Sentiment Recognition by Rule Extraction from Support Vector Machines

Joachim Diederich
James Cook University
600 Upper Thomson Rd.
Singapore 574421
Singapore
+65 6576 6856
joachim.diederich@jcu.edu.sg

Denise Dillon
James Cook University
600 Upper Thomson Rd.
Singapore 574421
Singapore
+65 6576 6880
denise.dillon@jcu.edu.sg

Abstract

Affective computation allows machines to express and recognize emotions, a core component of computer games. A natural way to express emotion is language, through text and speech; computational methods that accurately recognize emotion in text and speech are therefore important. Machine learning techniques such as support vector machines (SVMs) have been used successfully for topic detection in documents and speech as well as for the identification of authors/speakers. SVMs have also been used for emotion detection in written and spoken communication, although with mixed success. An impediment to emotion extraction by use of support vector machines is that, after learning, it is not quite clear what has been learned. For instance, a gamer may acoustically respond to a character with fear and the SVMs that observe user behaviour confuse the sentiment (fear) with the character (e.g. an in-game persona). Successful emotion identification by support vector machines requires methods that ensure the recognition of sentiments without any confusion with certain topics or characters. This paper provides an introduction to affective computation and rule extraction from support vector machines, a set of techniques used for emotion recognition in text.

Keywords

Emotion recognition, text classification, machine learning, support vector machines

1. Introduction

Recent interest has been shown in moving beyond the performance and efficiency focus of usability towards an awareness of the emotional impact of computers. Affective computing is a field of research exploring the means by which computers can be trained to recognise emotions. Computers that can recognise (and even express) emotions are important for a number of key application areas: interaction design (e.g. the use and creation of ambience) and human computer interaction (e.g. better help systems), computer supported cooperative work (e.g. better collaborative tools), Web search engines (e.g. rank pages with regard to emotional impact) and computer games that entertain and challenge.

The approach introduced here links affective computation with information retrieval and text mining. Building on recent successful projects on authorship attribution and topic detection by use of support vector machines, *this project aims at emotional content classification and ranking*. Similar to previous studies, full-text classification by use of all words in a document is performed. In this project, however,

the targets for supervised machine learning are descriptions of emotional states, and the experiments carefully separate emotion from topical information (facts and events), styles of a particular author and other content that is not affective by nature. The experimental design allows isolation of emotional content without contamination of other types of information. In addition, this study differentiates between *(1) affect expressed in a document* and *(2) the probable emotional response of the reader*.

2. Mood, Text Classification and Support Vector Machines

Cortes and Vapnik [1] introduce support vector machines (SVMs) which are based on the structural risk minimisation principle. Joachims [2] investigates SVMs that produce a *ranking* rather than the usual binary positive/negative decision. Tibianna [3] allows a user to manually rank a set of retrieval results based on relevancy. Once a required number of results are ranked, the set is downloaded, processed and presented to support vector machines for learning. Tibianna can actively reorder or discard search engine results based on the model it has learned. This provides a way of improving search results without requiring query refinement. Learning outcomes from experimental trials with Tibianna demonstrate the implications of using different pre-processing techniques and corpus sizes.

MyNewsWave [4] uses machine learning (including support vector machines) for a user-centred approach to full-text information retrieval as well as news delivery. The system uses knowledge sources such as WordNet [5] to refine keyword queries and learns user-preferences with regard to Web search. First steps have been made towards the assessment of incoming information with regard to the ‘mood’ or personal relevance to a user: DigiMood is a component of MyNewsWave that classifies web pages into mood categories.

Diederich, Kindermann, Leopold and Paass [6] apply support vector machines to the problem of authorship attribution. The authors performed a number of experiments with texts from a German newspaper. With nearly perfect reliability the SVM was able to reject non-authors and detected the target author in 60-80% of the cases. In a second experiment, Diederich et al. [6] investigated a more `content'-free summary of a text, including counts of grammatical tags combined with bigrams to capture morphologic details of language patterns. This resulted in slightly reduced performance. Author detection with SVM on full word forms is remarkably robust even if the author writes about a number of different topics such as sports, politics, or business, etcetera.

