My MOoD, a Multimedia and Multilingual Ontology Driven MAS: Design and First Experiments in the Sentiment Analysis Domain

Maurizio Leotta, Silvio Beux, Viviana Mascardi, and Daniela Briola

DIBRIS, University of Genova, Italy
maurizio.leotta@unige.it, silviobeux@gmail.com,
viviana.mascardi@unige.it, daniela.briola@unige.it

Abstract. In this paper we introduce the architecture of a *M*ultimedia and *M*ultilingual *O*ntology *D*riven Multiagent System (*My MOoD*) for classifying documents consisting of audiovisual and textual elements, according to classes described in a domain ontology. My MOoD will integrate software components devoted to the analysis of images, videos, and sound, with the multilingual text classifier based on BabelNet presented in this paper. All the integrated components will be wrapped by agents and will perform their classification based on a common domain ontology, which is a parameter of the multiagent system. Wrapper agents will interact in order to share the classification of the document's elements and agree on a coherent classification of the document as a whole, exploiting their background knowledge and reasoning capability to resolve ambiguities. Changing the ontology (and tuning or substituting the classifiers for dealing with the domain of interest) will allow the multiagent system to classify heterogeneous multimedia documents in whatever domain and for many different purposes. In the My MOoD instance discussed in this paper, the ontology (*sentiHotel*) describes the accommodation domain and the classification mines the sentiment of hotel reviews written in five different languages.

Keywords: Multimedia, Multilingual, Multiagent, Ontology, BabelNet

1 Introduction and Motivation

When it was born, at the beginning of the new millennium, sentiment analysis was conceived as a research area addressing text only, written in only one language. Because of the lack of multimedia social networks which were limited, at that time, to Friendster (2002), MySpace, LinkedIn and Hi5 (2003), Flickr and Facebook (2004), and the hardness of managing multilingual and multimedia objects, it is no surprise that the seminal works by Turney [34] and Pang et al. [26] published in 2002 had monolingual textual documents as their sole target. The well known article by Pang and Lee dating back to 2008 [25] defines opinion mining and sentiment analysis as areas dealing with the computational treatment of opinion, sentiment, and subjectivity in text, and even the most recent surveys on the topic [20, 35] do not consider the possibility to extract sentiments from objects rather than text. While the problem of multilingualism was addressed starting from 2007 [1, 12, 22], researchers drove their interest towards multimedia contents including images and video as valuable sources of opinions only in the last five years. Starting from 2010, visual sentiment analysis [6, 7, 31, 38, 39] emerged as an area complementing that of textual sentiment analysis, aiming at extracting the polarity conveyed by visual content, including movies. Multimodal sentiment analysis, taking spoken content into account [16, 30], is an even more recent approach.

Being able to extract and analyze emotions and opinions from multimedia, multimodal and multilingual social objects such as news, tweets, blogs, etc, would of course give great advantages, including economic ones. Strengthening the polarity of a written opinion because of images or spoken sentences that support it, or, on the other hand, making a deeper analysis when texts and videos referring to the same event seem to express different emotions, would give more precise and reliable (and hence, more precious and valuable) results. However, the complexity of each individual task involved in the multimedia and multilingual sentiment analysis process makes it so challenging that only a "divide et impera" approach, dividing the burden of the challenge among many intelligent, autonomous and cooperating entities, can work.

An intelligent software agent is a software component which is situated (receives sensory input from its environment and can perform actions that change the environment in some way), autonomous (acts without the direct intervention of humans or other agents and has control over its own actions and internal state), responsive (perceives its environment and responds in a timely fashion to changes that occur in it), pro-active (exhibits opportunistic, goal-directed behaviour and takes the initiative where appropriate) and social (interacts, when appropriate, with other artificial agents and humans in order to complete its own problem solving and to help others in their activities) [17]. A multiagent system, or MAS for short, is a system designed and implemented as several interacting agents. Quoting [17] again, *"multiagent systems are ideally suited to representing problems that have multiple problem solving methods, multiple perspectives and/or multiple problem solving entities"*. The problem of classifying different elements of a complex multimedia object, each of which may be a piece of text expressed in different languages, a fragment of audio or video track, an image, a manual sketch, and then combining these classifications to provide a coherent and meaningful classification of the object as a whole, requires to involve multiple problem solving entities (the classifiers) and to coordinate their outcomes in a non trivial way. A MAS is thus an extremely suitable approach for facing such a complex problem.

In this paper we present the design of the *My MOoD MAS* (My MOoD in the sequel), a general purpose *M*ultimedia and *M*ultilingual *O*ntology *D*riven multiagent system, and the first experiments to mine the polarity of multilingual texts exploiting the *SentiHotel* ontology. Although My MOoD is still in its design stage, we are confident that, once implemented, it will ensure the modularity, flexibility and scalability required for tackling a challenging task like multimedia and multilingual sentiment analysis.

The paper's structure is the following: Section 2 discusses the state of the art. Section 3 introduces the architecture of My MOoD. Section 4 describes the Multilingual Text Classifier. Section 5 describes the SentiHotel ontology. Section 6 discusses the results of the experiments carried out with hotel reviews in five languages. Section 7 concludes and highlights some directions for the future work.

