Straight Thinking Straight From The Net – On The Web-Based Intelligent Talking Toy Development

Rafal Rzepka, Shinsuke Higuchi, Michal Ptaszynski and Kenji Araki
Graduate School of Information Science and Technology
Hokkaido University
060-0814 Kita-ku, Kita 14 Nishi 9, Sapporo, Japan
Email: {kabura,shin_h,ptaszynski,araki}@media.eng.hokudai.ac.jp

***Abstract*—This paper introduces an early stage of a smart toy development project which combines several techniques to achieve a level of conversational skills and knowledge higher than currently available robots for children. We describe our ideas and achievements for three modules which we treat as the most important - topic unlimited talking engine, emotions recognizer and the moral behavior analyzer. We will also mention our novel evaluation method for freely speaking agents and possibilities of adding another module - an automatic joke generator.**

***Index Terms*—Intelligent systems, common sense, affect analysis, machine ethics**

I. Introduction

The Internet can help our intelligence and interaction in many ways. After a decade of using the Web for supporting our knowledge (information retrieval, web-mining, machine translation, e-learning, etc.) we can say that WWW has been used in almost every field of Natural Language Processing. The topic of the proposed session is on ”creative decision-making” and we are also interested in helping people to decide, however our standpoint is basically different from the field majority. Our goal is to make a toy, in our case it is going to be an implementation into a little and inexpensive humanoid[1] which could talk with children in a more mature manner than other talking toys as Ifbot or Kitty Robot[3] which are also rather rarely affordable by an average Japanese family.

A robot which can process language in deeper manner than greetings exchange and asking questions as ”what’s your name?” or ”how do you feel today?”. By gdeeperh we mean the way how the answers for these questions are treated and processed, what is their meaning for the interaction. Although these two questions above seem trivial, they are crucial, especially the latter, even if the first looks more important for the conversation flow. We put stress on emotion processing as we believe that the affect understanding is one of the most important factors in human behavior, learning and interaction. The way one feels is very subjective and hard to evaluate (we agree with Samsonovich[5] that modern science needs new evaluation methods for Artificial Intelligence) but we decided to tackle with this problem and use evaluator’s commonsense to recognize if the output is reasonable or not. Such methods are already being used[8], but they are not always fair, so we developed also a new evaluation method[6] which will be briefely described in this paper. We will explain what methods we use to retrieve associations from the net and how they are being used in the conversation, and then our ideas on basic ethical knowledge retrieval and processingwill be described.

[1]Currently we use MANOI PF01[1], however we are also testing even cheaper versions as MANOI AT01[2]

Fig. 1. Kyosho’s Manoi PF-01 Toy Robot

II. Basic Philosophy and Its Techniques

Most of our experiments are conducted on Japanese language and Japanese Internet resources as we confirmed that common sense behavior depends on cultural background[9] and what is obvious behavior for Chinese children can be very rare among Japanese and vice versa. Using English does not allow achieving very high cultural homogeneity. Two basic concepts we use are as follows.

A. Positiveness

A machine needs to be able to tell good from bad therefore we developed a scale resembling ideas of Jeremy Bentham[10]. We divided the affective reaction into 5 levels - negative, slightly negative, neutral, slightly positive and positive which are calculated by counting words, phrases or clauses with their neighbors containing ”emotional indicators”[11]. The simplest example of such specific opinion retrieval are *BEING HIT = unpleasant* and *BEING PRAISED = pleasant* where passive form of a verb forms a query with IF expressions to determine whether the search engine results are negative or positive.

B. Usualness

An intelligent talking toy also needs to be able to tell correct from wrong (of course to some extent). We assumed that if some combination of words or phrases does not exists in the Internet it can be treated as abnormality which is based on a popular n-gram frequency checkup but concentrates on Japanese particles which we find very useful for retrieving common sense knowledge. We assume that if something is usual, normally the system does not react or reacts the same way it usually does. At this moment the system is able to recognize abnormalities of situations represented by one actor, one object, one place and one action, as "a doggy swims in a toilet" or "mum can cut bread with fingers". Children have rich imagination so system is not programmed to say "it's not true" but to express surprise for example.

