IS IT OUT THERE? THE PERSPECTIVES OF EMOTIONAL INFORMATION RETRIEVAL FROM THE INTERNET RESOURCES

Rafal RZEPAKA†, Kenji ARAKI†, Koji TOCHINAI††
†Graduate School of Engineering, Hokkaido University
Kita-ku, Kita 13-jo, Nishi 8-chome, 060-8628 Sapporo,
††Graduate School of Business Administration, Hokkai-Gakuen University
Asahimachi 4-1-40, Toyohira-ku, 062-8605, Sapporo
Japan

Abstract

Many researchers tend to think of WWW as an enormous database, which unfortunately is full of informational garbage that makes the web mining or knowledge-bases creation difficult. However when we started to rummage through those seemingly useless for AI purposes personal homepages with very similar contents, we imagined the human brain cells might look exactly the same. Not only the stored pieces of semantic information are important but also the number of how many times such similar data was stored. We assumed that the Internet is an interesting material for retrieving a common sense, beliefs, opinions and emotional information for various types of agents. Without any sophisticated method our system is able to easily find out that in most cases being cold is not pleasant and cold beer almost always "sounds nice" or that one movie star is being loved and another hated. In this paper we introduce ideas for our project (GENTA – GENeral belief reTrieving Agent) and the results of initial experiments with implementing a primitive method of retrieving basic feelings towards human user’s utterances and applying this emotional information in Inductive Learning of the speech acts.

Key Words: affective computing, believable agents, web mining

1. Introduction

The earliest ideas for our project were born during observations of foreign students’ linguistic behavior while speaking Japanese language, which was not their mother tongue. Although the language abilities and cultural background happened to be very different, their conversations always seemed interesting, which agrees with intercultural will of communication theories [1]. We noticed several dependencies, for example that the keywords triggering the conversations are mostly of two types – topics that have clearly positive or negative emotional load as "a shower" or “the red tape” and that the conversational flow concentrates on what interlocutors have in common or what differs them. Even if quite obvious for human beings, these conditions for what we call interesting conversation are very difficult to be met by machines. First of all, computers have problems with recognizing what could be interesting and what would bore its conversational partner. Secondly, they have difficulties with spoken language, especially when there are plenty of grammatical mistakes not to mention the interlocutors who use the broken language. Therefore we looked for a conversational environment where the computer system could face-up to these challenges and chose IRC (Internet Relay Chat) because of its international character and "world simplicity" [2]. Although it is a multi-user environment we concentrated on one-on-one conversations, as the program does not handle multi-thread yet. We chose English language this time as the initial implementation was done in Japanese [3].

2. Basic assumptions

From the very beginning of human-computer interaction, the purpose of communication was almost always clear – the machine had to understand an order from the human user and help him some way. That approach is obviously being forced by the pressure that industry puts on the world of science. Even if the talk itself was a purpose of a program [4], it was always supposed to help the user somehow and to be socially useful. Pure chat for chat’s sake agents are not widely developed and scientifically neglected because of this low usability and problems with evaluating such systems. In our opinion, concentration on problem-solving or helping agents makes HMC (Human-Machine Conversations) research unbalanced and we argue against the importance of usability while developing programs that can communicate with human beings. We have noticed that nowadays approaches become more and more sophisticated, machines learn how to analyze metaphors and to stochastically use very large corpora but there is still too few agents based on affective computing [5][6][7], which can react not “store-clerk-human-way” but “human-way” if a user says “I’m sad” or just “I want to buy a dog”.

Since Damasio [8] has underlined the meaning of emotions in reasoning, the number of AI or cognitive science researchers who search for the methods of implementing emotions into machines grows rapidly. Virtual and physical architectures are built and agents or robots are taught to analyze emotional state as the
additional information for perceiving. In our approach we have assumed that additional is not enough and made it crucial, since we understand the human conversation as a cycle of exchanging emotion-based information. By that we do not mean that pure information exchange does not exist, we want to make a machine remember that its reaction depends mainly on the emotional load of an input. Hence our plan is to implement a kind of Pavlovian reactions into the computer’s conversation abilities.

For the simplicity of our system we also simplified an idea on the utterance structure. We imagined a human language interaction as an electron jumping through the layers shown in Fig. 1. It goes up when making an utterance and down when receiving it.

