Just Keep Tweeting, Dear: Web-Mining Methods for Helping a Social Robot Understand User Needs

Keisuke Takagi, Rafal Rzepka and Kenji Araki

Graduate School of Information Science and Technology
Hokkaido University, Sapporo, Japan
{eyu.8055,kabura,araki}@media.eng.hokudai.ac.jp

Abstract

An intelligent system of the future should make its user feel comfortable, which is impossible without understanding context they coexist in. However, our past research did not treat language information as a part of the context a robot works in, and data about reasons why the user had made his decisions was not obtained. Therefore, we decided to utilize the Web as a knowledge source to discover context information that could suggest a robot's behavior when it acquires verbal information from its user or users. By comparing user utterances (blogs, Twitter or Facebook entries, not direct orders) with other people's written experiences (mostly blogs), a system can judge whether it is a situation in which the robot can perform or improve its performance. In this paper we introduce several methods that can be applied to a simple floor-cleaning robot. We describe basic experiments showing that text processing is helpful when dealing with multiple users who are not willing to give rich feedback. For example, we describe a method for finding usual reasons for cleaning on the Web by using Okapi BM25 to extract feature words from sentences retrieved by the query word "cleaning". Then, we introduce our ideas for dealing with conflicts of interest in multiuser environments and possible methods for avoiding such conflicts by achieving better situation understanding. Also, an emotion recognizer for guessing user needs and moods and a method to calculate situation naturalness are described.

Introduction

Robotic research has been evolving rapidly for the last two decades, and it has become almost obvious that to achieve a better understanding of the real world, robots must discover it by themselves and learn to adapt to its changes. The importance of embodiment was confirmed, and visions of robots reading books in order to understand the world and people living in it were largely discarded. However, robots today still lack sensors that would let them experience rich stimuli a human does. The level of speech or image recognition is far from satisfying, which does not help robots to enter our lives more widely. We consider that it is time to combine the physical experience with written sources of the

Crowd Wisdom, such as those on the Internet, to help robots interpret meanings of and correlations among objects and actors they discover in a particular context. We believe this would also lead to developing machines that help us without being told that the help is needed. Some progress was made in the development of a guide robot that operates to find a person who potentially needs some services by understanding environment from accumulated trajectories (Satake 2009). We were interested in a multiuser environment; therefore, we proposed an algorithm where a robot learns social behavior using information on time and position of users evaluating the timing appropriateness of a cleaning task within the environment (Takagi 2010). However, the above-mentioned approaches did not use any language information as a part of the context, making the situation understanding very shallow. Most of human-robot interaction research concentrates on conversation; sophisticated robots are being used to talk to people, gesticulate, or follow faces, for example. This is obviously helpful for investigations on socialization or human reactions, but the robots tasks are rather artificial, e.g. guiding to places, delivering messages and so on. What we found challenging is to implement some communication skills that would actually allow an increase in the quality of a robots work while decreasing the user's burden. We also wanted to see how far we could go without sophisticated humanoid machines with many sensors. Therefore we chose a widely used robot vacuum cleaner, Roomba[1] and decided to test several open-source based natural language processing tools to improve the robot's work with a minimal need for communication.

Problems of The Cleaning Task

A floor-cleaning robot is too noisy to have a clear input from an on-board microphone and a camera is situated too low to have a natural angle view of a user for the face recognition. To tackle with the first problem, we set a Twitter[2] account for the vacuum-cleaner allowing a textual input and output, and for the user recognition during a direct physical input we used on-board RFID[3] reader and personal ID cards with RFID tags. The environment was the authors' laboratory