The significant learning capability of SVMs comes at a cost – an inherent inability to explain in a comprehensible form the process by which a learning result is reached. Hence, the situation is similar to neural networks where the apparent lack of an explanation capability has led to various approaches aiming at extracting symbolic rules from neural networks. For SVMs to gain a wider degree of acceptance it is desirable to offer an ‘explanation’ capability [7] in the form of rule extraction from SVMs. This technique is utilised in the experiments outlined below.

3. Emotion Classification from Text

The main problem in extracting emotional content from text, or lexical affect sensing, is that deeper meanings are not readily accessible through the surface level content. Many approaches enable access to surface level emotional content through identification of emotional keywords as defined in purpose-built dictionaries. However, emotional content beyond the surface level is much less accessible through such means. Fragopanagos and Taylor [8] give the example of the word 'accident', which is likely to provide a surface indication of negative emotion. However, the word itself does not sufficiently indicate the underlying meaning in the sentence 'I met my girlfriend by accident'.

Another problem for affect sensing is the question of whether emotions can be usefully classified into discrete units. Debate over this issue can be represented roughly by the difference between Russell's circumplex model of affect [9] [10] and the OCC model proposed by Ortony, Clore and Collins [11]. While Russell and colleagues propose a move away from the idea of defining affect as discrete sets of basic emotions, Ortony [12] proposes a consolidation of some twenty-two distinct emotion types into specialised categories of positive or negative reactions. There is perhaps not such a large divide between the two views; the circumplex model can also be described as bipolar clusters of positive and negative affect, albeit varying in levels of activation [13].

Existing approaches to the extraction of emotional content in texts include keyword spotting, lexical affinity and statistical natural language processing. The keyword spotting approach draws on categorical or dimensional systems of emotions as sources of classified affect words. For example, the Affective Lexicon [14] is a taxonomical system for classifying emotions, and Whissell's [15] [16] Dictionary of Affect in Language (DAL) is drawn from natural language samples. The DAL provides normative emotional ratings on two dimensions (Valence and Pleasantness, Arousal and Activation) for 8742 English words. The two dimensions are those that define Russell's [9] circumplex model of affect.

A third matter for consideration is the variation in the amount and type of affective content in different corpora. Osherenko [17] compared a movie review corpus of 84 human-rated film reviews with the Sensitive Artificial Listener (SAL) corpus of 27 annotated dialogues. Feature extraction was performed on both corpora for affect word features, average value features, and part-of-speech tags. Evaluation was performed using the SMO classifier, which is analogous to SVM. Osherenko noted differences in the classification of affective utterances between the two corpora on recall and precision measures, with superior results for the movie review corpus compared with the SAL corpus. These expected differences reflect the nature of the two corpora.

This type of corpus specificity is problematic. The lexical affinity approach to affect sensing in texts extends the keyword spotting technique by assigning probabilities for words having an affinity with a particular emotion. Because such probabilities are derived from specific corpora, the probabilities are biased towards the features and contexts inherent in that type of corpus [8]. A further source of

potential probability bias is cultural background. For instance, there are nuances of difference even between British and American corpora stemming from cultural particularities; Asian cultures are different again. Furthermore, culture and culture-specific value systems can strongly influence one's emotional experience and emotional expression. Contrasts between Eastern and Western cultural backgrounds, such as an Eastern emphasis on self-restraint, reflects the significant influence of culture on affective response [18].

Hancock, Landrigan and Silver [19] examined how people express and detect emotions during text-based communication, an environment that eliminates the nonverbal cues typically associated with emotion. The results from 40 interactions suggest that users relied on four strategies to express happiness versus sadness, including disagreement, negative affect terms, punctuation, and verbosity. Users readily distinguished between positive and negative valence emotional communicators in this text-based context.

Liu, Lieberman and Selker [20] analysed emotions at the sentence level by using background knowledge. Most interesting are their 'smoothing models'. After sentences have been annotated with one of the six basic emotions or 'neutral,' Liu et al. [20, p.129] applied various techniques aimed at smoothing the transition of emotions from one sentence to the next. These techniques operate on the assumption that the sentences in a story have coherence and, therefore, some affective continuity. Smoothing techniques include (1) decay ('surprised' is followed by two or more neutral sentences), (2) interpolation (a neutral sentence is between two angry sentences) and (3) global mood (larger sections such as a paragraph or even the entire story establish and preserve moods. Liu et al. [20, p. 129] also allow for Meta-Emotions which are composed of the six basic emotions. Examples are (1) Frustration – Repetition of low-magnitude anger (2) Relief – Fear followed by happy (3) Horror – Sudden high-magnitude fear and (4) Contentment – Persistent low-level happy.

