一段階高い感情解析を目指して：感情の文脈的適切性の概念と展望について

ミハウ・プタシンスキ
北見工業大学 情報システム工学科

Taking Affect Analysis One Step Higher: The Idea of Contextual Appropriateness of Emotions and Its Perspectives

Michal Ptaszynski
Department of Computer Science, Kitami Institute of Technology

Abstract
I present my research in Contextual Affect Analysis (CAA) for the need of future application in intelligent agents, such as conversational agents or artificial tutors. In agent-user discourse it is crucial that the artificial agent was able not only to detect user emotions, but also to verify towards whom they were directed and whether they were appropriate for the context of the conversation. I focus on verification of contextual appropriateness of emotions. I present the results of experiments and discuss implications and future directions in development of this method.

Introduction

Research on emotions in the fields of Artificial Intelligence and Natural Language Processing, like the ones described above has flourished rapidly through several years. Unfortunately, in much of such research contextuality of emotions is disregarded. Based only on behavioral approaches, methods for emotion recognition ignore the context of emotional expression. Therefore, although achieving good results in laboratory conditions, such methods are often inapplicable in real world tasks. For example, a system for recognition of emotions from facial expressions, assigning "sadness" when user is crying would be critically mistaken if the user was, e.g., cutting an onion in the kitchen. Similarly in language, not including con text in the processing could lead to various processing errors. For example, one can consider a system detecting happiness when user uses a word "happy". The system would be critically mistaken if the user actually said: "I'm not happy at all!" This shows that not considering at least grammatical context in the processing causes erroneous detection of opposite emotion. However, a deeper problem appears when the user said something like "I would be so happy if that bastard was hit by a car!" Here a grammatical context does not suffice correct processing and deeper context is required in the processing. As the above examples show, recognizing emotions without recognizing their context is incomplete and cannot be sufficient for real-world applications.

The outline of this paper is as follows. I firstly present the background for this research in which I describe the fields of Affect Analysis (AA) and Contextual Affect Analysis (CAA). Next I describe previously developed tools applied in the research. Further section describes the method for verifying whether the emotions expressed in conversations are appropriate to the context of the situation. Next section contains descriptions of the design of evaluation experiment and its results. In final section I present a discussion on further implication of context aware Affect Analysis. Lastly, I conclude the paper and propose some ideas to improve the described method.

Background

Affect Analysis

Text based Affect Analysis (AA) has been defined as a field focused on developing natural language processing techniques for estimating the emotive aspect of text [16]. For example, Elliott [17] proposed a keyword-based Affect Analysis system applying an affect lexicon (including words like "happy", or "sad") with intensity modifiers (words like "extremely", "somewhat"). Liu et al. [18] presented a model of text-based affect sensing based on OMCS (Open-Mind Common Sense), a generic common sense database, with an application to e-mail interpretation. Alm et al. [1] proposed a machine learning method for affect analysis of fairy tales. Aman and Szpakowicz also applied machine learning techniques to analyze emotions expressed on blogs [2].

There have also been several attempts to achieve this goal for the Japanese language. For example, Tsuchiya et al. [21] tried to estimate emotive aspect of utterances with a use of an association mechanism. On the other hand, Tokuhisa et al. [3] and Shi et al. [4] used a large number of examples gathered from the Web to estimate user emotions. Furthermore, Ptaszynski et al. [5] proposed a Web-based supported affect analysis system for Japanese text-based utterances.

Contextual Affect Analysis

Processing the context of emotions, or Contextual Affect Analysis (CAA) [6, 7, 8, 9, 10, 11, 12, 13, 14], is a newly recognized field. During its fifteen years of history, research on computer processing of emotions, or Affective Computing [15], was in great part focused on the recognition of expressions of user emotions. However, little research addressed the need for computing the context of the expressed emotions. In the age of information explosion, with an easy access to very large sources of data (such as the Internet), the time has come to finally address this burning need. My research is focused on only one type of emotion processing, affect analysis of text. The future challenge will be to develop methods for processing the context in more general meaning, making the machines aware of the sophisticated environment humans live in. It has been shown that CAA is a feasible task, although much further research in this matter needs to be done in the near future. In this paper, I focused in particular on applying context processing to text-based affect analysis. I did this in two ways.

Firstly, one of the common problems in the keyword-based systems for affect analysis is confusing the valence of emotion types, since the emotive expression keywords are extracted without their grammatical context. An idea aiming to solve this problem is the idea of Contextual Valence Shifters (CVS), words and phrases like "not", or "never", which change the valence of an emotion (positive/negative). As the first step towards contextual processing of emotions I applied CVS as a supporting procedure for affect analysis system for Japanese.

Secondly, I have developed a method making use of the wider context an emotion is expressed in. The method, using a Web mining technique, determines, whether the expressed emotion is appropriate for its context. It introduces an idea of Contextual Appropriateness of Emotions to the research on emotion processing. This idea adds a new dimension in emotion recognition, since it assumes that both positive and negative emotions can be appropriate, or inappropriate, depending on their contexts. The method is based on the assumption that the Internet can be considered as a database of experiences people describe on their homepages or weblogs. Since the context of emotions is formulated through collecting experiences, these experiences could be as well "borrowed" from the Internet [16].

In conclusions to this paper I present a discussion on future directions and applications of context processing within Affective Computing.

Affect Analysis Tools

In this section I describe all tools, methods and resources for basic AA used further in CAA tasks.

Table 1. Distribution of emotive expressions across emotion classes in Nakamura's dictionary, ordered by the number of expressions per class.

emotion class	nunber of expressions	emotion class	nunber of expressions
dislike	532	fondness	197
excitement	269	fear	147
sadness	232	surprise	129
joy	224	relief	106
anger	199	shame	65
		sum	**2100**

Emotive Expression Dictionary [18] is a dictionary developed by Akira Nakamura in a period of over 20-year time. It is a collection of over two thousand expressions describing emotional states collected manually from a wide range of literature. It was converted into an emotive expression database by Ptaszynski et al. [5, 19] in their research on affect analysis of utterances in Japanese. Nakamura's dictionary is a state-of-the art example of a handcrafted lexicon of emotive expressions. It also proposes a classification of emotions that reflects the Japanese language and culture the most appropriately. In particular, Nakamura proposes ten emotion types: 喜 ki/yorokobi (joy, delight; later referred to as **joy**), 怒 dō/ikari (**anger**), 哀 ai/aware (sorrow, sadness, gloom; later referred to as **sadness**), 怖 fu/kowagari (**fear**), 恥 chi/haji (shame, shyness, bashfulness; later referred to as **shame**), 好 kō/suki (liking, fondness; later referred to as **fondness**), 厭 en/iya (dislike, detestation; later referred to as **dislike**), 昂 kō/takaburi (**excitement**), 安 an/yasuragi (**relief**), and 驚 kyō/odoroki (surprise, amazement; later referred to as **surprize**). This classification is also applied in the lexicon itself. All expressions are classified as representing a specific emotion type, one or more if applicable. The distribution of separate expressions across all emotion classes is represented in Table 1.

ML-Ask

ML-Ask, or *eMotive eLement and expression Analysis* system is a keyword-based language-dependent system for automatic affect annotation on utterances in Japanese constructed by Ptaszynski et al. [5, 19]. It uses a two-step procedure:

1. Specifying whether an utterance is emotive, and
2. Recognizing the particular emotion types in utterances described as emotive.