---

[1] www.irobot.com

[2] www.twitter.com

[3] www.rfid.org

and 10 experiment participants could give the cleaning robot a positive or negative feedback on timing when it is preferable or not to do the cleaning. By using SVM, the robot was able to learn when and where it is better not to clean but the lab members' willingness for cooperation was lower than expected (which made us wonder how unreal are the experiments where users are asked to interact with a robot). People busy with their tasks would rather rely on the cleaner's common sense than to give feedback every time they are pleased or frustrated. In addition, there were some cases where users' evaluation conflicted during the experiments. As such situations affect the existing human relationships within a society (Sakamoto 2005), we decided to implement a Heider's Balance Theory we describe in next sections. To solve both problems – lack of common sense about cleaning and handling users' disagreement, we need to acquire information on users' motivations. To achieve this goal we propose several NLP methods for a robotic system that can access the Web as a knowledge source and discover context information for guessing its users' possible needs and motivations usually hidden behind simple statements. Such common sense knowledge about a task can be expanded to retrieving social relations, action consequences and emotions following them. Acquired separately for a given situation with different objects and actors it would surely help to improve the speed of robot's cognition. This paper is aiming at robot researchers who are not familiar with text-based methods and who could consider implementing them into their systems.

Balance Theory

Most HRI experiments being performed concentrate on one robot interacting with one or two persons simultaneously. However, as far as we know, there is no study on methods for cases where there are several users able to send conflicting feedback or input to a single robot. Is a decision by majority the best choice in such cases? What if most users do not know some important reasons a task should be performed? What if the higher position members' choice is ignored? To tackle these problems, we decided to utilize the balance theory, which was advocated by Fritz Heider and extended by his disciples (Insko 1984). The basic theory says that the balance is kept when three people (acknowledger (P), other person (O), and object (X)) harmonize with each other. However, when three people do not harmonize, the balance collapses, and the breaking of the balance is assumed to cause stress. Therefore, the involved persons tend to try to make an effort to keep overall balance, and to avoid possible stress (see Figure 1). Moreover, Cartwright expands the balance theory in the graph theory (Mitsunaga 1956). In short, the interpersonal relationships of a certain group can be explained by expanding triadic relations. Our past research showed that there were cases where users' feedback was different during our experiment, meaning that for one person cleaning in time T and at place P is preferable, and for another, it was not. We assume that in situations like this a robot should try to convince people or be able to make a compromise while choosing its behavior. As a method, it is possible to generate a message including

elements that can persuade conflicting people, and a robot could response to the users trying to explain the reason behind its decision. Some kind of persuasive factor might be preferable. For example, factors such as majority rule, difference in social position, or a reason for urgency could be considered. In addition, it seems possible that social positions could be learned from observing trends in compromises and explanations given by users. If the success rate of the persuasion increases in a case of a particular user, it can be said that the person has a higher position in the given society or that he or she is very popular. The ideal situation, though, is when a machine correctly guesses all dependencies automatically. To realize this goal we need to combine different kinds of knowledge. The following sections introduce possibilities arising from using Web text mining.
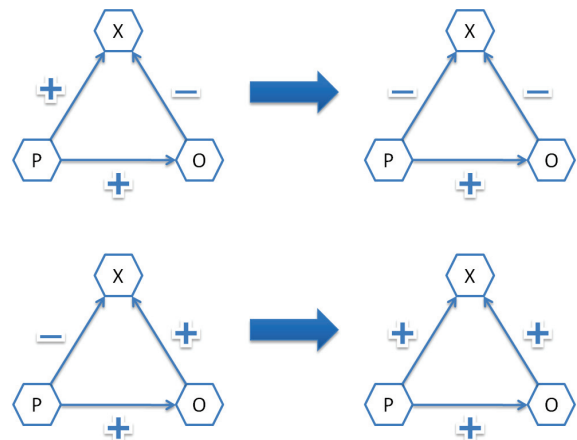

Figure 1: P: Person without given priority, O: Person who should be given priority, X: Robot. First situation: Robot performs a positive action for a person without priority and P feels dissatisfied. Second situation: Robot performs a positive action for a person who should be given priority. P is convinced.