III. Assuming Reasons of Human Behavior

We have been working with emotions on very instinctive "Pavlov's reaction" level[12]. We based our system on Bentham's view saying that whatever human does, it is for his pleasure. In case of children, the understanding of human emotional behavior is supposed to be easier (though we still have not confirmed this assumption experimentally), however, as children's reasoning is straightforward, it is more important to retrieve cases showing that good things have their dark sides. Sweets have very high Positiveness value but they can lead to decayed teeth which have very low Positiveness value. If there is a possibility that Positiveness will decrease or anything threaten the high Positiveness value even if actual state is not negative - adequate warnings should be performed. If the child says it is cold, the toy has to calculate if it is negative state for its user, and if so, check what kind of actions are usually performed in such case. Such commonsensical retrieval can become a safety valve which, after recognizing a threat, can lead to sending an alert to the child's parents. The toy should be also able to guess user's mental states, therefore we also work on affect recognition, which will be described in the "Affect Recognition" section.

IV. Web-supported Conversation Module

Our goal is to make a conversational agent which has no limitations of topic and as it is supposed to play with children, therefore we will have to deal with a very difficult contradiction in near future. Before that we had to develop an algorithm finding satisfactory semantic associations in as short period of time as possible. Describing this module shortly, the system uses co-occurrence rankings of verbs, nouns and adjectives to retrieve sets of words for utterance generation[13] within less than 10 seconds (which is still not quite satisfactory). For example if a child speaks of some cartoon character, character's items (nouns), descriptions (adjectives) and actions (verbs) are easily found in 500 top snippets of Google search engine results which the system uses. The retrieved words are combined to create simple sentences as "Mr. A's weapon is powerful" or "Mr. A rides a bike". As the combination can become a nonsense, the system queries the Google systems once again to see if the combination exists. If not, another combination is created. To make the output more natural, handcrafted modality expressions are being added randomly to create sentences as "Mr. A's weapon is powerful, isn't it?" or "Mr. A rides a bike, you know". Experiments showed that adding modality even in a random way doubles the impression of the system without modality expressions (though the evaluators were grown-up). In the next step we plan to use machine learning methods to make the toy acquire knowledge about the discourse and modality usage.

A. Main Parts of the Conversation Engine

In the current stage of development our Conversation Engine is very simple and easy to implement because only a morphological analyzer and the Web resources are needed to fill several handcrafted utterances templates. Here we explain the algorithm.

1) Associations Retriever: In the basic version of our conversational engine, all nouns, verbs and adjectives contained in a user utterance form a query which is sent to Google search engine. Then, the ranking of again most frequent nouns, verbs and adjectives is created from the first 500 snippets of the Google search results. This lets the system, to some extent, focus on a given context which is difficult in methods concentrating on one keyword automatically decided to be important, for example by calculating information amount[14] of all keyword candidates. For instance when a kid talks about a particular adventure of his favorite hero, the list of words associated to this topic will be used instead of just random association of the hero. No database is needed, no update is needed. If the adventure was shown on TV just the night before, the system will "know" about its as even adult animation fans widely comment online the latest happenings.

2) Utterance Builder: Next, the engine takes a set of top-ranked nouns, verbs and adjectives and generates an expression as "Adjective Noun Verb" (for instance "coward Superman escapes"). Then, the generated string becomes a query and the search engine checks its validity - if the exact match is less than 100 hits, another combination is created. When a possible candidate is found, interjections and sentence endings (for example "should" - Japanese is SOV type language) are added (randomly in the basic version) to create utterances like "Well, then you must Verb Adjective Noun" (or "Did you know that cowardly Superman escapes?"). Depending on the input length, this process takes from 5 to 15 seconds and depends on the search engine which is used. This is why we consider creating a database of the most common association to search them quickly when nothing unexpected happens. Superman flies and saves people but it is too obvious knowledge to retrieve it every time from the net. It is much faster to use if such "straight knowledge straight from the net" is cached.

