Detecting Emotion from Dialogs and Creating Personal Ambient in a Context Aware System

Lun-Wei Ku[1] and Cheng-Wei Sun[2]

[1] Institute of Information Science, Academia Sinica, Taipei, Taiwan
lwku@iis.sinica.edu.tw
[2] Department of Information Science and Engineering, National Yunlin University of Science and Technology, Yunlin, Taiwan
chengwei.kenny.sun@gmail.com

Abstract. This paper presents a personal ambient creation systme, IlluMe, which detects users' emotion from their chatting context in instant messages and then analyze them to recommend suitable lighting and music to create a personal ambient. The system includes a mechanism for recording users' feedback of the provided ambient to learn their preference. The aim of the proposed system is to link human language and emotion with the computer created environment seemlessly. To achieve this, we propose four apporaches to calculate emotion scores of words: Topical Approach, Emotional Approach, Retrieval Approach and Lexicon Approach. Natural language processing techniques such as normalization, part of speech tagging, word bigram utilization, and sentiment dictionaries lookup are incorporated to enhance system performance. Experiments results are shown and discussed, from which we find the system satisfactory and several future research directions are inspired.

Keywords: emotion detection, blog articles, instant messages, ambient creation, context aware system.

1 Introduction

Language is one of the major tools used by users to interact with computer interfaces, which includes texts, speech, facial expressions and body languages. To provide a satisfactory usage experience, context aware systems have tried to detect users' emotion when receiving commands [3], [8] and considered it in the proceeding process. To find the users' emotion, real time information like the facial expression or the speech utterance was gathered [2]. However, additional cameras and microphones became necessary. Some researchers used sensors to watch the heart beat and the body temperature of residents to know their current emotion for further respondence, but then users had to wear sensors and it was inconvenient. Instead of watching body signals, we postulate that the communication among people is one of the important factors to influence their emotions and from the content we can find hints over a certain period of time. Therefore, we hope to watch users' conversations and then detect their emotion.

N. Streitz and C. Stephanidis (Eds.): DAPI/HCII 2013, LNCS 8028, pp. 128–137, 2013.

In the natural language processing research community, emotion analysis has drawn a lot attention and the development of fundamental approaches as well as applications has been proposed [4], [9], [12]. In this research, clues from users' textual conversations were mined by these approaches to detect their psychological emotion state. Then the ambient of their personal working or living space will be changed as a feedback of the designed system. As a start, music and lightings were utilized. The design of the proposed system could be easily integrated into or applied to personal emotion management, self-care, wellness management, or any other human-machine interfaces which intend to consider users' emotion.

There are many ways to categorize emotions. Different emotion states were used for experiments in previous research [1]. To find suitable categories of emotions, we adopted the three-layered emotion hierarchy proposed by Parrott shown in Table 1 [7]. Six emotions are in the first layer, including love, joy, surprise, anger, sadness and fear. The second layer includes 25 emotions, and the third layer includes 135 emotions. Using this hierarchical classification enables the system the ability to categorize emotions from rough to fine granularities and degrade to the upper level when the experimental materials are insufficient. In addition, mapping categories in other researches to ours becomes easier with the hierarchy, and more information is provided to annotators when marking their current emotion.

We hope to find emotion from authors' aspect instead of readers' aspect from texts to fulfill our purpose. In this research, users' conversations were collected from the log of instant message software Yahoo! Messager. To automatically learn the accompanied emotion from a large dataset, texts containing emoticons in Yahoo! blog articles were utilized. Then statistical approaches were adopted and compared. As to relations of emotions and the music, most researchers looked for the emotions in songs or rhythms [10-11]. They classified music into different emotional categories and developed the system to tell what emotion a song might bring to a listener, which was from readers' aspect. However, if the aim is to create a comfortable ambient, what songs a person in a certain emotional state wants to listen to becomes the question. A happy user does not always enjoy happy songs, and vice versa, which makes the technology developed in the previous work not applicable. Learning and adapting to the personal preference become the aim in this research. This is also true for the lightings. To further help the system perform better, collecting users' feedback was realized by using smart phones as the personal controller.

