2009 International Conference on Information Management and Engineering

Towards Text-based Emotion Detection

A Survey and Possible Improvements

Edward Chao-Chun Kao
Chun-Chieh Liu
Institute of Information Systems and Applications
National Tsing Hua University
Hsinchu, Taiwan
edkao@cs.nthu.edu.tw
tustus28@gmail.com

Ting-Hao Yang
Department of Computer Science
National Tsing Hua University
Hsinchu, Taiwan
hueyletter@gmail.com

Chang-Tai Hsieh
Institute for Information Industry
Taipei, Taiwan
hsct@iii.org.tw

Von-Wun Soo
Department of Computer Science and Information Engineering
National University of Kaohsiung
Kaohsiung, Taiwan
soo@cs.nuk.edu.tw

***Abstract*—This paper presents an overview of the emerging field of emotion detection from text and describes the current generation of detection methods that are usually divided into the following three main categories: keyword-based, learning-based, and hybrid recommendation approaches. Limitations of current detection methods are examined, and possible solutions are suggested to improve emotion detection capabilities in practical systems, which emphasize on human-computer interactions. These solutions include extracting keywords with semantic analysis, and ontology design with emotion theory of appraisal. Furthermore, a case-based reasoning architecture is proposed to combine these solutions.**

Keywords-emotion detection; ontology; case-based reasoning

I. INTRODUCTION

Emotion detection is developed to provide cues for further human-computer interactions, so computers may stand as social actors to achieve more believable interactions [7][18]. Besides emotion detection from texts, much work has been done to detect users' emotion states from multimodal sources such as audio, gestures, and eye gazes over the last decade [6][8][10][15][22].

While multimodal interactions with computers have shown to be appealing, the most common way for people to interact with computer systems is still via texts. As web 2.0 emerges, more and more people have blogs to share their feelings with unspecified public. Though sharing pictures and videos has becoming popular, texts and blog articles still stand as an important role for expressing emotions. Compared to research of emotion detection in multimodal fields, emotion detection from text is still not mature and requires more improvements to be assembled as practical applications.

These improvements include better understanding about newly evolved vocabularies, incorporation of psychological theories to infer emotion behind texts, utilization of contextual knowledge, developing more advanced emotion detection methods that allow more categories of emotions and inference.

The research in this area gradually draws attentions because it provides a rich possibility in both theoretical and practical domains. It would be interesting to validate psychological theories via objective computer programs, and applications such as artificial pets may participate in blogs with the input of masters' emotion states from texts.

In this paper, we propose various directions to improve the capabilities of current systems of text-based emotion detection. However, firstly we present a survey of current methods, identifying their limitations and some initial approaches to improve their capabilities in Section II. Then we propose system architecture to further integrate these improvements in Section III. We conclude in Section IV.

II. THE SURVEY OF EMOTION DETECTION FROM TEXTS

Definitions about emotion, its categories, and their influences have been an important research issue long before computers emerged, so that the emotional state of a person may be inferred under different situations. Since Picard proposed the concept of affective computing in 1997 [17], the role of emotions in human-computer interactions has been gradually established [14], and this domain soon attracts interdisciplinary researchers from computer science, biology, psychology, cognitive science and so on. Following the trend, the computational research of emotion detection from texts emerged to determine human emotions from another point of view.

In its most common formulation, the emotion detection from text problem is reduced to finding the relations between specific input texts and the actual emotions that drives the author to type/write in such styles. Intuitively, finding the relations usually relies on specific surface texts that are included in the input texts, and other deeper inferences that will be formally discussed below. Once the relations can be determined, they can be generalized to predict others' emotions from their articles, or even single sentences.

At the first glance, it does not seem to involve so many difficulties. In real life, different people tend to use similar phrases (i.e. "Oh yes!") to express similar feelings (i.e. joy) under similar circumstances (i.e. achieving a goal); even they

978-0-7695-3595-1/09 $25.00 © 2009 IEEE
DOI 10.1109/ICIME.2009.113

native languages are different, the mapping of such phrases from each language may be obvious. More formally, the emotion detection from text problem can be formulated as follows: Let E be the set of all emotions, A be the set of all authors, and let T be the set of all possible representations of emotion-expressing texts. Let r be a function to reflect emotion e of author a from text t, i.e., r: $A \times T \rightarrow E$, and the function r would be the answer to our problem.