ML-Ask is based on the idea of two-part classification of realizations of emotions in language into:

1) *Emotive elements* or *emotemes*, which indicate that a sentence is emotive, but do not detail what specific emotions have been expressed. For example, interjections such as "whoa!" or "Oh!" indicate that the speaker (producer of the utterance) have conveyed some emotions. However, it is not possible, basing only on the analysis of those words, to estimate

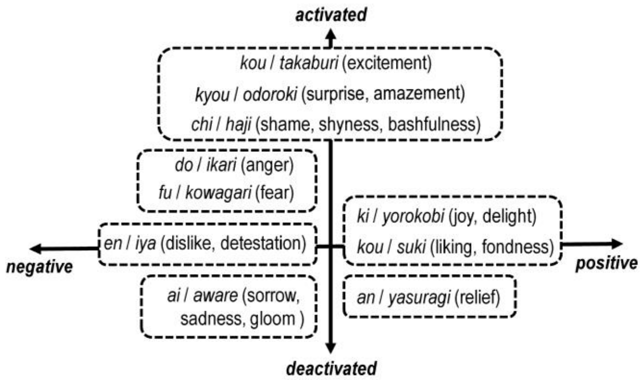

Fig. 1 Mapping of Nakamura's classification of emotions on Russell's 2D space.

precisely what kind of emotion the speaker conveyed. Ptaszynski et al. [19] include in emotemes such groups as interjections, mimetic expressions, vulgar language and emotive markers. The examples in Japanese are respectively: *sugee* (great! - interjection), *wakuwaku* (heart pounding - mimetic), *-yagaru* (syntactic morpheme used in verb vulgarization) and '!', or '??' (sentence markers indicating emotiveness). Ptaszynski et al. collected and hand-crafted a database of 907 emotemes. A set of features similar to what is defined by Ptaszynski et al. as emotemes has been also applied in other research on discrimination between emotive (emotional/subjective) and non-emotive (neutral/objective) sentences [2, 23, 30].
2) *Emotive expressions* are words or phrases that directly describe emotional states, but could be used to both express one's emotions and describe the emotion without emotional engagement. This group could be realized by such words as *aijou* (love - noun), *kanashimu* (feel sad, grieve - verb), *ureshii* (happy - adjective), or phrases such as: *mushizu ga hashiru* (to give one the creeps [of hate]) or *ashi ga chi ni tsukanai* (walk on air [of happiness]). As the collection of emotive expressions ML-Ask uses a database created on the basis of Nakamura's Emotive Expression Dictionary [18].

With these settings ML-Ask was proved to distinguish emotive sentences from non-emotive with a very high accuracy (over 90%) and to annotate affective information on utterances with a sufficiently high Precision (85.7% compared to human annotators), and satisfying, although not ideal Recall (54.7%) [19, 27]. To improve the system performance I also implemented Contextual Valence Shifters.

The idea of Contextual Valence Shifters (CVS) was first proposed by Polanyi and Zaenen [24]. They distinguished two kinds of CVS: negations and intensifiers. The group of negations contains words and phrases like "not", "never", and "not quite", which change the valence (also called polarity or the semantic orientation) of an evaluative word they refer to. The group of intensifiers contains words like "very", "very much", and "deeply", which intensify

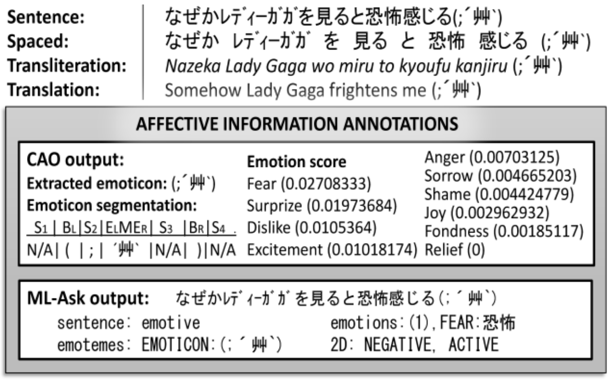

Fig. 2 Output examples for ML-Ask and CAO.

the semantic orientation of an evaluative word. ML-Ask fully incorporates the negation type of CVS with a 108 syntactic negation structures. Examples of CVS negations in Japanese are structures such as: *amari -nai* (not quite-), *-to wa ienai* (cannot say it is-), or *-te wa ikenai* (cannot [verb]-). In this paper I compared the performance of ML-Ask with and without (baseline) CVS improvement, within the evaluation of the procedure for verification of emotion appropriateness. As for intensifiers, although ML-Ask does not include them as a separate database, most Japanese intensifiers are included in the emoteme database. The system calculates emotive value, which is interpretable as emotional intensity of a sentence. It is calculated as the sum of emotemes in the sentence. The performance of setting the emotive value was evaluated on 84% comparing to human annotators [27]. Finally, the last distinguishable feature of ML-Ask is implementation of Russell's two dimensional affect space [33]. It assumes that all emotions can be represented in two dimensions: the emotion's valence (positive/negative) and activation (activated/deactivated). An example of negative-activated emotion could be "anger"; a positive-deactivated emotion is, e.g., "relief". The mapping of Nakamura's emotion types on Russell's two dimensions proposed by Ptaszynski et al. [19] was proved reliable in several research [6, 19, 26]. The mapping is represented in Figure 1. An example of ML-Ask output is represented in Figure 2.

CAO

CAO, or *emotiCon Analysis and decOding of affective information system* is a system for estimation of emotions conveyed through emoticons developed by Ptaszynski et al. [26]. Emoticons are sets of symbols widely used to convey emotions in text-based online communication, such as blogs. CAO extracts an emoticon from an input (a sentence) and determines specific emotion types expressed by it using a three-step procedure. Firstly, it matches the input to a predetermined raw emoticon database containing over ten thousand emoticons. The emoticons, which could not be estimated using only the database are

Table 2. Example of context n-gram phrases separation from an utterance. Grammar shortcuts: SUB = subject particle, GER = gerund, PRF = perfect form.

Original utterance **English translation**	*Aa, pasokon ga kowarete shimatta…* Darn, the PC has broken…	
longest n-gram **(here: hexagram)**	(1) *Aa* *pasokon* *ga* *koware-* *te* *shimau* [interjection] [noun] [SUB] [verb] [GER] [PRF]	
pentagram	(2) *pasokon ga koware te shimau*	
tetragram	(3) *Aa, pasokon ga kowareru*	
trigrams	(4) *pasokon ga kowareru*	(5) *koware te shimau*

Table 3. Hit-rate results for the eleven morphemes with the ones used in the Web mining technique in bold font.

morpheme	***-te***	***-node***	***-tara***	*-nara*	*-kotoga*	*-nowa*
result	**41.97%**	**7.20%**	**5.94%**	1.17%	0.35%	2.30%
morpheme	***-to***	***-kara***	**subtotal**	*-ba*	*-noga*	*-kotowa*
result	**31.97%**	**6.32%**	**93.40%**	3.19%	2.15%	0.30%

automatically divided into semantic areas, such as representations of "mouth" or "eyes", basing on the idea of kinemes, or minimal meaningful body movements, from the theory of kinesics [28, 29]. The areas are automatically annotated according to their co-occurrence in the database. The annotation is firstly based on eye-mouth-eye triplet. If no triplet was found, all semantic areas are estimated separately. This provides hints about potential groups of expressed emotions giving the system coverage of over 3 million possibilities. The performance of CAO was evaluated as close to 98% [26] which proved CAO as a reliable tool for the analysis of Japanese emoticons. In the annotation process CAO was used as a supporting procedure in ML-Ask to improve the performance of the affect annotation system and add detailed information about emoticons appearing in the text. An example of CAO output is represented in Figure 2.

