CONSCIENCE OF BLOGS: VERIFYING CONTEXTUAL APPROPRIATENESS OF EMOTIONS BASED ON BLOG CONTENTS

Michal Ptaszynski  Pawel Dybala  Wenhua Shi  Rafal Rzepka  Kenji Araki
Graduate School of Information Science and Technology, Hokkaido University
Kita-ku, Kita 14 Nishi 9, 060-0814 Sapporo, Japan
{ptaszynski,paweldybala,shibuka,kabura,araki}@media.eng.hokudai.ac.jp

ABSTRACT
This paper presents a method for estimating contextual appropriateness of speaker’s emotions. The method is using an affect analysis system to estimate the speaker’s emotions and a Web mining technique gathering from the Internet associations about emotional common-sense. The baseline of the Web mining technique, using all of the Web, is improved by restricting the query field to the contents of blogs as containing more evaluative information. A conversational agent equipped with this system could choose an appropriate conversational procedure. The proposed method is evaluated using two conversational agents. The use of blog contents improved the method in the aspect of quality as well as time of processing.

KEY WORDS
Computational Intelligence, Affect Analysis, Intelligent Agents, Natural Language Processing, Contextual Appropriateness of Emotions

1. Introduction
In recent years the research on computing different aspects of human intelligence has become an essential part of Computer Science and Artificial Intelligence research. One of the most important aspects of human intelligence is expressing and understanding emotions [1]. The field of AI embracing this subject, Affective Computing, initiated by Picard [2], focuses on recognizing the emotions of human-users in human-computer interaction. In the usual techniques developed for this task the emotions are recognized from: facial expressions [3], voice [4], or biometric data [5]. However, these methods, based on behavioral approaches, ignore the semantic and pragmatic context of emotions.

A field approaching emotion recognition from the semantic point of view is Affect Analysis - a field focused on developing natural language processing techniques for estimating the emotive aspect of text. The mainstream approach towards affect analysis is knowledge based syntactical analysis of particular utterance content [6]. However, recent introductions in our lives of technologies such as ubiquitous networks or wireless LAN, which made the Internet an indispensable everyday article, caused an increase of research using Internet resources to support emotion recognition by statistical approximation of numerous data gathered from the Web [7][8]. Within the last few years there were several attempts to retrieve information concerning human emotions and attitudes from the Internet [9]. However, until now there was no method capable to analyze affect with regard to a context and estimate whether an emotion conveyed in a conversation is appropriate for the particular situation.

Ptaszynski et al. [10], as the first one, presented a method 1) specifying what type of emotion was expressed and 2) determining whether the emotion is appropriate for the context it appears in. In the method he used Ptaszynski’s [6] system for affect analysis of utterances and Shi’s method for gathering emotive associations from the Web [11]. However, one of the problems with Shi’s method was gathering too much noise from the Web contents. In research such as the one by Abbasi et al. [9] it was proved that public Internet service, such as blogs or forums are a good material for affect analysis because of their richness in evaluative and emotive information. Therefore we restricted the query scope of Shi’s technique to mine not the whole Web, but only the contents of Yahoo! Japan - Blogs1 - a robust blog web service rich in information valuable and assumed to be more accurate for our research.

2. Aproach and Definitions
2.1 Emotions and Emotional Intelligence Framework
The idea of Emotional Intelligence (EI) was first officially proposed by Salovey and Mayer [1] who defined it as a part of human intelligence including the abilities to I) perceive emotions, II) integrate emotions to facilitate thoughts, III) understand emotions and to IV) regulate emotions to promote personal growth. In their EI Framework [1] they generally divide the first step, perception of emotions, into a) the ability to identify emotions and b) discriminate between accurate (or appropriate) and inaccurate (or inappropriate) expression of emotions, which is a key ability for the third step in EI Framework - understanding and analysis of emotional states.

However, as it is argued by Salovey and Mayer, recognizing emotions is merely the first basic step to acquire
the full scope of Emotional Intelligence (EI). Obtaining information about the emotion expressed by a user does not tell us anything about whether it is appropriate or not to express such an emotion in a given situation, and what actions should be undertaken as a reaction. As is argued by Solomon [12], e.g. anger can be appropriate when it is warranted (a direct offence) and inappropriate when it is unwarranted (scolding one’s child for ones own mistakes) and the reactions should be different for the two different contexts of anger. Or joy - appropriate during a small talk with friends, would be very inappropriate when expressed during a funeral. Not taking this fact into consideration in the grand majority of emotion processing research can drive us to the conclusion that such research assumes that positive emotions are always appropriate and negative ones - inappropriate.