In some early work, Gruendel et al. [21] reported limited success with using SVMs for the classification of emotions in documents. Gruendel et al. used ten standard emotional categories such as 'happiness', 'anger' etcetera but, most importantly, they did not distinguish between emotions expressed in a document and emotional responses of the reader. Consequently, Gruendel et al. manually classified their documents (that is, annotation has been performed by the researchers themselves) for machine learning. Their data set is very limited with no examples at all for 'disgust' and 2-21 positive as well as 16-92 negative examples for the other categories. Gruendel et al. reported limited agreements between assessors. Devitt and Ahmad [22] also found low levels of agreement among raters on a seven-point scale of semantic orientation of texts. They were exploring financial news text for positive and negative polarity consistent with human judgments.

Wu, Chuang and Lin [23] used emotion generation rules (EGRs) which were manually deduced to represent the conditions for generating emotion. The emotional state of each sentence can be represented as a sequence of semantic labels (SLs) and attributes (ATTs). SLs are defined as the

domain-independent features, while ATTs are domain-dependent. The emotion association rules (EARs) represented by SLs and ATTs for each emotion are automatically derived from the sentences. Finally, a separable mixture model (SMM) is adopted to estimate the similarity between an input sentence and the EARs of each emotional state. Only three emotional states, *happy, unhappy, and neutral,* are considered.

Ni, Xue, Ling and Yang [24] used machine learning methods for classifying informative and affective articles among weblogs. Ni et al. achieved about 92% on performance measures including precision, recall and F1.

In summary, it is interesting to note that even some of the more recent work is based on standard and very broad emotion categories such as ‘happy’, ‘sad’, ‘anger’ and their composites. This is a limitation that should be targeted in future research.

4. Experiments and Results

To illustrate the problem (and opportunity) faced by rule extraction from SVMs when applied to emotion classification, support vector machines are used to classify business news articles from the Persian Gulf with regard to emotional content. A total of 914 news articles are used for this experiment. In addition, experiments on authorship attribution (identifying the author of a text) and topic classification are performed.

The pre-processing includes two processes: text extraction and feature selection. Text extraction is performed by lexical analysis to strip all non-word annotations and to convert the text into a list of words or tokens. This step can be summarised as follows: (1) upper case letters are converted to lower case, (2) all words containing non-letter characters are removed including hyphenated words and words with an underscore, (3) all punctuations are replaced with space characters to be treated as token delimiters, (4) author identities are extracted, (5) the texts are converted to a ‘bag-of words’ (BOW) representation.

In addition, all class identifiers are removed from the articles in the BOW format before they are used to generate a fixed vocabulary – author names and words used for topic and emotion identification are removed (see the section below on the clustering process used to generate targets for supervised machine learning). After the text extraction process, a fixed length vocabulary is built from the set of all extracted news articles through a feature selection process. Firstly, stopword removal and stemming are performed on each extracted text. Secondly, document frequency thresholding is used to reduce further the feature vector space. Words occurring once only are removed.

After the vocabulary generation process, for each class (four authors, emotions and topics), the extracted texts are (1) labeled with the class and (2) mapped to an SVM data file in which each line represents a news article. Each row includes a label that indicates whether the article belongs to the

target class or not. The vocabulary and the class label of each article are used to map each extracted text in BOW representation to an SVM-feature.

Pre-processing of the text samples, including elimination of frequent words (using an edited list of the 6500 most frequent words in English) led to the development of lists of words that are low frequent, but included some words with topical or emotional content which are common. Simple clustering techniques are used to extract topic and emotion information from texts to perform supervised learning.

The following is a description of the algorithm used to identify topical and emotive information:

For all articles

1. Generate a ranked list of n words that are not in a stoplist (comprising an edited list of the 6500 most spoken words in English).
2. Apply cluster analysis to the ranked word lists extracted from the documents.
3. Identify words that are high-frequent in clusters.

