

# Language Learning for the Autonomous Mental Development of Conversational Agents

Jin-Hyuk Hong, Sungsoo Lim, and Sung-Bae Cho

Dept. of Computer Science, Yonsei University  
134 Sinchon-dong, Sudaemoon-ku  
Seoul 120-749, Korea

{hjinh, lss}@sclab.yonsei.ac.kr, sbcho@cs.yonsei.ac.kr

**Abstract.** Since the manual construction of our knowledge-base has several crucial limitations when applied to intelligent systems, mental development has been investigated in recent years. Autonomous mental development is a new paradigm for developing autonomous machines, which are adaptive and flexible to the environment. Language development, a kind of mental development, is an important aspect of intelligent conversational agents. In this paper, we propose an intelligent conversational agent and its language development mechanism by putting together five promising techniques; Bayesian networks, pattern matching, finite state machines, templates, and genetic programming. Knowledge acquisition implemented by finite state machines and templates, and language learning by genetic programming are developed for language development. Several illustrations and usability tests show the usefulness of the proposed developmental conversational agent.

## 1 Introduction

Mental development is a primary human characteristic, and vision, audition, behavior, and language are the principal ingredients (Johnson & Munakata, 2005; Weng, 2000). Matured people are usually better than the younger in these areas, since they are more mentally developed. Similarly, an intelligent system that has a richer and more sophisticated knowledge base might be better than other systems. It is, however, very difficult to construct any kind of knowledge base during the initial stage of the manual development of the system (Lauria, 2001). Manual construction is apt to make a system static and stiff toward its environment. Therefore, recent research on intelligent systems has focused on an autonomous mental development (AMD) mechanism. Like humans, this mechanism develops its mental capabilities through autonomous real-time interactions with its environment. This includes interactions with human supervisors. AMD, a new paradigm for developing autonomous machines, constructs an intelligent system that is based on incremental learning (Joshi & Weng, 2003) and is adaptive and flexible to the environment.

Linguistic sense is highly related to mental development. Since adults generally accumulate more knowledge and skill than children, they can generate more sophisticated and informative sentences (Clack, 2004). A conversational agent, a popular type of intelligent system, provides users with informative answers based on such a linguistic sense. Since experiments are commonly based on pattern-response pairs, the

level of the knowledge base determines the quality of the response. That is, if the knowledge-base of the conversational agent is composed of lots of qualified pattern-response pairs, the agent works well enough to respond to the user's queries. Autonomous Language Development (ALD) of a conversational agent is different from the conventional manual approach.

Computational implementations of learning mechanisms provide language acquisition researchers with important clues as to how the brain starts to use language (Plunkett, 1997). Moreover, these implementations promote the development of intelligent systems that interact with humans by using a realistic and scalable conversational interface. Lauria *et al.* proposed an instruction-based learning method for personal robots (Lauria, 2001), while Dominey and Boucher developed a system for language learning. Zhou proposed a robot learning method with GA-based fuzzy reinforcement learning (Zhou, 2003), while Jerbic *et al.* examined an autonomous agent based on reinforcement learning and an adaptive shadowed network. Joshi and Weng introduced the concept of AMD and studied the autonomous learning of speech production under reinforcement learning (Joshi & Weng, 2003).

## 2 The Developmental Conversational Agent

### 2.1 Conversational Agent Architecture

The proposed conversational agent is composed of three parts: language interpretation, language generation, and language development. Language interpretation infers the intention of the user's queries, while language generation generates answers that correspond to those queries. When it does not understand the query, language development learns the pattern by interacting with the human. Language development also