The simple definition of emotions employed by Nakamura [13], says that emotions are every temporary state of mind, feeling or emotional state evoked by experiencing different sensations. However, regarding the above, it is necessary to complement it with Solomons theorem that emotions are, first of all, engagements with the world. This makes implementing the EI framework crucial to acquire fully computational model of Emotional Intelligence and therefore human intelligence in general. The attempts to implement the EI Framework usually do not go beyond theory [14], and the few practical attempts eventually still do not go beyond the first basic recognition step [15] - recognition of emotions.

This paper presents an attempt to one step further. In our method after recognizing the affective states of a speaker we also check whether they are appropriate for the context they appear in. Although computing appropriateness of affective states is a novel task in the field, it is a key issue to fully incorporate the Emotional Intelligence Framework into machines.

2.2 Blogs as Generalized Human Conscience

The development of widely accessible wireless Internet brought to the light different kinds of activities unavailable till then. Today statistically every family has at least one personal computer and an access to the Internet [16]. Socialization of the Web as personal and social space caused a development of services to connect people through from all over the world. One of that kind service are blogs - open diaries in which people encapsulate their own experiences, opinions and feelings to be read and commented by other people. In this shape - blogs has became an indicator of general human commonsense and conscious, and therefore are came into focus of scientific fields such as opinion mining, or sentiment and affect analysis [18]. As we assume, the information hidden in blogs can be used for one other purpose, namely to help distinguish whether an emotion conveyed by a user during the interaction with an agent is appropriate for the situational context it is used in. As the blog service we use for mining the desirable information we use Yahoo! Japan - Blogs Internet service as one of the most popular and robust blog services in Japan.

3. Methods

3.1 Affect Analysis

For affect analysis we use Ptaszynski’s et al. [6] ML-Ask (Emotive Elements / Emotive Expressions Analysis System), which analyses the emotive content of utterances in two steps: 1) Analyzing the general emotiveness of utterances; 2) Recognizing the particular emotion types. The system is based on Ptaszynski’s idea of two-part analysis of realizations of emotions in language into: Emotive elements, indicating that emotions have been conveyed, but not detailing what specific emotions there are, e.g. interjections (sugee2 /great!), mimetics (wakuwaku /heart pounding/), or vulgarities (-yagaru /a vulgarisation of a verb/); Emotive expressions, parts of speech or phrases, that in emotive sentences describe emotional states, e.g. nouns (aijou /love/), verbs (kanashimu /feel sad/) or adjectives (ureshii /happy/);

The emotive element database was divided into interjections, mimetics, endearments, vulgarities and representations of non-verbal emotive elements, such as exclamation mark or ellipsis {907 items in total}. The database of emotive expressions contains Nakamura’s collection [13] {2100 items in total}. For a textual input utterance provided by the user, the system searches in order for: 1) emotive elements to determine the emotiveness (emotive / non-emotive); 2) emotive expressions in emotive utterances to determine the specific types of the conveyed emotions. The system uses Nakamura’s 10 type-classification of emotions said to be the most appropriate for the Japanese language: [joy, delight], [anger], [sorrow, sadness], [fear], [shame, shyness, bashfulness], [liking, fondness], [dislike, detestation], [excitement], [relief] and [surprise, amazement].

However, keyword-based extraction of emotive expressions caused misinterpretations in valence polarity determination. To solve this problem, we applied an analysis of Contextual Valence Shifters.

3.2 Contextual Valence Shifters

The idea of Contextual Valence Shifters (CVS) was proposed by Polanyi and Zaenen [17]. They distinguished 2 kinds of CVS: negations (changing the valence polarity of semantic orientation of an evaluative word) and intensifiers (intensifying the semantic orientation). Examples of CVS negations in Japanese are: -nai (not-), amari -nai (not quite-), -to wa ienai (cannot say it is-), or -te wa ikenai (cannot+[verb]-). Intensifiers are: totemo- (very-), sugoku-(a lot), or kiwame-te- (extremely).

To specify the emotion types after changing valence polarity with CVS, we apply Russell’s idea of the 2-

---

2 In this paper we use italics for expressions in Japanese
dimensional model of affect [19] which assumes that all emotions can be described in 2 dimensions: the emotion’s valence polarity (positive/negative) and activation (activated/deactivated). Nakamura’s emotion types were mapped on the 2-dimensional model of affect and their affiliation to one of the spaces determined. When a CVS structure is discovered, ML-Ask changes the valence polarity of the detected emotion. The appropriate emotion after valence changing is determined as the one with valence polarity and activation parameters different to the contrasted emotion (see Figure 1). The flow chart of the ML-Ask system including CVS procedure is presented on Figure 2, two examples of analysis are shown in Table 1.