V. AFFECT RECOGNITION

We can say that the affective computing field is now 10 years old[15] but it still seems to be neglected. For the assumptions described above we needed a robust emotion recognizer and the Emotive Elements Analysis Module (ML-Ask)[7] was created. The analysis of emotiveness in ML-Ask is based on finding emotive elements in an utterance. Top-down determined (basing on different research outcomes) databases of emotive elements in speech are used after being divided into interjections, emotive mimetics (gitaigo), endearments, vulgar vocabulary, which belong to lexical layer of speech, and symbols representing emotive elements from non-lexical layer of speech, like exclamation marks, syllable prolongation marks, etc. (the remaining problem is the sufficient method for correct automatic transcription while using voice system but as we decided to concentrate on knowledge first, now we mainly work on text dialogue simulations). For classifying emotions we chose Nakamura's[16] classification of 9 basic emotions[2]. The recognition of emotions existence reached 93% accuracy and the classification was evaluated as correct in 46% of cases.

A. Using Web-mining Support

In the second stage of development we supported the method with a Web mining technique (see Fig. 2). In every case the Web mining improved the accuracy of the system in extracting the specific types of emotions, although some differences in commonsense levels for the two extraction routes checked. There is a need for further experiments on a larger evaluation material determine the most accurate extraction route. Experiments on different variants of the method showed, that it is more effective to keep only the emotions that achieved the three best results, rather than extract all of the emotions. Furthermore, since the system's procedures are activated gradually, improving accuracy in earlier stages clearly prognosticates improving of the emotion types extraction. The tools for morphological analysis used in the system are also not perfect, which decreases the system's real accuracy. However improving the extraction of n-gram phrases from the queried sentence will surely eliminate these difficulties. In the near future we also plan to apply Russel's two-dimensional model of emotions[19] to reduce the ambiguities in databases of emotive elements. The system is capable to be used in real-time applications, since an approximate time of processing one utterance is 0.143 s. Only delays, if they appear, arise from problems with potential network connection or search engine troubles. What we were not satisfied with most, were existing evaluation methods, therefore we proposed a new one briefely described below.

B. Our Novel Double-Standpoint Evaluation Method

This is a method, where the evaluated system is watched from two different standpoints: recognitive (the first person evaluation) and commonsensical (the third person evaluation). In practice this means that a system is first evaluated by users and to confirm whether the results were adequate, another evaluation is performed by an objective group of third party evaluators. In the present case, to judge the accuracy of the system we performed the evaluation basing on Ptaszynski's corpus of utterances with 90 utterances[6]. The corpus is tagged by authors of the utterances in the same way, as the system's procedure - they first determine whether an utterance is emotive. If it is, they set the emotive value (0-5) and describe the specific emotion types conveyed. This will provide us the evaluation from the first - recognitive standpoint. To broaden the evaluation, the second, commonsense standpoint is applied. By tagging the same utterances by the third party human evaluators (10 people on average) and calculating the system's unanimity with them, we check the commonsense level of the evaluated method.

[2]That is: ki / yorokobi (joy, delight), do / ikari (anger), ai / aware (sorrow, sadness), fu / kowagari (fear), chi / haji (shame, shyness, bashfulness), kou / suki (liking, fondness), iya / iyodomi (dislike, detestation), kou / takaburi (excitement), an / yasuki (relief) and kyou / odoroki (surprise, amazement).

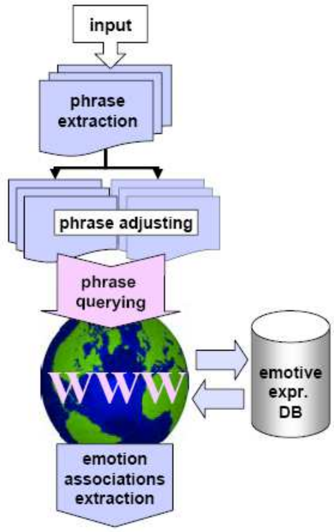

Fig. 2. Flowchart of the web-mining technique we used

VI. POSSIBILITIES OF USING HUMOR MODULE

Everybody knows that humor is healthy and it has aslo been already proven that humor helps in task-oriented human-computer interaction and computer-mediated communication[20]. We also know that robots make jokes even funnier[22]. This is why we decided for our toy to have at least some basic joking ability though we know that even the basic level is a very difficult task when it comes to make somebody laugh. There were several attempts to write puns-generating programs[21], but they first find possibly funny pair of words (e.g. "cereal" and "serial") and then generated a very narrow context (e.g. 'fibers" and "murderer") to prepare a one-liner / riddling pun output ("What do you call a murderer that has fiber? A cereal killer"). Unfortunately, in case of freely talking agent such approach is almost useless - decision for generating a pun would most probably mean to change the topic of conversation what is against the basic rules of communication.