The central problem of emotion detection systems lies in that, though the definitions of E and T may be straightforward from the macroscopic view, the definitions of individual element, even subsets in both sets of E and T would be rather confusing. On one hand, for the set T, new elements may add in as the languages are constantly evolving. On the other hand, currently there are no standard classifications of "all human emotions" due to the complex nature of human minds, and any emotion classifications can only be seen as "labels" annotated afterwards for different purposes.

As a result, before seeking the relation function r, all related research firstly define the classification system of emotion classifications, defining the number of emotions. Secondly, after finding the relation function r or equivalent mechanisms, they still need to be revised over time to adopt changes in the set T. In the following subsections, we will present a classification of emotion detection methods proposed in the literature, based on how detection are made. Although they can all be classified into content-based approaches from the point of view of information retrieval, their problem formulation differs from each other:

1. Keyword-based detection: Emotions are detected based on the related set(s) of keywords found in the input text;
2. Learning-based detection: Emotions are detected based on previous training result with respect to specific statistic learning methods ;
3. Hybrid detection: Emotions are detected based on the combination of detected keyword, learned patterns, and other supplementary information;

Besides these emotion detection methods that infer emotions at sentence level, there has been work done also on detection from paragraphs or articles [1][9]. For example, though each sentence in a blog article may indicate different emotions, the article as a whole may tend to indicate specific ones, as the overall syntactic and semantic data could strengthen particular emotion(s). However, this paper focuses on detection methods with respect to single sentences, because this is the foundation of full text detection.

A. Keyword-based Methods

Keyword-based methods are the most intuitive ways to detect textual emotions. To approximate the set T, since all the names of emotions (emotion labels) are also meaningful texts, these names themselves may serve as elements in both sets of E and T. Similarly, those words with the same meanings of the emotion labels can also indicate the same emotions. The keywords of emotion labels constitute the subset EL in set T, where EL also classifies all the elements in E. The set EL is constructed and utilized based on the assumption of keyword independence, and basically ignores the possibilities of using different types of keywords simultaneously to express complicated emotions.

Keyword-based emotion detection serves as the starting point of textual emotion recognition. Once the set EL of emotion labels (and related words) is constructed, it can be used exhaustively to examine if a sentence contains any emotions.

However, while detecting emotions based on related keywords is very straightforward and easy to use, the key to increase accuracy falls to two of the preprocessing methods, which are sentence parsing to extract keywords, and the construction of emotional keyword dictionary. Parsers utilized in emotion detection are almost ready-made software packages, whereas their corresponding theories may differ from dependency grammar to theta role assignments. On the other hand, constructing emotional keyword dictionary would be naval to other fields [20]. As this dictionary collects not only the keywords, but also the relations among them, this dictionary usually exists in the form of thesaurus, or even ontology, to contain relations more than similar and opposite ones. Semi-automatic construction of EL based on WordNet-like dictionaries is proposed in [2] and [24].

As was observed in [21], keyword-based emotion detection methods have three limitations described below.

1) Ambiguity in Keyword Definitions

Though using emotion keywords is a straightforward way to detect associated emotions, the meanings of keywords could be multiple and vague. Except those words standing for emotion labels themselves, most words could change their meanings according to different usages and contexts, and it is just not feasible to include all possible combinations into the set EL. Moreover, even the minimum set of emotion labels (without all their synonyms) could have different emotions in some extreme cases such as ironic or cynical sentences.

2) Incapability of Recognizing Sentences without Keywords

Keyword-based approach is totally based on the set of emotion keywords. Therefore, sentences without any keywords would imply they do not contain any emotions at all, which is obviously wrong. For example, "I passed my qualify exam today" and "Hooray! I passed my qualify exam today" should imply the same emotion (joy), but the former without "hooray" could remain undetected if "hooray" is the only keyword to detect this emotion.

3) Lack of Linguistic Information

Syntax structures and semantics also have influences on expressed emotions. For example, "I laughed at him" and "He laughed at me" would suggest different emotions from the first person's perspective. As a result, ignoring linguistic information also poses a problem to keyword-based methods.

In summary, keyword-based methods should also detect not only the existence of keywords, but also their linguistic information to detect emotions more accurately.

B. Learning-based Methods

Researchers using learning-based methods attempt to formulate the problem differently. The original problem that determining emotions from input texts has become how to classify the input texts into different emotions. Unlike keyword-based detection methods, learning-based methods try to detect emotions based on a previously trained classifier, which apply various theories of machine learning such as support vector machines [26] and conditional random fields [23], to determine which emotion category should the input text belongs.