Table 1. Emotion Categories (Tertiary Emotion Not Listed)

Primary Emotion	Secondary Emotion
Love	Affection, Lust, Longing
Joy	Cheerfulness, Zest, Contentment, Pride, Optimism, Enthrallment, Relief
Surprise	Surprise
Anger	Irritation, Exasperation, Rage, Disgust, Envy, Torment
Sadness	Suffering, Sadness, Disappointment, Shame, Neglect, Sympathy
Fear	Horror, Nervousness

In this paper, we described the interfaces and functions of the context aware system IlluMe but focused more on its core technology, the emotional analysis. We hope the proposed techniques can help to develop future systems considering users' emotions.

2 System Description

The potential working area for IlluMe is home or a small space. The system was designed to fit in with the modern people's life style: programs are installed in users' personal computer and smart phone. The smart phone functions as the remote control and the music player, while all setting signals are sent out from the personal computer. The smart phone and the personal computer communicate through the wireless network. The only additional hardware requirement is the lighting set. Now many smart phones are functioned with instant messages. In that case, the personal computer is not necessary and the system is obviously more convenient.

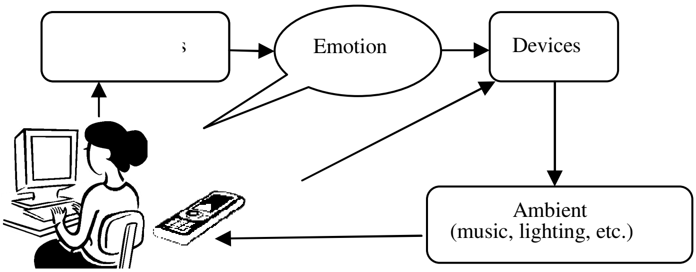

Fig. 1. System Illustration of IlluMe

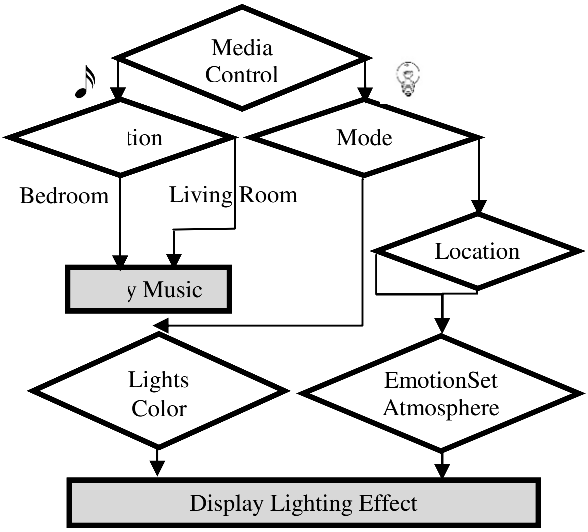

Fig. 2. Operation Flowchart

Figure 1 demonstrates the proposed system IlluMe designed for a small space personal environment. Detailed system framework can be found in Ku's research [6]. We expect that this system could interactively respond to users' personal status by providing a feeling of the companion. We view the IlluMe system as a realization of detecting emotions from users' textual conversations and then responding the best ambient accordingly. System features include an *Emotion Detection Switch* which detects users' current emotion according to messenger logs once a preset time period; an *Auto Ambient Setting* which sets the current ambient by a specific combination of a song and a light group which corresponds to the emotion or represents a special atmosphere; a *Manual Ambient Adjustment* which provides a user interface to change the settings of music and lightings from smart phones; a *Personal Preference Learning Mechanism* which records the new settings, learns the preference and then performs the user adaptation; the *Unlimited Melodies and Rich Light Colors* where songs are added by users and 65,536 lighting colors are provided; *Instant State Update* which watches the users' input from messenger when the software is on and changes the music and lighting according to the detected emotion to make users feel like the environment is interacting with them. Figure 2 shows the operational flow of the user interface.