Survey on Robot Priorities To use the balance theory in practice, the robot must be able to persuade people with lower priority. To test what factors potential users regard as more important, we performed a simple survey on six situations the social robot could face. The online questionnaire we set stated: *Two people or groups (A and B) live with a robot within one living space. Sometimes A and B give conflicting orders to the robot. Which do you think should be given priority to in the following cases?*
C1) A: Person of higher social rank (e.g. manager) gives an order to the robot [HiRank]

B: Person of lower social rank (e.g. employee) gives an order to the robot [LoRank]
C2) A: High emergency level order is given (e.g. somebody is coming soon, clean the flat / a baby is sleeping, don't clean the flat) [HiEmer]

B: Low emergency level order is given (e.g. somebody is coming tomorrow, clean the flat, I am reading, don't clean the flat) [LoEmer]

C3) A: Majority of users demand an action [Majority]

    B: Minority of users demand an action [Minority]

C4) A: Person of higher social rank asks for a low emergency task [HiRank+LoEme]

    B: Person of lower social rank asks for a high emergency task [LoRank+HiEmer]

C5) A: Bigger group of lower social rank requests a task [LoRank Major]

    B: Smaller group of higher social rank requests a task [HiRank Minor]

C6) A: Bigger group requests low emergency task [Major+LoEmer]

    B: Smaller group requests high emergency task [Minor+HiEmer]

Thirty people participated in the experiment by grading their answers to the above questions on a seven-point semantic differential scale (see Table 1). The answers to cases C1-3 show that high emergency level is the most important factor for giving priority. A significant difference was confirmed between C2 and C1, C2 and C3 ($p<0.01$). The answers to C4-6 show that high emergency level is seen to be more important than higher social rank, and the will of the majority, by most potential users. A significant difference was confirmed between C4 and C5, and C6 and C5 ($p<0.01$). Interestingly, the answers to C4 (Person of higher social rank asks for low emergency task) showed a significant difference between Japanese and foreign respondents to the survey ($p<0.1$). It can be presumed that this is because Japanese people generally attach greater importance to social status.

Helpful Text Processing Methods

We work on Japanese language, however all the tools and methods we describe are accessible for most major natural languages. In addition to using usual Japanese WWW resources, we also collected logs of Japanese Twitter via Twitter Streaming API. From a random sample of Twitter's public timeline in real time, Japanese tweets are being distinguished by using the Perl Lingua::LanguageGuesser. Thus, the time of the tweet, user name, and the tweet are written to the file and matched to the physical data of users' actions inside the lab. An overview of our approach is shown in Figure 2.

Method 1: Task Knowledge Collection from the Web

Development of robots using SNS resources has already started. Mavridis et al (Mavridis 2009) have utilized a connection between a face-detecting conversational robot and Facebook; however, the robot's purpose is only to socialize, by remembering and exchanging information. There are also examples of tweeting by people who pretend existing machines, for instance NASA's Mars Phoenix or humanoid Robonaut 2, which shows the interest in robots talking about their work. In Japan there are several robots able to display Twitter messages, such as from InterRobot, from NEC (Paperoch Tsubuyaki-chu), but they are purely entertainment
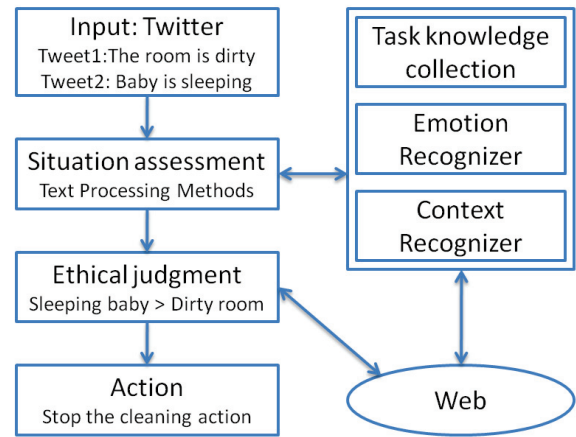

Figure 2: An overview of our approach.

