AUTOMATIC GENERATION METHOD FOR SPOKEN SENTENCES BY GENETIC PROGRAMMING AND ITS APPLICATION TO EMOTION ORIENTED DIALOG SYSTEMS

TAKUMI ICHIMURA[1)], KAZUYA MERA[1)], KENNETH J. MACKIN[2)], TOSHIYUKI YAMASHITA[3)], MAKOTO YOSHIE[1)], AND TERUAKI AIZAWA[1)]

1)Faculty of Information Sciences, Hiroshima City University,
3-4-1, Ozuka-higashi, Asaminami-ku, Hiroshima. JAPAN 731-3194
2)Fujitsu Limited, 3-9-18 Shinyokohama, Yokohama 222-0033, Japan
3) Graduate School of Engineering, Tokyo Metropolitan Institute of Technology
6-6 Asahigaoka, Hino, Tokyo 191-0065, Japan

Abstract

We have already developed an emotional interface system with facial expressions applying Emotion Generation Calculation. However, we found that users required an interface that supported natural spoken sentence generation in order to carry out a smooth dialog between human and computers. In this paper we propose a spoken sentence generation system using Genetic Programming (GP), which can generate various expressions while keeping the predetermined rules of the conversations. To verify the effectiveness of the proposed method, we applied the method to a www-based health care system.

Keywords

ADF-GP, Emotion Generation Calculation, Case Frame Representation, Japanese

1 Introduction

As computers have come into wide use, not only the younger generation but also the elderly have come to use the computer. Various methods have been previously developed in order to provide interfaces between human and computers. Many interfaces target realizing natural language communication from computers. Many of these researches have focused on displaying appropriate facial expressions. Besides verbal messages, human face-to-face communication usually involves nonverbal messages such as facial expressions, vocal inflection, speaking rate, pauses, hand gestures, body movements, and posture. Such human interfaces must include an emotional interface that can display the human face with some feeling depending on the conversation situation.

The function, which we must take into account to develop the interface, is based on natural learning processing. However, mainstream in the research works is the techniques of translating from a language into another one. The developed techniques are greatly contributed into artificial intelligence; however, the system is not able to grasp the real meaning of sentences under the conversation, so that we feel a sense of incongruity in the generated sentences by the techniques because of a monotonous utterance sentence generation. In a traditional technique, the sentence generation is processed as a sequential procedure so that a correct utterance sentence is generated certainly in the quality and grammar. However, only the same expression is always generated on the other hand in a similar situation. In human's utterance sentence generation processing, a lot of expressions can be used by each occasion feelings and other person's knowledge, even if the expressions are the same content.

We propose the utterance sentence generation system by Genetic Programming (GP)[1][2], which can generate various expressions with keeping the predetermined rule of the conversations. In this research, we limit usage to the system that aims to treat Japanese. Because the restriction to the word order and the omission of case element is comparatively small in Japanese, the variation increases large effectively. Therefore, GP, which is an effective means as the evolution algorithm, can search variation by the genetic search. Besides, the research, which the variety of the utterance sentence is considered, is the Japanese utterance sentence generative system by Genetic Algorithm by Araki. He developed the multi-stage decision method for production of new translation examples.

Furthermore, we have developed the interface into the web-based health care diagnostic system for quality of life(QOL)[3]. There is a tendency that they wants to speak their surroundings in the conversation even when we asked a question related to QOL. If we request only its answer, he/she feels so stress and hates to think about the question itself. Therefore, it is necessary to generate the utterance sentences in the

0-7803-7087-2/01/$10.00 © 2001 IEEE

interface. These sentences are not only the questions but also the contents related to the subject's utterances. In this paper, we try to include a generation method of utterance sentences by ADF-GP (Automatically Defined Function Genetic Programming) [4] into the emotional interface to reply various sentences to the subject's utterances.

This method is based on the Machine Translation Method using Inductive Learning [5]. Based on some generating sentence rules, this method consisted of the 3 following parts; a generating utterance part, a learning part, and a feedback part calculating the fitness function. A chromosome is a pair of question sentence and its reply utterance. The feedback part makes a decision of propriety for the generated sentence from the generating utterance part. The learning part makes a great variety of sentences by genetic operators; crossover, mutation, selection.

In this paper, the section 2 provides an overview of GP and a guideline for its application to natural language processing. In the section 3 we apply the ADF-GP into generated partial tree to maintain the sentence's meaning itself. Successively, The section 4 describes the experimental results related to utterance sentences generation by GP. The section 5 explains WWW-based Analytical System of Health Service needs among Healthy Elderly and discuss the capability of its application. Finally, we give a conclusive discussion in the section 6.