However, comparing the satisfactory results in multi-modal emotion detection [13], the results of detection from texts drop considerably. The reasons are addressed below:

1) Difficulties in Determining Emotion Indicators

The first problem is, though learning-based methods can automatically determine the probabilities between features and emotions, learning-based methods still need keywords, but just in the form of features. The most intuitive features may be emoticons [], which can be seen as author's emotion annotations in the texts. The cascading problems would be the same as those in keyword-based methods.

2) Over-simplified Emotion Categories

Nevertheless, lacking of efficient features other than emotion keywords, most learning-based methods can only classify sentences into two categories, which are positive and negative. Although the number of emotion labels depends on the emotion model applied, we would expect to refine more categories in practical systems.

C. Hybrid Methods

Since keyword-based methods with thesaurus and naïve learning-based methods could not acquire satisfactory results, some systems use a hybrid approach by combining both or adding different components, which help to improve accuracy and refine the categories. The most significant hybrid system so far is the work of Wu, Chuang and Lin [21], which utilizes a rule-based approach to extract semantics related to specific emotions, and Chinese lexicon ontology to extract attributes. These semantics and attributes are then associated with emotions in the form of emotion association rules. As a result, these emotion association rules, replacing original emotion keywords, serve as the training features of their learning module based on separable mixture models. Their method outperforms previous approaches, but categories of emotions are still limited.

D. Summary and Conclusions

As described in this section, much research has been done over the past several years, utilizing linguistics, machine learning, information retrieval, and other theories to detect emotions. Their experiments show that, computers can distinguish emotions from texts like humans, although in a coarse way. However, all methods have certain limitations, as described in the previous subsections, and they lack context analysis to refine emotion categories with existing emotion models, where much work has been done to put them computationalized in the domain of believable agents. On the other hand, applications of affective computing would expect more refined results of emotion detection to further interact with users. Therefore, developing a more advanced architecture based on integrating current approaches and psychological theories would be in a pressing need.

III. PROPOSING A NEW SYSTEM ARCHITECTURE

Emotion detections from texts, as described in Section II, can be modified and integrated to further extend their capabilities. Our version of integrated architecture includes extracting semantic information with semantic analysis, designing ontology based on emotion models, and adopting new keywords with case-based reasoning. This architecture is proposed to aim at providing systematic understanding of textual input and better flexibility to different domains. In the remainder of this section, we describe each component of the proposed architecture in Fig. 1, and related research for developing them as well.

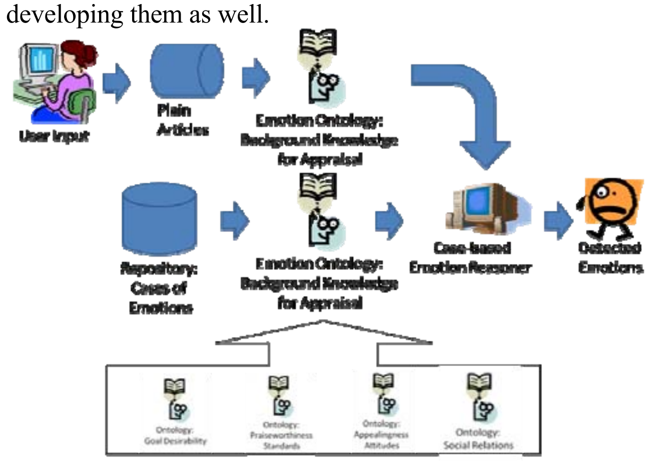

Figure 1. The Proposed System Architecture

A. Semantic Analysis

Semantic analysis utilizes techniques of natural language processing such as statistic-based parsing to retrieve more information about sentences, such as dependency trees [5][19][27]. While current standard of keyword and attribute extractions is rule-based and ad-hoc, semantic analysis is proposed to extract keywords based on the linguistic information of sentences [3][4][27].

While statistic-based natural language parsers are many, the choice of parsers should be based on whether one could extract the desired information, which includes persons, actions, time, place, and objective. A more realistic example would be given in the next subsection. Performing semantic analysis can not only extract the keywords, but also their tenses, relations with the subjective/objective, negations, and so on. Refer [25] for more examples.

B. Emotion Models

Emotion models determine the background knowledge we need to associate the results of semantic analysis to possible emotions. Although only the final classification of emotion categories are adopted in related work of emotion detection from texts, we argue that not only the results are important, but also how they are derived, and both of them should be taken into account when constructing a system of emotion detection from text. Moreover, some emotion models already predefine needed knowledge to appraise events.

