EUROPEAN PROJECT SPACE

on Intelligent Technologies, Software engineering, Computer Vision, Graphics, Optics and Photonics

Rome, Italy

February, 2016

Paolo Dell'Olmo, Marco Brambilla and Maria Raposo (Eds.)

Sponsored and Organized by

INSTICC

Published by

SCITEPRESS
SCIENCE AND TECHNOLOGY PUBLICATIONS

European Project Space on Intelligent Technologies, Software engineering, Computer Vision, Graphics, Optics and Photonics

Rome, Italy
February, 2016

Paolo Dell'Olmo, Marco Brambilla and Maria Raposo (Eds.)

Edited by Paolo Dell'Olmo, Marco Brambilla and Maria Raposo

Printed in Portugal
ISBN: 978-989-758-206-6
Depósito Legal: 417003/16

Table of Contents

Papers

Mixed Emotions. Social Semantic Emotion Analysis for Innovative Multilingual Big Data Analytics Markets

Carlos A. Iglesias[1], J. Fernando Sánchez-Rada[1], Paul Buitelaar[2] and Francesco Danza[3]

[1]Intelligent Systems Group, Universidad Politécnica de Madrid, ETSI Telecomunicación, 28040 Madrid, Spain
[2]National University of Ireland, Galway (NUIG), Insight Centre for Data Analytics, IDA Business Park, Lower Dangan, Galway, Ireland
[3]Expert System, Via Virgilio, 48/H, Staircase 5, 41123 Modena, Italy
cif@dit.upm.es
http://www.gsi.dit.upm.es

Abstract. MixedEmotions is a European R&D project that aims at developing innovative multilingual multimodal Big Data analytics applications that analyze a more complete emotional profile of user behavior using data from mixed input channels: multilingual text data sources, A/V signal input (multilingual speech, audio, video), social media (social network, comments), and structured data. Commercial applications (implemented as pilot projects) are in Social TV, Brand Reputation Management and Call Centre Operations. Making sense of accumulated user interaction from different data sources, modalities and languages is challenging and has not yet been explored in fullness in an industrial context. MixedEmotions implements an integrated Big Linked Data platform for emotion analysis across heterogeneous data sources, different languages and modalities, building on existing state of the art tools, services and approaches that will enable the tracking of emotional aspects of user interaction and feedback on an entity level.

1 Introduction

Emotion analysis is central to tracking customer and user behavior and satisfaction [1], which can be observed from user interaction in the form of explicit feedback through email, call center interaction, social media comments, etc., as well as implicit acknowledgment of approval or rejection through facial expressions, speech or other non-verbal feedback. In Europe specifically, but increasingly also globally, an added factor here is that user feedback can be in multiple languages, in text as well as in speech and audio-visual content. This implies different cultural backgrounds and thus different ways to produce and perceive emotions in everyday interactions, beyond the fact of having specific rules for encoding and decoding emotions in each language.

Making sense of accumulated user interaction from different (mixed) data sources, modalities and languages is challenging and has not yet been explored in fullness in an industrial context. Commercial solutions exist but do not address the multilingual aspect in a robust and large-scale setting and do not scale up to huge data volumes that need

to be processed, or the integration of emotion analysis observations across data sources and/or modalities on a meaningful level, i.e. keeping track of entities involved as well the connections between them (who said what? to whom? in the context of which event, product, service?).

The European R&D project MixedEmotions[1] implements an integrated Big Linked Data platform for emotion analysis across heterogeneous data sources, languages and modalities, building on existing state of the art tools, services and approaches that will enable the tracking of emotional aspects of user interaction and feedback on an entity level. The MixedEmotions platform provides an integrated solution for:

- Large-scale emotion analysis and fusion on heterogeneous, multilingual, text, speech, video and social media data streams, leveraging open access and proprietary data sources, exploiting also social context by leveraging social network graphs.
- Semantic-level emotion information aggregation and integration through robust extraction of social semantic knowledge graphs for emotion analysis along multidimensional clusters The platform will be developed and evaluated in the context of three cross-domain Pilot Projects that are representative of a variety of data analytics markets: Social TV, Brand Reputation Management, Call Centre Operations.