Web Mining Technique for Emotion Association Extraction

To verify the appropriateness of the speaker's affective states I applied Shi et al.'s [4] Web mining technique for extracting emotive associations from the Web. Ptaszynski et al. [5] already showed that ML-Ask and Shi's technique are compatible and can be used as complementary means to improve the emotion recognition task. However, these two methods are based on different assumptions. ML-Ask is a language based affect analysis system and can recognize the particular emotion expression conveyed by a user. On the other hand, Shi's technique gathers from the Internet large number of examples and derives from this data an approximated reasoning about what emotion types usually associate with the input contents. Therefore it is more reasonable to use the former system as emotion detector, and the latter one as a verifier of naturalness, or appropriateness of user emotions.

Shi's technique performs common-sense reasoning about which emotions are the most natural to appear in the context of an utterance, or in other words, which emotions should be associated with it. Emotions expressed, which are unnatural for the context (low or not on the list) are perceived as inappropriate. The technique is composed of three steps: 1) extracting context phrases from an utterance; 2) adding causality morphemes to the context phrases; 3) cross-referencing the modified phrases on the Web with emotive lexicon and extracting emotion associations for each context phrase.

Phrase Extraction Procedure

An utterance is first processed by MeCab, a tool for tokenization and part-of-speech analysis of Japanese [25]. Every element separated by MeCab is treated as a unigram. All unigrams are grouped into larger groups of n-grams preserving their word order in the utterance. The groups are arranged from the longest n-gram (the whole sentence) down to all groups of trigrams. N-grams ending with particles are excluded, since they gave too many ambiguous results in pre-test phase. An example of phrase extraction is presented in Table 2.

Morpheme Modification Procedure

On the list of n-gram phrases the ones ending with a verb or an adjective are then modified grammatically with causality morphemes. This is performed in line with linguistic argument that Japanese people tend to convey emotive meaning after causality morphemes [37]. Shi et al. [4] independently confirmed this argument experimentally. They distinguished eleven emotively stigmatized morphemes for the Japanese language using statistical analysis of Web contents and performed a cross reference of appearance of the eleven morphemes with the emotive expression database using the Google search engine. This provided the results (hit-rate) showing which of the eleven causality morphemes were the most frequently used to express emotions. For the five most frequent

Table 4. Examples of n-gram modifications for Web mining.

Original n-gram	*pasokon ga koware te shimau*	*/causality morpheme/*
n-gram phrase adjusting (morpheme modification)	*pasokon ga koware te shimat -te*	*/ -te /*
	pasokon ga koware te shimau -to	*/ -to /*
	pasokon ga koware te shimau -node	*/ -node /*
	pasokon ga koware te shimau -kara	*/ -kara /*
	...	...

Table 5 Example of emotion association extraction from the Web and its improvement by blog mining procedure.

Sentence: *Konpyūta wa omoshiroi desu ne.* (Computers are so interesting.)

Extracted emotion types	**Baseline: Type extracted / all extracted types (Ratio)**	**Extracted emotion types**	**Blogs: Type extracted / all extracted types (Ratio)**	
fondness	79/284(0.287)	fondness	601/610(0.985)	
surprise	30/284(0.105)	excitement	1/610(0.001)	[rejected
excitement	30/284(0.105)	fear	1/610(0.001)	as
fear	29/284(0.102)	relief	1/610(0.001)	noise]
...	...	...	...	

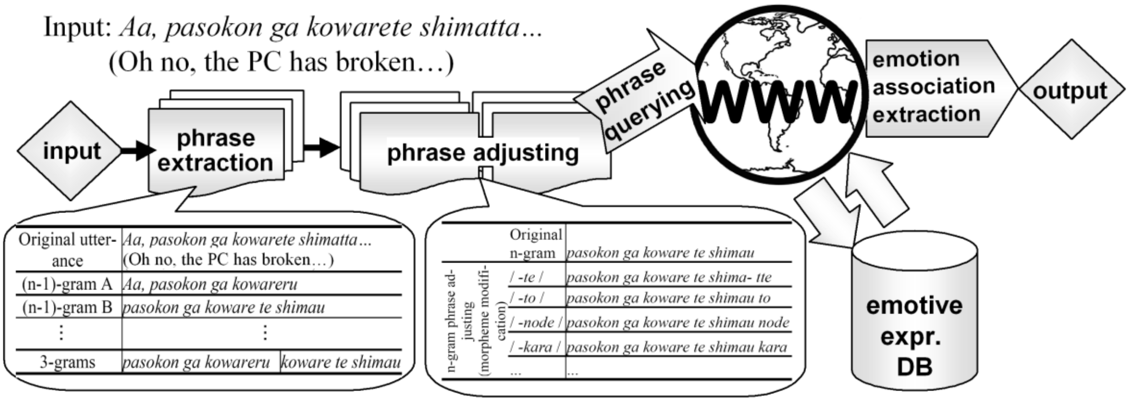

Fig. 3 Flow chart of the Web mining technique.

morphemes, the coverage of Web mining procedure still exceeded 90%. Therefore for the Web mining they decided to use those five ones, namely: *-te, -to, -node, -kara and -tara* (see Table 3). An example of morpheme modification is presented in Table 4.

Emotion Association Extraction Procedure

In this step the modified n-gram phrases are used as a query in Google search engine and 100 snippets for one morpheme modification per query phrase is extracted. This way a maximum of 500 snippets for each queried phrase is extracted. These are cross-referenced with emotive expression database (see Figure 3). The emotive expressions extracted from the snippets are collected, and the results for every emotion type are sorted in descending order. This way a list of emotions associated with the queried sentence is obtained. It is the approximated emotive commonsense used further as an appropriateness indicator. An example of emotive association extraction is shown in Table 5.

Blog Mining

The baseline of the Web mining method, using Google to search through the whole Web, was gathering a large amount of noise. To solve this problem I made two modifications. Firstly, I added a heuristic rule stopping the search if any emotions were found using the longer n-grams. This changed the method from Recall-oriented to Precision-oriented by assuring the extraction of only the closest emotive associations. It also speeds up the extraction process. Secondly, since, as mentioned before, people convey on blogs their opinions and emotions, I restricted the mining to blog contents to assure extraction of more accurate emotive associations. The blog mining procedure performs the query first on the public blogs from *Yahoo!Japan-Blogs* (blogs.yahoo.jp). The paragraphs of each blog containing query phrases are co-referenced with emotive expression database to gather the emotive associations. If no information was gathered from the blog contents, the same search is performed with the baseline conditions - on the whole Web. An example of improvement is presented in Table 5.

Method for Verification of Contextual Appropriateness of Emotions

As one of the recent advances in affect analysis, it was shown that Web mining methods can improve the performance of language-based affect analysis systems [3, 4, 22]. However, in these methods, although the results of experiments appear to be positive, two extremely different approaches are mixed, the language/keyword based approach and the Web mining based approach. The former, in which the information provided by the user in input is matched to the existing lexicons and sets of rules, is responsible for recognizing the particular emotion expression conveyed by the user at a certain time. The latter one is based on gathering from the Internet large numbers of examples and derives from these an approximated reasoning about what emotions usually associate with certain contents. Using the Web simply as a complementary mean for the language based approach, although achieving reasonable results, does not fully exploit the potential lying in the Web [16].

Here I present a method utilizing these two approaches in a more effective way. The method is capable to analyze affect with regard to a context and estimate whether an emotion conveyed in a conversation is appropriate for the particular situation. In the method I used previously developed systems for affect analysis (ML-Ask and CAO described in section 3). Next, I used a method for gathering emotive associations from the Web developed by Shi et al. [4].