2 A GP to natural language processing

The process of the utterance sentence generation is roughly divided into two stages of "what-to-say" and "how-to-say." As for "what-to-say", the answered content is decided. As for "how-to-say", the expression of the content of the answer is decided. The expression technique of the utterance sentence of "how-to-say" is paid to attention in this research, although both "what-to-say" and "how-to-say" for smooth conversation sentence generation are important. After the content of utterance is decided by "what-to-say," the proposed method is implemented. In this method, the utterance sentences in the case frame representation are given as an input. This expression is described in the Lisp language as shown in Figure 1. Genetic programming operators vary the input case frame representation and the variation is given to the sentence of the omission of the word order and each element etc. The applied genetic operators are limited only to the "G-mutation" and the "G-inversion". The "G-crossover" is not applied because the content of the utterance changes into different one if it is activated.

2.1 GP description

First, we present briefly an overview of genetic programming.

2.1.1 The initial structure

The set of possible structures in genetic programming is the set of all possible components of functions that can be composed recursively from the set of functions and the set of terminals.

The initial structures consist of the individuals in the initial population represented in the form of LISP S-expression is started a randomly generated rooted and point-labeled tree with ordered branches represented in the S-expression.

2.1.2 G-crossover

The crossover operation for genetic programming, G-crossover, creates variation in the population to produce new offspring that consists of partially inherited from each parent. The G-crossover operation starts with two parental S-expressions and produces two offspring in S-expressions. Figure 2 shows an example of two parents in LISP S-expression.

2.1.3 G-mutation

The mutation operation for genetic programming, G-mutation, introduces random changes in structures in the population as in Figure 3.

2.1.4 G-inversion

The inversion operator for the conventional genetic algorithms is applied to individuals with relatively high fitness. Then, the inversion operator for genetic programming, G-inversion, aids in the establishment of a genetic linkage between combinations of alleles that perform well together. Figure 4 shows the G-inversion operator's behavior.

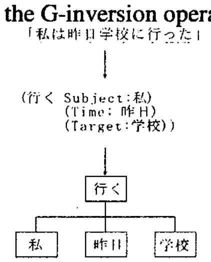

Figure 1 Lisp programming representation

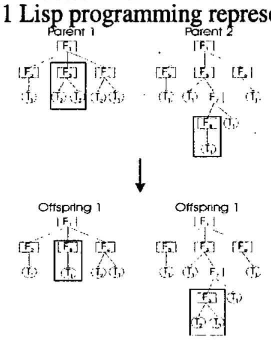

Figure 2 The G-crossover

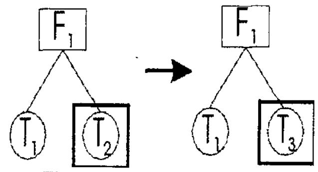

Figure 3 The G-mutation

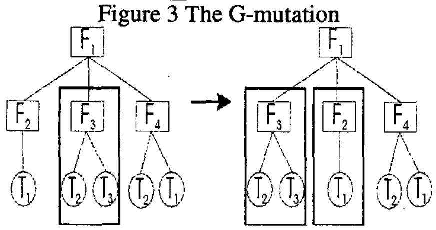

Figure 4 The G-inversion

2.2 Sentence generation by GP[1][2]

2.2.1 G-mutation

In the G-mutation, we consider that four kinds of operators between the terminal and non-terminal node; "Mutation from terminal node to non-terminal node," "Mutation from terminal node to terminal node," "Mutation from non-terminal node to terminal node," and "Mutation from non-terminal node to non-terminal node."

Mutation from terminal node to non-terminal node

This means the generation of a new sub tree in GP. Moreover, the sub tree means the subordination clause in the case frame representation. Then, in this paper, we recollect the event where the word that generates the sub tree is a subject from the conceptual knowledge and we add the subordination clause as shown in Figure 5. The case element of non-modifier is excluded from the case frame of the subordination clause added as a sub tree.

Mutation from terminal node to terminal node

This means the word of each element is paraphrased to other words. However, because the structure for the case frame representation has a meaning, it is not possible to replace it in disorder. Then, in this paper, we replace the word, which becomes an object with a pronoun, or assume it as zero-pronoun, that is, it will be eliminated as shown in Figure 6.

Mutation from non-terminal node to terminal node

This process is the inverse of "Mutation from terminal node to non-terminal node" and it can be achieved by deleting the sub tree or the subordination clause.

