Expert Systems with Applications 42 (2015) 7375–7385

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

The role of idioms in sentiment analysis

Lowri Williams [a], Christian Bannister [a], Michael Arribas-Ayllon [b], Alun Preece [a], Irena Spasić [a,*]

[a] School of Computer Science & Informatics, Cardiff University, 5 The Parade, Cardiff CF24 3AA, UK
[b] School of Social Sciences, Cardiff University, King Edward VII Avenue, Cardiff CF10 3WT, UK

ARTICLE INFO

Article history:
Available online 28 May 2015

Keywords:
Emotion recognition
Sentiment analysis
Natural language processing
User-generated content
Tagging

ABSTRACT

In this paper we investigate the role of idioms in automated approaches to sentiment analysis. To estimate the degree to which the inclusion of idioms as features may potentially improve the results of traditional sentiment analysis, we compared our results to two such methods. First, to support idioms as features we collected a set of 580 idioms that are relevant to sentiment analysis, i.e. the ones that can be mapped to an emotion. These mappings were then obtained using a web-based crowdsourcing approach. The quality of the crowdsourced information is demonstrated with high agreement among five independent annotators calculated using Krippendorff's alpha coefficient ($\alpha = 0.662$). Second, to evaluate the results of sentiment analysis, we assembled a corpus of sentences in which idioms are used in context. Each sentence was annotated with an emotion, which formed the basis for the gold standard used for the comparison against two baseline methods. The performance was evaluated in terms of three measures – precision, recall and *F*-measure. Overall, our approach achieved 64% and 61% for these three measures in two experiments improving the baseline results by 20 and 15 percent points respectively. *F*-measure was significantly improved over all three sentiment polarity classes: Positive, Negative and Other. Most notable improvement was recorded in classification of positive sentiments, where recall was improved by 45 percent points in both experiments without compromising the precision. The statistical significance of these improvements was confirmed by McNemar's test.

1. Introduction

The proliferation of user-generated content (e.g. product reviews) on the Web 2.0 provides opportunities for many practical applications that require consumer opinion (e.g. market research) as an alternative or a supplement to more traditional qualitative research methods such as surveys, interviews and focus groups. However, the sheer scale of text data acquired from the Web poses challenges to qualitative analysis. Text mining has emerged as a potential solution to the problems of information overload associated with reading vast amounts of text originating from diverse sources. In particular, sentiment analysis (or opinion mining) aims to automatically extract and classify sentiments (the subjective part of an opinion) and/or emotions (the projections or display of a feeling) expressed in text (Liu, 2010; Munezero, Montero, Sutinen, & Pajunen, 2014). Most research activities in this domain have focused on the problem of sentiment classification, which classifies an opinionated text segment (e.g. phrase, sentence or paragraph) in terms of its polarity: positive, negative or neutral (e.g. Aue & Gamon, 2005; Bethard, Yu, Thornton, Hatzivassiloglou, & Jurafsky, 2004; Breck, Choi, & Cardie, 2007).

Features used to support sentiment analysis include terms, part of speech, syntactic dependencies and negation (Pang & Lee, 2008). Most commonly, opinionated words that carry subjective bias are used in a bag-of-words approach to classify opinions (e.g. Attardi & Simi, 2006). Opinionated words can be utilized from lexicons such as SentiWordNet (Esuli & Sebastiani, 2006), WordNet-Affect (Valitutti, Strapparava, & Stock, 2004) and NRC word–emotion association lexicon (Mohammad & Turney, 2010). Dynamic calculation of word polarity (or semantic orientation) based on its statistical association with a set of positive and negative paradigm words is an alternative to predefined lexicons of opinionated words (Turney & Littman, 2003). Other features explored in sentiment analysis include more complex linguistic models based on lexical substitution, *n*-grams and phrases (Dave, Lawrence, & Pennock, 2003). Using an *n*-gram graph based method to assign sentiment polarity to individual word senses, experiments implied that figurative language (i.e. the language which digresses from literal meanings) not only conveys sentiment, but actually drives the polarity of a sentence (Rentoumi, Vouros, Karkaletsis, & Moser, 2012). Although the value of phrase-level features in sentiment

* Corresponding author. Tel.: +44 29 2087 0320; fax: +44 29 2087 4598.
E-mail addresses: WilliamsL10@cardiff.ac.uk (L. Williams), BannisterCA@cardiff.ac.uk (C. Bannister), Arribas-AyllonM@cardiff.ac.uk (M. Arribas-Ayllon), PreeceAD@cardiff.ac.uk (A. Preece), i.spasic@cs.cardiff.ac.uk (I. Spasić).

http://dx.doi.org/10.1016/j.eswa.2015.05.039
0957-4174/

analysis has been acknowledged (Socher et al., 2013; Wilson, Wiebe, & Hoffmann, 2009), few approaches have extensively explored idioms as features of this kind (e.g. Thelwall, Buckley, & Paltoglou, 2012). Nonetheless, the error analysis of sentiment classification results often reveals that the largest percentage of errors are neutral classifications when no opinionated words are present or when idioms are used to express sentiment (Balahur et al., 2010).

