FUNDACION Dr.J.Mas

To be published in Proceedings of ICONIP2000(International Conference On Neural Information Processing), 2000 Nov. 14-18, Korea

A revised study and endorsed by the Foundation Dr.J.Mas,
for the importance of the contribution of this report.

Simulation of Human Language Acquisition Process by Brain Like Memory System

Takashi Omori
Graduate School of Engineering
Hokkaido University
omori@complex.eng.hokudai.ac.jp

Takayuki Shimotomai
Graduate School of Engineering
Tokyo Univ. of Agriculture & Technology
taka@omlab.base.tuat.ac.jp

Abstract

The acceleration phenomenon of infant word acquisition cannot be explained by simple neural learning mechanism of brain. To explain it, we applied memory model PATON for word meaning acquisition, and modified the model to include SRN to enable syntax learning. The combination of PATON meaning representation and SRN syntax representation enabled explanation of the syntax based word acquisition acceleration. The computer simulation has shown rapid acquisition of new word within a few presentations.

1. Introduction

The understanding of human language acquisition is, not only for scientific interest, expected to be a breakthrough for higher-level intelligent machine realization. For this purpose, we focus on infant language acquisition process and try to reproduce the same process by brain like neural network model. Then we aim to understand the process by comparing the modeled with real infant one.

Word acquisition, syntax acquisition, and pragmatics acquisition are main paradigm of language development research. In this study, we focus on infant primary word acquisition process through ungrammatical conversation with parents, and try to model acceleration of the acquisition process. As the vocabulary spurt phenomenon is not seen other than human, the study has possibility to reveal essential difference on symbol processing between human and other animals.

For the syntax acquisition, Elman has applied Simple Recurrent Network (SRN) for learning of English syntax [1][2]. However, SRN requires a lot of time for additional learning and do not represent meaning other than statistics of word use. It does not satisfy requirement of infant model that accelerate word learning speed by experience.

On the other hand, Omori et al. proposed a memory model of sensory signal based concept representation and its manipulation that is possible for animal brain [3]. PATON model is consisted of two layered memory network and a control system that control the memory manipulation by sequential attention (Fig.1). It can represent multi-modal concept and its operation of recognition, recollection and association with context. As the result, the model realizes inference looking macroscopic behavior using the concept.

For the language representation and acquisition, Omori and Nishizaki applied PATON model on answer retrieval behavior acquisition through un-grammatical conversation by reinforcement learning [4]. They also applied the model to the acquisition of functional word "what", concept of "color" or "shape" and their instance words through two words sentence [5]. However, the order of word, syntax, is given apriori in the study. Current PATON model can not represent syntax. And more, its learning is based on probabilistic search and can not

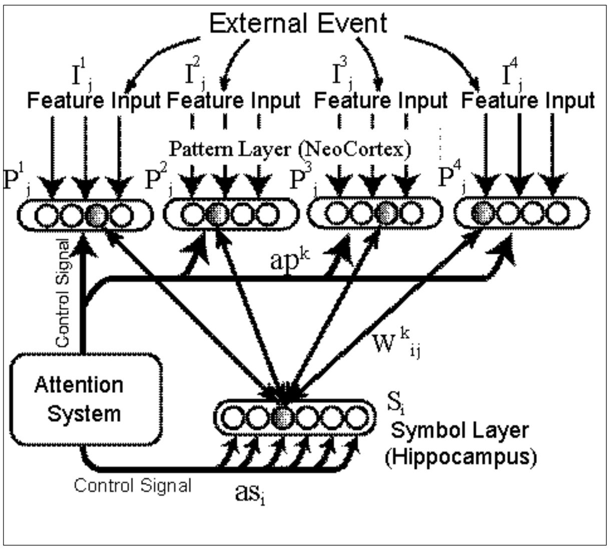


Fig.1 Structure of PATON model.
The two layer structured memory network consists of sensory attribute areas that encode variety of sensory modality and a symbol layer that associates them. The attention system operates on them to control their behavior.

explain the acceleration.

In this study, we propose a model of word acquisition acceleration based on combination of PATON concept manipulation and SRN syntax learning. The SRN part of the model learns syntax from given sentences, and new learning action in PATON is triggered by the syntax situation. Then, learning time for new word in the sentences is evaluated. The simulation result shows that the addition of simple rule over conventional rule drastically accelerates learning of new word.

2. PATON model

In brain sensory processing system, each modality signal is preprocessed independently, and an outer world event is represented as a set of features in modality specific associative areas. Those features are then converge to hippocampus area and form an association memory that corresponds to the outer event. The process corresponds to the structure of the memory network in Fig.1. And more, we assume bi-directional connection between the attribute areas and the symbol layer that form association between multiple modalities.

On the other hand, the old areas in brain, such as limbic system, have plenty of projection to neo-cortex that injects neuro-modulators. In cortex, it is known also that areas are activated/inactivated depending on task. It is relevant that there are some mechanisms that control activity of functional modules in cortex.

From this observation, we introduce an attention that dynamically control activity of cortical areas including internal processing ones. In PATON model, the attention controls state update and output gain of each area, and the set of control signal forms a vector that designate behavior of whole model circuit of the moment. We also assume that the attention vector changes sequentially. As the result, the attention vector and its sequence decides signal flow from input to output for a task. If we change the attention depending on a recognition result, the system becomes a universal circuit that changes its structure depending on task and situation.

The basic behavior of PATON model that is realized by the attention vector includes (1) recognition of input, (2) recollection of attribute that is associated to a recognition, and (3) association between memorized items. The specific feature of PATON model is the ability that changes its circuit that joins the operation of the moment by attention. Omori et al. showed that the behavior is compatible with finite state automaton from the theory of associative memory.