Mutation from non-terminal node to non-terminal node

This can be achieved by processing "Mutation from non-terminal node to terminal node" after the processing of "Mutation from terminal node to terminal node" as shown in Figure 7.

2.2.1 G-inversion

The position of each element is arranged by G-inversion. We become accustomed our ears to the word position of subject-object-verb in basic Japanese. However, there is no strict definition in each element except above positions. Moreover, even the rule that there is an object after the subject is not kept in the spoken language. Japanese can be called a language with a comparatively free in word position from these observations. Therefore, we meet that various surface expressions can be obtained without ruining the meaning by applying the change to the word order. In this research, G-inversion changes the word order of each element in Figure 8.

However, there is always a position of the predicate at the end because the predicate is different in the case frame representating the hierarchy from other case elements.

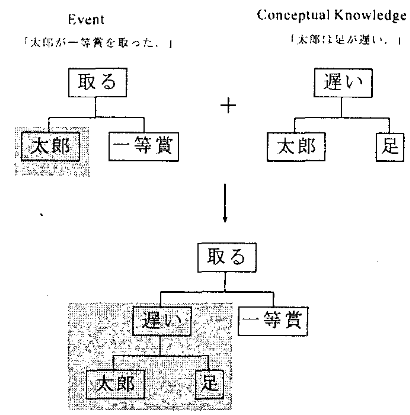

Figure 5 Mutation from terminal node to non-terminal node

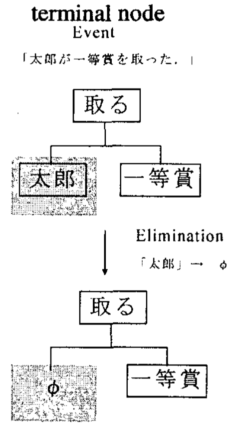

Figure 6 Mutation from terminal node to terminal node

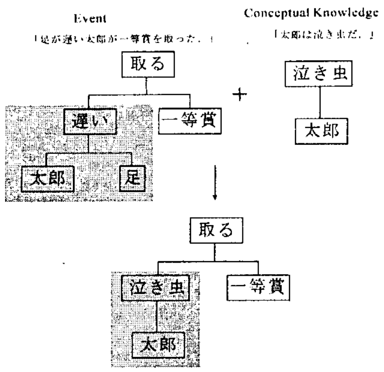

Figure 7 Mutation from non-terminal node to non-terminal node

3 An utterance sentences generation method by ADF-GP

In the variation of the natural-language structure, when the inheritance across sub trees is implemented, some sentences as a generated child have different meaning from original one[6]. Especially, when the genetic operators are activated unrestrictedly, such a phenomenon is remarkable. Consequently, this causes the situation with an increase of a fatal gene. To avoid such a situation, we apply GP's operators to only each sub tree.

To realize this aim, it is very effectively to use the ADF-GP method[4]. This method becomes possible to adopt genetic operators without an essential structure of the case frame structure. In this research, the case element, which contains the subordination clause, is defined as a unit of ADF-GP. The argument is a modifier and a subordination clause. Figure 8 shows an example of tree representation by the ADF-GP method.

3.1 Fitness Evaluation

The adaptation of a variety of case frame representation generated with GP is calculated based on Japanese grammatical rules and the contents of conversation and the audience person's preference information in the following.

*Appropriateness of pronoun and omission.
*The order of presenting each element.
*Length of utterance sentence.
*No declaration of what the audience person feels unpleasant.

The evaluation function takes a real value between 0.0 and 1.0. As the fitness value is larger the utterance sentence becomes a natural sense.

3.1.1 Appropriateness of pronoun and omission[7]

In case of omitting a part of speech or change the word of sentence into a pronoun, it is necessary to correspond to general common sense and the history of conversation. However, because to make a presumption from general common sense is very difficult, only the corresponded one to the history of conversation is discussed in this paper.

First, it is necessary to include the omitted word in the history of conversation. When the omitted word is not included in the history of conversation, we judge it as a bad omission, and then the fitness value is decreased 0.1. Otherwise, the fitness value is increased 0.1.

3.1.2 The order of presenting each element

There are generally the following tendencies in the structure of the sentence in a Japanese sentence [6].