Furthermore, I checked several versions of the method to optimize its procedures. Firstly, I checked two versions of ML-Ask, with and without Contextual Valence Shifters. Secondly, I checked two versions of the Web mining technique, one performing the search on the whole Internet and the second one searching only through blogs.

Method Description

Affect Analysis

As the first step of the method for verification of contextual appropriateness of emotions, I used the two affect analysis systems described in section 3 (ML-Ask and CAO). The affect analysis provides information on whether an utterance was emotive or not, and what type of emotion was expressed in the utterance. In a conversation between a user and an agent, the affect analysis is performed on each user utterance in user-agent conversation. For every emotive utterance with specified emotion type a Web mining technique is used as a verifier of emotion appropriateness.

Web Mining

In the second step a list of emotive associations is obtained from the Web. This is done with the use of the Web mining technique described in section 3.4. The list contains emotion types that associate with the sentence contents. The emotion types that correlate the most strongly appear on the top of the list. Emotion types with weaker correlations appear lower on the list. Emotion types that do not appear on the list at all are considered as the ones with no correlation with the sentence contents. As the rule of thumb I assumed that the emotion types appropriate for the sentence contents (or context in general) should appear in approximated 50% of all results. An example of such a list is represented in table 5. It shows that when the Web mining is based on the whole Web, emotions considered as appropriate include fondness, and two more (surprise and excitement). However, when the Web mining is limited to blogs the emotion type extraction is more precise and the result (only fondness) is more accurate. Context phrases with less but frequent emotion types extracted are considered as more straightforward and unambiguous. Context phrases with numerous but less frequent emotion types are considered as more ambiguous.

Assessment and Verification of Contextual Appropriateness of Emotion

The final step is to use the two kinds of information (affect analysis and Web mining) in CAEV [1] procedure for assessing and verifying contextual appropriateness of the expressed emotions. The information obtained by affect analysis systems and the Web mining technique described above is combined as follows. When ML-Ask discovers an emotive utterance and the emotion types are successfully specified, the Web mining technique begins the process of verification of whether the expressed emotions are appropriate to be used in such context. The emotions commonly associated with the context of the utterance are extracted from the Web and the emotions reported by ML-Ask from the user utterances are compared to the Web mining results. If an emotion type detected by ML-Ask appears on the list of emotions frequently associated with the context (approx. 50% of the whole extracted list), the emotion expressed by the user is determined to be appropriate for the context. In such situations, a conversational agent equipped with this system could choose a dialog strategy that **sympathizes** with the user. Two hypothetical examples are presented 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!

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.

[1] Abbreviation for "Contextual Appropriateness of Emotions Verification".

However, if the verification procedure indicates that the expressed emotion is inappropriate for the context, the agent could undertake different measures, e.g., **helping user manage his/her emotions**, for example by changing the focus of the conversation from the object of emotion to the expressed emotion itself, or proposing an appropriate emotion. Two hypothetical examples are shown below.

Positive-inappropriate emotion:
User: I'd be happy if that bastard was hit by a car!
`ML-Ask: joy; Web mining: fear, sadness`
Agent: Are you sure that is what you are really feeling?

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!

Two-dimensional Model of Affect in CAEV Procedure
According to Solomon [17], people sometimes misunderstand the specific emotion types, but they rarely misunderstand their generally perceived valence. One could, e.g., confuse anger with irritation, but not admiration with detestation. Therefore, I checked if at least the general features matched even when specific emotion types did not match perfectly with the emotive associations. By general features I refer to those proposed by Russell [33] in the theory of the two-dimensional model of affect (valence and activation). Using the mapping of Nakamura's emotion types on Russell's model I checked whether the emotion types tagged by ML-Ask and CAO belonged to the same space, even if they did not perfectly match the emotive associations gathered from the Web.

Evaluation Experiment

Experiment Description

To test the method, I performed an evaluation experiment on two non-task-oriented conversational agents. The first agent, Modalin [38], is a simple conversational agent which generates responses by 1) using Web-mining to gather associations to the content of user utterance; 2) making propositions by inputting the associations to the prepared templates; and 3) adding modality to the basic propositions to make the utterance more natural. The second agent, Pundalin [21], based on the first one, generates a humorous response to user utterance every third turn. The humorous response is a pun created by using user input as a seed to gather pun candidates from the Web and inputting the most frequent pun candidates into pun templates. The choice of the agents was deliberate. They differed only in one feature - the humorous responses in the latter one. Humor processing is considered to be one of the most creative human activity and therefore difficult task in Artificial Intelligence [20]. Therefore if appropriateness verification is done correctly, it should be easier to perform on the non-humor-equipped agent.

There were 13 participants in the experiment, 11 males and 2 females. All of them were university undergraduate students. The users were asked to perform a 10-turn conversation with both agents. No topic restrictions were made, so that the conversation could be as free and human-like as possible.

In the experiment I used the chat logs of users with Modalin and Pundalin. All conversations were analyzed by ML-Ask and CAO. For the conversations which contained emotional expressions the Web mining procedure was carried out to determine whether the emotions expressed by the user were contextually appropriate. I compared four versions of the method: 1) ML-Ask and Web mining baseline; 2) ML-Ask supported with CVS, Web mining baseline; 3) ML-Ask baseline and blog mining; 4) both improvements (affect analysis supported with CVS and blog mining). Then a questionnaire was designed to evaluate how close the results were to human thinking. One questionnaire set consisted of one conversation record and questions inquiring what were: 1) the valence (answer to the question: "Were the emotions expressed in this sentence positive or negative?" [choice of three: POS/NEG/Don't know]) and 2) the specific type of emotions conveyed in the conversation (answer to the question: "What was the specific emotion expressed in this sentence?" [free choice]), and 3) whether they were contextually appropriate (answer to the question: "Were the emotions appropriate for the context of the conversation? [choice of three: YES/NO/Don't know] If not, which emotion would be appropriate?" [free choice]). Every questionnaire set was filled by 10 people (undergraduate students, but different from the users who performed the conversations with the agents).

For every conversation set I calculated how many of the human evaluators confirmed the system's results. The evaluated items were: A) specific emotion types determination; and B) general valence determination accuracies of affect analysis systems; and the accuracy of the method as a whole (affect analysis verified by Web mining) to determine the contextual appropriateness of C) specific emotion types and D) valence.

Then I checked how many people agreed with each other and with the results given by the system. Since every questionnaire set was evaluated 10 times, a number of agreements for each evaluated item (A-D) in all twenty evaluated cases could be from 0 (nobody agreed with the system) to 10 (all people agreed with the system). When comparing the four versions of the

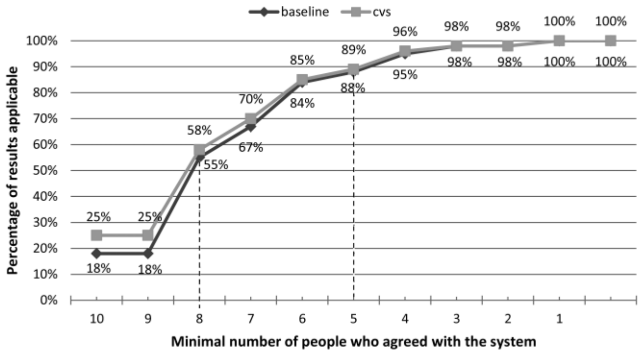

Fig. 4 Visualization of the distribution of agreements for both versions of affect analysis procedure in determining about emotion types.