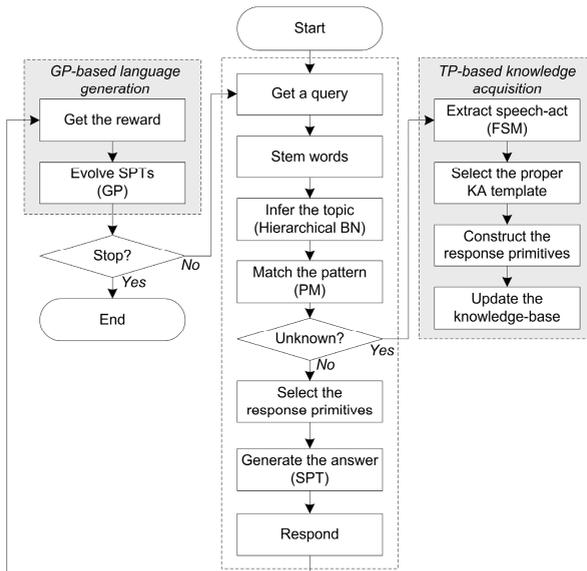


Fig. 1. The process of the proposed conversational agent

evolves its linguistic sense for generating answers by using genetic programming. Fig. 1 shows the management of the input query. A pattern is analyzed and accumulated into the knowledge base, while the answer is used to evaluate the genetic programming that produces the sentence plan trees (SPTs).

### 1) Two-stage inference for dialogue management

At first, words that infer the user's intention are extracted from the input query. This inference is determined by two methods: the Bayesian network and keyword matching. The Bayesian network infers the user's intention, while keyword matching selects a proper pattern-response pair for generating the corresponding answer. Analyzing queries in these stages makes it feasible to infer the detailed intention of the user and to model the context of the conversation. Furthermore, dividing the knowledge base improves the scalability and portability of the conversational agent.

**Topic Inference using the Bayesian Network:** The Bayesian network is used to infer the topic of a user's query and model the context of dialogue. This leads to a definition of the scope of the dialogue. Since Bayesian networks are based on graph theory, they are effective in inference and representation. In this paper, the Bayesian network is hierarchically constructed with three levels based on function: keyword, concept, and topic. The keyword level consists of words related to topics in the domain. The concept level is composed of the entities or attributes of the domain, while the topic level represents entities whose attributes are defined.

**Response Selection using Keyword Matching:** Once the topic is selected for an input query, keyword matching (using the knowledge base associated with the topic) is performed to find the corresponding response. When there are many scripts, performance declines because of the redundancy of information. In this paper, we divide the scripts into several groups based on their topics. This reduces the number of scripts to be compared. A script is stored as an XML form. A set of candidate scripts are sequentially matched to find an appropriate response. The matching scores are calculated by the F-measure, which is a popular measurement in text classification. When there is a corresponding pattern-response pair, language generation is used to generate the answer.

### 2) Answer generation using the sentence plan trees (SPTs)

Once a pattern-response pair is selected, language generation is applied to the corresponding response primitives. In language generation, a SPT is selected and filled with response primitives to construct an answer. A SPT is a binary tree with leaves labeled by pre-defined templates of simple sentences (SSs), and with interior nodes labeled with joint operators (JOs) that combine two sentences (Ratnaparkhi, 2002). We define five types of JOs (suitable for the Korean language) to combine 2 sentences, as follows.

Sentence A = subject (s1) + response primitive (t1) + verb (v1).

Sentence B = subject (s2) + response primitive (t2) + verb (v2).

- **JO 1:** Combine sentences A and B by using the conjunction 'and'. The result is 's1 t1 v1, and s2 t2 v2.'
- **JO 2:** Combine two sentences that have the same subject (s1 = s2). The result is 's1 t1 v1 and t2 v2.'

- **JO 3:** Combine two sentences that have the same subject ( $s1 = s2$ ) and the same verb ( $v1 = v2$ ). The result is ‘ $s1\ t1\ t2\ v1.$ ’
- **JO 4:** Combine two sentences that have the same response primitive ( $t1 = t2$ ) and the same verb ( $v1 = v2$ ) by making a proper subject ( $s3$ ). The result is ‘ $s3\ t1\ v1$ ’ where  $s3$  is the subject that includes  $s1$  and  $s2$ .
- **JO 5:** Combine two sentences that have the same subject ( $s1 = s2$ ) and different verbs but can be replaced by a verb  $v3$ , which includes both  $v1$  and  $v2$ . The result is ‘ $s1\ t1\ t2\ v3.$ ’

## 2.2 Autonomous Language Development in the Conversational Agent

### 1) Dialogue-based knowledge acquisition

For the developmental conversational agent, knowledge acquisition is accomplished (as shown in Fig. 1) through dialogue. It first analyzes the pattern of the input query and then constructs the response through interaction with a human. The pattern of a script is composed of the topic and a set of words. The topic is obtained by the Bayesian network, while the words are extracted by preprocessing.