The platform is being developed and evaluated in the context of three cross-domain Pilot Projects that are representative of a variety of data analytics markets: Social TV, Brand Reputation Management, Call Centre Operations. Each of the companies involved in the pilot projects have specific innovation objectives as follows:

Pilot Project I (Social TV): Expert System S.p.A. aims at extending its current Expert System COGITO API suite with added features for Social Multimedia Analytics namely in the COGITO Media API. MixedEmotions will enable the inclusion of entity-level emotion and sentiment analysis (program, actor, sportsperson), related contents analytics (liked content, mentioned entities), multimedia media analytics (emotion and sentiment analysis on online videos), and Social TV (related social networks activity while watching TV programs). This pilot is developed in cooperation with the partner Deutsche Welle, who is a German public broadcaster that provides multimedia contents for this pilot.

Pilot Project II (Brand Reputation Management): Paradigma Digital seeks to enhance their Online Reputation Tool in four main areas: enhance real-time capabilities, adding multimodal input would dramatically enhance the quality and accurateness of generated reputation reports, multilingual data analysis and report generation to increase internationalisation of the product. As all of these are central to MixedEmotions, its outcomes will enable Paradigma to provide better reputation management services to its customers.

Pilot Project III (Call Centre Operations): Phonexia offers solutions for call centre operations which will benefit from the MixedEmotions platform in the following ways: enable the implementation and use of optimization algorithms for speech recognition in the context of big data MixedEmotions (huge streams of speech recordings

[1] http://mixedemotions-project.eu

in mid-size and large call centres), advanced extraction of emotions from speech, and enhanced multilingual processing.

Pilot Project I-III: SindiceTech helps companies build "Enterprise Knowledge Graphs" to quickly leverage multiple information sources in big data, high variety environments. MixedEmotions will enable the inclusion of emotion as knowledge in an Enterprise Knowledge Graph and exploiting this in use cases such as Social TV, Brand reputation Management and Call Centre Operations.

MixedEmotions will apply a technology readiness level approach to assess the maturity of the technologies and solutions adopted. This approach, in conjunction to the Customer Development methodology used, will stimulate the process of technology transference in the project and the readiness of its further commercialisation.

Table 1. Language and Modalities of the pilots.

Pilot	Media, modalities	Languages covered
SocialTV	A/V, Social Media	English, German, Spanish
Brand Reputation Management	Social Media, text, A/V	English, Spanish
Call Center Operation	Speech, Audio, Social Media	Czech, English

2 Linked Data Model Built upon Previous Results: The Eurosentiment Project

The work presented here is partly based on an earlier work [2] developed within the European R&D Eurosentiment project[2].

The Eurosentiment model proposes a linked data approach for sentiment and emotion analysis, and it is based on the following specifications:

- Marl [3] is a vocabulary designed to annotate and describe subjective opinions expressed on the web or in information systems
- Onyx [4] is built on the same principles as Marl to annotate and describe emotions, and provides interoperability with Emotion Markup Language (EmotionML) [5]
- Lexicon Model for Ontologies (lemon) [6] defines a lexicon model based on linked data principles which has been extended with Marl and Onyx for sentiment and emotion annotation of lexical entries
- NLP Interchange Format (NIF) 2.0 [7] which defines a semantic format and API for improving interoperability among natural language processing services

The way these vocabularies have been integrated is illustrated in the example below, where we are going to analyse the sentiment of an opinion ("Like many Paris hotels, the rooms are too small") posted in TripAdvisor. In the Eurosentiment model, *lemon* is used to define the lexicon for a domain and a language. In our example, we have to generate this lexicon for the hotel domain and the English language[3]. A reduced lexicon for Hotels in English (le:hotel_en) is shown in Listing 1.1 for illustration purposes. The

[2] http://eurosentiment.eu/

[3] The reader interested in how this domain specific lexicon can be generated can consult [8].

lexicon is composed of a set of lexical entries (prefix lee). Each lexical entry is semantically disambiguated and provides a reference to the syntactic variant (in the example the canonical form) and the senses. The example shows how the senses have been extended to include sentiment features. In particular, the sense small_1 in the context of room_1 has associated a negative sentiment. That is, "small room" is negative (while small phone could be positive, for example).