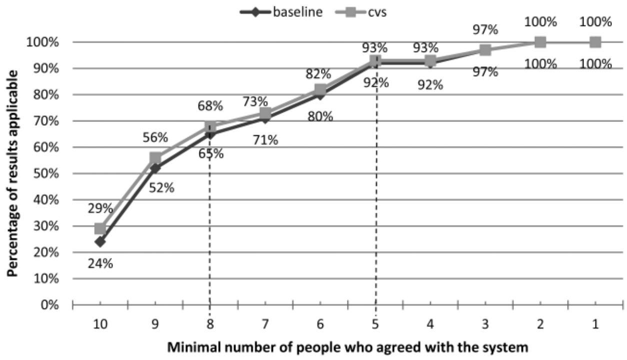

Fig. 5 Visualization of the distribution of agreements for both versions of affect analysis procedure in determining about emotion valence.

method described above, I assumed the better version of the system is the one which achieved more agreements with a larger number of evaluators. I took into the consideration all types of agreement conditions, from ideal (all 10 people agreed) to the smallest one (at least one person agreed with the system). I focused the most on two of the ten agreement conditions. Firstly, since the idea of appropriateness is based on the rule of thumb, I checked how much of the cases could be considered as positive examples if at least half of the people (five) agreed (medium condition). Secondly, as a more severe condition, I checked how much of the cases could be considered as correct examples if eight and more people (80%) agreed with the system (grand majority). Moreover, to verify the strength of the agreements I independently calculated Fleiss' multi-rater kappa [31] between all human evaluators for all sets of the results.

Results

Evaluation of Affect Analysis Procedure

The first part of the evaluation process consisted in evaluation of affect analysis procedure. The results were as follows. For all possible agreements of the system with the evaluators about all evaluation sets (200 possible cases of agreement) baseline version of affect analysis obtained 110 (55%) agreements about determining emotion type and 126 (63%) agreements about determining valence. The strength of agreements in this setting was κ=0.66 and κ=0.68, respectively, which indicates substantial agreement for both sets according to Landis and Koch interpretation [32]. As for the affect analysis procedure upgraded with CVS (ML-Ask last part of the procedure), there were 120 agreements (60%) for emotion types and 138 (68%) for valence determination. The strength of these sets of agreements was κ=0.66 and κ=0.68, respectively. As for the distribution of the agreements, most of the results for emotions types (over 50% of all actual agreements) were enclosed in a group where at least 8 people agreed with the system (grand majority conditions). Similarly, most of the results for valence were enclosed in a group where at least 9 people agreed with the system (nearly ideal, rigorous conditions). Ideal conditions (all agree) appeared from 18% to 29% of cases. However, the grand majority of the results (over 80%) were enclosed in a group where at least 6 people agreed with the system. The conditions including medium (at least 5 people) and more relaxed conditions enclosed from nearly 90% and above. The results for both versions of affect analysis procedure are represented on Figure 4 (for emotion type determination) and Figure 5 (for valence determination).

Evaluation of CAEV Procedure

Secondly I checked the results for the determination of emotion appropriateness by the CAEV procedure. The results were as follows. For all possible agreements of the system with the evaluators about the evaluation sets (200 possible cases of agreement) baseline version of CAEV procedure obtained 69 (35%) agreements about determining the appropriateness of emotion type and 85 (43%) agreements about determining the appropriateness of valence. The strength of agreements in this setting was κ=0.652 and κ=0.677, respectively, which indicates substantial strength of agreements. As for the version of CAEV procedure with affect analysis upgraded with CVS, there were 78 agreements (39%) for emotion type- and 94 (47%) for valence based appropriateness determination. The strength of these sets of agreements was κ=0.642 and κ=0.667, respectively. As for the version of CAEV procedure with Web mining restricted to blogs, there were 81 agreements (41%) for emotion type and 95 (48%) for valence-based appropriateness determination. The strength of these sets of agreements was κ=0.667 and κ=0.643, respectively. Finally, for the version of CAEV procedure with both improvements (ML-Ask upgraded with CVS and Web mining restricted to blogs), there were 90 agreements (45%) for emotion type- and 104 (52%) for valence-based determination. The strength of these sets of agreements was κ=0.643

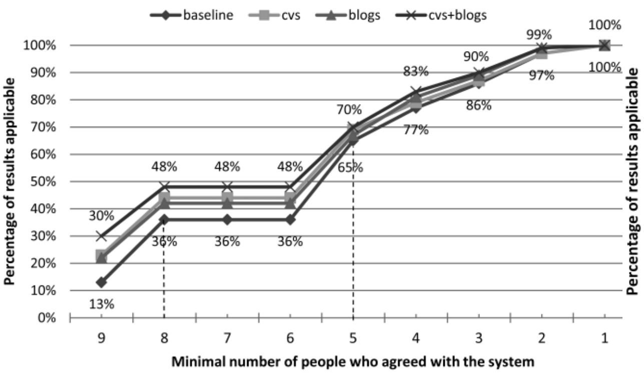

Fig. 6 Visualization of percentage of results encapsulated for each condition, from "at least 9" to "at least 1" (for emotion type determination).

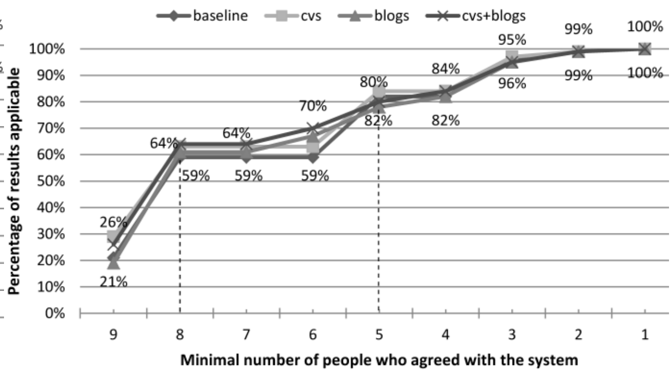

Fig. 7 Visualization of percentage of results encapsulated for each condition, from "at least 9" to "at least 1" (for valence determination).

and κ=0.633, respectively.

As for the distribution of the agreements, the majority of the results (over 50% of all actual agreements) for determining about appropriateness of emotion types were enclosed in a group where at least 5 people agreed with the system (medium conditions). For determining about valence appropriateness, most of the results were enclosed in a group where at least 8 people agreed (grand majority conditions). The results which enclosed at least 80% of agreements oscillated for emotion types verification around groups where at least three to four people agreed. For valence verification the groups enclosing at least 80% of agreements oscillated from at least 4 to at least 5. Although there were no cases with ideal conditions, the best version of the system (both improvements) encapsulated with the use of grand majority condition (at least 8 people) 48% of results for emotion types and 64% for valence.

The results are represented Figure 6 (for emotion type determination) and Figure 7 (for valence determination). Some of the successful examples are represented in Table 6.

Both improvements, the one with CVS procedure and the one limiting the query scope in the Web mining procedure to search only through blog contents, positively influenced the performance of the Contextual Appropriateness of Emotion Verification procedure, in all of the cases for both of the agents. The improvement was noticeable both on the level of specific emotion types and of valence, and also for the result of both agents taken together as well as separately.

The most effective version of the method was the one with both improvements applied, by which the system's performance (number of agreements with evaluators) was generally improved for all considered cases. For example, for the grand majority condition (at least 8 people agreed) the results were improved from 36% to 48% (emotion types) and from 59% to 64% (valence), with the highest score achieved in conversations with Modalin (75% under the "grand majority" condition).