Table 2: Extracted feature words

Feature word	Score of Okapi BM25
souji (cleaning)	11.0050506147475
yogore (dirt)	7.22616194426290
hokori (dust)	7.02808877680245
HOKORI (dust)	5.96823084475785
kitanai (dirty)	5.62837247255482
gomi (garbage)	5.26486776818813

devices. Awarehanger (Tajima 2010) is a gadget that actually helps its user to know the state of his or her laundry on a hanger by sending e-mail or tweeting. The authors also present their tweeting Post Box (called Mail Tweeter). The systems cannot learn or retrieve additional knowledge, for example about appropriate sending hours. The knowledge about possible reasons for a given task can be collected from the Web. By using the wildcard exact match query expression *-node souji-shita ("I cleaned because *") we have collected sentences for our vacuum cleaner explaining reasons (expressions following "because"). Then, feature words (nouns and adjectives) from these sentences are extracted by using Okapi BM25 (Robertson 1995), which is a ranking function to rank matching documents according to their relevance to a given search query. The higher the occurrence rate of the word in the sentences, and the lower the hit number of the word using Yahoo Web API, the score of Okapi BM25 rises. Table 2 shows the feature words that scored high in Okapi BM25. An extracted word is set to become the next query word. Consequently, new sentences including these feature words can be easily discovered.

Results Over 9 days, we collected 41,295 entries from Japanese Twitter as data for the experiment. 22 sentences were judged to describe a cleaning situation, and the judgment was performed when two or more of three graduate students (members of our laboratory) agreed with each other. Next, the feature words were retrieved from the log and 36 sentences were found. Table 3 shows examples of discov-

Table 1: Results of the survey on priorities (Max4A = Maximum priority for A, Prio4A = Priority for A, Low4A = Low priority for A, Hard = Hard to say, Low4B = low priority for B, Prio4B = Priority for B, Max4B = Maximum priority for B)

A side	B side	Max4A	Prio4A	Low4A	Hard	Low4B	Prio4B	Max4B
HiRank	LoRank	19.23%	42.31%	7.69%	30.77%	0.00%	0.00%	0.00%
HiEmer	LoEmer	57.69%	38.46%	0.00%	3.85%	0.00%	0.00%	0.00%
Majority	Minority	11.54%	23.08%	34.62%	30.77%	0.00%	0.00%	0.00%
HiRank+LoEmer	LoRank+HiEmer	0.00%	0.00%	3.85%	15.38%	11.54%	50.00%	19.23%
LoRank Major	HiRank Minor	7.69%	7.69%	19.23%	34.62%	11.54%	15.38%	3.85%
Major+LoEmer	Minor+HiEmer	0.00%	0.00%	0.00%	0.00%	23.08%	57.69%	19.23%

ered sentences. 15 of 36 sentences were judged as a correctly recognized "cleaning situation" by two or more of three judges. Table 4 shows examples of incorrect sentences from discovered sentences. Additionally, Table 5 shows examples of undiscovered sentences.

Error Analysis The incorrect sentences clearly show the problems one has to tackle while dealing with text. For example "I will clean the bath tub" does not apply to floor cleaning. To solve this problem it is usually enough to set up a set of objects, places and most of all actions a given robot can do to filter out the actions it cannot perform (such method will be described later). It will then prevent Roomba from assuming it could clean a bath tub. "It got so clean that I can't recognize it" describes the result of cleaning. In that case, it is necessary to consider the tense of the verb ("it got"). On the other hand, such sentences can be used by an emotion recognizer to guess if the consequence of cleaning (becoming clean or nice) is a positive or negative experience. In the sentence "It is dirty because of my habit of taking notes on the cover" the word "dirty" is used in a slightly different meaning. There were 7 undiscovered sentences, which were associated with a dirty floor, but the feature word is not included. To solve this problem, common sense concerning the concept of dirtiness" is needed. In addition to a simple Web search, the common sense ontology called ConceptNet (Havasi 2007) can be used (however, it is not well developed for other languages than English), as well as WordNet (Fellbaum 1998) for increasing or expanding search queries, for example by using synonyms, hypernyms or hyponyms.