The method in step 2 includes a model-based approach to hierarchical clustering. The method includes a two-step strategy to determine the number of clusters. The model associated with a cluster covers both numerical and categorical attributes and constitutes a mixture of Gaussian and multinomial models. The distance between two clusters is defined as a decrease in log-likelihood caused by merging of the two clusters under consideration. The process continues until a stopping criterion is met. As such, determination of the best number of clusters is automatic.

The emotion categories extracted by this process are ‘boom’, ‘confident’, ‘regret’ and ‘demand’. Pedagogical rule extraction from SVMs as outlined in Barakat and Diederich [25] is applied to the trained SVMs. The procedure for rule extraction is as follows:

1. Divide data into two or more sets
2. Train SVM on a subset of data A
3. Get SVM predictions on subset B
4. Combine inputs from subset B with SVM predictions
5. Train a symbolic machine learning system on the new data set
6. Obtain rules from the symbolic machine learning system (in this case a decision tree learner and a classification/regression tree).

Since the decision tree learning system cannot efficiently deal with high-dimensional input spaces, the first 200 features are used for See5 learning only. The following rules (Table 1) were extracted in one run from an SVM trained on an emotion identification problem. The learning result for this SVM is acceptable (leave-one-out cross-validation result: error: 8.6%, precision: 94.9%, recall: 70%).

Table 1. Rules extracted from an SVM trained on an emotion classification problem.

Rule 1:	suppli > 0 estat > 0.032 -> Target emotion
Rule 2:	close > 0.02 estat > 0.032 -> Target emotion
Rule 3:	estat <= 0.032 -> Other emotion
Rule 4:	estat > 0.032 -> Other emotion
Default class:	Other emotion

5. Discussion

The rule sets obtained from the SVM trained on the emotion classification problem are intriguing: Rule quality is very high, rule extraction is consistent and the learning results are good. However, *none of the antecedents expresses emotion in any way at all! ('suppli' refers to supplies and 'estat' to estate: the method discovered regret in the 'real estate' market due to 'over supplies').* Even though all SVM and decision tree learning results are acceptable (with exception of the 'extreme' k-fold cross-validation), the rules *appear* not to be linked to the task at hand: classify documents into categories that express the emotion of a text. Since it is one of the objectives of rule extraction to explain 'how' classification is realized by an SVM, the question must be asked to what extent the above rule sets help to provide an answer.

It is very clear that in future studies an approach similar to Liu's et al. [20] method which analyses emotions at the sentence or even phrase level needs to be applied. This method allows appropriate handling of negations and, once emotions have been determined at this level, the mood for an entire paragraph or text can be compiled.

Textual communication in computer games is brief, often at the phrase or sentence level. We aim at a method that recognises emotions in game situations and, based on this assessment, reinforces or weakens affect to enhance the experience for the user. In the context of movie reviews, very good results have been achieved due to the obvious emotional reaction reviewers have to many films. It is very well possible that positive results can be achieved when the method is applied to computer games.

Acknowledgement

The rule extraction from SVM experiment has been performed in collaboration with Insu Song, Aqeel Al Ajmi, Jihan Zhu, Imran Fanaswala and Mark Pedersen.