The results were generally better in Modalin. This confirms the assumption I made earlier. I assumed that since humor is one of the most creative human activities, the appropriateness verification should be more difficult to perform in humorous conversations and easier in non-humorous conversations.

In almost all cases the results which changed after the improvements were statistically significant on a 5% level. The only version in which the change of the results was not significant was the baseline method compared to only CVS improvement (P value = 0.1599). Improving the system with blog mining,

Table 6 Three examples of the results provided by the emotion appropriateness verification procedure (CAVP) with a separate display of the examples showing the improvement of the procedure after applying CVS.

Part of conversation in Japanese (English translation)	**ML-Ask output**	**Web mining**	**CAEV**
USER: *Konpyūta wa omoshiroi desu ne.* (Computers are so interesting!)	**positive** [joy]	**positive** [joy]	**appropriate**
SYSTEM: *Sore wa oishii desu ka.* (Is it tasty?) [about instant noodles]	x	x	x
USER: *Oishii kedo, ore wa akita kana.* (Its tasty, but I've grown tired of it.)	**negative** [dislike]	**negative** [dislike]	**appropriate**

Part of conversation in Japanese (English translation)	**ML-Ask baseline**	**ML-Ask +CVS**	**Web Mining**	**CAEV**
SYSTEM: *Sore wa omoshiroi tte* (It's so funny!) [about conversation]	x	x	x	x
USER: *Sore hodo omoshiroku mo nakatta yo.* (It wasn't that funny.)	positive [joy]	**negative** [dislike]	**negative** [fear], [sad]	**appropriate**

when compared to both - baseline version of the method and with CVS, were statistically significant (P value = 0.0274) and, what is the most important, the results of the version fully improved were the most significant of all (P value = 0.0119).

Implications for Future Research

Computational Conscience
Expressing and understanding emotions is one of the most important cognitive human behaviors present in everyday communication. In particular, Salovey and Mayer [40] showed that emotions are a vital part of human intelligence, and Schwarz [41] showed, that emotional states influence the decision making process in humans. If the process of decision making is defined as distinguishing between good and bad, or appropriate and inappropriate, the emotions appear as an influential part of human conscience. The thesis that emotions strongly influence the development of human conscience was proved by Thompson and colleagues [42] who showed, that children acquire the conscience by learning the emotional patterns from other people. The significance of the society was pointed out also by Rzepka et al. [39], who defined the Internet, being a collection of other people's ideas and experiences, as an approximation of general human common sense. Since conscience can be also defined as a part of common sense (moral common sense in particular), this statement can be expanded further to saying that the Web can also be used to determine human conscience. The need for research in this matter, was pointed out inter alia by Rzepka et al. [70], who raised the matter not of creating an artificial human being, as it is popularly ventured in Artificial Intelligence research, but rather an intelligent agent in the form of a toy or a companion, designed to support humans in everyday life. To perform that, the agent needs to be equipped, not only in procedures for recognizing phenomena concerning the user, in which emotions play a great role, but it also needs to be equipped with evaluative procedures distinguishing about whether the phenomena are appropriate or not for a situation the user is in. This is an up to date matter in fields such as Roboethics [37], Human Aspects in Ambient Intelligence [36], and in Artificial Intelligence in general. In my research I performed that by verifying emotions expressed by the user with a Web mining technique for gathering an emotional common sense, which could be also defined as an approximated vector of conscience. I understand that the idea of conscience is far more sophisticated, however, when defined narrowly as the ability to distinguish between what is appropriate and what is inappropriate, my method for verifying contextual appropriateness of emotions could be applied to obtain simplified conscience calculus for machines. I plan to develop further this idea and introduce it as a complementary algorithm for the novel research on discovering morality level in text utterances presented by Rzepka and colleagues [34, 35].

Irony and Sarcasm Processing
As for recent developments regarding the idea of emotion appropriateness, the detailed analysis of specific examples has brought us to the following further observations. Some emotionally loaded sentences in which the emotions expressed are inappropriate to the context can be interpreted as ironical or sarcastic. The irony effect comes from the fact that the part of the sentence which introduces the context is connoted with an opposite emotion type one would expect. For example, a sentence such as below can be interpreted as introducing an irony effect.

(1) *I feel so sorry for you winning the Golden Globe.*

This is due to the fact that in humans natural linguistic processing of the context phrase "winning a Golden Globe" comes with different emotional load (happiness, satisfaction, etc.) than the expression that appears, "to feel sorry for someone", which is associated with sadness, or pity. This incongruence forces an additional inexplicit interpretation, which leads to the ironical effect. Similar effect can be obtained with emoticons. Example (2) shows this kind of effect.

(2) *Yeah, I feel so happy for you winning the Golden Globe.* (-_-;)

The incongruence here does not appear between context phrase and emotional expression, but between the whole sentence (positive attitude) and the emoticon added on the end of the sentence (expressing negative attitude, such as boredom, etc.). The irony effect will not be forced if on the end of the sentence an emoticon expressing positive attitude will be added (example (3)) or no emoticon will be added (example (4)).

(3) *Yeah, I feel so happy for you winning the Golden Globe.* ヽ(^o^)ノ

(4) *Yeah, I feel so happy for you winning the Golden Globe.*

The above observations indicate that the idea of emotion appropriateness could contain a potential to help explain irony, which remains an under-explained phenomenon in linguistic research and an almost untouched topic in Computational Linguistics and Artificial Intelligence research.

Conclusions

In this paper I presented my work on Contextual Affect Analysis (CAA) with a future goal of its application in conversational agents such as agent companions or artificial tutors. I applied two systems for affect analysis and a Web mining technique to develop a method for estimating contextual appropriateness of emotions. The method is composed of two parts, a language based affect analysis procedure utilizing the two affect analysis systems (used as an emotion detector), and a Web mining technique for extracting from the Internet lists of emotional associations considered as a generalized emotive common sense (used as an emotion verifier). I checked the performance of four versions of the method. The affect analysis procedure is compared with and without Contextual Valence Shifters. As for the Web mining technique, two versions are compared: one, using all of the Internet resources and the second one improved by restricting the search scope to the contents of blog documents. The improvements positively influenced the results and were statistically significant. I also observed that emotion appropriateness was more difficult to determine in conversations containing humorous responses (puns).

The method provides the conversational agent with computable means to determine whether emotions expressed by a user are appropriate for the context they appear in. A conversational agent equipped with this method could be provided with hints about what communication strategy would be the most desirable at a certain moment. For example, a conversational agent could choose to either sympathize with the user or take precautions and help the user manage his/her emotions.

As for the future work, I plan to focus on deepening the understanding of emotions by bootstrapping the context phrases. For example, in a sentence “I’m so depressed since my girlfriend left me...” the context phrase would be “my girlfriend left me”. The Web mining procedure provides for such phrases a list of appropriate emotions. However, using similar Web mining procedure I plan to go further and find out the reason for an emotion object to happen. For example, to find out “why girls leave their boyfriends?”. An answer for this question, found in the Internet, could be, e.g., “because boys are not sporty enough”, or “because boys have no money”. Next asked question could be, e.g., “why boys have no money?”, etc. Sufficient accuracy in such bootstrapping method would provide a deeper knowledge about the causality of experiences. When applied in artificial tutor or a companion agent in general, this would help providing hints about predictable undesirable consequences of user activities.

Acknowledgements

This research has been supported by: a Research Grant from the Nissan Science Foundation (years 2009-2010), The GCOE Program founded by Japan’s Ministry of Education, Culture, Sports, Science and Technology (years 2009-2010), (JSPS) KAKENHI Grant-in-Aid for JSPS Fellows (Project Number: 22-00358) (years 2010-2012), and (JSPS) KAKENHI Grant-in-Aid for Scientific Research (Project Number: 24600001) (years 2012-present).