![Utterance Structure Diagram](image)

**Fig. 1.** The idea of the utterance structure before simplifying

We assume that the emotional layer includes intentions that motivate to start, to continue, to change or to stop a conversation flow. We imagine that if anyone puts an electron to any area of semantic layer, for example by saying “a kingdom for a soda!” and an interlocutor who does not know what this cliché exactly means hears it, the electron jumps down to the emotional layer instead of wandering along the complicated semantic network. The interlocutor feels that “kingdom” and “soda” are not abusive, “soda” even sounds nice. The nicer it sounds to him, the more probability of choosing it for a conversation topic, or as we call it triggering keyword, appears. An aggressive reaction becomes less possible in such case.

Guided by above-mentioned assumption, we decided to realize a method that would allow us to simplify semantic layer processing by limiting choices in lexical and emotional layers. We also made an assumption based on Fletcher division of blind people’s mental imagery [9] and we believe that machines could gain not deficient or inefficient but different imagination. Our idea is that the electron having fewer choices in the emotional and lexical layers may naturally decrease the possibilities of exploring the semantic layer as we illustrate in Fig. 2. Since we are interested also in the aspects of modeling an artificial imagination, we decided to base our system on pure written word-level conversation without audio-visual stimuli, although we used their simplified substitute, as they are necessary for obtaining the basic emotional information. This substitute is the usage of emoticons (facemarks) widely exploited during keyboard-based chat, which suggest if given utterance was, for instance, an irony or supposed to be a joke. For example, if the utterance from Fig. 2. is done with a smiley “:)” the electron should not hit emotional layer at fear spot because it would not be natural behavior.

![Emotional Layer Diagram](image)

**Fig. 2.** The idea of the utterance structure after simplifying

### 3. Preliminary experiments

In order to see what kind of information could be retrieved from the Internet resources and used for general belief and emotion processing, we made several tests with different searching engines. For retrieving common emotions we were observing the hits numbers of the results of searched frames as:

- “I am afraid of” {N} (for nouns causing fear)
- “I always wanted a” {V} and “I always wanted to” {V} (for what people may dream about) or
- “Usually people {V} at {N}” (for retrieving verbs for building a script of given noun).

We discovered that it is possible to retrieve thousands of sentences that could be processed for what we call emotional common-sense statistics. The more specific given frame was, for instance “I always wanted to be” instead of “I always wanted to”, the less ambiguous sentences were retrieved and results as “I always wanted to do it” were apparently decreasing.

As we noticed in previous works [3], the Internet can also provide basic semantic tagging by counting frames including prefixes (or particles as in Japanese). For example if a parser receives an unknown word, let’s say “Sapporo”, it can easily declare that it is a place by comparing result numbers of given searching frames:

- “I talked to [Sapporo] and”, “I felt [Sapporo] then”, “I ate [Sapporo] with”, “I went to [Sapporo] by”, etc. Such frames must be chosen very carefully and one should use stop words as “and”, “then”, “with” or “by” to avoid counting expressions as “I talked to Sapporo Council”. Searching engines also provide misspell detection, what can be useful in fast typing chat environments with non-native speakers.

Unfortunately automation of such searching requires usage of crawlers (web robots) and complicated filtering which still takes too much time for keeping up with interactive conversation processing. Therefore we decided to limit the retrieved information only to the
simple test if something is liked or disliked and how common it is.

4. Description of our system

Agreeing with Groenendijk et al. [10] that conversational information could be divided into two kinds—information about the world and discourse information—we concentrated on following two parts of our system—General Belief, used interchangeably with Common Belief, and Conversation Keeper, which is based on the Inductive Learning from humans’ chat conversations.

4.1. General Belief

By this expression we mean a mixture of common sense based on retrieved opinions and simplified environmental knowledge. System’s knowledge of a user is assumed as almost none — GENTA does not know the nationality, age or sex of its conversational partner; he or she is not necessarily native speaker of English. After a greetings exchange, system waits for user’s initial utterance and if it is not done starts conversation using learning data of "Conversation Keeper" which will be described later. While detecting speech act, which also will be explained below, GENTA tries to guess the leading keyword(s) from the first user’s utterance since the domain of conversation is still unknown. Next, GENTA searches the Internet for the whole utterance and its grammatically connected parts previously parsed by a parser [11] trying to establish what can be associated with given verbs, nouns, noun phrases, adjectives or conditional expressions concentrating on feelings-based opinions. For example when input is "Do you like playing soccer when it rains?" GENTA counts how many sentences "I like playing soccer when it rains", "I love playing soccer", "I hate playing soccer", and "I love when it rains, I hate when it rains", etc. appear on the Web. This lets our system achieve “own”, which is assumed as general or common, opinion about playing soccer and when it rains. Then, paraphrasing Shannon’s information theory [12], we assume that the keyword with less frequency is more interesting for interlocutors and GENTA chooses playing soccer for a leading topic (27268 vs. 33985 hits). The system believes that the discourse should be continued in this “semantic direction”. But before that, “Conversation Keeper” must establish which linguistic behavior (dialog act) will be proper for a reply, which is our current task.