We adopted the concept of collaborative filtering to design the function of personal ambient learning. In the early stage of using IlluMe, it proposes the most frequently selected settings, that is, the choice of a group of people in the specific emotional state. If the user is connected to the Internet, the user experience will be transferred back to the servers to help suggest a better ambient to other users.

The user experience can be optimized because of design of using smart phones. As the users update the settings, the system knows their preference. In the later stage, the learning function is able to consider the preference of both the individual and the group to create a unique ambient for each user.

3 Emotion Analysis

The emotion analysis that IlluMe performed is to find the emotions that texts in messenger logs bear in order to create a comfort ambient by sound and lighting accordingly. To achieve this, the system needs to understand the Internet language first, and then detect emotions and categorize them. The system works on the Chinese chatting environment and analyzes Chinese texts to detect emotions. Two dictionaries, the Chinese sentiment dictionary NTUSD [14] and the Chinese emotion dictionary [13], were adopted for detecting emotions. The former categorized sentiment words into positive and negative, while the latter into eight emotion types: awesome, heartwarming, surprising, sad, useful, happy, boring, and angry. Notice that these eight emotion types appeared in Yahoo! News Taiwan in the year 2008 and not all of them were general emotion states. Therefore, we tried to align Lin's emotion categories with those in Parrott's emotion hierarchy before using his dictionary.

Messenger logs were used as the source to detect emotions. We collected texts from Yahoo! Messenger and MSN Messenger logs of 8 annotators. When the installed collecting program in their computers was on, it ran as a service and continuously logged their messages. Whenever there was at least one new message, once an hour the collecting program would pop up the menu and ask them to annotate the current emotion together with the preferred settings of the music and lighting. There were 3,290 songs, 15 emotional lighting colors and 6 atmospheres for selection. A total of 150 records are annotated for experiments.

Before the sentiment analysis, some preprocessing steps were performed. In addition to the segmentation and part of speech tagging [16], which are the common preprocessing steps when utilizing Chinese texts, messenger logs and sentiment dictionaries were first transformed into zhuyin [15], a Chinese phonetic symbol set, before looking for emotions to avoid the mismatch caused by this popular type of creative use of writing systems.

3.1 Learning Emotional Scores of Words

The emoticon sentences, i.e., there is at least one emoticon in these sentences, were treated as the learning materials and from them the emotional score of each word was calculated. The learned emotional scores of the words in the messenger log were accumulated to determine the emotion class of the log. Four approaches were proposed: *Topical Approach*, *Emotional Approach*, *Retrieval Approach* and *Lexicon Approach* [5]. *Topical Approach* utilized the concept of *tf* • *idf* score (term frequency multiplied by inversed document frequency) and distributed it to 40 emoticon classes by the probability of observing the emoticon sentences in each emoticon class over all emoticon sentences; In *Emotional Approach* the emoticon sentences of the same emoticon class were concatenated into one document and scores of words were calculated as in *Topical Approach* (the mapping is shown in Table 2); *Retrieval Approach* took the current sentence for judgment as a query and found the most similar 10 sentences (P@10, precision at ten) to determine the emotion it bore; *Lexicon Approach* looked up words from the Chinese emotion dictionary [13] and calculates their emotional scores [14]. *Topical Approach* and *Retrieval Approach* consider the importance of words in the query sentence and the emoticon sentences, while *Emotional Approach* calculates the "emoticonal" tendency of the words in the query sentence.

Table 2. The Mapping of the Emotion Class and Emoticon Classes

Emotion	Emoticon
Love	7(love), 8(shy), 10(kiss)
Joy	1(smile), 4(happy), 13(smug), 18(laugh)
Surprise	11(surprise)
Angry	12(angry)
Sadness	17(cry), 37(sign)
Fear	15(worried)

3.2 Experimental Results and Discussions

To evaluate the performance of the emotion detection in messenger logs, 10-fold experiments were performed. The results of four approaches for emotion detection were listed in Table 3. The best result of emotion detection among four approaches was generated by *Topical Approach*, while *Emotional Approach* performed the worst. After looking over the emotional scores, we found that the unsatisfactory performance of *Emotional Approach* was caused by the concatenation of the emoticon sentences of the same class. This process made forty very large documents so that term frequency became the dominate factor and deteriorated the performance.