References

1. Alm, C. O., Roth, D. & Sproat. R. (2005). Emotions from text: machine learning for text based emotion prediction. In Proc. of HLT/EMNLP, (pp. 579-586).
2. Aman, S. & Szpakowicz, S. (2007). Identifying Expressions of Emotion in Text. In Proceedings of the 10th International Conference on Text, Speech, and Dialogue (TSD-2007), LNCS 4629, (pp. 196-205), Berlin Heidelberg, Springer-Verlag.
3. Tokuhisa, R., Inui, K. & Matsumoto, Y. (2008). Emotion Classification Using Massive Examples Extracted from the Web. In Proceedings of Coling 2008, (pp. 881-888).
4. Shi, W., Rzepka, R. & Araki, K. (2008). Emotive Information Discovery from User Textual Input Using Causal Associations from the Internet [In Japanese]. In Proceedings of FIT2008, (pp. 267-268).
5. Ptaszynski, M., Dybala, P., Shi, W., Rzepka, R. & Araki, K. (2009). A System for Affect Analysis of Utterances in Japanese Supported with Web Mining. Journal of Japan Society for Fuzzy Theory and Intelligent Informatics, 21(2), 30-49.
6. Ptaszynski, M., Dybala, P., Shi, W., Rzepka, R. & Araki, K. (2009). Towards Context Aware Emotional Intelligence in Machines: Computing Contextual Appropriateness of Affective States’, In Proceedings of Twenty-first International Joint Conference on Artificial Intelligence (IJCAI-09), (pp. 1469-1474).
7. Ptaszynski, M., Dybala, P., Shi, W., Rzepka, R. & Araki, K. (2009). Shifting Valence Helps Verify Contextual Appropriateness of Emotions. The IJCAI-09Workshop on Automated Reasoning about Context and Ontology Evolution (ARCOE-09), In Working Notes of Twenty-first International Joint Conference on Artificial Intelligence (IJCAI-09), (pp. 19-21).
8. Ptaszynski, M., Dybala, P., Shi, W., Rzepka, R. & Araki, K. (2009). Conscience of Blogs: Verifying Contextual Appropriateness of Emotions Basing on Blog Contents. In Proceedings of The Fourth International Conference on Computational Intelligence (CI 2009), (pp. 1-6).
9. Ptaszynski, M., Dybala, P., Shi, W., Rzepka, R. & Araki, K. (2010). Contextual Affect Analysis: A System for Verification of Emotion Appropriateness Supported with Contextual Valence Shifters, International Journal of Biometrics, 2(2), 134-154.
10. Zhang, L. (2010). Exploration of Metaphorical and Contextual Affect Sensing in a Virtual Improvisational

Drama. Transactions on Edutainment IV, LNCS 6250, 105-116, Berlin Heidelberg, Springer-Verlag.
11. Zhang, L. (2010). Exploitation of Contextual Affect-Sensing and Dynamic Relationship Interpretation. ACM Computers in Entertainment, 8(3), Article No. 18.
12. Zhang, L. (2011). Affect Sensing and Contextual Affect Modeling from Improvisational Interaction. International Journal of Computational Linguistics. 1(4), 45-60.
13. Zhang, L. (2011). Exploitation in Context-sensitive Affect Sensing from Improvisational Interaction. Transactions on Edutainment 8, LNCS 6872, p.12-23.
14. Zhang, L. (2011). Affect Sensing in Metaphorical Phenomena and Dramatic Interaction Context. In Proceedings of the Twenty-second International Joint Conference on Artificial Intelligence (IJCAI). (pp. 1903-1908).
15. Picard, R. W. (1997). Affective Computing. Cambridge: The MIT Press.
16. Rzepka, R. & Araki, K. (2007). Consciousness of Crowds - The Internet As a Knowledge Source of Human's Conscious Behavior and Machine Self-Understanding. In Proceedings of AAAI Fall Symposium on AI and Consciousness: Theoretical Foundations and Current Approaches, (pp. 127-128).
17. Solomon, R. C. (1993). The Passions: Emotions and the Meaning of Life. Indianapolis: Hackett Publ.
18. Nakamura, A. (1993). Kanjō hyōgen jiten [Dictionary of Emotive Expressions] (in Japanese), Tokyo: Tokyodo Publishing.
19. Ptaszynski, M., Dybala, P., Rzepka, R. & Araki, K. (2009). Affecting Corpora: Experiments with Automatic Affect Annotation System - A Case Study of the 2channel Forum -, In Proceedings of The Conference of the Pacific Association for Computational Linguistics 2009 (PACLING-09), (pp. 223-228).
20. Boden, M. A. (1998). Creativity and artificial intelligence. Artificial Intelligence, 103(1-2), 347-356.
21. Dybala, P., Ptaszynski, M., Higuchi, S., Rzepka, R. & Araki, K. (2008). Humor Prevails! -Implementing a Joke Generator into a Conversational System. In Proceedings of Proceedings of the 21st Australasian Joint Conference on Artificial Intelligence, LNAI 5360, (pp. 214-225).
22. Ptaszynski, M., Dybala, P., Shi, W., Rzepka, R. & Araki, K. (2008). Disentangling emotions from the Web. Internet in the service of affect analysis. In Proceedings of KEAS'08, (pp. 51-56).
23. Wiebe, J., Wilson, T. & Cardie, C. (2005). Annotating expressions of opinions and emotions in language. Language Resources and Evaluation, 39(2-3), 165-210.
24. Zaenen, A. & Polanyi, L. (2006). Contextual Valence Shifters. In J. G. Shanahan, Y. Qu, J. Wiebe (Eds.), Computing Attitude and Affect in Text, (pp. 1-10). Dordrecht, The Netherlands: Springer Verlag.
25. Kudo, T. (2001). MeCab: Yet Another Part-of-Speech and Morphological Analyzer, Retrieved from http://mecab.sourceforge.net/
26. Ptaszynski, M., Maciejewski, J., Dybala, P., Rzepka, R. & Araki, K. (2010). CAO: Fully Automatic Emoticon Analysis System, In Proc. of the 24th AAAI Conference on Artificial Intelligence (AAAI-10), (pp. 1026-1032).
27. Ptaszynski, M., Dybala, P., Shi, W., Rzepka, R. & Araki, K. (2009). A System for Affect Analysis of Utterances in Japanese Supported with Web Mining, Journal of Japan Society for Fuzzy Theory and Intelligent Informatics, 21(2), 30-49.
28. Birdwhistell, R. L. (1952). Introduction to kinesics: an annotation system for analysis of body motion and gesture, Louisville, Kentucky: University of Louisville.
29. Birdwhistell, R. L. (1970). Kinesics and Context, Philadelphia: University of Pennsylvania Press.
30. Wilson, T. & Wiebe, J. (2005). Annotating Attributions and Private States. In Proceedings ofthe ACL Workshop on Frontiers in Corpus Annotation II, (pp. 53-60).
31. Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. Psychological Bulletin, 76(5), 378-382.
32. Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. Biometrics, 33, 159-174.
33. Russell, J. A. (1980). A circumplex model of affect. J. of Personality and Social Psychology, 39(6), 1161-1178.
34. Rzepka, R., Masui, F. & Araki. K. (2009). The First Challenge to Discover Morality Level In Text Utterances by Using Web Resources, In Proceedings of The 23rd Annual Conference of the Japanese Society for Artificial Intelligence, (Article no.: 2L1-1).
35. Komuda, R., Ptaszynski, M., Momouchi, Y., Rzepka, R. & Araki, K. (2010). Machine Moral Development: Moral Reasoning Agent Based on Wisdom of Web-Crowd and Emotions, International Journal of Computational Linguistics Research, 1(3), 155-163.
36. Treur, J. (2007). On Human Aspects in Ambient Intelligence, In Proceedings of The First International Workshop on Human Aspects in Ambient Intelligence, (pp. 5-10).
37. Yamashita, Y. (1999). Kara, Node, Te-Conjunctions which express cause or reason in Japanese (in Japanese), Journal of the International Student Center, 3, 1-14, Hokkaido University.
38. Higuchi, S., Rzepka, R. & Araki, K. (2008). A Casual Conversation System Using Modality and Word Associations Retrieved from the Web. In Proceedings of EMNLP 2008, (pp. 382-390).
39. Rzepka, R., Ge, Y. & Araki, K. (2006). Common Sense from the Web? Naturalness of Everyday Knowledge Retrieved from WWW. Journal of Advanced Computational Intelligence and Intelligent Informatics, 10(6), 868-875.
40. Salovey, P. & Mayer, J. D. (1990). Emotional intelligence, Imagination, Cognition, and Personality, 9, 185-211.
41. Schwarz, N. (2000). Emotion, cognition, and decision making. Cognition & Emotion, 14(4), 433-440.
42. Thompson, R. A., Laible, D. J.&Ontai, L. L. (2003). Early understandings of emotion, morality and self: Developing a working model, Advances in child development and behavior, 31, 137-171.

