Annotations for Opinion Mining Evaluation in the Industrial Context of the DOXA project

Patrick Paroubek[1], **Alexander Pak**[1], **Djamel Mostefa**[2]

[1]LIMSI-CNRS, BP 133
91403 Orsay cedex, France
{alexpak,pap}@limsi.fr

[2] Evaluation and Language resources Distribution Agency (ELDA)
55-57 rue Brillat-Savarin, 75013 Paris, France
mostefa@elda.org

Abstract

After presenting opinion and sentiment analysis state of the art and the DOXA project, we review the few evaluation campaigns that have dealt in the past with opinion mining. Then we present the two level opinion and sentiment model that we will use for evaluation in the DOXA project and the annotation interface we use for hand annotating a reference corpus. We then present the corpus which will be used on DOXA and report on the hand-annotation task on a corpus of comments on video games and the solution adopted to obtain a sufficient level of inter-annotator agreement.

1. Introduction

Along with an interest for incorporating emotions in technological devices, the recent years have seen the emergence of automatic opinion and sentiment analysis methods (B.Pang and L.Lee, 2008) particularly in the image management and survey business. Opinions are carried over various media, the press, web sites, radio, television etc. They are a spontaneous source of information, which is updated daily and provides the means to draw quickly an image of the perception that the public entertain with respect to some service, product or major actor of the entertainment or political scene. Survey and analysis of these information sources provide a company with a better knowledge of its customers. They give the means to anticipate new demands, to ensure their fidelity and to reduce attrition risks.
The DOXA[1] project aims at specifying and developing components, resources and services which will allow to :

- Automatically detect topics addressed in large volumes of texts in French and in English,
- Automatically detect feelings and opinions expressed within large volumes of texts in French and in English,
- Automatically detect relations between feelings and opinions expressed and the topics concerned by these feelings and opinions,
- Transform extracted information from texts into structured information to combine this new information with structured information, associated with texts and their authors, to deduct synthetized and exploitable knowledge, by using techniques of data analysis,
- Integrate the components of texts and data analysis into a new version of the INFOM@GIC's platform (services oriented) to build three applications dedicated to "opinion watch" , "consumers and citizens intelligence" , "customer loyalty and churn" for the end-users of the project : OpinionWay, EDF and Meetic.

The applications developed for end-users will help to survey in dynamic, quantitative and qualitative ways:

- the positioning of consumers, customers and users,
- the relationships they maintain with the universes about which they express themselves,
- the trends or evolutions of these universes.

They will help to improve both decision-making (On-Line Analytic Processing, segmentation, scoring, etc.) and operational processes (profiling), by integrating enriched knowledge into these processes.
In the next section, we will make a rapid survey of the various models we have found in the literature in relation with opinion analysis, and we will draw a picture of their relative positions based on the information dimensions that they consider, as far as it is possible to provide an integrated view based on their widely varying characteristics. This will serve us to locate in the landscape the model of (Y.Yannik-Mathieu, 1991) which was used as starting point for our opinion model in DOXA. Then we will have a second state of the art section, but this time devoted to a rendering of the evaluation activities for opinion mining. Once the background picture has been set we will see how both previous topics are addressed in the context of DOXA with first a presentation of opinion model that will be used for annotating the evaluation corpus and second, a presentation of the annotation guidelines and toolkit.

[1]DOXA is a project (DGE nº 08-2-93-0888) supported by the numeric competitiveness center CAP DIGITAL of Île-de-France region which aims among other things at defining and implementing an OSA semantic model for opinion mining in an industrial context. See http://www.projet-doxa.fr

2. Opinion Mining and Sentiment Analysis (OSA) Models

OSA models vary greatly in their orientation. They may be either oriented toward discovering expression of opinion based on more or less rational considerations, judgments or appreciations, either oriented toward the modeling and representation of the expression of the sentiment/emotions that one entertains about an object or an issue. They vary also greatly in the number of dimensions that they use to represent opinion or sentiments and in the granularity of their semantics.