2 State of the Art

Multilingual sentiment analysis. In [22], Mihalcea et al. explore methods for generating subjectivity analysis – namely identifying when a private state is being expressed and identifying attributes of that private state including who is expressing the private state, the type(s) of attitude being expressed, about whom or what the private state is being expressed, the intensity of the private state, etc. [37] – in a target language L by exploiting tools and resources available in English. Given a bilingual dictionary or a parallel corpus acting as a bridge between English and the selected target language L, the methods can be used to create tools for subjectivity analysis in L. Experiments are carried out with Romanian. Ahmad et al. [1] classify sentiments within a multilingual framework (English, Arabic, and Chinese) following a local grammar approach. Domain-specific keywords are selected by comparing the distribution of words in a domain-specific document to the distribution of words in a general language corpus. Words less prolific in a general language corpus are considered to be keywords. Denecke [12] introduces a methodology based on lexical resources for sentiment analysis available in English (SentiWordNet, http://sentiwordnet.isti.cnr.it/) for determining polarity of text within a multilingual framework. The method is tested for German movie reviews selected from Amazon and is compared to a statistical polarity classifier based on n-grams. The paper by Boiy and Moens [5] describes machine learning experiments with regard to sentiment analysis in blog, review and forum texts found on the World Wide Web and written in English, Dutch and French. The proposed approach combines methods from information retrieval, natural language processing and machine learning. An automated sentiment analysis on multilingual user generated contents from various social media and e-mails is described in [33]. The sentiment analysis is based on a four-step approach including language identification for short texts, part-of-speech tagging, subjectivity detection and polarity detection techniques. The prototype has been tested on English and Dutch. More recently, the paper [3] presents an evaluation of the use of machine translation to obtain and employ data for training multilingual sentiment classifiers. The authors demonstrate that the use of multilingual data, including that obtained through machine translation, leads to improved results in sentiment classification and that the performance of the sentiment classifiers built on machine translated data can be improved using original data from the target language. The languages explored by the authors are Turkish, Italian, Spanish, German and French. Finally, the paper [14] describes the adoption of meta-learning techniques to combine and enrich existing approaches to single and cross-domain polarity classification based on bag of words, n-grams or lexical resources, adding also other knowledge-based features. The proposed system uses the BabelNet multilingual semantic network [24] to generate word sense disambiguation and vocabulary expansion-derived features. Being based on BabelNet, the system can cope with multilingual documents. By now its evaluation has been carried out on a monolingual dataset, the Multi-Domain Sentiment Dataset (version 2.0, http://www.cs.jhu.edu/ mdredze/datasets/sentiment/). Evaluating the polarity classification approach in other languages is part of the authors' future work.

Ontology driven sentiment analysis. One of the first papers on ontology-based sentiment classification is [29], where the ontology was used to classify and analyze online product reviews by providing lexical variations and synonyms of terms that could be met in the reviews. In [10], Chaves and Trojahn present Hontology, a multilingual ontology for the hotel domain. Hontology has been proposed in the context of a framework for ontology-driven mining of Social Web sites contents. Comments are annotated with concepts of Hontology, which are manually labeled in Portuguese, Spanish and French. Hontology reuses concepts of other vocabularies such as Dbpedia.org and Schema.org. The work on Hontology was further expanded in [11]. ArsEmotica [4] is a software application for associating the predominant emotions with artistic resources of a social tagging platform. A rich emotional semantics (i.e., not limited to a positive or a negative opinion) is extracted from tagged resources through an ontology driven approach. The ArsEmotica Ontology (AEO [27]) is based on Plutchik's model [28] and incorporates, in a unifying model, multiple ontologies which describe different aspects of the connections between media objects (e.g., the ArsMeteo artworks, http://www.arsmeteo.org/), persons and emotions. In particular, it includes an ontology of emotions which have been linked, via `owl:sameAs`, to the corresponding emotions in DBpedia. Furthermore, it incorporates an ontology of artifacts, derived from the alignment of a domain ontology obtained from the DB of the ArsMeteo on line portal, with the OMR (Ontology for Media Resources, http://www.w3.org/TR/mediaont-10/). The paper [19] proposes the deployment of original ontology-based techniques towards a more efficient sentiment analysis of Twitter posts. The novelty of the proposed approach is that posts are not simply characterized by a sentiment score, as is the case with machine learning-based classifiers, but instead receive a sentiment grade for each distinct notion in the post. The proposed architecture aims at providing a more detailed analysis of post opinions regarding a specific topic.

Multiagent systems for sentiment analysis. While we know many papers dealing with agents which show emotions and sentiments, we are aware of only two papers where agents are used to analyze sentiments of documents. In [2] a MAS exploiting machine learning classification for analyzing the sentiment of product features in different social media sources is presented. The MAS exploits different agents to deal with different kind of information from different social media networks. Agents communicate and interact with each other to learn new information. Kechaou et al. [18] describe a MAS based on a thorough linguistic analysis which enables to resolve the ambiguities and complexities of the natural evaluative language and to strengthen, as well as consolidate, the results achieved at the various analysis stages. We are not aware of other agent-based approaches to sentiment analysis.

Comparison. While exploiting an ontology for driving the sentiment analysis is far from being an original idea and the papers on this topic are much more than those that we mentioned in this section, exploiting a MAS for that purpose seems to have received little attention by the research community. The two MASs we are aware of address textual documents only, and written in only one language. From this point of view our proposal, albeit preliminary, seems to be an original one. With respect to the existing literature on multilingual sentiment analysis, our work is among the few ones that perform an evaluation involving five languages, hence demonstrating the actual multilingualism and flexibility of the approach. As far as the adopted tools are concerned, the work closer to our is [14] for the heavy exploitation of BabelNet.