<table>
<thead>
<tr>
<th>Example</th>
<th>Japanese</th>
<th>Emotem</th>
<th>English Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>Kyo wa nante kimochi ii hi nanda !</td>
<td>noun:THEM ptl:ML nante MX:joy noun:SUBJ ML:nanda ML:!</td>
<td>Translation: Today is such a nice day!</td>
</tr>
<tr>
<td>(2)</td>
<td>Akirame cha ikenai yo!</td>
<td>MX:dislike CV:cha CVS:cha-ikenai{—joy} ML:yo ML:!</td>
<td>Translation: Don’t cha give up!</td>
</tr>
</tbody>
</table>

Figure 1. Grouping Nakamura’s classification of emotions on Russell’s space.

### 3.3 Web Mining

As a verifier of appropriateness of the speaker’s affective states recognized by ML-Ask we apply Shi’s Web mining technique for extracting emotive associations from the Web based on causality [11]. The technique is made up of three stages: a) extracting n-gram phrases from an utterance (phrases contain n-grams of the length from one whole utterance to trigrams); b) modification of phrases ending with adjectives and verbs by the use of causality morphemes to perform semantically deeper Web mining; and c) extraction of emotion associations from the Web using the prepared n-gram phrases as query phrases in Google search engine and cross-referencing them with the emotive expressions database.

### 3.4 Blog Mining

The baseline method for Web mining developed by Shi [11] is gathering emotive associations from the whole Web using Google search engine. However, in this form the general commonsense arousing from the mining results included a large amount of noise. To solve the noise problem we restricted the mining to the Yahoo! Japan - Blogs service. This operation is also assumed to provide more accurate emotive information since, as mentioned above, people use blogs mostly to convey information on their own opinions and emotions.

The Blog mining procedure performs the same kind of query first on all of the public blogs from Yahoo! Japan - Blogs. The paragraphs of each blog containing query phrases are co-referenced with the database of emotive expressions to gather the emotive associations approximated in the general commonsense used as an indicator of appropriateness. In a case when no information was gathered from the blog contents, the mining procedure performs the same search with the baseline conditions - on the whole Web, but with the use of Yahoo search engine instead of Google. A flow chart of the baseline method for Web mining and supported with Blog mining is presented in Figure 3. A result example is presented in Table 2.
Table 2. Example of emotion association extraction from the Web and its improvement by blog mining procedure.

<table>
<thead>
<tr>
<th>Extracted emotion type</th>
<th>Baseline (Type extracted / all extracted types(Ratio))</th>
<th>Blogs (Type extracted / all extracted types(Ratio))</th>
</tr>
</thead>
<tbody>
<tr>
<td>liking</td>
<td>79/284 (0.287)</td>
<td>603/610 (0.985)</td>
</tr>
<tr>
<td>surprise</td>
<td>30/284 (0.105)</td>
<td>1/610 (0.001) [noise:discarded]</td>
</tr>
<tr>
<td>excitement</td>
<td>30/284 (0.105)</td>
<td>1/610 (0.001) [noise:discarded]</td>
</tr>
<tr>
<td>fear</td>
<td>29/284 (0.102)</td>
<td>1/610 (0.001) [noise:discarded]</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

4. Contextual Appropriateness of Emotion Verification (CAEV) Procedure

The two described above systems are combined as follows. When ML-Ask discovers an emotive utterance and successfully specifies an emotion type, the Web mining technique verifies whether the expressed emotion type is appropriate for the context. The emotions associating with the context of the utterance in a common sense manner are extracted from the Web and the emotions discovered by ML-Ask in user’s utterance are compared to the Web mining procedure results. If the two match, the emotions expressed by the user are determined as appropriate. In such situations, a conversational agent equipped with this method can choose a dialogue strategy to sympathize with the user (see two examples below).

Positive-appropriate emotion:
User: I’m so happy I passed the exam!
ML-Ask: joy; Web mining: joy, excitement
Agent: Yeah! That’s great! [sympathy]

Negative-appropriate emotion:
User: I hate him for making a fool of me in front of everyone.
ML-Ask: dislike;
Web mining: anger, dislike
Agent: You have a reason to be angry. [empathy]

However, if the verification procedure indicates that the expressed emotion is inappropriate for the context, the agent could undertake different measures e.g., helping the user to manage their emotions (see two examples below).