According to (A.Esuli and F.Sebastiani, 2006), opinion mining consists both in searching for the opinions or sentiments expressed in a document and in acquiring new methods to automatically perform such analysis. The authors mentioned three main activities of the field:

- A1 developing language resources for opinion mining, e.g. building a lexicon of subjective terms;
- A2 classifying text according to the expressions of opinion contained;
- A3 extracting from text opinion expressions, taking into account the relationship that links the expression of opinion (the words expressing the opinion) to the source (the author of the opinion statement) or to the target of the expression of opinion (the object the opinion is about) (S.-M.Kim and E.Hovy, 2006).

To build our synthetic view of the various models we will make use a set of general "features", each one broadly associated with a particular information dimension. The previous definition of the activities associated with opinion mining, provides us with the four main features that we will use in our description of the various models, namely:

1. the *opinion marker*, i.e. the language items expressing an opinion (A1 & A3),
2. the *opinion polarity*, the more or less positive impression felt when one reads the opinion expression (A2),
3. the *source*, the (possibly indirect) reference to the beholder of the opinion (A3),
4. and the *target*, the reference to the object/issue/person about which an opinion is expressed (A3).

Among the other features that we will use to organize our presentation of the various models for opinion mining, we have:

- the *intensity*, i.e. the relative strength of an expression,
- the *theme/topic*, whether the models makes use of a representation of the topic addressed, in the document containing an expression of opinion,
- the *information*, the more or less factual aspect of the opinion expression,
- the *engagement*, the relative implication that the opinion holder is supposed to have in supporting his opinion expression.

Listing the features sets of all the models we have encountered and putting them into relation yielded a graph that is too complex to be easily displayed because of the numerous links. So we decided to sort our presentation features according to an arbitrary order based on the intuitive importance one would accord to a given feature if it were missing from an opinion statement. In our mind, an opinion statement which would mention only the *intensity* of an opinion without giving any indication of its *polarity* should be considered less informative for opinion mining. As a result, we put *polarity* before *intensity* in our arbitrary ordering and following the same train of thought, we have afterward: the *target*, the *information*, the *engagement* and and the *source*. Putting the *source* last may seem strange, but very often the source is not explicitly mentioned in a document, since the source is the author. Then we sorted the different models of opinion, first according to the number of "features" they display and second according to the relative position of their features in our arbitrary ordering. For instance a model having only the attribute *polarity* would be judged more generic than a model which would have both *polarity* and *target*. With this considerations in mind, the twenty different models organize themselves into a quasi linear sort. From the most generic to the most specific model, we have identified six levels in the hierarchy of models in Figure 1. The first level of our hierarchy lists authors who have not proposed any attribute in particular, but have addressed the subject of opinion and sentiment in language. They are associated in our representation to the most generic (top) attribute *OSA model*. Level 2 shows authors who do not have any *polarity* in their model and level 3 those who did not address *Intensity*, and so on. Then we have used the same methodology at each level to refine our hierarchy. At level 1, we find the models of (R.Quirk et al., 1985), (J.Kamps et al., 2004) and (S.Berthard et al., 2004). They have defined other attributes of opinion expression, like *polarity, intensity, target, information* etc. (R.Quirk et al., 1985) have introduced the notion of *private state* which regroups all the expressions of subjectivity like emotions, opinions, attitudes, evaluations etc. This notion is also present in the model of (J.Wiebe et al., 2005), (B.Pang and L.Lee, 2008). The models of (K.Dave et al., 2003), (P.Turney, 2002), (A.Harb et al., 2008), (S.Somasundaran et al.,), (S.-M.Kim and E.Hovy, 2006) and (V.Stoyanov et al.,) are located at level 2. The models of (T.Mullen and N.Collier, 2004), (V.Stoyanov et al.,) and (H.Yu and V.Hatzivassiloglou, 2003) were considered more specific than those of level 2 because they stressed the importance of *target* and *source* for opinion mining. The work of (Y.Yannik-Mathieu, 1991) is characterized by a categorization of verbs expressing feelings. The model of (J.R.Martin and P.R.R.White, 2005) deals with evaluative aspects. The authors have mentioned three subtypes of evaluation, characterized by their respective attributes which are: *attitude*, *engagement* and *graduation*. *Attitude* refers to values returned by judgement from one or more sources and can be associated to emotional responses. Its three subtypes are: *judgement*, *affect* and *appreciation*. *Engagement* explicits the position, the implication of the source with respect to its expression of opinion. It is one of the main character-