easy task, not only because of the technological challenges the identification and extraction technologies face with natural language processing, but also due to size of the social web and the amount of resources that need to be processed. As a result, the employed technologies must be both accurate in the results that they produce, and computationally efficient in order to be exploited in a commercial environment.

In this paper we present a real world application which applies natural language processing on the large scale, aiming to detect opinion polarity about a vast collection of individuals, companies and products in the Greek Web, as harvested by the larger search engine in Greece. Commercialised under the brand name “PaloPro”, this application is the first commercial automated platform for reputation management in Greece, driven by the co-operation of two companies: Intellitech¹, responsible for the linguistic analysis, and Palo² which harvests the Greek Web and commercialises the final product. We will try to present an overview of the “PaloPro” application and the challenges we are facing regarding the linguistic technologies employed for detecting named-entities and opinions about these entities in the context of a large-scale, real-world application for the less linguistically resourced Greek language.

The rest of the paper is organised as follows: In section 2 the application “PaloPro” is presented, while section 3 presents “OpinionBuster”, which is responsible for recognising named-entities (section 3.1) and detecting polarity for each recognised entity mention (section 3.2). An empirical evaluation with the help of two manually annotated “gold” corpora is presented in section 4, along with an evaluation on two specific entities, while section 5 concludes this paper.

2 A Real-World Application for Large-Scale Reputation Management: “PaloPro”

PaloPro is a subscription service which aggregates all news, blog posts, discussions and videos in Greek through a simple, friendly and useful tool for monitoring and analysis, in effect a Reputation Management System. The user has the opportunity to view in real-time, the source of the buzz, the parameters that affect the positive, negative or neutral reputation towards an organization and, ultimately, the overall polarity sentiment and trend on the Web. This is achieved by gathering and processing all references through natural language technologies that extract entities and opinions about these entities. Being a commercial

¹ Intellitech Digital Technologies PC: <http://www.intellitech.gr>

² Palo Digital Technologies Ltd: <http://www.palo.gr>