3 MAS Architecture

IndianaMAS [21] is a project funded by the Italian Ministry for Education, University and Research, MIUR, spanning from March 2012 to February 2015. It integrates intelligent software agents, ontologies, multilingual natural language processing, sketch and image recognition techniques to develop a technology platform for the digital preservation of rock carvings. The IndianaMAS platform has been conceived as a general, scalable and flexible holonic MAS, namely a MAS consisting of components which are at the same time "part" of a bigger MAS (the MAS that contains them), as well as independent MASs [15]. Classification of texts, images and sketches is driven by an ontology [8] named *Indiana Ontology*, modeling information about Mount Bego's prehistoric rock art. For reaching the same objectives as the IndianaMAS project, but in a different domain, the ontology can be changed with any other ontology from any other domain, keeping the general MAS architecture almost unchanged: image and sketch recognition algorithms must of course be modified in order to recognize images and sketches in the domain of interest; classifiers for audio and video tracks must be added if the input documents contain elements of this kind; multilingual text classification, instead, requires limited or no tuning at all as the only assumption it makes is the existence of an ontology modeling those concepts according to which the classification must be performed.

Given the raising importance of sentiment analysis in social and expressive media, we investigated how to move from the IndianaMAS for the rock art domain to a more general MAS for classifying multimedia and multilingual documents consisting of text, sketches, drawing, images, but also video and audio tracks. The result of our investigation is the My MOoD MAS shown in Figure 1.

Our research is currently targeted to the ontology-driven classification of multilingual textual documents only: the Multilingual Text Classifier Agent MUTCA wrapping the Multilingual Text Classifier in Figure 1 is highlighted for this reason. Since emotions can be extracted from audiovisual content as well, as witnessed by the literature on visual and multimodal sentiment analysis, and since - although, to the best of our knowledge, not yet addressed by the research community - it should be possible to extract the polarity of manual sketches exploiting techniques similar to those described in [9], agents for classifying movies, audio tracks, images and manual sketches have been included in the My MOoD architecture as well.

The interaction among these different agents and holons will allow My MOoD to correctly interpret multimedia contents also in case of ambiguous classifications. Consider for example an image with a woman wearing an elegant long white dress, whose expression is clearly touched. If the agent devoted to image recognition can extract concepts like "woman", "elegant" "white dress", and "moved" from the picture, and the domain of interest deals with religious celebrations including "wedding" and "funeral", the correct classification could be both of them: in Hindu tradition, in fact, white is the standard color for funerals and the woman might be a related of the deceased, whereas in Western cultures a woman dressed in white attending a religious celebration is likely to be the bride. If the textual caption of the image says nothing about the event, but states that it took place in New Delhi, then intelligent agents able to reason about all the information extracted from the document, including geographical data, can agree that the picture shows a Hindu funeral.

The analyzed documents can be stored – temporarily or permanently – into an internal DataBase together with their classification. The DataBase can also store aggregated results. The user of My MOoD can perform queries on the stored data. Queries will be based on the ontology, which is the core of the system and the driver of the domain modeling and of the document classification.

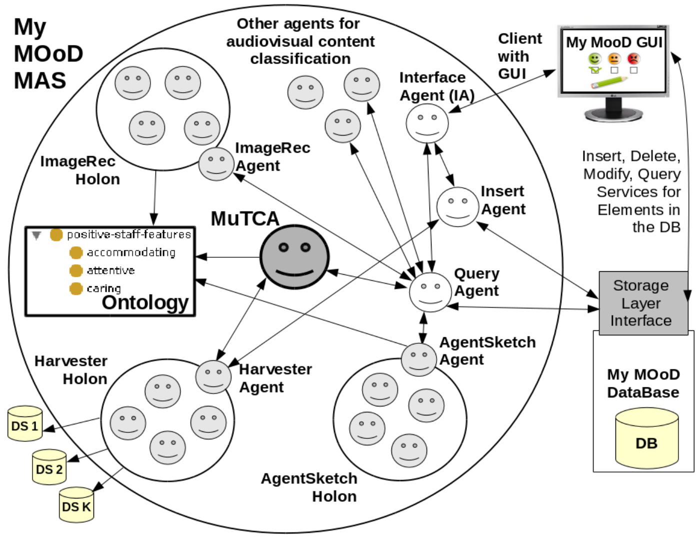

Fig. 1. My MOoD architecture (for sake of clarity, not all the arrows modeling control and data flow between components are shown).

The main components of My MOoD will be:

– The ontology, which structures the domain of interest of the project. In IndianaMAS the domain was that of Mount Bego rock art modeled by the Indiana ontology, while in this paper it is that of opinions about hotels modeled by the SentiHotel ontology.
– The Multilingual Text Classifier Agent MuTCA.
– The other components that we implemented and tested in IndianaMAS and that we will reuse, after the required tuning and integration, namely:
  - the AgentSketch holon, for interpreting manual drawings and sketches;
  - the ImageRec holon, for recognizing and classifying images;
  - a holon for searching the Web to retrieve documents that meet the user's needs and requirements;
  - the Interface, Insert and Query Manager Agents, for providing an interface between the MAS and the user and for offering operations over data in the DataBase.
– Additonal agents for classifying other kinds of digital objects.
– The internal DataBase, to store multimedia documents (or their references/URLs) that have been classified.
– The Web Interface, to let users perform operations on data (look for new data on the web, store retrieved data, analyze and query them).

4 Ontology-driven Multilingual Text Classifier

Classifying a document can be defined as the task of assigning it to one or more classes or categories. For instance, we might want to classify a text w.r.t. a set of geographical, historical, and topic classes (e.g., understanding whether a text is about the neolithic rock art in France, as we did in the IndianaMAS project). Our Multilingual Text Classifier (TEXTCLASS in the sequel), designed and implemented to face such a classification task, takes in input (1) an ontology whose classes model the domain of interest and whose names are expressed in any language from a predefined set[1] and (2) a document containing the text to classify written in any language from the above set. It returns a classification of the text w.r.t. the ontology taken in input. The classification performed by the TEXTCLASS is multilingual and exploits BabelNet and WordNet.