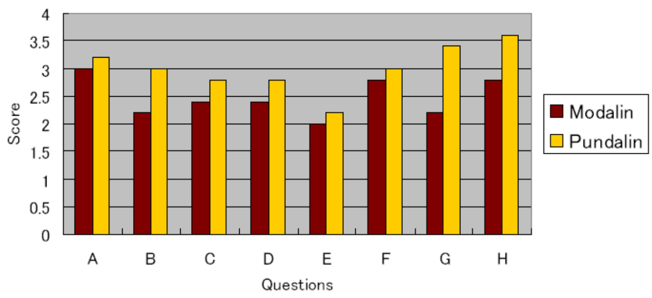

Fig. 3. User's evaluation - results for Modalin (non-humor equipped system) and Pundalin (humor-equipped system) for detailed questions: A) Do you want to continue the dialogue? B) Was the system's talk grammatically natural? C) Was the system's talk semantically natural? D) Was the system's vocabulary rich? E) Did you get an impression that the system possesses any knowledge? F) Did you get an impression that the system was human-like? G) Do you think the system tried to make the dialogue more funny and interesting? and H) Did you find system's talk interesting and funny? Answers were given in a 5-point scale.

TABLE I
EXAMPLE CONVERSATION WITH THE DIALOGUE MODULE

user	*Nani-ka suki-na tabemono aru?* (What food do you like ?)
system	*Maa, tabemono-wa oishii-desu.* (Well, food tastes good.)
user	*Saikin-wa o-sake-mo suki-ni natte kita-n-da.* (Recently, I've become quite a alcohol lover)
system	*Demo, sake-wa yowai-no-yo-ne.* (But I canft drink much, oh no.)
user	*Yowai-n-da. Sugu yocchau-no?* (Oh, I see, you get loaded quickly?)
system	*Maa, sore-wa sugo-wa-ne.* (Well, it's amazing.)
user	*Boku-wa wari-to o-sake-wa tsuyoi hou-nan-da.* (In my case I guess I can drink quite a lot)
system	*Iya, sore-wa tsuyoi-n-de.* (Ooh, strong, one can say.)

We need an algorithm that can generate a pun fitting the context, not changing the context to fit a pun. Therefore we decided to collaborate with Dybala, the main designer of PUNDA project[23]. He used our conversation engine (calling it a Modalin as it uses modality) as a baseline system and add his algorithm to equip the baseline with joking ability (he called it Pundalin as a mix of Modalin and PUNDA project). The detailed information on the system and experiments will be published later this year, however we will introduce some partial results here to show that artificial sense of humour is useful.

VII. Towards a Moral Behavior Extractor

Designing a system which has basically no limitation of conversation topic is very difficult. Even in cases of information kiosks or automatic museum guides[26] which are rather task-oriented systems, irrelevant and very surprising input is very common[27]. Children's fantasy or will to find out things can lead to unexpected conversations.

Tab. 1 (language errors translated as close as possible) shows a dialog example between an adult tester and the system. One can easily notice that if the interlocutor was a child, not adult, the conversation could look very similar. This brings the ethical problem of "unlimited topic conversations". As we aim at children and "smarter" toy than other toys, we need a mechanism that follows moral rules even in a basic level in order to make the potential buyers (parents) buy it. The simplest approach is a parental advisory approach where specific keyword list is just banned and the robot reacts to them by simply not reacting to them. However, new toys, to be really new generation of toys, must be educational but still trusted by children. We assume that having the right view on sensitive topics is better than avoiding them. On the other side listing all of such topics is also very difficult task. For those reasons we are working on automatic methods for retrieving simple moral rules from the Internet[24].

A. Schankian Scripts Retrieval

The fundamental idea is to use classic thoughts as of Schank[25] to retrieve and calculate common behavior patterns which combined with Positiveness calculation give the system information about what consequences may a given action bring. For example neutral word "escaping" becomes negative while inside of "robbery script" or positive in "picnic script".

B. Causal Rules Retrieval

This works on the same basis as Scripts Retrieval but uses several Japanese "if" forms which have abilities to categorize causal dependencies. In this case Usualness of single happenings becomes more important - if a Script cannot be created, it can be made from single causalities generalized semantically by using a thesaurus. If there is not enough data for "quarreling with professor" maybe "quarreling with teacher" could bring possible consequence patterns.