We consider the dialogue act of an input query to select a proper answer template that is used to collect the response primitives. An automaton extracts a dialogue act, and then 30 automata are designed for 30 dialogue acts, as shown in Table 1. A subsumption architecture is adopted to select one dialogue act for per query.

**Table 1.** Dialogue acts defined in this paper

Category	Dialogue act
Question	Ability, Description, Fact, Location, Method, Miscellaneous, Obligation, Reason, Time, WhatIf, Who, Acquisition, Comparison, Confirmation, Cost, Direction, Example, More, Possession
Statement	Act, Fact, Message, Miscellaneous, Possession, Status, Want, Cause, Condition, Feeling, Time

Each dialogue act has corresponding answer templates. This template technique is useful to represent static information in situations where variables change dynamically. In the answer template, “class” means the dialogue act, and “question” is a sentence to collect the response primitive. “Requirement (information collected from the user)” is the response primitive for constructing the answers. We define 64 templates: three templates for each question dialogue act, and six positive/negative templates for each statement dialogue act. Finally, the response is completed with the template and saved as the response part of the script. The pattern-response pair is generated with the patterns and the responses.

### 2) GP-based language generation

Genetic programming, as proposed by John R. Koza, is used to generate the adaptive structure of answers to the domain (Koza, 1994). An individual represented as an SPT denotes a structure of answers. As mentioned before, the leaf node of an SPT contains an SS that corresponds to the response primitive. The parent nodes represent JOs. SPTs can evolve by genetic programming when humans are involved.

An individual exposed to genetic programming is represented as an SPT that consists of the joint operator set  $\{JO1, JO2, JO3, JO4, JO5\}$  and the response primitive

set  $\{r_1, r_2, \dots, r_n\}$  where  $n$  is the number of response primitives. The grammar for the SPT is:  $G = \{V = \{EXP, OP, VAR\}, T = \{JO1, JO2, JO3, JO4, JO5, r_1, r_2, \dots, r_n\}, P, \{EXP\}\}$ , where the rule set  $P$  is as the following:

$$\begin{aligned} EXP &\rightarrow EXP\ OP\ EXP\ | \ VAR \\ OP &\rightarrow JO1\ | \ JO2\ | \ JO3\ | \ JO4\ | \ JO5 \\ VAR &\rightarrow r_1\ | \ r_2\ | \ \dots\ | \ r_n \end{aligned}$$

Crossover and mutation are employed to generate diverse sentence structures. Crossover randomly selects and changes sub-trees from two individuals, while mutation changes a sub-tree into a new one. These processes are conducted according to predefined probabilities. Through the evolution in real time, the system develops its linguistic sense for generating sentences.

### 3 Experimental Results

#### 3.1 Implementation

To show the usefulness of the proposed method, we have developed an artificial secretary system using MFC. It consists of a main window for displaying information, an input text box, and the avatar system with a speech generation engine. When a user types a query, the avatar provides the answer in speech with a corresponding action. Q-avatar<sup>1</sup> is employed as the avatar system, while ‘Voiceware<sup>2</sup>’, a solution for speech recognition and generation, is used to provide user a realistic and convenient interface.

The target domain of the genetic programming of the proposed method is travel planning, and Table 2 describes six response primitives of the domain. When the user’s query arrives, it extracts the response primitives from the query to collect information for planning a train trip. There are six SPTs based on the number of response primitives collected. If there is no information on travel, the agent uses the

**Table 2.** Response primitives used by a travel planning agent

Response primitive		Simple Sentence
Departure Location	Question	Where are you leaving from?
	Statement	You leave [dLocation].
Arrival Location	Question	Where are you going?
	Statement	You are going to [aLocation].
Departure Date	Question	What day do you leave?
	Statement	You are leaving on [Date].
Departure Time	Question	What time do you leave?
	Statement	You leave at [dTime].
Arrival Time	Question	What time do you want to arrive?
	Statement	You arrive at [aTime].
Seat Type	Question	What kinds of seats do you want?
	Statement	You’ll take [tGrade].