Method 2: Emotion Recognizer and Related Algorithms

Even if we discover reasons or consequences of a given action, the robot will have problems with interpreting them. To challenge this problem, we use textual affective computing. At our lab we have developed emotion recognition algorithms for textual resources for almost 10 years; however, they have been used only for text processing. But if a robot is capable of reading entries on its user's actions and states, a whole new range of possibilities opens. As mentioned above, an autonomous robot can discover what emotions are hiding underneath humans' needs, what they are afraid of and what makes them happy. We use Nakamura's (Nakamura 1993) ten categories of emotions (joy, fondness, anger, surprise, gloom, excitement, dislike, shame/bashfulness, fear and relief) not only to determine the most common emotive associations (Shi 2008), but also to solve moral problems (Rzepka 2009), to implement a very basic algorithm for self-judging an agent's own sense of humor (Dybala 2010) or even categorize automatically recognized emoticons (Ptaszynski 2010). First, the emotion recognizer achieved about 45% accuracy, but after adding the Web-mining method, it soon increased to over 60%. However, obtaining information about the emotion expressed by a user does not tell us everything[4]. We do not know whether it is, e.g., appropriate for a given situation, and what actions could and should be undertaken as a reaction. In the next section we introduce our approach to this problem.

Method 3: Context Appropriateness Recognizer

From early childhood, a human learns various types of knowledge: about the physical world, social rules and abstract concepts. We carry an enormous common sense knowledge base within our skulls, but using it entirely is impossible even for such efficient processors as our brains. Although being bombarded with information while perceiving the world around us, we shadow out irrelevant data and focus on the situation we face. Today we know that Broca's area, a part of our brain responsible for language understanding, also plays an important role in ignoring irrelevant input (Haxby 1994). We can observe how important context fixation is when evaluating commonsense knowledge. The human judges' opinions vary depending on how rich his or her imagination or experience is. But in real life, context awareness limits the obvious possibilities to a minimum. When you observe Laika sitting inside Sputnik, your "a dog smells another dog" association becomes shadowed and "dogs help people" becomes stronger. When our robot vacuum cleaner processes the verb "to clean", it should not analyze knowledge of cleaning teeth, desktops, cars or (as the above example showed) bath tubs. We noticed that it is easier to use (and evaluate) common sense knowledge by setting contextual restraints for retrieving concepts. Therefore we simply decided to limit it to one environment, and for the first trial

---

[4]As shown in Table 6, the affect recognizer without a deeper understanding abilities count also sentences describing children excited about breaking eggs or wrongly interpreting speech acts (lack of a filter for "want to" retrieved "relief" sentences for "a lot of garbage" query in the last example')

Table 3: Discovered sentences

Original	Translation
Sate, souji demo siyo kana	Well, I probably should clean up
Hadazamui node, honjitsu no osouji oyasumi desu.	No cleaning today. It's way too cold
Heya souji shita ga nattoku dekin	I cleaned the room but I'm not satisfied
Heya ga seidai ni kitanai	The room is incredibly dirty
Mairumu kitanasa MAX	The dirtiness of my room is at the maximum
Kitto... hokori kusai	I bet it stinks dusty here
Gomi ooina soreni shitemo	Oh my, it's a lot of garbage