WordNet[2] [13, 23] is the main resource for lexical knowledge upon which BabelNet is based. WordNet groups English words into sets of synonyms called synsets. A label that indicates the part of speech (e.g., n means noun) and sense number is associated with each word in the synset. Words are assigned sense numbers based on frequency of use in semantically tagged corpora. Senses in WordNet are generally ordered from most to least frequently used, with the most common sense numbered 1 . Frequency of use is determined by the number of times a sense is tagged in the various semantic concordance texts. To make an example, a sysnset can be of the form:

$$\{play_{1n}, drama_{1n}, dramatic\_play_{1n}\}$$

WordNet also provides a textual definition (gloss) for each synset. The major weakness of WordNet is that it available for English only; BabelNet was born to overcome this limitation.

BabelNet[3] [24] is a very large multilingual semantic network, based on the automatic mapping of concepts onto WordNet and Wikipedia[4], the largest multilingual Web encyclopedia. The result is an "encyclopedic dictionary", in which words (Babel Senses) in different languages (BabelNet 3.0 supports 271 languages including all European languages, most Asian languages, and even Latin) are grouped into sets of synonyms called Babel Synsets. Each Babel Synset has different features like shorts definitions (glosses) in many languages harvested from both WordNet and Wikipedia, and many relations in the semantic network provided by WordNet (e.g., hypernymy and hyponymy, meronymy and holonymy, antonymy and synonymy, etc.).

Given an ontology o and a document d to classify, TEXTCLASS identifies the classes in o which d belongs to. For instance, in case of a geographic ontology, TEXTCLASS associates with each document (for example, a tourist guide) the geographical place(s) that it describes.

The strengths of TEXTCLASS are the following: (1) it is able to classify documents described in several languages (2) using ontologies in different languages; (3) the languages used in the documents and in the ontologies can be different; (4) there is no need to state in advance the languages of the ontologies and documents, as TEXTCLASS can automatically recognize them[5]; (5) the documents' format can be either plain text or pdf; and (6) documents can be classified w.r.t. different ontologies in a single step

[1] In theory we could cope with any of the languages supported by BabelNet; in practice, if we want to apply a stemming stage to words as we actually do for obtaining acceptable results, we can manage only those for which the Porter stemmer is implemented. If a stemmer is not available, stemming could even be avoided, but we would expect poor results without it.

[2] http://wordnet.princeton.edu/

[3] http://babelnet.org/

[4] https://www.wikipedia.org/

[5] The automatic language recognition feature is currently implemented for texts but not for ontologies; extending it to ontologies would be straightforward.

(provided that all the ontologies are described in the same language). In the following, to keep the description simpler, we describe the functioning of TEXTCLASS when only one ontology is used.

More in detail, TEXTCLASS (1) extracts the text T, i.e. a list of words, from d (if d is not on the computer, downloads the document from the URL); (2) detects the language l used in T; (3) translates each word $w \in T$ into the language of the ontology using BabelNet and WordNet; (4) classifies T w.r.t. o; and finally (5) returns the classification.

The current prototype of TEXTCLASS integrates one module for each step above and has been developed in Java on the Ubuntu 14.04.1 Linux platform.

Extracting Text from Document (Module 1). This module is devoted to extracting the text T (a list of words) contained in document d.

T = extractText(URL or FilePath of d)

Implementation Details: The document can be provided to TEXTCLASS in two ways: (1) it could be already saved in a local directory (e.g., /home/user/text/sample.pdf) or (2) it could be available online (e.g., http://site/sample.pdf). In the latter case, the file is downloaded in a temporary folder by using the copyURLToFile(...) method provided by the org.apache.commons.io.FileUtils library. Then, the file is read by using different methods depending by the file type. TEXTCLASS currently supports txt and pdf files.
In both cases the file is opened and its textual content loaded, cleaned (i.e., substituting all the occurrences of multiple white spaces or non visible characters such as tab and newline with a single white space) and assigned to a list of String T that is provided to Module 2.

Detecting the Language of the Text (Module 2). This module is devoted to recognizing the language L_T used in text T extracted from document d. This step is necessary because the following modules need to know the language of the document. TEXTCLASS adopts a naive Bayes with character n-gram for fast language detection.

L_T = detectLanguage(T)

Implementation Details: TEXTCLASS employs the Language-Detection library[6] that is able to detect, with a precision greater than 99%, 53 languages making use of naive Bayesian filters. In particular, TEXTCLASS analyses the text T provided by the previous module and, depending on its length, calls the language detector library using different profiles. Indeed, in case of very short texts (few words), it is recommended to use specific profiles rather than the standard ones. To speed up the language detection, TEXTCLASS avoids to provide the complete text of the document to the language detector, given that, potentially, TEXTCLASS could be required to classify documents long tens or hundreds pages. From our experiments, we noticed that using the first 100 words of the text (e.g., about 500-800 characters) provides very good results in terms of both precision and performance.

Translating Text (Module 3). The main goal of this module is to translate each word of the text T into the language L_o used to describe the ontology (L_o is an information associated with each ontology). For each word $w \in T$, two steps are performed.

[6] https://code.google.com/p/language-detection/

- First, all the synsets containing the word w are retrieved. Note that w is supposed to belong to the language L_T. Obviously w can appear in more than one synset.
  For instance, in case of an Italian text containing the word "pulito", the BabelNet function getSynsets is called with the parameters L_T=Language.IT and w="pulito", and returns a set of synsets S. Indeed, "pulito" in Italian has different meanings[7], including for instance: (1) free from dirt or impurities[8], (2) characterized by freedom from troubling thoughts (especially guilt)[9].
- Second, given L_o the target language used in the ontology, all the words associated with each synset $s \in S$ in the language L_o are retrieved by means of the BabelNet function getSenses. In the case of the word w="pulito" and L_o=Language.EN, we obtain several translations including: clear, clean, neat, uncontaminated, orderly, elegantly, untarnished, untainted, unstained, stainless, unsullied.

