Common Belief Retrieval for Talking Agent
Based on Primitive WWW Search

Rafal Rzepka*1  Kenji Araki*1  Koji Tochinai*1

*1 Hokkaido University, Engineering Dept., Signal Processing and Language Information Lab.

In our project (GENTA - GENeralt belief reTrieving Agent), we try to build a conversational agent, which uses Web resources not only to enrich its linguistic abilities but also to use those resources as a material for creating agent's "own" opinions and beliefs what could lead to automatic personality creation. This paper introduces our ideas of how to simulate and use such "artificial personality" during Human Machine Conversations (HMC) in "unlimited chat", by which we mean spontaneous opinion exchange without any given domain.

1. Introduction

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 what was forced by pressure that industry puts on scientists. Even if the talk itself was a purpose of a program [Wetzenbaum 1966], it was 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 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 "understand" metaphors and use stochastically very large corpora but there is still lack of common-sense background letting a machine react "human-way" if a user says "I'm sad" or just "I want a dog". We claim that system which can answer differently than "Then you must buy a pet at XYZ - write down their address..." may be also demanded by a part of society [Bell 1999][Gustafson 1999], for instance by foreign languages students who do not have much opportunity to speak with foreigners, as here, in Japan.

Creating a small database for one domain is very laborious and such databases are needed for task-solving systems using dialog interfaces, although they do have to be essential for chatting agents. In our approach we assumed that Web resources might resemble a human brain, which consists of a large amount of information that is crucial for statistical analysis even if never used directly. For instance, normal corpora may include a sentence "I love dogs" but only Web resources can inform the system that usually people prefer dogs than cats. Being influenced by Vico's [Vico 1725] and Malinowski's [Malinowski 1922] classic works we also assume that:

a) Creating a simulation of imagination is more important than particular knowledge about new topic,

b) Conversational system without domain limitations should be based on intercultural will of communication regardless of knowledge differences.

In this paper, we shortly describe our ideas of how to realize above-mentioned assumptions and represent results of initial experiments for retrieving data giving us prospects of satisfying achievements as our project is still in its preliminary stage.

2. Main features of GENTA system

Agreeing with Groenendijk, Stokhof and Veltman [Groenendijk 1995] 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 inductive learning from humans' chat conversations.

2.1 "General Belief"

By this expression we mean a mixture of common sense based on retrieved opinions and simplified environmental knowledge

(1) Initial Environmental Knowledge

Because its informal character and "world simplicity" [Rzepka 1999] we chose IRC (Internet Relay Chat) for experiments. Although it is multi-user environment we concentrated on one-on-one conversations, as the program does not handle multi-thread yet.

System's knowledge of a user is assumed as 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 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.

(2) Main topic detection

While detecting speech act, which 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. First, GENTA searches Internet for the whole utterance and its grammatically connected parts previously parsed by link-parser [Sleator 1993] 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 [Shannon 1948], we assume that the keyword with less frequency is more interesting for interlocutors and GENTA chooses playing soccer for a leading topic (27.268 vs. 33.985 hits). 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.

2.2 “Conversation Keeper”

Taking last decade research results into consideration [Groenendijk 1995] [Alexanderson 1996] [Jurafsky 1997] [Kreutel 1999] we decided to find such a webmining and Inductive 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 a first step, before creating the real dialog manager, NLG module, etc. we decided to confirm that our system is able to learn from above-word level information. 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+Positivity+Usualness) OD-P5/PU-cond-PU:US

which means that it was a Demand of Opinion consisting of two positive and usual expressions connected by subordinate clause conjunctors (SCC) “cond” (conditional clause). What is characteristic for our method, even if 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.

3. Inductive Learning

GENTA system has an ability to learn from human spontaneous conversations. We use the Inductive Learning method [Araki 2000] to predict which utterance should be used and to make new rules. The system represents dialog discourse as connected DAPU strings:


which are divided into double rules

(A1 B1) (B1 A2) (A2 B2) (B2 A3) (A3 B3) …

stored in a Dictionary. When new input is done, GENTA parses the utterance to DAPU string by recognizing a dialog act determinator, which are words 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 - 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 positivity and usualness. For example if not known 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 dictionary.

4. Experiment and evaluation

(1) Method

In our opinion, existing standards [Minker 1998] 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 third person evaluates naturalness of strings when system chose different dialog act, as there are more possibilities than one.

(2) Results

Two non-native English speakers took part in our experiment. There was no given topic of conversation but emoticons (face marks characteristic for IRC) were prohibited, as they still cannot be parsed. Subjects made 128 turns and they talked mostly of sports. GENTA’s dictionary was empty in initial phase and we thought the system only one determinator for every dialog act and only one conjunctor for every kind of subordinate clauses.

By comparing user B and GENTA’s DAPU strings we understood that:

• The systems started to use learned data early as the chat was mostly question – answer style, but finally less than half (37.5%) of dialog acts were chosen the same way by human 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 79.1% the same as human user's.

• 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 expression.


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 experiment.

4. Conclusions and Future Work

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. 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. 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. What seems promising for future tasks, our program 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.

Thus there is a need of experimenting with different parsers and of creating mechanisms, which allow GENTA learn other things from 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 is also interesting from the sociological point of view, since GENTA can become a “mirror personality” of an average computerized English-speaker, or average Japanese speaker as in his original version [Rzepka 2001], what could make it much more interesting conversation partner than its predecessors.

References