Retrieval Approach was better than *Emotional Approach* but worse than *Topical Approach*. Instead of distributed the *tf • idf* score to 40 emotion classes like *Topical Approach*, *Retrieval Approach* utilized it to rank sentences for voting on the emotion class. As a result, we can say that considering the composite important words to find the emotion class performs better than letting similar sentences to determine.

Lexicon Approach was different from the other three in that it did not calculate scores based on emoticon sentences. Its performance was the second among all. The advantage of using lexicons was that we could find words not appearing in the emoticon sentences and hence would still be able to know the emoticon class of sentences, even though there were no previously seen words in them. However, having fixed lexicon set was also its disadvantage. When there were many emoticon sentences so that scores of various words were learned in *Topical Approach*, *Lexicon Approach* suffered from the limited lexicons.

Table 3 shows that all approaches tended to perform unsatisfactory for emotion class Love, Angry and Fear. For Angry and Fear, the insufficiency of emoticon sentences was one causing factor of the low performance. Moreover, these two emotion classes were seldom selected by annotators. Logs of these classes might be related to specific events represented by special word compositions instead of a certain subjective words.

Generally for all approaches, we found several causes of mis-categorization. First, some infrequent words (i.e. less than 10 times) had unreasonably high scores. Second, the polarity of a sequence of words might be different from that of its composite words. Third, positive words in negative sentences (and vice versa) caused noise when determining the polarity. We hence proposed corresponding enhancement approaches in the next section.

Table 3. Performance of Emotion Detection in Message Logs

Approach→ Metric↓	Topical	Emotional	Retrieval	Lexicon
Love	0.000	0.000	0.000	0.000
Joy	0.850	0.238	0.438	0.325
Surprise	0.000	0.000	0.000	1.000
Angry	0.000	0.000	0.000	0.000
Sadness	0.103	0.000	0.103	0.026
Fear	0.000	0.000	0.000	0.000
Macro-Avg	0.159	0.040	0.090	0.225
Micro-Avg	0.480	0.127	0.260	0.187

3.3 Performance Enhancement Approaches

According to the observations mentioned in section 3.2, we tried several approaches to improve the performance. Three of them made some progress and are listed below.

Accumulated Probabilistic (*AcuProb*). As mentioned, *Topical Approach* considers the inversed document frequency (*idf*) in emoticon classes to give the same word different score in each class, but the numbers of documents in these classes vary a lot. Emoticon class *laugh* () contains 131,148 emoticon sentences, which is the largest quantity, while the smallest class *secret* () contains only 5,855 sentences. This causes infrequent words to obtain unreasonable high scores. Therefore, we removed the *idf* part in the calculation process but added a normalization factor to it to make the new score more like a normalized probabilistic one. Scores of words were summed up (accumulated) to generate the final score for the sentence for judgment.

Word Bigram with Specific Parts of Speech (*BiPos*, *NBiPos*). From the previous experiments, we found that very often only a set of words, a two-word set (bigram) is the most commonly seen, can express a complete emotional concept. Therefore, we extracted four word bigram classes which bear emotion most often according to their parts of speech and treated them as one word when calculating emotional scores (*Bi-Pos*). These word bigrams are

- *Vt + Vi.* A transitive verb is followed by an intransitive verb, such as 喜歡 (like to) 飛行 (fly) and 不利 (harm) 辦事 (work);
- *ADV + V.* An adverb is followed by a verb, such as 很 (very) 偷懶 (be lazy) and 更 (more) 用力 (use power);
- *N + Vi.* A noun is followed by an intransitive verb, such as 壽命 (lifetime) 增加 (increase);
- *Vt + N.* A transitive verb is followed by a noun, such as 打擊 (fight) 犯罪 (crime).