T' = translateText(T, L_T, L_O) $\qquad$ T' is a set containing a list for each $w \in T$.

If no translation is found for w, the corresponding list contains only w, otherwise it contains all the computed translations. Each list is also associated with how many times the original word w was found in the text T.
Implementation Details: Since such operations are repeated for each word and are time consuming (the BabelNet indexes have a total size of about 30GB), we execute a pre-processing step that consists in (1) removing all the stop words[10] from text T[11] obtaining a cleaned text in L_T, and (2) searching the translations of each word only once even if it is repeated multiple times in the original text.

Assigning Weights to the Ontology Nodes (Module 4). In this phase, the ontology nodes are labeled with weights in order to consider the frequency of the corresponding terms in the text. In detail, each node in o is compared with all the elements (i.e., words) of all the lists in T'. Every time a match is found, the label containing the weight of the node is increased by the value associated with the list containing the matching word.

o_W = assignWeights(o) $\qquad$ o_W is the weighted ontology.

Implementation Details: the ontology that drives the classification is expressed in OWL and is navigated and manipulated by means of the Jena Java framework[12]. To increase the probability of finding a match between the words translated by BabelNet and the words used for labeling the ontology nodes, we reduce the inflected words (both in the ontology and in the list) to their word stem. For this purpose we adopt the Snowball framework[13] by Martin Porter, that contains specific stemming algorithms for 16 languages.

Generating the Final Classification (Module 5). In this phase, all the nodes in o_W are visited and those with a weight greater than 0 are inserted into the result list. The

[7] http://BabelNet.org/search?word=pulito&lang=IT
[8] http://BabelNet.org/synset?word=bn:00099776a&details=1&orig=pulito&lang=IT
[9] http://BabelNet.org/synset?word=bn:00099807a&details=1&orig=pulito&lang=IT
[10] Stop words are words which are filtered out before or after processing of natural language data, http://en.wikipedia.org/wiki/Stop_words
[11] We used the lists of stop words included in the BabelNet API (24 languages supported). Each list typically includes from one to several hundreds of stop words.
[12] https://jena.apache.org/
[13] http://snowball.tartarus.org/texts/introduction.html

list is ordered in decreasing order of weight. For instance, when classifying texts using a geographic ontology, TEXTCLASS can return the following result [[Liguria, 25],[Italy, 12],[Nice, 4],[France, 2]]. This result could be interpreted as: the text T describes something located in Liguria (an Italian region) but also, to a lesser extent, something that concerns Nice, a French city near the border with Italy. We would obtain such a result if, for example, the text was centered around Monte Beigua, located in Liguria, whose name has the same root as Mount Bego, located in France, whose petroglyphs are studied by archaeologists working in Nice.

C = classification(o_W) C is the final classification list.

5 Modeling Opinions in the Accommodation Domain: the SentiHotel Ontology

In order to test the behavior of TEXTCLASS with an ontology different from the *Indiana* one, we developed an ontology of opinion words in the accommodation domain that integrates the four emotional branches of *WordNet Affect* [32] (`positive-emotion`, `negative-emotion`, `neutral-emotion` and `ambiguous-emotion`) and added to them about 400 opinion words (as subclasses) based on 30 positive reviews and 30 negative reviews retrieved from [36] and carefully analyzed by the authors to devise the most frequent concepts expressing positive/negative feelings. The ontology was manually developed in OWL Lite using Protégé 3.4.8 (http://protege.stanford.edu/) and is publicly available from http://www.disi.unige.it/person/MascardiV/Download/sentiHotel.owl.

The dataset described in [36] is available to the community and contains reviews from *Tripadvisor* (and other sources, that we did not used because out of scope), which we used to create our ontology and to test the Multilingual Text Classifier, as described in Section 6.

The domain depended opinion words, each mapped into an OWL Class, are divided into a `negative-accommodation` branch divided into eight sub-trees (`negative-experience-causes`, `negative-experience-consequences`, `negative-experience-features`, `negative-food-features`, `negative-location-features`, `negative-price-features`, `negative-room-features`, `negative-staff-features`) and containing 270 classes, and a `positive-accommodation` branch divided into six sub-trees (`positive-experience-features`, `positive-food-features`, `positive-location-features`, `positive-price-features`, `positive-room-features`, `positive-staff-features`) and containing 100 classes. We created no branches for neutral and ambiguous words.

The negative branch is larger than the positive one because reviewers use many different terms to express negative emotions, including impolite and slang words, while they use almost the same terms (“splendid”, “wonderful”, “amazing”, ...) in the positive ones.

In the negative branch, we added two more sub-trees related to the experience, modeling the causes and the consequences of the bad experience: in these branches we added some terms, not essentially “emotions related”, that are often found in negative reviews (for example “refund”, or “broken”). In Figure 2 the reader can see the trees structure of positive and negative branches, with some examples of the terms under

Fig. 2. My MOoD ontology (part of).

each sub-tree: due to space limitation, we cannot describe the complete ontology. The interested reader can retrieve it from the web.

6 Experiments with Hotel Reviews in Five Languages

The research question we tackled to evaluate the effectiveness of TEXTCLASS is:

> **RQ**: *Is* TEXTCLASS *able to classify documents in English, Italian, Spanish, French and German w.r.t. the opinion/sentiment they describe using the Senti-Hotel ontology?*

The metrics used to answer RQ is the number of documents correctly classified over the total number of documents.

Data set. We conducted our preliminary evaluation of TEXTCLASS over a sample of multilingual reviews from TripAdvisor[14]. In particular, we focused on classifying reviews in English, Italian, Spanish, French and German.