The way these vocabularies have been integrated is illustrated in the example below, where we are going to analyse the sentiment of an opinion ("Like many Paris hotels, the rooms are too small") posted in TripAdvisor. In the Eurosentiment model, *lemon* is used to define the lexicon for a domain and a language. In our example, we have to generate this lexicon for the hotel domain and the English language[4]. A reduced lexicon for Hotels in English (le:hotel_en) is shown in Listing 1.1 for illustration purposes. The lexicon is composed of a set of lexical entries (prefix lee). Each lexical entry is semantically disambiguated and provides a reference to the syntactic variant (in the example the canonical form) and the senses. The example shows how the senses have been extended to include sentiment features. In particular, the sense small_1 in the context of room_1 has associated a negative sentiment. That is, "small room" is negative (while small phone could be positive, for example).

```
lee:sense/small_1 a lemon:Sense;
lemon:reference "01391351";
lexinfo:partOfSpeech lexinfo:adjective;
lemon:context lee:sense/room_1;
marl:polarityValue "-0.5"^^xsd:double;
marl:hasPolarity marl:negative.

le:hotel_en a lemon:Lexicon;
lemon:language "en";
lemon:topic ed:hotel;
lemon:entry lee:room, lee:Paris, lee:small.

lee:room a lemon:LexicalEntry;
lemon:canonicalForm [ lemon:writtenRep "room"@en ];
lemon:sense [ lemon:reference wn:synset-room-noun-1;
lemon:reference dbp:Room ];
lexinfo:partOfSpeech lexinfo:noun.

lee:Paris a lemon:LexicalEntry;
lemon:canonicalForm [ lemon:writtenRep "Paris"@en ];
lemon:sense [ lemon:reference dbp:Paris;
lemon:reference wn:synset-room-noun-1 ];
lexinfo:partOfSpeech lexinfo:noun.

lee:small a lemon:LexicalEntry;
lemon:canonicalForm [ lemon:writtenRep "small"@en ];
lemon:sense lee:sense/small_1;
lexinfo:partOfSpeech lexinfo:adjective.
```

Listing 1.1. Sentiment analysis expressed with Eurosentiment model

[4] The reader interested in how this domain specific lexicon can be generated can consult [8].

The Eurosentiment model uses NIF in combination with Marl and Onyx to provide a standardised service interface. In our example, let us assume the opinion has been published at http://tripadvisor.com/myhotel. NIF follows a linked data principled approach so that different tools or services can annotate a text. To this end, texts are converted to RDF literals and an URI is generated so that annotations can be defined for that text in a linked data way. NIF offers different URI Schemes to identify text fragments inside a string, e.g. a scheme based on RFC5147 [9], and a custom scheme based on context. In addition to the format itself, NIF 2.0 defines a REST API for Natural Language Processing (NLP) services with standardised parameters. An example of how these ontologies are integrated is illustrated in Listings 1.2, 1.3 and 1.4.

```
<http://tripadvisor.com/myhotel#char=0,49>
rdf:type nif:RDF5147String , nif:Context;
nif:beginIndex "0";
nif:endIndex "49";
nif:sourceURL <http://tripadvisor.com/myhotel.txt>;
nif:isString "Like many Paris hotels, the rooms are too small
   ";
marl:hasOpinion <http://tripadvisor.com/myhotel/opinion/1>.
```

Listing 1.2. NIF + Marl output of a service call http://eurosentiment.eu?i=Like many Paris hotels, the rooms are too small.

```
<http://tripadvisor.com/myhotel/opinion/1>
rdf:type marl:Opinion;
marl:describesObject dbp:Hotel;
marl:describesObjectPart dbp:Room;
marl:describesFeature  "size";
marl:polarityValue "-0.5";
marl:hasPolarity: http://purl.org/marl/ns#Negative.
```

Listing 1.3. Sentiment analysis expressed with Eurosentiment model.

```
<http://eurosentiment.eu/analysis/1>
rdf:type marl:SentimentAnalysis;
marl:maxPolarityValue "1";
marl:minPolarityValue "-1";
marl:algorithm "dictionary-based";
prov:used le:hotel_en;
prov:wasAssociatedWith http://dbpedia.org/resource/UPM.
```

Listing 1.4. Sentiment analysis expressed with Eurosentiment model.

In the context of the MixedEmotions project, these vocabularies are being extended to provide coverage of the social context of users, so that we can analyse sentiments and emotions based on the social interactions of the users. In addition, MixedEmotions aims at providing a reference architecture for fostering the use of these vocabularies so that emotion applications and services can seamless interoperate, and language resources can be easily reused or repurposed.