An extension which incorporatee the negation, such as no, not, but, cannot, etc., with the word bigram was also tested (*NBiPos*).

Noise Elimination by Sentiment (*EliNoise*). As we have sentiment and emotion dictionaries, their entries were utilized to eliminate noise. Forty emoticon classes were put on the arousal-valence sentiment plane [13] according to the emotions they bear as shown in Figure 3. Words whose polarities were different from that of the classes were treated as noise. Therefore, those words found in positive dictionaries were eliminated in the sentences of negative emoticon classes before calculating scores, and vice versa.

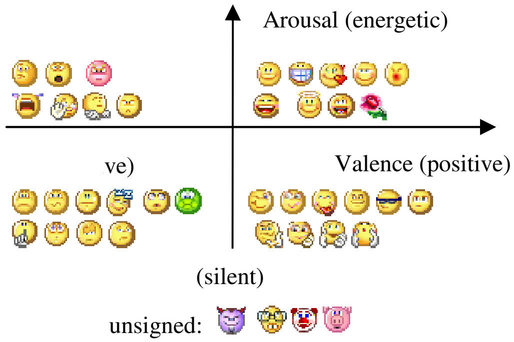

Fig. 3. Emoticons on the Sentiment Plane

To treat each emoticon class as of equal importance, we evaluated the performance by the Macro Average metric here. We compared the performance of the enhancement approaches on *Topical Approach*. Experiment results are shown in Table 4.

Table 4. Performance of enhancement approaches

Approach	Topical	AcuProb	BiPos	NBiPos	EliNoise
Macro-Avg	0.159	0.136	0.158	0.164	0.159
t-test (α=0.05;*dof*=39)	N/A	passed	Passed	passed	Failed

From *Macro Average*, we found that the enhancement was not obvious as we had a total of 1,540,163 emoticon sentences. Among all improvement approaches, only the *Macro Average* of *NBiPos* was increased. We further performed a t-test on the "emoticonal analysis" over 40 emoticon classes. The t-test showed that when we compared the performance before the enhancement (*Topical Approach*) with that after the enhancement, *AcuProb*, *BiPos*, and *NBiPos* all significantly improved the system. After taking a detail look at the *EliNoise* approach, we found that though it did not improve the performance as a whole, it did significantly improve the performance for 28 categories (and decreased the others). In other words, sentences of some emoticon classes are more controversial than others so that there is more noise in these sentences. Therefore, we can say that selecting enhancement approaches depends on the emotions we want to deal with for that *AcuProb* and *BiPos* improve the performance of most categories. Moreover, *NBiPos* improves the performance from both macro and micro perspectives.

Bellegarda reported that his best f-measure was 0.340 also for 6 categories. Notice that his work analyzed from the reader's perspective, while our work analyzed from the author's perspective. The emotion analysis from author's perspective was generally considered more difficult than from the reader's perspective as what a user felt might not be consistent with what he/she wrote in instant messages. Therefore, though Bellegarda's experiments and experiments in this paper were done on different datasets and evaluated by different metrics, we believe the performance reported by this paper was comparable.

4 Conclusion and Future Work

Through the work we aim to apply the language technology to redefine the concept of a personal space. Via the proposed interface, the space is enabled the capability to observe human emotion, and create corresponding consoling ambient according to the residents' different status. The emotion analysis technique equips the space with the interaction ability with the residents. The instant interior lightings and music change expressed with constructed ambient and residents give feedbacks, which complete a new form of "conversation".

For the mentioned interface and functions, the core technology which detects users' emotion is the most critical part. We proposed several approaches to enhance the performance of the system, and showed satisfactory results. Along with the developed technology, a good communication between computer controlled devices and users is feasible. Moreover, several further applications utilizing system components become easier to implement.

Continuing collecting annotated materials and user feedbacks for learning, and then performing a long term experiment to develop good learning approaches is the future plan. Conversations from the Internet, such as Facebook, blog feedbacks, or line could be sources to gather various materials for advanced emotion detection. Making the system components real products like the home lighting system, the intelligent table lamp, or the music album promoter will be the next research direction.