For English reviews we chose Wang TripAdvisor Data Set [36]; this Data Set is composed by more than 12000 Json files each of which contains about 10 TripAdvisor reviews with different information about them (e.g., review text, overall score, ID). From this dataset we randomly chose 455 English reviews with a balanced distribution of different overall scores (i.e., we have a similar number of positive and negative reviews).

For Italian, Spanish, French and German reviews, we randomly selected 25 reviews for each language, 5 for each value of the overall score (from 1 to 5), resulting in a total of 100 reviews.

Procedure. To answer our RQ we proceeded as follows:

- We selected the positive-review and negative-review sub-trees of the SentiHotel ontology. Such sub-trees play respectively the role of positive o_P and negative o_N ontologies during the classification performed by TEXTCLASS.
- For each review, we executed TEXTCLASS and recorded the classification w.r.t. the ontologies o_P and o_N. In particular, we recorded the number of different positive and negative elements in the ontologies o_P and o_N that match at least one word in the text of the review. For instance, for a review we can find that m_P=12 is the total number of matches in the positive ontology o_P while m_N=4 is the total number of matches in the negative ontology o_N.
- For each review, we computed the normalized classification C_{norm} in order to fit the range [1,5] (i.e., the same used by the TripAdvisor's reviews). The formula used is $C_{norm} = 5 - ((4 * m_N)/(m_P + m_N))$. In the previous example we obtain $C_{norm} = 4.00$. We have no cases in which m_P=m_N=0. In the other cases, the formula correctly returns 3 when m_P=m_N, 5 when m_N=0 and 1 when m_P=0.
- We classify each review as positive if $C_{norm} >= Tr$, negative otherwise. We initially set Tr to 3, which is the C_{norm} value returned when m_P=m_N, namely when there are as many negative opinion words as the positive ones. Higher values should indicate a positive polarity and lower values a negative one. As discussed below, the results obtained with Tr equal to 3 were not satisfactory, so we empirically devised another threshold, 3.4, giving better results.

[14] http://www.tripadvisor.com/

– For each review, we compared our classification (i.e., computed as shown in the previous step) with the overall score provided by the real user and recorded in the dataset together with the review. The classification is correct when: (1) we classified a review as positive and the user provided a score $>= 3$, (2) we classified a review as negative and the user provided a score < 3. In the other cases the classification is wrong.

Results. Table 1 reports the data used to answer RQ. For each dataset (i.e., set of reviews in a specific language) and for each overall score (i.e., the number [1,5] assigned by the users), it reports the number of correctly classified reviews and the corresponding percentage over the total number of reviews. In the last columns, we report aggregate results over all the five datasets.

Table 1. TEXTCLASS Classification Results (threshold = 3)

Overall Score	Reviews EN			Reviews IT			Reviews FR			Reviews ES			Reviews DE			Reviews		
	Correctly Classified		Total	Correctly Classified		Total	Correctly Classified		Total	Correctly Classified		Total	Correctly Classified		Total	Correctly Classified		Total
	N	%		N	%		N	%		N	%		N	%		N	%	
5	74	91,4	81	5	100,0	5	5	100,0	5	4	80,0	5	5	100,0	5	93	92,1	101
4	145	90,1	161	5	100,0	5	5	100,0	5	4	80,0	5	5	100,0	5	164	90,6	181
3	77	82,8	93	5	100,0	5	4	80,0	5	4	80,0	5	5	100,0	5	95	84,1	113
2	16	31,4	51	3	60,0	5	2	40,0	5	2	40,0	5	0	0,0	5	23	32,4	71
1	39	56,5	69	2	40,0	5	4	80,0	5	4	80,0	5	4	80,0	5	53	59,6	89

Concerning the reviews with evaluation 5 (i.e., very good) or 4 (good), we can see that TEXTCLASS is able to provide, most of the times, a correct classification. In particular, in case of overall score = 5 and considering all the languages employed in the five datasets, TEXTCLASS correctly classifies the 92.1% of the reviews. In three cases, IT, FR, and DE the classification is perfect. Similarly, TEXTCLASS correctly classifies the 90.6% of the reviews with overall score = 4, and in the cases of IT, FR, and DE the classification is completely correct.

Conversely, TEXTCLASS is not able to classify correctly the reviews with evaluation 1 (i.e., very bad) or 2 (bad). Indeed, respectively only in the 59.6% and 32.4% of the cases it produce a correct results. From the data reported in Table 1, it is evident that the result of the classification is unbalanced, and tends to favor positive ratings.

We reported also the classification returned for the reviews with overall score = 3. They express a judgment that obviously is neither positive nor negative. Thus, a binary classification (i.e., positive vs negative) cannot be used for classifying such kind of reviews. But, for such reviews, we expect TEXTCLASS to behave as a classifier which assigns a review to one of the two classes (positive and negative) with a probability of 50%, while, by adopting the threshold Tr=3, this is not true (see the 84.1% reported in the table). Thus we searched for a threshold value that allows to obtain, for the overall score = 3, a results as close as possible to 50%. Such threshold value is 3.4.

Table 2 reports the results of the classification performed using Tr=3.4. Concerning the reviews with evaluation 5 (i.e., very good) or 1 (very bad), we can see that TEXTCLASS is able to provide, most of the times, a correct classification. In particular, in case of overall score = 5 and considering all the languages employed in the five datasets, TEXTCLASS correctly classifies the 83.2% of the reviews. In three cases, IT, FR, and DE the classification is perfect. Similarly, TEXTCLASS correctly classifies the

92.1% of the reviews with overall score = 1, and in the cases of IT, FR, and ES the classification is completely correct.