Table 4: Incorrect sentences

Original	Translation
Ofuro souji shite kuru	I will clean the bath tub
Souji suru to michigaeru youni kirei ni natta	It got so clean that I can't recognize it
Hyoushi ni memo suru kuse ga aru kara kitanai	It is dirty because of my habit of taking notes on the cover
Gomi sute ni ittekuru	I will go to throw away the garbage

we chose a "house" (rooms, kitchen, bathroom, etc.) as an environment, furniture and utensils as objects, and family members (plus the robot) as actors. Then we performed an experiment for the automatic discovery of uncommon behavior. We assumed that the labels are already acquired and a machine can name all significant places (we picked up 9 nouns), items (37 nouns) and actions (21 verbs). We prepared an algorithm for creating random acts" from places, objects and actions. Its basic output was "ACTION with ITEM at PLACE" and this set was sent as an exact query to the Yahoo Japan Blog search engine, together with the shorter queries action in a place" and action with a tool", to find frequencies of particular n-grams. Differences between them were used to find the most uncommon parts of an action (eating ice-creams is natural, but uncommon if eaten in a bathroom). In order to use this method with different robots it is enough to limit the verbs (functions) and nouns (objects and places) that the robot is capable of dealing with.

Recognition Experiment First, the system itself evaluated the permutations, and 100 natural and 100 unnatural outputs were shown to 3 graduate school subjects. The scoring was done by comparing frequencies of exact matches on the Web, and 1 point was given if the students agreed, 0 when they disagreed and 0.5 when it was difficult to decide. The results were that three human evaluators agreed with system's judgments in 77.08% of cases. We consider that this is the first step toward robots knowing what they and their users can/should and cannot/should not do, or when a request is probably impossible to fulfill. This knowledge can be used in the reverse way, e.g. a machine could discover common places or actions for particular objects: "bring a fork" would not cause a stream of detailed questions about its location, a robot could just ask in which drawer in the kitchen cupboard do you keep forks?". Also, speech recognizers could use the "impossible action" discovery for guessing a correct object or action, as there is a high probability that a user's input was correct but the recognition process failed.

Conclusions

This paper aimed to introduce several natural language processing methods that can help to discover users' possible needs using the combination of physical context information and Web knowledge. We use Internet resources to guess user requests before they are made, to understand context better and to collect sets of reasons and consequences for a particular behavior - not only from users' direct feedback, but also from his or her daily tweets or blog entries, or even from the whole Wisdom of Web Crowd". Such knowledge could be useful for solving conflicts of interest in environments with several users and only one robot. We described several opportunities for and the preliminary effects of proposed methods. We showed that sentences containing reasons for cleaning can be retrieved from the Web and the feature words can be extracted by Okapi BM25 scoring. The feature words were retrieved from the log of Japanese Twitter, and 15 of 36 sentences were judged as being correct situations for cleaning. We also suggested further readings, hoping to spark a fruitful discussion between web-mining and robotics researchers. We strongly believe that by tweeting or writing blogs we all are already unconsciously" creating a collective intelligence which will increasingly help robots to help us.

Future Work

As future work, we plan to combine methods described in this paper for solving user conflicts in multi-person single-robot environments. By automatic retrieval of context-dependent common sense and emotions, we hope to achieve better situation understanding, which could allow avoiding tiresome confirmations about the robot's decisions. We also intend to continue our work on each method to improve their performance.

Table 5: Undiscovered sentences

Original	Translation
Heya katazuketeta kedo, amarini atusugite chudan	Though I cleaned the room, I've stopped because it is too hot
Tamago wareta	The egg broke
Yokei ni chirakatta	It got even more cluttered
Johanshin no zenbu no kawa ga...	I have peeling skin from sunburn all over my upper body...
...hiyake de mukete kite heyaju nukegara darake	...so my cast-off skin is all over the room

Table 6: Examples of affect analysis (only the sum of top recognitions greater than 50% is considered as a correct output)

Phrase	Related Emotions
It got so clean that I can't recognize it	joy 65%
The store was much too dirty	dislike 37%, anger 37%
The atelier was a mess	anger 91%
The egg broke	surprise 23%, anger 19%, excitement 15%
It's a lot of garbage	surprize 49%, relief 24%

References

Satake, S., Kanda, T., Glas, D.F., Shiomi, M., Ishiguro, H. and Hagita, N. 2009. The guide robot that provide information efficiently by understanding environment from accumulated trajectories. In *Proceedings of Interaction09*, pp. 173-180.