3 MixedEmotions Architecture

The MixedEmotions project is centered around the development of a Big Linked Data platform for emotion recognition and analysis that will enable faster, more accurate and integrated analysis of emotion in multiple use cases, across application domains, languages and modalities. The MixedEmotions platform will build on existing technologies and approaches available from the industry and academic partners. The platform will be developed and tested in the context of the three pilot projects mentioned previously: Social TV, Brand Reputation Management, Call Centre Operations.

MixedEmotions will build on existing infrastructure. Paradigma Digital provides a Big Data platform that will form the core of the MixedEmotions platform, extending it with the large scale solutions for semantic information integration (Enterprise Knowledge Graphs) provided by SindiceTech. Basic services for emotion recognition from speech are provided by Phonexia (PSIP), as well as for emotion recognition and analysis from text by Expert System (Expert System COGITO API). More advanced algorithms for multilingual, multimodal emotion recognition and analysis are provided by the academic partners UP (emotion recognition from speech, audio, video), Brno University of Technology (BUT) and Insight Centre for Data Analytics, National University of Galway (NUIG) (emotion recognition from text in multiple languages). Academic partner Technical University of Madrid (UPM) provides advanced algorithms for the analysis of social context of emotion from social media. All of these algorithms will be integrated into the MixedEmotions platform as services. Academic partners NUIG, UPM and BUT will also add existing approaches for entity linking and semantic relatedness on unstructured data to the SindiceTech Enterprise Knowledge Graphs platform.

The pilot projects will use large-scale open, linked and proprietary data sources as described in detail above.

The picture below shows the basic architecture of the MixedEmotions platform as well as the envisioned data value chain arising from the integrated approach taken.

The MixedEmotions platform will provide a number of content analytic tools, such as emotion recognition from text, audio and video, which can use other services of the platform, such as speech transcription and translation. In addition, the platform provides semantic analysis of the user contributions in social networks, which advanced services for topic extractions, suggestion mining and social network analysis. Apart from providing emotion analysis of the different sources, the project aims at researching on fusion model of the different modalities in order to improve fusion prediction.

4 Outlook

The MixedEmotions project aims at providing a Big Linked Data platform for multimodal emotion recognition, that exhibits a novel emotion fusion approach and contributes to the Linguistic Linked Open Data Cloud. The platform will be defined based on business requirements and validated in commercial use cases. The MixedEmotions project is leading the W3 Community Group Linked Data Model for Sentiment and Emotion Analysis[5] with the aim of reaching a consensus in the sentiment analysis community.