References

[1] Cortes, C.; Vapnik, V.: Support-Vector Networks. Machine Learning, 20(3), 273-297, 1995.
[2] Joachims, T.: Optimizing Search Engines Using Clickthrough Data. ACM Conference on Knowledge Discovery and Data Mining, 133-142, 2002.
[3] Heyer, C.; Diederich, J.: Tibianna: A Learning-Based Search Engine with Query Refinement. Thom, J.; Kay, J. (Eds.): Proceedings of the Australian Document Computing Symposium. Sydney 16 December 2002. 105-108. Sydney: The University of Sydney, 2002.
[4] Heyer, C.; Madden, J., Hollingsworth, K.; Heydon, P.; Bartlett, K.; Diederich, J.: MyNewsWave: User-centered Web search and news delivery. Thom, J.; Kay, J. (Eds.): Proceedings of the Australian Document Computing Symposium. Sydney 16 December 2002. 121-124. Sydney: The University of Sydney, 2002. ISBN 1 86487 525 9.
[5] Moldovan, D.; Mihalcea, R.: Using WordNet and Lexical Operators to Improve Internet Searches. IEEE Internet Computing, 4(1), 34-43, 2000.
[6] Diederich, J.; Kindermann, J.; Leopold, E.; Paass, G.: Authorship Attribution with Support Vector Machines. Applied Intelligence, 19(1) (July/Aug), 2003.
[7] Diederich, J. (Ed.), Rule Extraction from Support Vector Machines. Series: Studies in Computational Intelligence, Vol.80. Berlin, Heidelberg, New York: Springer Verlag, 2008.
[8] Fragopanagos, N.; Taylor, J. G.: Emotion recognition in human-computer interaction. Neural Networks, 18, 389-405, 2005.
[9] Russell, J. A.: A circumplex model of affect. Journal of Personality and Social Psychology, 39(6), 1161-1178, 1980.
[10] Posner, J.; Russell, J. A.; Peterson, B. S.: The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology. Development and Psychopathology, 17, 715-734, 2005.
[11] Ortony, A.; Clore, G. L.; Collins, A.: The Cognitive Structure of Emotions. Cambridge: Cambridge University Press, 1988.
[12] Ortony, A.: On making believable emotional agents believable. In R. Trappl, P. Petta, & S. Payr (Eds.). Emotions in humans and artifacts (pp. 189-211). Cambridge, MA: MIT Press, 2003.
[13] Carroll, J. M.; Yik, M. S. M.; Russell, J. A.; Feldman Barrett, L.: On the psychometric principles of affect. Review of General Psychology, 3(1), 14-22, 1999.
[14] Ortony, A.; Clore, G. L.; Foss, M. A.: The referential structure of the Affective Lexicon. Cognitive Science, 11, 341-364, 1987.
[15] Sweeney, K.; Whissell, C.: A dictionary of affect in language: 1. Establishment and preliminary validation. Perceptual and Motor Skills, 59, 695-698, 1984.
[16] Whissell, C.: A comparison of two lists providing emotional norms for English words (ANEW and the DAL). Psychological Reports, 102, 597-600, 2008.
[17] Osherenko, A.: Affect sensing using lexical means; comparison of a corpus with movie reviews and a corpus with natural language dialogues. Proceedings of the workshop on Emotional Corpora, pp. 64-67, 2006. Accessed February 6th, 2009 from http://www.lrec conf.org/lrec2006/IMG/pdf/programWSemotion-LREC2006-last1.pdf
[18] Markus, H. R.; Kitayama, S. Culture and the self: implications for cognition, emotion, and motivation. Psychological Review, 98(2), 224-253, 1991.
[19] Hancock, J.T.; Landrigan, C.; Silver, C.: Expressing Emotion in Text-based Communication. CHI 2007, 929-932, San Jose, California, USA, 2007.
[20] Liu, H.; Lieberman, H.; Selker, T.: A Model of Textual Affect Sensing using Real-World Knowledge. Proceedings of the 2003 International Conference on Intelligent User Interfaces, 125-132. Miami, Florida, 2003.
[21] Gruendel, H.; Naphtali, T.; Wiech, C.; Gluba, J.M.; Rohdenburg, M.; Scheffer, T.: Clipping and Analyzing News Using Machine Learning Techniques. Proceedings of the International Conferencce on Discovery Science, 2001.
[22] Devitt, A; Ahmad, K.: Sentiment polarity identification in financial news: a cohesion-based approach. Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, pp. 984-991, 2007.
[23] Wu, C.H.; Chuang, Z.J.; Lin, Y.C.: Emotion Recognition from Text Using Semantic Labels and Separable Mixture Models. ACM Transactions on Asian Language Information Processing, 5(2), 165-182, 2006.
[24] Ni, X.; Xue, G.R.; Ling, X. Yu, Y.; Yang, Q.: Exploring in the Weblog Space by Detecting Informative and Affective Articles. WWW 2007 / Track: Industrial Practice and Experience, 281-290. Banff, Alberta, Canada, 2007.
[25] Barakat, N., Diederich, J.: Eclectic rule extraction from support vector machines. International Journal of Computational Intelligence, 2 (1), 59-62, 2005.