4.2. Conversation Keeper

Taking last decade research results into consideration [13-20] we decided to find such a combination of web-mining and learning methods that would help us to create a dialog system not requiring big initial data prepared by hand and does not need sophisticated modules. As the first step, before creating the real dialog manager, NLG module, etc. we decided to confirm that our system is able to learn from word-level information. As we already mentioned, we assumed that peculiarity and emotional load of given expressions could support intention recognition, which is one of the most important tasks of human discourse management. Therefore we divided General Beliefs here into two above word-level values that we call positivity and usualness, which are also measured by counting above mentioned string frequencies upon WWW. We prepared dialog act tags, as handing or demanding of information, opinion and reason; advising, warning, greeting and nodding. GENTA can automatically declare usualness and positivity for utterances, as in example:

Do you like playing soccer when it rains?

becomes a DAPU string (Dialog + Act + Positiveness + Usualness)

OD P5U4 cond P3U5

which means that it was a Demand of Opinion consisting of two positive and usual expressions connected by subordinate clause conjunctor (SCC) “cond” (conditional clause). What is characteristic for our method, even if the positivity of expression seems to be doubtful (most Web pages creators like when it rains what does not have to mean that most human beings would state so) it does not disturb the process since the opinion stays logical.

Values of usualness and positivity are calculated by comparing frequencies of ("I don’t like ..." / "I hate ...") and ("I like ..." / "I love ...") searching frames. The frequency thresholds are different depending how long the searched string is.

4.3. Inductive Learning

GENTA system has an ability to learn from human spontaneous conversations. We use the Inductive Learning method [14] to predict which utterance should be used and to make new rules while talking. The system represents dialog discourse as associated DAPU strings

OH P5U2 : OD P5 U3 : ND : ...
(A1 B1 A2 B2 A3 B3 ...)
and usualness. For example, if unknown dialog act determinator appears, our system decides the most probable tag and unless a user cancels computer’s output by using one of canceling expressions as “???” or “What are you talking about?” and so on, new rule is created in the rule dictionary (Tab. 4.). If an emoticon is detected, positiveness value is being decreased or increased depending on the kind of a facemark.

Fig. 3. A simple example of the learning process from the examples already parsed for positiveness, usualness and DAPU strings.

attributed to every dialog act tag. For example should determines advising tag. If subordinate clause conjunctor (SCC) is detected, both clauses are parsed into DAPU strings and they become an individual element for learning. When there is more than one sentence during one turn, GENTA confirms if they are of the same dialog act. If not, the input is divided - the last rule is changed and a new one is created, for example:

A1: Do you care?
B1: Well, I don’t care. What about ya?
A2: Me neither, man!


Learning concentrates on dialog acts tagging and conjunctors, and their coexistence with positiveness

5. Experiment and evaluation

5.1. Method

In our opinion, existing standards [21] in HMC evaluation, which concentrate on semantic quality of output, are not useful when evaluating spontaneous chat where discourse is more important than quality of content. Since this time we experiment only with above-word information supported dialog acts we prepared our own evaluation method:

- There are two human interlocutors A and B.
- They converse through IRC channel, which is monitored by our system (G).
- G listens to A’s utterances and proposes its own answers (as DAPU strings).
- DAPU strings of B’s utterances are compared with G’s ones.
- Afterwards the third person evaluates naturalness of strings when system chose different dialog act, as there are more possibilities than one.
5.2. Results

As we are not particular about language perfection, two non-native English speakers took part in our experiment. There was no given topic of conversation. Subjects made 128 turns and they talked mostly of sports. GENTA’s dictionaries were empty in initial phase and we taught the system only one determinator for every dialog act and only two-three basic conjunctors for every kind of subordinate clauses. We decided empty dictionaries to see when system starts to learn and because we want GENTA to retrieve what is needed from the Internet. The Web as a corpus is changing constantly and creating dictionaries for positiveness or usualness would not reflect those changes. By comparing user B and GENTA’s DAPU strings we understood that:

- The systems started to use learned rules already in the eight turn, as the chat was mostly question – answer style but finally less than half (37.5%) of dialog acts were chosen the same way by a user. Although 81.25% of those different ones were evaluated as natural by human being.
- Positiveness (5 grade scale: 1-negative, 2-slightly negative, 3-neutral, 4-slightly positive, 5-positive) of system’s output that had the same dialog act tag as human’s was in 59.1% the same as the user’s. By the same we mean three levels: positive, indifferent, negative.
- Usualness (5 grade scale: 5-very usual, 4-usual, 3-slightly peculiar, 2-peculiar, 1-very peculiar) of system’s output was in only 20.8% the same as human user’s, because all parser errors due to misspells were detected as the most peculiar expressions.

Positiveness and usualness were compared only in cases where dialog acts were the same in human and machine outputs, as dialog act choice influences those two values significantly. Because these two values were depending on the Internet connection speed (about 15-20 seconds for 1 calculation), computer’s propositions were given with growing time difference. Even though, it had no influence on the results of our experiment.

5. Considerations

We have described a new approach to WWW statistical information usage in dialog system, which is able to achieve information that is not obvious to the machine without using logic programming and other sophisticated methods. For the reason that it is the preliminary stage of our project, we could evaluate our idea only indirectly as grammatically built sentences were not outputted by the system. However the results are convincing enough to continue walking upon chosen path – even if the system was not guessing interlocutor’s intentions properly, it proposed its own dialog acts, which were not against the logical flow of conversation. Too few turns made learning material inefficient but it is quite difficult to evaluate GENTA system before implementing further modules that will lead to generating understandable outputs. The following example of this difficulty:

A: They prefer drinking beer. (IH P5U4)
G: P5U4 ex (that could be: “a beer!!!”)
B: C’mon, they’re watching games too. (IH P4U3)

It seems obvious that it is too early for evaluating our system as a talking system at this stage and it is rather impossible to see its abilities in context management. Although we can foresee that such outputs will be useful when we add the knowledge retrieval and representation modules.

6. The perspectives and future work

We believe that tuning up the parsing methods and increasing data for learning will help to achieve better results in future. We made a first step to the creation of an agent that should be able to chat about any topic with proper human-like reactions.

What seems promising for future tasks, our program was based only on automatically retrieved knowledge of common opinion and peculiarity of users utterance, which could be used in many interesting ways, as manipulating GENTA’s “personality” for example by decreasing its positiveness when, for instance, the weather is bad. Other idea is that one could use retrieved information as a model imagination of an interlocutor. We want GENTA to know what probably his partner thinks while saying for instance “I need a girlfriend”.

Thus there is a need of experimenting with different parsers and of creating mechanisms, which allow GENTA learn other things from the Internet - the biggest and rapidly growing database in the world and try to apply those methods to commonly explored areas as for example qualitative spatial reasoning. It must be also able to answer “wh-questions”, so we plan to concentrate on implementing substitution of imagination which should be an elastic plan retrieval mechanism supported by commonsense libraries created through search frames as I always (verb) when it rains or usually people buy (noun) when they want to (verb + noun), and also on automatic creation of such frames.

Our method could be also interesting from the sociological point of view, since GENTA can become a “mirror personality” of an average wired English-or Japanese speaker as in his original version [2], what could make it much more interesting conversation partner than its predecessors. By using searching frames as “computers never”, “computers can” we can also model a basic knowledge base for a machine that could be conscious of its possibilities. We also understood

\[ 2 \text{ As the Link Parser [11], which we use, was made for proper English input we need to concentrate on input preprocessing method, which resembles chat language-proper language translation. We stick to this parser because of its good ability to quick connecting and labeling strings for Web searching.} \]
that adopting GENTA to other languages would be limited only to translating the search frames.

We strongly believe that with constantly improving computer and network abilities, the Internet will become the main source of any kind of knowledge for future AI systems. We also predict that millions of private homepages sharing users' feelings and opinion will be crucial information helping the machine to "understand" what we usually think and what we normally know. It is out there but we still have to learn how to dig it out and use properly.

References