Among all emotion models, the OCC model [16], which includes 22 emotion categories, also explains their derivation methods and supporting domain knowledge. For example, joy is defined as "one achieves his/her goals", and its negation is distressed, where one does not achieve his/her goals. In this example, the supporting domain knowledge is required to identify keywords or phrases that could become "goals". Other required domain knowledge in the OCC model are "moral standards" (defining righteous / sinful actions), "appealing attitudes" (defining whether one likes an object, including other people), and "social relations" (defining positive / negative relations).

Incorporating emotion models into semantic analysis would yield a more systematic way to analyze different textual inputs, whereas emotion detection systems can be used to validate the correctness of these models. While the needed emotion categories in practical systems may differ from those in emotion models, a mapping table between two emotion classification systems should be sufficient to solve this issue.

C. *Case-Based Reasoning*

Finally, case-based reasoning is proposed to adopt new words and usages due to the rapid evolving nature of languages, especially on the Internet. In a case-based reasoning system, the input will be compared with every case stored in the case repository, and the distances between the input and every case of repository will be calculated with algorithms to estimate the degree of similarity. Therefore, the results of previous two components can be stored as different cases in case-based reasoning systems. When a new keyword or usage is born, a new case consisting of this new word should be sufficient to cover it, without the need to retrain the whole classifier. Furthermore, an input case can have multiple emotions in case-based reasoning system, since we can list the emotions in the most similar cases, according to the similarity. This also covers the problem of multiple emotions [12].

However, the design of estimation algorithm, its input data format, and the quality of cases in the repository (which is annotated manually beforehand) will be the factors to affect the accuracy.

IV. CONCLUSIONS

Emotion detection is an important research field in affective computing over the last decade as much research has been done to detect facial, audio, and gestural emotions. The detected emotions would stand as important clues for advanced human-computer interactions. On the other hand, emotion detection from texts draws less attention.

However, even in the era of web 2.0, text-based input is still the most common way for humans to interact with computers, and thus emotion detection from texts should be refocused as an important research issue in affective computing.

In this paper, we surveyed existing research of emotion detection and reviewed the limitations to improve detection capabilities, describing a proposal of integrated system architecture. These improvements include identification of newly-evolved vocabularies, systematic emotion ontology based on OCC model as background knowledge, and collaborative method to detect multiple emotions in the form of case-based reasoning.

We hope that researchers in related fields may find the survey presented in this paper useful and advance the discussion in emotion detection from texts. As the future work, we will implement this system based on actual blog corpus to validate our claims.

ACKNOWLEDGMENT

This study is conducted under the "III Innovative and Prospective Technologies Project" of the Institute for Information Industry which is subsidized by the Ministry of Economy Affairs of the Republic of China (Taiwan).

REFERENCES

[1] J. C. Ang, "Prosodic cues for emotion recognition in communicator dialogs," M.S. thesis, University of California Berkeley, San Francsico, 2002.

[2] D. B. Bracewell, "Semi-Automatic Creation of an Emotion Dictionary Using WordNet and its Evaluation," Proc. IEEE conference on Cybernetics and Intelligent Systems, IEEE Press, Sep. 2008, pp. 21-24, doi: 10.1109/ICCIS.2008.4670735.

[3] A. C. Boucouvalas, "Real time text-to-emotion engine for expressive internet communications," in Emerging Communication: Studies on New Technologies and Practices in Communication, G. Riva et al, Eds. IOS Press, pp. 305-318.

[4] A. C. Boucouvalas and X. Zhe, "Text-to-emotion engine for real time internet communication," Proc. International Symposium on CSNDSP 2002, Jul. 2002, pp. 164-168.

[5] Z.-J. Chuang and C.-H. Wu, "Emotion recognition from textual input using an emotional semantic network," Proc. International Symp. Spoken Language Processing (ICSLP 02), 2002, pp. 177-180.

[6] J. F. Cohn and G. S. Katz, "Bimodal expression of emotion by face and voice," Proc. 6th ACM international conference on Multimedia: Face/gesture recognition and their applications, ACM, 1998, pp. 41-44, doi:10.1145/306668.306683.

[7] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J. Taylor, "Emotion recognition in human-computer interaction," IEEE Signal Processing Magazine, vol. 18(1), Jan. 2001, pp. 32-80, doi: 10.1109/79.911197.

[8] L. C. De Silva and P. C. Ng, "Bimodal emotion recognition," Proc. 4th IEEE International Conference on Automatic Face and Gesture Recognition, IEEE Computer Society, 2000, pp. 332-335, doi: 10.1109/AFGR.2000.840655.

[9] L. Devillers, I. Vasilescu, and L. Lamel, "Annotation and detection of emotion in a task-oriented human-human dialog corpus," Proc. ISLE Workshop on Dialogue Tagging for Multi-Modal Human-Computer Interaction, Dec. 2002.