[5] https://www.w3.org/community/sentiment

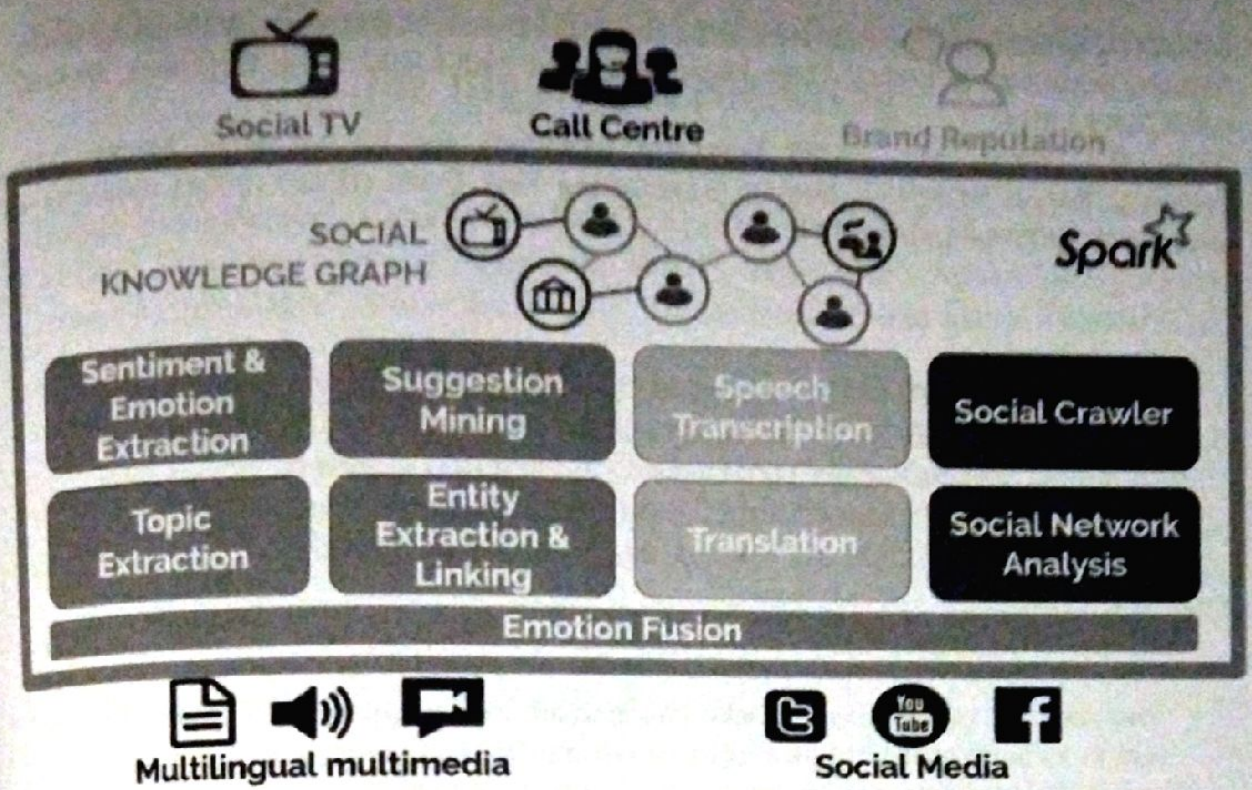

Fig. 1. MixedEmotions architecture.

Acknowledgements. This research has been partially funded by European Union through the projects MixedEmotions (Grant Agreement no: 141111) and Eurosentiment FP7 Project (Grant Agreement no: 296277) and by the Spanish Ministry of Economy though the project SEMOLA (TEC2015-68284-R).

References

1. Peters, L.D., Pressey, A.D., Greenberg, P.: The impact of CRM 2.0 on customer insight. Journal of Business & Industrial Marketing 25 (2010) 410–419
2. Buitelaar, P., Arcan, M., Iglesias, C.A., Sánchez-Rada, J.F., Strapparava, C.: Linguistic linked data for sentiment analysis. In: 2nd Workshop on Linked Data in Linguistics (LDL-2013): Representing and linking lexicons, terminologies and other language data, Pisa, Italy (2013)
3. Westerski, A., Iglesias, C.A., Tapia, F.: Linked opinions: Describing sentiments on the structured web of data (2011)
4. Sanchez-Rada, J.F., Iglesias, C.A.: Onyx: Describing emotions on the web of data. In: ESSEM@ AI* IA, Citeseer (2013) 71–82
5. Schröder, M., Baggia, P., Burkhardt, F., Pelachaud, C., Peter, C., Zovato, E.: Emotionml an upcoming standard for representing emotions and related states. In DMello, S., Graesser, A., Schuller, B., Martin, J.C., eds.: Affective Computing and Intelligent Interaction. Volume 6974 of Lecture Notes in Computer Science. Springer Berlin Heidelberg (2011) 316–325
6. Buitelaar, P., Cimiano, P., McCrae, J., Montiel-Ponsoda, E., Declerck, T.: Ontology lexicalization: The lemon perspective. In: Workshop at 9th International Conference on Terminology and Artificial Intelligence (TIA 2011). (2011) 33–36
7. Hellmann, S., Lehmann, J., Auer, S., Brümmer, M.: Integrating nlp using linked data. In: The Semantic Web–ISWC 2013. Springer (2013) 98–113
8. Vilari, G., Buitelaar, P., Negi, S., Pereira, B., Arcan, M., Coughlan, B., Sánchez-Rada, J.F., Iglesias, C.A.: Generating [illegible] Sentiment Lexicons from

Legacy Language and Semantic Resources. In: th International Workshop on Emotion, Social Signals, Sentiment & Linked Open Data, co-located with LREC 2014, Reykjavik, Iceland, LREC2014 (2014)
9. Wilde, E., Duerst, M.: URI Fragment Identifiers for the text/plain Media Type. Technical Report RDF5147, Internet Engineering Task Force (IETF) (2008) Available at https://tools.ietf.org/html/rfc5147.