Table 2. TEXTCLASS Classification Results (threshold = 3.4)

	Reviews EN			Reviews IT			Reviews FR			Reviews ES			Reviews DE			Reviews		
Overall Score	Correctly Classified		Total	Correctly Classified		Total	Correctly Classified		Total	Correctly Classified		Total	Correctly Classified		Total	Correctly Classified		Total
	N	%		N	%		N	%		N	%		N	%		N	%	
5	65	80,2	81	5	100,0	5	5	100,0	5	4	80,0	5	5	100,0	5	84	83,2	101
4	116	72,0	161	5	100,0	5	5	100,0	5	4	80,0	5	5	100,0	5	135	74,6	181
3	49	52,7	93	1	20,0	5	2	40,0	5	3	60,0	5	1	20,0	5	56	49,6	113
2	36	70,6	51	4	80,0	5	5	100,0	5	3	60,0	5	1	20,0	5	49	69,0	71
1	63	91,3	69	5	100,0	5	5	100,0	5	5	100,0	5	4	80,0	5	82	92,1	89

As expected, in cases of reviews that does not express a sharp judgment (i.e., overall score 4 and 2) the correctness of the classification performed by TEXTCLASS is slightly worse. In particular, in case of overall score = 4 and 2, and considering all the languages employed in the five datasets, TEXTCLASS correctly classifies respectively the 74.6% and 69% of the reviews. Also in this cases, TEXTCLASS is able, for certain languages, to perform an exact classification (i.e., IT, FR, and DE when overall score = 4, and FR when overall score = 2). Only in the case of the reviews in DE, with score = 4, the TEXTCLASS is not able to carry out a good classification.

It is interesting to note that the classification of reviews written in French is always correct (obviously excluding overall score 3), and similar results are achieved for Italian reviews.

Obviously the choice of the threshold is subjective but its value has been selected in order to balance the results on the reviews with overall score = 3 and not for achieving the best possible classification. A deeper investigation will be necessary to understand if 3.4 is a good value for the threshold also for larger data sets involving more languages, or if further tuning is required.

To summarize, with respect to the research question RQ we can say that, using an appropriate threshold that we empirically set to 3.4, TEXTCLASS is able to classify correctly the majority of the reviews in all the five considered languages. The preliminary evaluation reported in this paper shows the feasibility of the approach implemented by TEXTCLASS, even if further investigations are required to refined and fine-tune both the approach and the tool.

7 Conclusions and Future Work

In this paper we have discussed the design of My MOoD and our first experiments with TEXTCLASS. Although My MOoD is not implemented yet, the implementation and the correct functioning of IndianaMAS, which shares with My MOoD the architecture and – to some extent – the purpose, makes us very confident in the possibility to actually build and use My MOoD for multilingual and multimedia sentiment analysis.

With respect to TEXTCLASS, our first implementation neglects many aspects that should improve its analysis capability. In particular, we do not deal with negation which is a well known threatening aspect for carrying out a correct analysis of a document's polarity. To make an example, in the first version of the SentiHotel ontology we included `timely` as a positive feature of the staff. However many negative reviews contained complaints about the time required to obtain some services. These reviews were tagged with `timely` which was considered as a positive feature, thus resulting into wrong classifications. We plan to add a document pre-processing stage during which negated words and sentences will be recognized and changed into their antonyms. This stage will be of course language-dependent and for this reason requires time and special care.

In our future work we also plan to investigate the reason behind the unbalanced classification obtained when using the standard threshold Tr=3. For instance, this could depend from the kinds of positive/negative words used in the reviews, from the coverage of the positive/negative words by BabelNet or from the inability of TEXTCLASS of managing negations. However, we plan to better validate the effectiveness of the approach implemented by TEXTCLASS through a cross-validation (aka, leave-one-out) procedure. To this aim we will use a leave-one out cross validation with k datasets. We will split the original datasets into $k-1$ datasets used for training the threshold Tr and one dataset used for testing the effectiveness of TEXTCLASS employing such threshold, with the testing dataset rotated so as to test TEXTCLASS on each of the k available datasets.

References

1. K. Ahmad, D. C. D, and Y. Almas. Multi-lingual sentiment analysis of financial news streams. In *Proc. of CAASL2*, pages 1–12. Linguistic Society of America, 2007.
2. M. Almashraee, D. M. Diaz, and R. Unland. Sentiment classification of on-line products based on machine learning techniques and multi-agent systems technologies. In *Proc. of ICDM 2012*, pages 128–136, 2012.
3. A. Balahur, M. Turchi, R. Steinberger, J. M. P. Ortega, G. Jacquet, D. Küçük, V. Zavarella, and A. E. Ghali. Resource creation and evaluation for multilingual sentiment analysis in social media texts. In *Proc. of LREC-2014*, pages 4265–4269. ELRA, 2014.
4. M. Baldoni, C. Baroglio, V. Patti, and P. Rena. From tags to emotions: Ontology-driven sentiment analysis in the social semantic web. *Intelligenza Artificiale*, 6(1):41–54, 2012.
5. E. Boiy and M. Moens. A machine learning approach to sentiment analysis in multilingual web texts. *Inf. Retr.*, 12(5):526–558, 2009.
6. D. Borth, T. Chen, R. Ji, and S. Chang. Sentibank: large-scale ontology and classifiers for detecting sentiment and emotions in visual content. In *Proc. of ACM MM 2013*, pages 459–460, 2013.
7. D. Borth, R. Ji, T. Chen, T. M. Breuel, and S. Chang. Large-scale visual sentiment ontology and detectors using adjective noun pairs. In *Proc. of ACM MM 2013*, pages 223–232, 2013.
8. D. Briola, V. Deufemia, V. Mascardi, L. Paolino, and N. Bianchi. Ontology-driven processing and management of digital rock art objects in indianamas. In *Proc. of EuroMed 2014*, LNCS, pages 217–227. Springer, 2014.
9. G. Casella, V. Deufemia, V. Mascardi, G. Costagliola, and M. Martelli. An agent-based framework for sketched symbol interpretation. *J. Vis. Lang. Comput.*, 19(2):225–257, 2008.
10. M. Chaves and C. Trojahn. Towards a multilingual ontology for ontology-driven content mining in social web sites. In *Proc. of ISWC 2010, Volume I*, 2010.
11. M. S. Chaves, L. A. de Freitas, and R. Vieira. Hontology: A multilingual ontology for the accommodation sector in the tourism industry. In *Proc. of KEOD 2012*, pages 149–154, 2012.
12. K. Denecke. Using sentiwordnet for multilingual sentiment analysis. In *Proc. of ICDE 2008*, pages 507–512, 2008.
13. C. Fellbaum, editor. *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press, 1998.