Ackowledgements. Research of this paper was partially supported by National Science Council, Taiwan, under the contract NSC101-2628-E-224-001-MY3.

References

1. Bellegarda, J.R.: Emotion Analysis Using Latent Affective Folding and Embedding. In: Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text, Los Angeles, pp. 1–9 (2010)
2. Busso, C., Deng, Z., Yildirim, S., Bulut, M., Lee, C.M., Kazemzadeh, A., Lee, S., Neumann, U., Narayanan, S.: Analysis of Emotion Recognition using Facial Expressions, Speech and Multimodal Information. In: Proceedings of ACM 6th International Conference on Multimodal Interfaces (ICMI 2004), State College, PA (2004)
3. Conti, N., Jennett, C., Celdran, J., Sasse, A.: When Did My Mobile Turn Into A 'Sellphone'? A Study of Consumer responses to Tailored Smartphone Ads. In: Proceedings of the 26th Annual BCS Interaction Specialist Group Conference on People and Computers (HCI 2012), pp. 215–220 (2012)
4. Das, D.: Analysis and Tracking of Emotions in English and Bengali Texts: A Computational Approach. In: Proceedings of the International World Wide Web Conference (WWW 2011), Ph. D. Symposium, pp. 343–347 (2011)
5. Ku, L.-W., Sun, C.-W.: Calculating Emotional Score of Words for User Emotion Detection in Messenger Logs. In: Proceedings of the 2012 IEEE 13th International Conference on Information Reuse and Integration (IEEE IRI 2012), EMRITE Workshop, pp. 138–143 (2012)

6. Ku, L.-W., Sun, C.-W., Hsueh, Y.-H.: Demonstration of IlluMe: Creating Ambient According to Instant Message Logs. In: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (ACL 2012), pp. 97–102 (2012)
7. Parrott, W.: Emotions in Social Psychology. Psychology Press, Philadelphia (2001)
8. Roast, C., Zhang, X.: Exploring the Motivations Involved in Context Aware Services. In: Proceedings of the 26th Annual BCS Interaction Specialist Group Conference on People and Computers (HCI 2012), pp. 274–279 (2012)
9. Sarwar, B., Karypis, G., Konstan, J., Riedl, J.: Item Based Collaborative Filtering Recommendation Algorithms. In: Proceedings of the International World Wide Web Conference (WWW 2001), pp. 285–295 (2001)
10. Yang, Y.-H., Chen, H.H.: Ranking-Based Emotion Recognition for Music Organization and Retrieval. IEEE Transactions on Audio, Speech, and Language Processing 19(4) (2011)
11. Zbikowski, L.M.: Music, Emotion, Analysis. Music Analysis. Blackwell Publishing Ltd., Oxford (2011)
12. Zheng, V.W., Cao, B., Zheng, Y., Xie, X., Yang, Q.: Collaborative Filtering Meets Mobile Recommendation: A User-centered Approach. In: Proceedings of Twenty-Fourth National Conference on Artificial Intelligence, AAAI 2010 (2010)
13. Lin, K.H.-Y., Yang, C., Chen, H.-H.: Emotion Classification of Online News Articles from the Reader's Perspective. In: Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence, pp. 220–226 (2008)
14. Ku, L.-W., Chen, H.-H.: Mining Opinions from the Web: Beyond Relevance Retrieval. Journal of American Society for Information Science and Technology. Special Issue on Mining Web Resources for Enhancing Information Retrieval 58(12), 1838–1850 (2007)
15. Su, H.-Y.: The Multilingual and Multi-Orthographic Taiwan-Based Internet: Creative Uses of Writing Systems on College-Affiliated BBSs. Journal of Computer-Mediated Communication 9(1) (2003), `http://jcmc.indiana.edu/vol9/issue1/su.html`
16. CKIP (Chinese Knowledge Information Processing Group): The Content and Illustration of Academica Sinica Corpus. (Technical Report no 95-02/98-04). Taipei: Academia Sinica (1995/1998)