1) The word which catches the wide range comes earlier than the other words. The sentence becomes a structure which the modification fills with non-intersection condition.
2) The order of the time, the place, the subject, the object, and the end of sentence is standard as a word position.
3) A long clause comes ahead of a short clause.
4) The clause, which contains context directive, comes before a clause not so.
5) The district of the topic presentation comes to the head of the sentence. For example, -Japanese words; "ha", "mo", "nado"-. The "ha" is comes before the "mo."
6) As for the place, the word position of the target after starting point is standard.
7) The word position of the clause which comes the left is different in the operation predicate and the state predicate.
8) There are a lot of word positions as shown in "Partial sentence concerning the content, other person, and the predicate."

Among these descriptions, we assume the increase factors of fitness function based on 2), 3), 5), and 6). Other factors are not mentioned in this research. Then, we obtained the following situations;

1) Near the word position of "the time, the place, the subject, the object, and he end of sentence", the fitness shows better.
2) The clause with the subordination clause (sub tree) is put ahead.
3) Puts the case element of the target after the starting point.
4) In the operation predicate, the place is put from the subject on the back. In the state predicate, the subject is put from the place on the back.

If each factor is filled, the fitness value is increased 0.1. Otherwise, the fitness value is decreased 0.1. For the first factor, the positions of two case elements except the end of sentence are investigated, if the position is corresponded to the above mentioned, the fitness function is increased 0.017(=0.1/6). Otherwise, the fitness function is decreased 0.017.

「太郎がテレビを壊した．」

↓

「彼がテレビを壊した．」

「太郎がテレビを…．」

Figure 8 the omission of unpleasant words

3.1.3 Length of utterance sentence

The whole contents of very long conversation might not be able to be understood for human. We collected the conversation corpuses with the aged by the real conversations. Then, the lengths of 432 utterances except chimes were examined among these corpuses and the average length of the utterance sentences and decentralization of them were obtained. As a result, the fitness function is decreased 0.1 if eight independence words or more are included while uttering. Otherwise, the fitness function is increased 0.1 if seven pieces of less independent words are includes.

3.1.4 No declaration of what the audience person feels unpleasant[8]

We have already proposed the Emotion Generating Calculations[1], which generates a subject's emotion using his/her favorable values for the event in the case frame form. Human occurs various emotions from the content when he/she hears the conversation sentences. If other person's *preference information* is understood beforehand, we can forecast and evade unpleasant feelings' occurring from the utterance. However, our method aims that do not vary the content of utterance sentences but change the expressions. Then, when we can be forecast to occur unpleasant feelings from the utterance sentences, we consider that the word which becomes the factor of the unpleasantness includes in the case frame is omitted or make it pronoun in Figure 8. Then, by the method the fitness value is increased 0.1, in case of the generation of pleasant emotion from the utterance by EGC, otherwise the fitness value is decreased 0.1.

4 Experimental Results

In this section, we show the variation of the utterance sentence by ADF-GP to the case frame representation of the content of the following utterance in Figure 9.

"Taro took the first prize by the marathon on Sunday." (1)

(取る[take] (Time: 日曜[Sunday])
(Place: マラソン[marathon])
(Subject: 太郎[Taro])
(Object: 一等賞[first prize]))

Figure 9 the case frame representation

(a) (取る[take] (Time: 日曜[Sunday])
(Place: マラソン[marathon])
(Subject: ADF0)
(Object: 一等賞[first prize]))

(ADF0 (Non-modifier, subordination clause)
(太郎[Taro]
(遅い[slow] Subject: (φの) 足[foot])))

(b) (取る[take] (Time: 日曜[Sunday])
(Place: マラソン[marathon])
(Subject: φ)
(Object: 一等賞[first prize]))

(c) (取る[take] (Object: 一等賞[first prize])
(Place: マラソン[marathon])
(Subject: 太郎[Taro])
(Time: 日曜[Sunday]))

Figure 10 the case frame representation by GP

(a)日曜，マラソンで足の遅い太郎が一等賞を取った．
(b)日曜，マラソンで一等賞を取った．
(c)一等賞をマラソンで太郎が日曜に取った．

Figure 11 the utterance sentences by GP

There is an utterance sentence of "Taro is very pleased" as a situation before the utterance (1), and the EGC assumes that the audience has a good feeling to Taro.

For this case frame representation, we apply the ADF-GP method once. As a result, we obtained the new case frame representations in Figure 10. The obtained Japanese sentences corresponded to case frame representations in Figure 10 are shown in Figure 11.

In (a), GP is applied to the part of the subject and the subordinate clause is generated. The fitness values are determined as follows.

f(The order of presenting each element) =+0.017*6=+0.1
f(subordinate clause)=-0.1
f(operation predicate)=-0.1
f(independent word)=0.1
f(pleasant)=+0.1

Therefore, 0.5+0.1+(-0.1)+(-0.1)+(0.1)+(0.1)=0.6

For (b) and (c), the fitness value are 0.7 and 0.532 respectively in the same way.