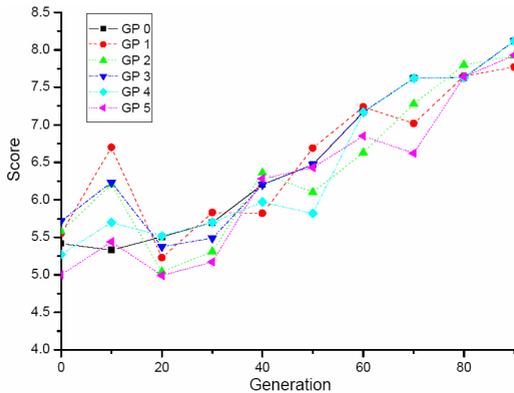
<sup>1</sup> <http://www.qavatar.com>

<sup>2</sup> <http://www.voiceware.co.kr>



**Table 3.** Results of the usability test for autonomous linguistic mental development model

		Manual construction		Proposed method	
		Experts	Novices	Experts	Novices
Pattern-response pairs generated		50	50	44	44
File size		28 KB	30 KB	24 KB	24 KB
Construction time		20 min.	1 hour	5 min.	8 min.
Accuracy	Training	92%	84%	100%	96%
	Test	84%	82%	88%	86%
Feedback (0~5)		2	1	4	4



**Fig. 3.** Score distribution through evolution

Fig. 3 presents the changes of the average score as the generation grows. GP  $n$  means the STP whose number of response primitives given is  $n$ . We limited the generation to 90 steps because more steps cause overfitting and a decrease in diversity. As a result, the score increased rapidly during the first 10 generations and decreased between the 10th and 20th generations, then increased continuously after the 20th generation.

The system used the responses primitives of departure location, arrival location, and departure data. The initial sentence was disorderly and complex, while the sentence obtained through evolution became more refined.

- **Generation 0:** You leave Seoul on October 10th and what time do you want to arrive? And what time do you leave? And you are going to Pusan.
- **Generation 30:** What kind of train do you want to take and what time do you want to leave? And you leave Seoul on October 10th.
- **Generation 60:** What time do you leave Seoul? And you are going to Pusan.
- **Generation 90:** What time do you leave Seoul for Pusan?

## 4 Conclusion

We have proposed a conversational agent with autonomous language development. Since the manual construction of intelligent systems presents several problems, it was

necessary to propose a developmental conversational agent that might adapt itself to the environment. In language development, we addressed the autonomous increasing of the knowledge base and linguistic sense obtained by genetic programming. In this paper, various illustrations and usability tests demonstrated the usefulness of the proposed conversational agent.

Language is very important in mental development, and people improve their linguistic sense of understanding queries and generating sentences. Research on autonomous language development might not only be useful to demonstrate various cognitive language models in cognitive science, but also provide a direction in engineering to construct an effective communication method between humans and intelligent systems. In our future work, we will develop an intelligent system that includes autonomous mental and language development.

## Acknowledgement

This research was supported by Brain Science and Engineering Research Program sponsored by Korean Ministry of Commerce, Industry and Energy.

## References

- Clack, E. (2004). How language acquisition builds on cognitive development. *Trends in Cognitive Science*, 8, 472-478.
- Johnson, M., & Munakata, Y. (2005). Processes of change in brain and cognitive development. *Trends in Cognitive Sciences*, 9, 152-158.
- Joshi, A., & Weng, J. (2003). Autonomous mental development in high dimensional context and action spaces. *Neural Networks*, 16, 701-710.
- Koza, J. (1994). *Genetic programming, Automatic discovery of reusable programs*, The MIT Press.
- Lauria, S., et al. (2001). Personal robots using natural language instruction. *IEEE Intelligent Systems*, 16, 38-45.
- Plunkett, K. (1997). Theories of early language acquisition. *Trends in Cognitive Science*, 1, 146-153.
- Ratnaparkhi, A. (2002). Trainable approaches to surface natural language generation and their application to conversational dialog systems. *Computer Speech and Language*, 16, 435-455.
- Weng, J., et al. (2000). Autonomous mental development by robots and animals. *Science*, 291, 599-600.
- Zhou, C. (2002). Robot learning with GA-based fuzzy reinforcement learning agents. *Information Science*, 145, 45-68.