C. To Pretend or Not To Pretend?

In our approach, the toy's "self" can be set as "average man" (the idea of *consciousness of crowds*, unfortunately we still have no idea how to limit it to the "average child") or "average robot". The latter uses "a robot" as the keyword to find what is possible for robots and what is not. The method is very naive but based on, in our opinion, very fair idea - especially for a conversational agent. When a user invites an agent for a picnic, it's algorithm can guess that "its kind" is not capable of such task, though there is a big amount of noise because of online science-fiction stories. However, recognizing real world stories from unreal stories by statistical methods seems to be possible and we are working on it. Giving the list of robots' functions or even its type could also help because some current robot toys can play soccer and others cannot. In case of toy which is supposed to be trusted by children the "not pretending" method is probably fairer and should help the adolescent user become more tolerant for handicapped beings. We have already suggested that a machine should be "conscious" enough to forecast danger for us[24]. There are many ways to use systems that can analyze a given situation, guess the previous actions, possible plans and goals, then estimate the consequences and how a user will feel about it. From very simple house devices (adjective "cold" in "home" context brings "possible flu" alert and "close the window" as the best action proposal) the toy can gather information and react properly also to protect a child which current toys are not capable of.

VIII. Conclusion

In this paper we outlined the basic ideas for our project and our latest achievements while building a toy which can talk to children about anything but knowing good from bad or wrong. As this is ongoing project we describe wider only the parts where experiments were performed and the evaluation was done - conversation module, affect recognition and we also partially introduced some results of system improvement due to using artificial sense of humor. We claim the results are promising and the performance should be easily increased by combining methods we simultaneously work on. We must add that we are running preliminary experiments on web-mining and so far we can say that simple causal knowledge can be retrieved from the WWW and the Positiveness of the action results can be also calculated. We are still far away from the universal ethical algorithm, but by this paper we are suggesting that WWW-retrieved "common consciousness" and "average self" could be an option which becomes easier and easier to implement along with development of systems using large scale text-mining and affection analysis. For now most of the WWW-based techniques concentrate on making us, human beings, smarter. What we propose is to use the Internet resources to make machines smarter and, what should follow, think straight and be more human-like.

IX. Future Work

At this moment seven people work on the toy project and even more modules (language learning and acquisition[17], remembering and forgetting[18], humor processing[23], etc.) are going to be combined within next year. The most challenging task awaiting us are experiments with children which are very difficult subjects and for example any generation failures, time delays, sound recognition errors etc. may be disastrous for the tests.

Acknowledgment

The authors would like to thank the Nissan Science Fundation for the financial support and the feedback on current expectations regarding robots mentioned for everyday life. We are especially grateful to Mr. Takashi Sunda from Mobility Laboratory, Nissan Research Center.

References

[1] http://www.kyosho.com/jpn/products/robot/pf01/pf01.html

[2] http://www.kyosho.com/jpn/products/robot/at01/at01.html

[3] http://www.business-design.co.jp/

[4] P. D. Turney. *Thumbs up? thumbs down? semantic orientation applied to unsupervised classification of reviews*. Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL'02), Philadelphia, Pennsylvania, USA. July 8-10, pp. 417-424, 2002

[5] A. V. Samsonovich, G. A. Ascoli, H. Morowitz, M. L. Kalbfleisch. *A Scientific Perspective on the Hard Problem of Consciousness* Proceedings of AGI-08 Workshop on the Sociocultural, Ethical and Futurological Implications of Artificial General Intelligence. Memphis, TN, March, 2008

[6] M. Ptaszynski, P. Dybala, R. Rzepka and K. Araki. *Double Standpoint Evaluation Method for Affect Analysis Systems* CD-ROM Proceedings of The 22nd Annual Conference of the Japanese Society for Artificial Intelligence, 2008

[7] M. Ptaszynski, P. Dybala, R. Rzepka and K. Araki. *Effective Analysis of Emotiveness in Utterances based on Features of Lexical and Non-Lexical Layer of Speech* Proceedings of the NLP'08 Conference, pp 170-174, Tokyo, March, 2008