5 Application to emotional interaction system

An increase of elderly persons requires not only medical cases but also health services. The Japanese society has already built up the medical insurance and medical care systems. However, the health service needs in so-called healthy elderly are not analyzed. We developed the www-based analytical system of health service needs of elderly. The basic questionnaire consisted of 50 items. These items related to QOL, -Quality of Life-, life-satisfaction, life-style, mental stress, and social concern. The subject tried to answer these questionnaire sheets in the homepage and these answers are sent to website.

Successfully, the system checks if the questionnaire were fulfilled all conditions and were stored to implement a reasoning. Based on such calculations, the system presented the analytical results and health counseling comment to the subject in the browser.

The system was also developed to analyze the population-based health service needs to the official health center. The health service for the elderly persons will be offered on the basis of these results. The system is expected to classify the health service needs of the elderly in his/her home.

Although the personal computer diffusion in Japan reaches up to about 30% according to the statistical data of Economic Planning Agency of Japan, not a few people have difficulties with the use of computer. Especially, we often see that the elderly tends to think the use of the computer to be difficult. The distance between the conversation and the usability of the computer causes this kind of problem. That is, they hope the computer equipped with human-like interfaces, which enable an easy conversation like the greeting and so on. Besides verbal messages, human face-to-face communication usually nonverbal messages such as facial expressions, vocal inflection, speaking rate, pauses, hand gestures, body movements, postures, and so on. In this study, to improve this weak point of the www-based health service system, we developed the dialog system with the generation of utterance sentences.

Figure 12 shows the interface of our developed dialog system on web site.

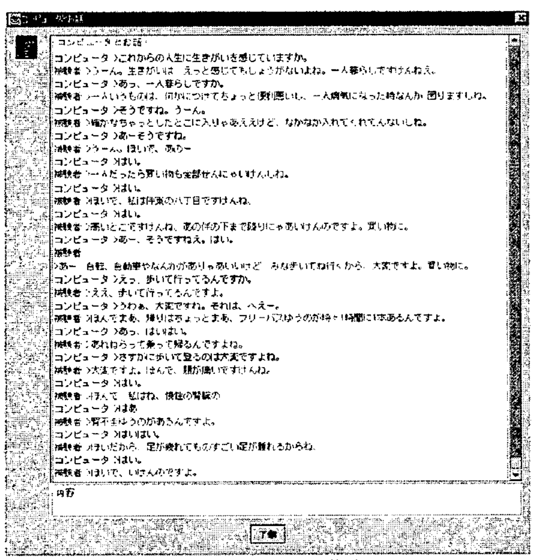

Figure 12 the developed dialog system

6 Conclusions

The proposed method gives a pair of the conversation, which is an initial group to GP, is actually given based on the conversation with elderly peoples. The conversation includes the background such as individual daily life. Therefore, we consider that the system needs the algorithm by which an individual background is acquired from the conversation automatically and we are planning to develop it.

References

[1]J.Koza, "Genetic Programming", The MIT Press(1992)

[2]H.Iba, "Genetic Programming", Tokyo Denki University Publishers (1996)(in Japanese)

[3]Takumi ICHIMURA, Kazuya MERA and Toshiyuki YAMASHITA, "Construction of a dialog system with emotions for elderly persons by neural networks", Proc. of Intl. Conf. on IEEE SMC, pp.3594-3599 (2000)

[4]J.Koza, K.E.Kinner(ed.), et al., "Scalable Learning in Genetic Programming Using Automatic Function Definition", Advances in Genetic Programming, pp.99-117(1994)

[5]K.Araki et al., "Application of Genetic Algorithms for Examples-Based Machine Translation Method Using Inductive Learning and Its Effectiveness", Information Processing Society of Japan, Vol.37, No.8, pp.1565-1579 (1996)

[6]S.Nagao, "Natural Language Processing", Iwanami Publishers (1996)

[7]Mitsuko Yamura-Takei, Miho Fujiwara and Teruaki Aizawa (2001). Resolving and Generating Zero Anaphora in Japanese: a Language Learning Aid Perspective. Pacific Associations for Computational Linguistics (PACLING) (to appear in 2001)

[8] Kazuya Mera, Shinji Kawamoto, Kenji Ono, Takumi Ichimura, Toshiyuki Yamashita, and Teruaki Aizawa, "A learning method of individual's taste information", Proc. of KES2001(to appear in 2001)