14. M. Franco-Salvador, F. L. Cruz, J. A. Troyano, and P. Rosso. Cross-domain polarity classification using a knowledge-enhanced meta-classifier. Under review, 2015.
15. C. Gerber, J. H. Siekmann, and G. Vierke. Holonic multi-agent systems. Technical Report DFKI-RR-99-03, Deutsches Forschungszentrum für Künztliche Inteligenz, 1999.
16. H. Gunes and B. Schuller. Categorical and dimensional affect analysis in continuous input: Current trends and future directions. *Image Vision Comput.*, 31(2):120–136, 2013.
17. N. R. Jennings, K. P. Sycara, and M. Wooldridge. A roadmap of agent research and development. *Autonomous Agents and Multi-Agent Systems*, 1(1):7–38, 1998.
18. Z. Kechaou, M. B. Ammar, and A. M. Alimi. A multi-agent based system for sentiment analysis of user-generated content. *IJAIT*, 22(2), 2013.
19. E. Kontopoulos, C. Berberidis, T. Dergiades, and N. Bassiliades. Ontology-based sentiment analysis of twitter posts. *Expert Syst. Appl.*, 40(10):4065–4074, 2013.
20. B. Liu and L. Zhang. A survey of opinion mining and sentiment analysis. In *Mining Text Data*, pages 415–463. Springer, 2012.
21. V. Mascardi, D. Briola, A. Locoro, V. Deufemia, L. Paolino, N. Bianchi, H. de Lumley, D. Grignani, D. Malafronte, and A. Ricciarelli. A holonic multi-agent system for sketch, image and text interpretation in the rock art domain. *IJICIC*, 10(1):81–99, 2014.
22. R. Mihalcea, C. Banea, and J. Wiebe. Learning multilingual subjective language via cross-lingual projections. In *Proc. of ACL 2007*, pages 976–983. ACL, 2007.
23. G. A. Miller. Wordnet: A lexical database for english. *Commun. ACM*, 38(11):39–41, 1995.
24. R. Navigli and S. P. Ponzetto. Babelnet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artif. Intell.*, 193:217–250, 2012.
25. B. Pang and L. Lee. Opinion mining and sentiment analysis. *Found. Trends Inf. Retr.*, 2(1-2):1–135, Jan. 2008.
26. B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up? sentiment classification using machine learning techniques. *CoRR*, cs.CL/0205070, 2002.
27. V. Patti, F. Bertola, and A. Lieto. Arsemotica for arsmeteo.org: Emotion-driven exploration of online art collections. In *Proc. of FLAIRS 2015*, 2015.
28. R. Plutchik. The nature of emotions. *American Scientist*, 89(4), 2001.
29. J. Polpinij and A. K. Ghose. An ontology-based sentiment classification methodology for online consumer reviews. In *Proc. of IEEE/WIC/ACM WI-IAT'08*, pages 518–524, 2008.
30. B. Schuller, A. Batliner, S. Steidl, and D. Seppi. Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge. *Speech Communication*, 53(9-10):1062–1087, 2011.
31. S. Siersdorfer, E. Minack, F. Deng, and J. S. Hare. Analyzing and predicting sentiment of images on the social web. In *Proc. of MM 2010*, pages 715–718, 2010.
32. C. Strapparava and A. Valitutti. Wordnet affect: an affective extension of wordnet. In *Proc. of LREC-2004*. ELRA, 2004. ACL Anthology Identifier: L04-1208.
33. E. Tromp and M. Pechenizkiy. Senticorr: Multilingual sentiment analysis of personal correspondence. In *Proc. of ICDMW '11*, pages 1247–1250. IEEE Computer Society, 2011.
34. P. D. Turney. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In *Proc. of ACL 2002*, pages 417–424, 2002.
35. G. Vinodhini and R. M. Chandrasekaran. Sentiment analysis and opinion mining: A survey. *IJARCSSE*, 2(6):282–292, June 2012.
36. H. Wang, Y. Lu, and C. Zhai. Latent aspect rating analysis without aspect keyword supervision. In *Proc. of ACM SIGKDD 2011*, KDD '11, pages 618–626, New York, NY, USA, 2011. ACM.
37. T. A. Wilson. Fine-grained subjectivity and sentiment analysis: Recognizing the intensity, polarity, and attitudes of private states. Doctoral Dissertation, University of Pittsburgh, 2008.
38. Q. You, J. Luo, H. Jin, and J. Yang. Robust image sentiment analysis using progressively trained and domain transferred deep networks. In *Proc. of AAAI 2015*, page 10, 2015.
39. J. Yuan, S. Mcdonough, Q. You, and J. Luo. Sentribute: image sentiment analysis from a mid-level perspective. In *Proc. of WISDOM 2013*, page 10, 2013.