Takagi, K., Rzepka, R. and Araki, K. 2010. Method for Social Behavior Acquisition by Robots Using Feedback from the Environment and Users In *Proceedings of IWMST 2010*, pp. 164-166.

Sakamoto, D. and Ono, T. 2005. Sociality of Robot: Robots Construct or Collapse the Human Relations In *IPSJ SIG Notes. ICS*, pp. 15-20.

Insko, C.A. 1984. Balance Thory, The Jordan Paradigm, and The Wiest Tetrahedron In *Advances in Experimental Social Psychology, Berkowitz, L. (Ed)*, Vol.18, pp.89-140, Academic Press.

Mavridis, N., Datta, C., Emami, S., Tanoto, A., BenAbdelkader, C. and Rabie, T. 2009. FaceBots: robots utilizing and publishing social information in facebook In *Proceedings of the 4th ACM/IEEE international conference on Human Robot Interaction*, pp.273-274.

Tajima, N., Tsukada, K. and Shiio, I. 2010. AwareHanger: Context-aware hanger to inform users of laundry states In *IPSJ SIG Technical Report*, Vol.2010-UBI-27 No.2 (in Japanese).

Mitsunaga, N., Miyashita, T., Ishiguro, H., Kogure, K. and Hagita, N. 2006. Robovie-IV: A Communication Robot Interacting with People Daily in an Office In *Proceedings of IROS 2006*, pp.5066-5072.

Cartwright, D. and Harary, F. 1956. Structural balance A generalization of Heiders theory In *Psychological Review*, Vol. 33, pp. 277-293.

Havasi, C., Speer, R. and Alonso, J. 2007. ConceptNet 3: a Flexible, Multilingual Semantic Network for Common Sense Knowledge In *Proceedings of Recent Advances in Natural Languages Processing*, pp. 277-293.

Robertson, S.E., Walker, S., Jones, S., Hancock-Beaulieu, M.M. and Gatford, M. 1995. Okapi at TREC-3 In *Proceedings of the 3rd Text Retrieval Conference*, pp. 109126.

Fellbaum, C. 1998. WordNet: An Electronic Lexical Database Cambridge, MA: MIT Press.

Nakamura, A. 1993. Kanjo hyogen jiten [Dictionary of Emotive Expressions] (in Japanese). Tokyodo.

Shi, W., Rzepka, R. and Araki, K. 2008. Emotive Information Discovery from User Textual Input Using Causal Associations from the Internet. In *Proceedings of FIT 2008*, pp.267-268 (in Japanese).

Ptaszynski, M., Maciejewski, J., Dybala, P., Rzepka, R. and Araki, K. 2010. CAO: A Fully Automatic Emoticon Analysis System Based on Theory of Kinesics. In *IEEE Transactions on Affective Computing*, Vol. 1, No. 1, January-June, pp. 46-59.

Ptaszynski, M., Maciejewski, J., Dybala, P., Rzepka, R. and Araki, K. 2010. CAO: A Fully Automatic Emoticon Analysis System Based on Theory of Kinesics. In *IEEE Transactions on Affective Computing*, Vol. 1, No. 1, January-June, pp. 46-59.

Rzepka, R., Komuda, R. and Araki, K. 2009. Bacteria Lingualis In The Knowledge Soup - A Webcrawler With Affect Recognizer For Acquiring Artificial Empathy In *Proceedings of The AAAI 2009 Fall Symposium on Biologically Inspired Cognitive Architectures (BICA-09), Washington, D.C., USA*

Haxby, J. V., Horwitz, B., Ungerleider, L. G., Maisog, J. Ma., Pietrini, P. and Grady, C. L. 1994. The functional organisation of human extrastriate cortex: a PET-rCBF study of selective attention to faces and locations. In *J. Neurosci. IR*, 6336-6353.