[8] S. Tsuchiya, E. Yoshimura, H. Watabe and T. Kawaoka. *The Method of the Emotion Judgement Based on an Association Mechanism* Journal of Natural Language Processing, Vol. 14, No. 3, The Association for Natural Language Processing, 2007

[9] Y. Ge, R. Rzepka and K. Araki. *Automatic scripts retrieval and its possibilities for soft science support applications* Proceedings of Intelligent Information Systems 2005 - New Trends in Intelligent Information Processing and Web Mining Springer's Advances in Soft Computing, pp 51-58, Gdansk, 2005

[10] J. Bentham. *An Introduction to the Principles and Morals of Legislation* London: T. Payne, 1789

[11] R. Rzepka, K. Araki, K. Tochinai. *Is it out there? the perspectives of emotional information retrieval from the internet resources* Proceedings of the Artificial Intelligence and Applications Conference, pp. 22–27, Malaga, 2003

[12] R. Rzepka, K. Araki, K. Tochinai. *Ideas for the web-based affective processing* Proceedings of the Seventh Multi-Conference on Systemics, Cybernetics and Informatics, volume XIV of "Computer Science", 376–381, Orlando, Florida, 2005

[13] S. Higuchi, R. Rzepka and K. Araki. *Dialog System Using Modality and Associations Retrieved From The Web* (in Japanese) Proceedings of the NLP'08 Conference, pp 175-178, Tokyo, March, 2008

[14] C. E. Shannon. *A mathematical theory of communication* Bell System Technical Journal, vol. 27, pp. 379-423 and 623-656, July and October, 1948.

[15] R. W. Picard, *Affective Computing* The MIT Press, Cambridge, 1997

[16] H. Nakamura. *Kanjo Hyogen Jiten - Dictionary of Emotive Expressions* Tokyodo Publishing, Tokyo, 2004

[17] D. Hasegawa, R. Rzepka and K. Araki. *Evaluation of Connectives Acquisition in a Humanoid Robot Using Direct Physical Feedback* Springer-Verlag Lecture Notes in Artificial Intelligence (LNAI) 4830, pp. 664-668, Berlin-Heidelberg, 2007

[18] M. Ptaszynski, K. Sayama. *The idea of dynamic memory management system based on a forgetting-recalling algorithm with emotive analysis* Language Acquisition and Understanding (LAU) Technical Report, pp 12-16, Sapporo, 2007

[19] J. A. Russel *A circumplex model of affect* Journal of Personality and Social Psychology, 39(6):1161-1178, 1980

[20] J. Morkes, H.K. Kernal, C. Nass. *Effects of humor in task-oriented human-computer interaction and computer-mediated communication: A direct test of srct theory* Human-Computer Interaction 14(4), 395–435, 1999

[21] K. Binsted, *Using humour to make natural language interfaces more friendly* Proceedings of the AI, ALife and Entertainment Workshop, Intern. Joint Conf. on Artificial Intelligence, 1995

[22] J. Sjobergh, K. Araki, *Robots Make Things Funnier* Proceedings of LIBM'08, 46–51, 2008

[23] P. Dybala, R. Rzepka and K. Arak,. *Dajare Generating Support Tool - Towards Applicable Linguistic Humor Processing* Proceedings of the NLP'08 Conference, pp 701-704, Tokyo, March, 2008

[24] R. Rzepka and K. Araki. *Consciousness of Crowds - The Internet As a Knowledge Source of Human's Conscious Behavior and Machine Self-Understanding* "AI and Consciousness: Theoretical Foundations and Current Approaches", Papers from AAAI Fall Symposium, Technical Report, pp.127-128, Arlington, USA, November, 2007.

[25] R. Schank and R. Abelson *Scripts, Plans,Goals, and Understanding* Hillsdale, NJ: Erl-baum, 1977.

[26] S. Kopp, L. Gesellensetter, N. Kramer and I. Wachsmuth. *A Conversational Agent as Museum Guide-Design and Evaluation of a Real- World Application* In Panayiotopoulos et al., (Eds.), Intelligent Virtual Agents, LNAI 3661, pp. 329-343.

[27] J. Gustafson and L. Bell, *Speech technology on trial: Experiences from the August system* Natural Language Engineering, 1 (1), pp. 1-15, 2000