Idioms are often defined as multi-word expressions, the meaning of which cannot be deduced from the literal meaning of constituent words, e.g. the idiom *a fish out of water* is used to refer to someone who feels uncomfortable in a particular situation. To distinguish idioms from related linguistic categories such as formulae, fixed phrases, collocations, clichés, sayings, proverbs and allusions, the following properties need to be considered (Nunberg, Sag, & Wasow, 1994):

1. *Conventionality*: Their meaning cannot be (entirely) predicted from the constituent words considered independently.
2. *Inflexibility*: Their syntax is restricted, i.e. idioms do not vary much in way they are composed.
3. *Figuration*: Idioms typically have figurative meaning stemming from metaphors, hyperboles and other types of figuration.
4. *Proverbiality*: Idioms usually describe a recurrent social situation.
5. *Informality*: Idioms are associated with less formal language such as colloquialism.
6. *Affect*: Idioms typically imply an affective stance toward something rather than a neutral one.

The last property emphasizes the importance of idioms in sentiment analysis as it implies that an idiom itself may often be sufficient to determine the underlying sentiment. There are two requirements for idioms to be effectively utilized in sentiment analysis methods: (1) Idioms need to be recognized in text, and (2) the associated sentiment needs to be explicitly encoded.

The inflexibility property (see property 2 above) makes the first requirement feasible. Lexico-syntactic patterns can be used to model idioms computationally and recognize their occurrences in text. A lot of the idioms are frozen phrases such as *by and large*, which can be recognized by simple string matching. Syntactic changes such as inflection (e.g. verb tense change) are often seen in idioms (Yusifova, 2013). Such linguistic phenomena can be modeled by regular expressions, e.g. *spill[s|t|ed] the beans*. More complex idioms have variables for open argument places (Jackendoff & Pinker, 2005) (e.g. *put someone in one's place*), which can still be modeled by means of lexico-syntactic patterns (e.g. *put NP in PRN's place*) and recognized in a linguistically pre-processed text. Less often, idioms are "syntactically productive", i.e. they can be changed syntactically without losing their figurative meaning, e.g. *John laid down the law* can be passivized to *the law was laid down by John* while retaining the original figurative interpretation that John enforced the rules (Gibbs & Nayak, 1989). Transformational grammars have been suggested as a framework to handle more complex syntactic changes such as nominalization (e.g. *you blew some steam off* vs. *your blowing off some steam*) (Fraser, 1970).

In Polish, a highly inflected language, idioms were recognized using a cascade of regular expressions and their effect on sentiment analysis results was evaluated on a corpus of product and service reviews, where idioms were found to occur rarely (Buczynski & Wawer, 2008). In English, despite the obvious need for regular expressions, idioms are usually recognized using a lexicon-based approach, which can only recognize those idioms that are syntactically unproductive or frozen. For example, (Shastri, Parvathy, Abhishek, J., & R., 2010) used a dictionary of idioms (e.g. *at a snail's pace*) in order to recognize them in text and map them to their abstract meaning (e.g. *slow*), which is then utilized to infer the sentiment. In another lexicon-based approach (Beigman Klebanov, Burstein, & Madnani, 2013), idiom recognition was further limited to 46 noun–noun compounds (e.g. *glass ceiling*). The use of their sentiment profiles was found to improve the performance of sentiment classification on a corpus of test-takers essays. Our own study aims to go beyond a lexicon-based approach to recognition of English idioms and use regular expressions instead. The added overhead of handcrafting regular expressions allowed us to explore a much wider set of idioms (beyond the *low-hanging fruits*) as part of sentiment analysis.

Assuming that idioms can be identified in text automatically, we need additional knowledge about the underlying sentiment in order to utilize them as features of sentiment analysis. While idioms have been extensively studied across many disciplines (e.g. linguistics, psychology, etc.), thus far there is no comprehensive knowledge base that systematically maps idioms to sentiments. This is the main reason why idioms have been underrepresented as features used in sentiment analysis approaches with few exceptions (e.g. Xie and Wang (2014) describe a set of 8160 Chinese idioms). Due to the subjective nature of the problem, multiple annotations are required in order to either determine the prevalent sentiment associated with an idiom or use a fuzzy logic approach to represent the sentiment with a degree of truthfulness and falsehood. To support this task, a web-based crowdsourcing approach can be used to efficiently collect a large amount of information relevant for sentiment analysis (Greenwood, Elwyn, Francis, Preece, & Spasić, 2013). We used crowdsourcing to systematically map 580 English idioms to 10 emotion categories, which represents the largest lexico-semantic resource of this kind to utilize in sentiment analysis.

The purpose of this paper is to study the effect of idioms on sentiment analysis. The study was designed as follows (see Fig. 1): (1) collect a set of idioms that can be mapped to sentiments, (2) map individual idioms to sentiments in order to support them as features for sentiment analysis, (3) assemble a corpus of sentences in which these idioms are used, (4) annotate sentences in the corpus with sentiments in order to create a gold standard for sentiment analysis, (5) implement a sentiment analysis approach that incorporates idioms as features, and (6) compare the evaluation results against a traditional sentiment analysis approach using the gold standard created in (4).

2. Data collection

2.1. Idioms

Idioms pose considerable difficulties for English language learners. Failure to understand idioms in context significantly affects one's understanding of language in a variety of personal and professional situations (Nippold & Martin, 1989). It is therefore not surprising that most syllabi for English as a second language pay special attention to studying idioms (Liu, 2003). As a result, there is an abundance of teaching material dedicated to the study of idioms. In this study, we relied upon an educational web site – *Learn English Today* (Learn English Today, 2013), which organizes idioms by themes, many of which can be mapped to emotions either directly (e.g. *Happiness/Sadness*) or indirectly (e.g. *Success/Failure*). We focused specifically on emotion-related idioms, as these are anticipated to have a substantial impact on sentiment analysis. We selected 16 out of a total of 60 available themes, listed in Table 1 together with a number of associated idioms. A total of 580 idioms were collected.