【自己紹介・略歴】

2010 年９月　　北海道大学大学院情報科学研究科 博士課程修了（博士学位取得）

2010 年１０月　北海道大学情報科学研究科において特任研究員

2010 年１１月　北海学園大学ハイテク・リサーチ･センターにおいて日本学術振興会外国人特別研究員

2012 年１１月　北見工業大学情報システム工学科において非常勤研究員

【主な研究業績】

[1] **"Towards Computational Fronesis: Verifying Contextual Appropriateness of Emotions"**, Michal Ptaszynski, Michal Mazur, Pawel Dybala, Rafal Rzepka, Kenji Araki and Yoshio Momouchi, *International Journal of Distance Education Technologies*, Special Issue on Emotional Intelligence for Online Learning, 2012. (to appear)

[2] **"Affect Analysis in Context of Characters of Narratives"**, Michal Ptaszynski, Hiroaki Dokoshi, Satoshi Oyama, Rafal Rzepka, Masahito Kurihara, Kenji Araki and Yoshio Momouchi, *Expert Systems With Applications*, Vol. 40, Issue 1 (2013). p. 168-176, Elsevier, 2013. (**Impact factor 2.455**) doi: 10.1016/j.eswa.2012.07.025 (**Impact factor 2.455**)

[3] **"Part-of-Speech Tagger for Ainu Language Based on Higher Order Hidden Markov Model"** Michal Ptaszynski and Yoshio Momouchi, *Expert Systems With Applications*, Vol. 39, Issue 14 (2012), pp. 11576–11582, Elsevier, 2012. doi: 10.1016/j.eswa.2012.04.031 (**Impact factor 2.455**)

[4] **"In the Service of Online Order: Tackling Cyber-Bullying with Machine Learning and Affect Analysis"** Michal Ptaszynski, Pawel Dybala, Tatsuaki Matsuba, Fumito Masui, Rafal Rzepka, Kenji Araki, and Yoshio Momouchi, *International Journal of Computational Linguistics Research*, Vol. 1, Issue 3, pp. 135-154, 2010.

[5] **"CAO: A Fully Automatic Emoticon Analysis System Based on Theory of Kinesics"** Michal Ptaszynski, Jacek Maciejewski, Pawel Dybala, Rafal Rzepka and Kenji Araki, *IEEE Transactions on Affective Computing*, vol. 1, no. 1, pp. 46-59, Jan.-June 2010 **[2011 IEEE Sapporo Section Encouragement Award]**

[6] **"An Automatic Evaluation Method for Conversational Agents Based on Affect-as-Information Theory"** Michal Ptaszynski, Pawel Dybala, Rafal Rzepka and Kenji Araki, *Journal of Japan Society for Fuzzy Theory and Intelligent Informatics*, Special Issue on Emotions, 知能と情報(日本知能情報ファジィ学会誌)特集「感情」, Vol. 22, No. 1 (February), pp. 73-89, 2010.

[7] **"Contextual Affect Analysis: A System for Verification of Emotion Appropriateness Supported with Contextual Valence Shifters"** Michal Ptaszynski, Pawel Dybala, Wenhan Shi, Rafal Rzepka and Kenji Araki, *International Journal of Biometrics*, Vol. 2, No. 2, 2010, pp. 134-154, Inderscience Enterprises.

[8] **"A System for Affect Analysis of Utterances in Japanese Supported with Web Mining"** Michal Ptaszynski, Pawel Dybala, Wenhan Shi, Rafal Rzepka and Kenji Araki, *Journal of Japan Society for Fuzzy Theory and Intelligent Informatics*, Special Issue on Kansei Retrieval, 知能と情報(日本知能情報ファジィ学会誌)特集「感性検索」, Vol. 21, No. 2 (April), pp. 30-49 (194-213), 2009.

【（GCOE で行った研究内容）】

北海道大学の博士後期課程中，特に GCOE では，自然言語処理分野の視点から見た感情表現をコンピュータに理解させ，それを基にユーザの感情状態を推定する研究を行ってきた．人工知能における感情（喜怒哀楽など）の研究（Affective Computing）では顔の表情や音声変動から感情認知を行う試みが多い．一方，言語の感情表現は社会的かかわりを表現しているが，その研究はまだ初期段階である．そこでまず文章内の感情解析システム ML-Ask の開発を行った．ML-Ask では，ユーザの入力文を手作業で収集した感情要素・感情表現のデータベースに照らし，順番にマッチングを行う．感情要素がマッチングできた文では感情的コンテキストが決定される．さらに感情表現のデータベースを参照し，抽出された感情を話者の感情状態とする．ML-Ask は大規模なテキストデータに感情タグ付けを自動的に付与することができる．しかし，日本語を豊富に含む大規模テキストデータであるインターネット（ブログ，チャットルーム，掲示板など）には顔文字など一般の辞書に存在しない表現が頻繁に使用されている．その処理を行うために顔文字解析システム CAO の構築を行った．CAO システムは入力文から顔文字を抽出し，それらが表す感情を推定する．まずは，インターネットから 1 万以上の顔文字を抽出し，自動的に感情のグループ分けを行った．さらに，顔文字を「口」や「目」などを表す部分に自動的に分け，システムのカバレッジ（顔文字の組合せ数）を 3 百万以上に拡大した．CAO システムの性能は 97%を超えた．これらのツールを利用しインターネット上で感情表現や感情文の働きに関する研究をさらに進めた．Web マイニング手法を用いて，認知した話者の感情状態が会話の場面に合っているかどうかを計算する手法を提案した．本手法では，文中の感情の種類・極性を判断した後，その文に出現した感情の原因フレーズをインターネットで検索し，それと頻繁に出現する感情表現を ML-Ask と CAO の結果と照合し，一致した場合に文中の感情が文脈に適していると判断する．これらのシステムおよび手法を対話エージェントに応用することで，ユーザの感情が理解でき適宜反応ができるロボットの開発に貢献ができると考えられる．