Positive-inappropriate emotion:
User: I’d be happy if that bastard died for cancer!
ML-Ask: joy; Web mining: fear, sadness
Agent: Is that what you really feel? [counselling]

Negative-inappropriate emotion:
User: I won the prize, but I feel so bored.
ML-Ask: dislike, depression;
Web mining: excitement, joy
Agent: You should be happy! [consolation]

4.1 Two-dimensional Model of Affect in CAEV Procedure

As is argued by Solomon [12] although people sometimes misattribute the specific emotion types of other people, they rarely misattribute their valence. One could, for example, confuse anger with irritation, but it is unlikely to confuse admiration with detestation, or love with hate. Therefore we decided to check if, in cases where the specific emotion types conveyed in an utterance do not perfectly match the emotive associations, at least their general features matched. By general features we mean the two ones proposed by Russell [19] in the theory of the two-dimensional model of affect - valence and activation.

5. Experiment Setting

To evaluate the method we performed an experiment. In the experiment we used two different conversational agents.

- Modalin is a non-task oriented keyword-based conversational agent, which uses modality to enhance Web-based propositions for dialogue. The agent was developed by Higuchi et al. [20].

- Pundalin is a non-task oriented conversational agent created by combining Modalin with Dybala’s Pun generating system PUNDA [21]. Pundalin therefore is a humor-equipped conversational agent using puns to enhance the communication with a user.

13 users (undergraduate students from different departments) performed a ten turn conversation with both agents. All conversations were then analysed by ML-Ask. 6 out of all 26 conversations contained no specified emotional states and were excluded from the further evaluation process. For the conversations containing sentences which ML-Ask described as emotive and specified the emotion types, the Web mining procedure was carried out to determine whether the emotions expressed by the user were contextually appropriate.

We compared two versions of the method: with the baseline Web mining procedure and with restricted search scope, to blog contents, as described in section 3.4. The difference in results appeared in 5 conversation sets. Next, the results were stored and a questionnaire was designed to evaluate how close they were to human thinking. One questionnaire set consisted of one conversation record and three questions inquiring 1) what was the valence of emotions conveyed in emotive utterances; 2) what were the specific emotion types conveyed in the conversation; and 3) whether they were contextually appropriate. Every questionnaire set was filled out by 10 people (undergraduate students different from the users who performed the conversations with the agents). The conversations where differences in results appeared for the two compared procedures, were evaluated separately for each version of the
Web mining method. For every conversation set we calculated how many people of the human evaluators agreed with the system’s results.

The evaluated items were: A) general valence determination and B) specific emotion types determination accuracies in ML-Ask; and the accuracy of the system as a whole to determine the contextual appropriateness of C) specific emotion types and D) valence. The results for A) and B) are provided in Table 3. The results for C) and D) are given in Table 3 and Figure 4: 1)-4).

![Figure 4: Results of CAEV Procedure evaluation in estimating appropriateness of: valence - 1), 3) and specific emotion types - 2), 4). Summarization in four - 3), 4) and two - 1), 2) groups of results.](image)

### Table 3. The number of people that agreed with the system when analysing one agent at a time for evaluated items A)-D); summarized results for both agents separately (upper) and together (lower); the results compared for the baseline and blog-mining versions of the system.

<table>
<thead>
<tr>
<th></th>
<th>Modalin</th>
<th>Pundalin</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of people</td>
<td>10-7 ppl.</td>
<td>6-4 ppl.</td>
</tr>
<tr>
<td><strong>A</strong></td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td><strong>B</strong></td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td><strong>C</strong></td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td><strong>D</strong></td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td><strong>Overall results</strong></td>
<td><strong>Summary (10-4 ppl.agreed)</strong></td>
<td><strong>Baseline system</strong></td>
</tr>
<tr>
<td>No. of people</td>
<td>10-7 ppl.</td>
<td>6-4 ppl.</td>
</tr>
<tr>
<td><strong>A</strong></td>
<td>14</td>
<td>4</td>
</tr>
<tr>
<td><strong>B</strong></td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td><strong>C</strong></td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td><strong>D</strong></td>
<td>8</td>
<td>4</td>
</tr>
</tbody>
</table>

### Table 4. Three examples of the results provided by the emotion appropriateness verification procedure (CAEV) with a separate display of the examples showing the improvement of the procedure after applying CVS.