[10] L. Devillers, L. Lamel, and I. Vasilescu, "Emotion detection in task-oriented spoken dialogues," Proc. 2003 international Conference on Multimedia and Expo - Volume 3 (ICME 03), IEEE Computer Society, Jul. 2003, pp. 549-552.

[11] K. Forbes-Riley and D. J. Litman, "Predicting emotion in spoken dialogue from multiple knowledge sources," Proc. Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL 04), 2004, pp. 201-208.

[12] J. T. Larsen, A. P. McGraw, J. T. Cacioppo, "Can People Feel Happy and Sad at the Same Time," Journal of Personality and Social Psychology, Vol. 81, 2001, pp. 684-696.

[13] C. M. Lee, S. S. Narayanan, and R. Pieraccini, "Combining Acoustic and Language Information for Emotion Recognition," Proc. 7th International Conference on Spoken Language Processing (ICSLP 02), 2002, pp.873-876.

[14] C. Maaoui, A. Pruski, and F. Abdat, "Emotion recognition for human-machine communication," Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 08), IEEE Computer Society, Sep. 2008, pp. 1210-1215, doi: 10.1109/IROS.2008.4650870.

[15] T. Nguyen, and I. Bass, "Investigation of Combining SVM and Decision Tree for Emotion Classification," Proc. 7th IEEE international Symposium on Multimedia (ISM 05), IEEE Computer Society, Dec. 2005, pp. 540-544, doi:10.1109/ISM.2005.72.

[16] A. Ortony, G. Clore, and A. Collins, The Cognitive Structure of Emotions, Cambridge Press, 1988.

[17] R. W. Picard, Affective Computing. Cambridge, MA: MIT Press, 1997.

[18] R. W. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: analysis of affective physiological state", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23(10), Oct. 2001, pp. 1175-1191, doi: 10.1109/34.954607.

[19] P. Subasic and A. Huettner, "Affect Analysis of Text Using Fussy Semantic Typing," IEEE Transactions on Fussy System, vol. 9(4), 2001, pp. 483-496.

[20] T. Wilson, J. Wiebe, and R. Hwa, "Just how mad are you? Finding strong and weak opinion clauses," Proc. 21st Conference of the American Association for Artificial Intelligence (AAAI 04), AAAI Press, Jul. 2007, pp. 761-769, doi:10.1.1.5.2078.

[21] C.-H. Wu, Z.-J. Chuang, and Y.-C. Lin, "Emotion Recognition from Text Using Semantic Labels and Separable Mixture Models," ACM Transactions on Asian Language Information Processing (TALIP), vol. 5, issue 2, Jun. 2006, pp. 165-183, doi:10.1145/1165255.1165259.

[22] T. Yanaru, "An emotion processing system based on fuzzy inference and subjective observations," Proc. 2nd New Zealand Two-Stream international Conference on Artificial Neural Networks and Expert Systems, IEEE Computer Society, Nov. 1995, pp. 15-20, doi: 0.1109/ANNES.1995.499429.

[23] C. Yang, K. H.-Y. Lin, and H.-H. Chen, "Emotion classification using web blog corpora," Proc. IEEE/WIC/ACM International Conference on Web Intelligence. IEEE Computer Society, Nov. 2007, pp. 275-278, doi: 10.1109/WI.2007.50

[24] J. Yang, D. B. Bracewell, F. Ren, and S. Kuroiwa, "The Creation of a Chinese Emotion Ontology Based on HowNet", Engineering Letters, vol. 16(1), Feb. 2008, pp. 166-171.

[25] J. Yang, D. B. Bracewell, F. Ren, and S. Kuroiwa, "A semantic analyzer for aiding emotion recognition in Chinese," in Lecture Notes of Artificial Intelligence 4114, D.-S. Huang, K. Li, and G. W. Irwin, Eds. Springer, Berlin Heidelberg, 2006, pp. 893-901, doi: 10.1007/11816171_112.

[26] Z. Teng, F. Ren, and S. Kuroiwa, "Recognition of Emotion with SVMs," in Lecture Notes of Artificial Intelligence 4114, D.-S. Huang, K. Li, and G. W. Irwin, Eds. Springer, Berlin Heidelberg, 2006, pp. 701-710, doi: 10.1007/11816171_87.

[27] J.-M. You, and K.-J. Chen, "Automatic Semantic Role Assignment for a Tree Structure," Proc. 3rd SIGHAN Workshop on Chinese Language Processing, 2004, pp. 109-115.