Available online at www.sciencedirect.com

ScienceDirect

Computer Speech and Language 28 (2014) 1–6

COMPUTER SPEECH AND LANGUAGE

www.elsevier.com/locate/csl

Preface

Computational approaches to subjectivity and sentiment analysis: Present and envisaged methods and applications

Abstract

Recent years have witnessed a surge of interest in computational methods for affect, ranging from opinion mining, to subjectivity detection, to sentiment and emotion analysis. This article presents a brief overview of the latest trends in the field and describes the manner in which the articles contained in the special issue contribute to the advancement of the area. Finally, we comment on the current challenges and envisaged developments of the subjectivity and sentiment analysis fields, as well as their application to other Natural Language Processing tasks and related domains.

Keywords: Subjectivity analysis; Sentiment analysis; Multilingual resources; Social Media mining; Chat analysis

1. Introduction

Recent years have witnessed a surge of interest in computational methods for affect, ranging from opinion mining, to subjectivity detection, to sentiment and emotion analysis. These methods typically focus on the identification of private states, such as opinions, emotions, sentiments, evaluations, beliefs, and speculations in natural language. While subjectivity classification labels text as either subjective or objective, sentiment classification adds an additional level of granularity by further classifying subjective text as either positive, negative or neutral, which is then further refined by emotion analysis by identifying the presence of emotions such as joy, anger, or fear.

In computational linguistics, the automatic detection of affect in texts is becoming increasingly important from an applicative point of view. Consider for example the tasks of opinion mining, market analysis, or natural language interfaces such as e-learning environments or educational/edutainment games. For instance, the following represent examples of applicative scenarios in which affective computing could make valuable and interesting contributions:

- *Sentiment analysis*. Text categorization according to affective relevance, opinion exploration for market analysis, etc., are examples of applications of these techniques. While positive/negative valence annotation is an active area in sentiment analysis, a fine-grained emotion annotation could also contribute to the effectiveness of these applications.
- *Computer assisted creativity* The automated generation of evaluative expressions with a bias on certain polarity orientation is a key component in automatic personalized advertisement and persuasive communication.
- *Verbal expressivity in human–computer interaction* Future human–computer interaction is expected to emphasize naturalness and effectiveness, and hence the integration of models of possibly many human cognitive capabilities, including affective analysis and generation. For example, the expression of emotions by synthetic characters (e.g., embodied conversational agents) is now considered a key element for their believability. Affective words selection and understanding is crucial for realizing appropriate and expressive conversations.

0885-2308/$ – see front matter © 2013 Published by Elsevier Ltd.
http://dx.doi.org/10.1016/j.csl.2013.09.003

The articles contained in this special issue are in their majority the extended versions of the best articles (as reviewed by the Program Committee) that have been presented at the 3rd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA 2012) – http://gplsi.dlsi.ua.es/congresos/wassa2012/. The event was organized in conjunction to the 50th Annual Meeting of the Association for Computational Linguistics, on July 12, 2012, in Jeju, Republic of Korea.

This edition has again shown that both the academic, as well as the industry communities have a great interest in the topics covered by the workshop. The large number of submissions and the tough review process ensured a high quality of the papers selected for presentation at the event. As such, the articles that are contained in this special issue regard important challenges in subjectivity and sentiment analysis areas, advancing the state of the art in these fields. They cover topics such as: multilingual sense subjectivity, multilingual sentiment analysis, lexicon and corpora building for subjectivity and sentiment analysis, emotion detection and contextuality and the applications of subjectivity and sentiment analysis to the detection of social and psychological phenomena. The types of texts on which the analysis is done vary from newspaper articles, to blogs and Social Media messages, thus exploring also the challenges posed by the structure of each of these text types.

In the following section, we present some of the most relevant work that has been recently conducted in subjectivity and sentiment analysis and describe the research trends in this field.

2. Recent trends in subjectivity and sentiment analysis

Subjectivity and sentiment analysis have been very active research topics in the past decade. Different authors have proposed methods to tackle the tasks from various types of texts and with diverse applications. Although much has been achieved in the field, some aspects still require additional tackling and further efforts are needed to expand the results to the multilingual and cross-lingual settings and new, informal types of texts. The articles contained in this special issue represent an advancement to the state of the art in this sense, as they deal with approaches to the above-mentioned issues. Their efforts are inscribed in the recent trends in the field, represented by related work described in the following subsections.

2.1. Multilingual subjectivity and sentiment analysis

The first aspect on which subjectivity and sentiment analysis (SSA) requires further efforts is related to the analysis of multilingual texts. The manual annotation of resources is a tedious and costly task, therefore very few task-specific corpora and dictionaries for SSA exist. In order to overcome this problem, researchers have proposed methods to adapt existing resources and tools for sentiment analysis developed for English to build resources in other languages. In this sense, lexicons, annotation schemes and annotated corpora were transferred to new languages using bilingual dictionaries, monolingual and multilingual bootstrapping or machine translation.

Existing English subjectivity and sentiment lexicons were transferred into Chinese (Ku et al., 2006), Romanian (Mihalcea et al., 2007), and Italian (Esuli and Sebastiani, 2011). In an early work, Kim and Hovy (2006) use a machine translation system and subsequently employ a subjectivity analysis system that was developed for English to create subjectivity analysis resources in other languages. Ahmad et al. (2007) use the topical distributions in different languages to detect important sentiment phrases in a multilingual setting, starting from the idea that words with a lower frequency are more representative of the topic and searching for sentiment-related terms around those. Inui and Yamamoto (2011) employ machine translation and, subsequently, sentence filtering to eliminate the noise obtained in the translation process, based on the idea that sentences that are translations of each other should contain sentiment-bearing words that have the same polarity. Mihalcea et al. (2007) propose a method to learn multilingual subjective language via cross-language projections. They use the Opinion Finder lexicon (Wilson et al., 2005) and use two bilingual English-Romanian dictionaries to translate the words in the lexicon. Another approach was proposed by Banea et al. (2008). To this aim, the authors perform three different experiments – translating the annotations of the MPQA corpus, using the automatically translated entries in the Opinion Finder lexicon and the third, validating the data by reversing the direction of translation. In a further approach, Banea et al. (2008) apply bootstrapping to build a subjectivity lexicon for Romanian, starting with a set of 60 words which they translate and subsequently filter using a measure of similarity to the original words, based on Latent Semantic Analysis (LSA) Deerwester et al. (1990) scores. Yet another approach to mapping subjectivity lexicons to other languages is proposed by Xiaojun (2009), who uses co-training to classify