<table>
<thead>
<tr>
<th>Part of conversation in Japanese</th>
<th>ML-Ask</th>
<th>Web Mining</th>
<th>CAEV</th>
</tr>
</thead>
<tbody>
<tr>
<td>(English translation)</td>
<td>positive</td>
<td>positive</td>
<td>appropriate</td>
</tr>
<tr>
<td>USER: Sore wa omoshiroi desn nu. (Computers are so interesting!)</td>
<td>[joy]</td>
<td>[joy]</td>
<td></td>
</tr>
<tr>
<td>SYSTEM: Sore wa oishii desu ka. (Is it tasty?) [about instant noodles]</td>
<td>negative</td>
<td>negative</td>
<td>appropriate</td>
</tr>
<tr>
<td>USER: Oishii kedo, ore wa akita kana. (It's tasty, but I've grown tired of it.)</td>
<td>[dislike]</td>
<td>[dislike]</td>
<td></td>
</tr>
<tr>
<td>SYSTEM: Sore wa omoshiroi tie (It's so interesting!) [about conversation]</td>
<td>negative</td>
<td>negative</td>
<td>appropriate</td>
</tr>
<tr>
<td>USER: Sore hodo omoshirokru mo nakatta yo. (It wasn’t that interesting.)</td>
<td>[dislike]</td>
<td>[sad]</td>
<td></td>
</tr>
</tbody>
</table>

### 6. Results

The usual number of evaluators in research on affect analysis and neighbour fields is five, three, and sometimes even one person [8]. In this perspective ten people employed in our evaluation makes it satisfactory. In summarization of the results we assumed that at least four people out of ten have to confirm the system's results. This is a fair ratio when we consider that it means that at least four people of ten provided exactly the same results as the system. This is a difficult thing to achieve in emotion research, because of the many ambiguities that go along with expressing and perceiving emotions.

The survey provided many positive results. Firstly, in most cases ML-Ask’s results on affect analysis were confirmed by humans. ML-Ask supported with the CVS procedure acquired 90.0% of accuracy in emotion valence recognition and 85% in specific emotion types recognition.

Since one of the agents was using humorous responses we also checked whether the jokes influenced the human-computer interaction. Most of the emotions expressed in the conversations with Pundalin were positive (67%) whereas for Modalin most of the emotions were negative (75%), which confirms that users tend to be positively influenced by the use of jokes in conversational agents [21].

Changing the query scope in the Web mining procedure to blog contents improved the performance of the appropriateness verification procedure both on the level of valence and specific emotion types. The performance was improved from 55.0% to 60.0% for the former and from 50.0% to 60.0% for the latter. The results were statistically significant on a 5% level (p-value = 0.0274). Some of the successful examples are shown in Table 4.

### 7. Conclusion and Future Work

In this paper we introduced a method for verifying contextual appropriateness of emotions conveyed in conversations. The method presents a novel approach towards user emotion estimation in human-computer interaction. Most of the similar methods are focused on discovering certain
emotion types, or more often, only the general valence of the emotions. This ignores the complexity of and strong context dependence of emotions. However, to create a machine capable to communicate with a user on a human level, there is a need to equip it with an Emotional Intelligence Framework [1]. The method described in this paper takes a step towards the practical implementation of such a framework by providing machine computable means for verifying whether an emotion conveyed in a conversation is contextually appropriate. In the proposed method a system for affect analysis is used to recognize user emotions and a Web mining technique, based on blog content searching, to verify their appropriateness in a particular context.

The affect analysis method was evaluated at 85.0% accuracy in determining the specific emotion types and 90.0% accuracy in determining valence of the emotions. The accuracy in determining the contextual appropriateness was significantly improved from 50.0% to 60.0% for particular emotion types and from 55.0% to 60.0% for valence. The results, although not perfect, are very encouraging and we will continue research on contextual appropriateness of emotions. An agent equipped with our system can be provided with hints about what communication strategy would be the most desirable at a certain moment. For example, a conversational agent can choose to either sympathize with the user or take precautions and help them manage their emotional states.

The theory of Emotional Intelligence, to which we refer in this paper, is a quickly developing field of research. It frequently delivers new discoveries about the structures and functions of emotions and therefore should be in focus of researchers attempting to develop means for computing human intelligence. We proved that computing emotions in context is a feasible task. Although the system as a whole is still not perfect and its components (ML-Ask and the Web mining technique) need further improvement, there have been seen a significant improvement by restricting Web mining to the contents of Yahoo!Japan-Blogs. This research also defines a new set of goals for Affective Computing, by proposing computational means for verification of contextual appropriateness of emotional states in conversation. This is a key task on the way to full implementation of Emotional Intelligence in machines and therefore is a valuable research in the field of Artificial Intelligence in general.

Acknowledgments

This research was partially supported by a Research Grant from the Nissan Science Foundation and The Global Centers of Excellence Program founded by Japan’s Ministry of Education, Culture, Sports, Science and Technology.

References