PATON model is also applied to self-organizing acquisition of environmental map in moving robot and planning of its route path in the map [6]. The path planning is typically thought of as a symbolic function in engineering sense, and the fact that the function is realized by brain like memory model PATON implies a possibility that so called symbolic functions might be realized as continuous computation in brain. Then, how about language that is often said to be symbolic? This is the next interest.

3. Word acquisition by PATON model

3-1. Meaning representation by PATON

Elemental manipulation unit of PATON is a memory that is an association of sensory input in a moment. It is similar to episodic memory. Contrary, each word in language represents a concept. In general, episode is composed of a set of concept that forms a scene. Concept is not always associated to all of sensory modality. To extract concept from episode, we have to segment an event from back grounding events, and then cut off unnecessary attribute from the event. In PATON model, the process is modeled as a selective learning of the connection between the attribute areas and the symbol layer. It is realized by a combination of an attention vector and learning action[4]. Experimental evidence suggests that infant unconsciously select attributes depending on

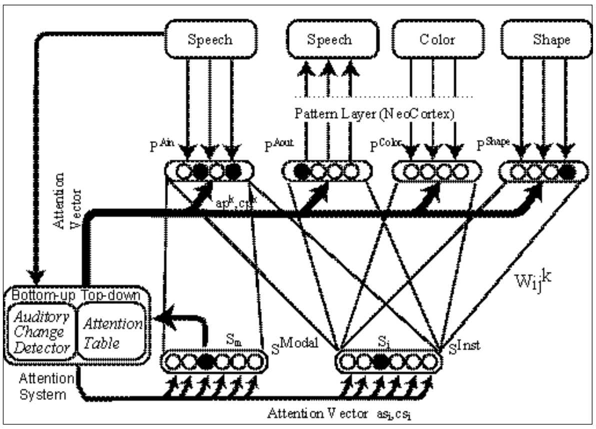
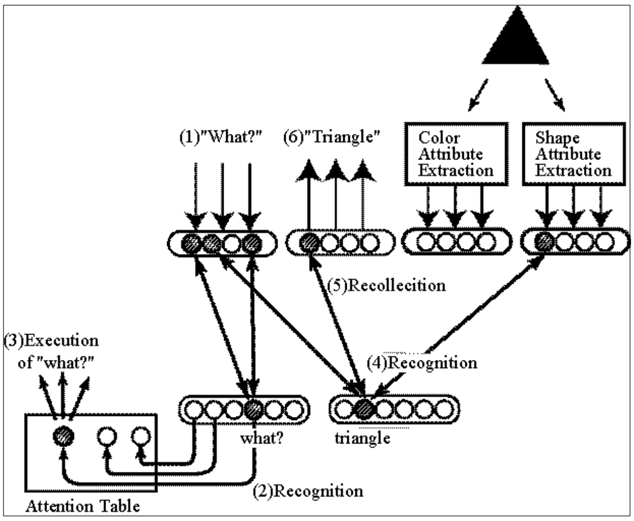
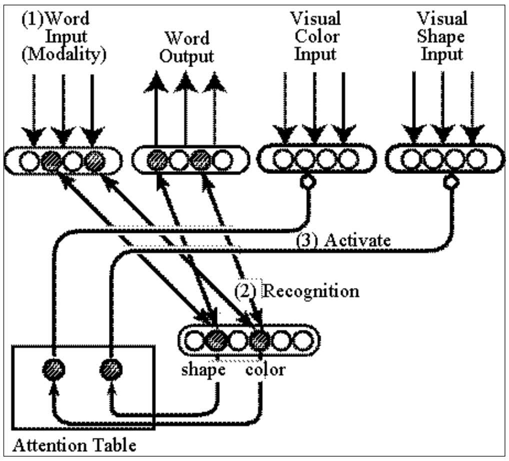


Fig.2 Addition of recognition driven attention driver. By reflective recognition of word input, an attention vector that is associated to the word is generated. The attention may work immediately, longer, or in temporal sequence



(a) Meaning of functional word “what”.
Given the word input “what”(1), a symbol cell recognize it and fires(2). An attention vector(3) that designates “recognize non verbal input on the moment (4) and generate pronunciation action of its associated name(5)(6)” is generated.



(b) Meaning of words “shape” and “color”.
Given the word “shape”(1), for example, it is recognized (2) and the activity generates a contextual attention that increases output gain of shape representing attribute area (3). The attention affects on next coming word driven action conduction, that is the context effect.

Fig.3 Word meaning representation as internal action

context when they learn new word [7].

The question here is how the memory is manipulated to realize some kind of problem solving or answer retrieval. Suppose a situation that a teacher and a child are looking at the same object, and the teacher asks to the child “What this?”, ”What color?” or “What shape?”. The child tries to answer. At initial stage, the child does not know what is the meaning of the word “what” and can not answer. After plenty of try and error, the child learns the meaning of “what” as an action of “recognize presented object, and say its name”. The same thing applies on words “color” and “shape”. Children acquire action to recognize color or shape by those words.

When we think just from this task, meaning of functional word “what” is same as the action explained. For two word sentence of “Color what?”, we can explain the answering mechanism by the combination of color information focusing action by word “color”, and recognition and reply action by following word “what” with the color context. That is, we can realize primary language understanding by activation of memory search and selective recollection of found memory with modification from context word. It is our problem to give interpretation of adult level higher intelligence [1]. Of course children acquire more general and essential word meaning with development. But it is natural to think that the acquisition of more general word use is rather later process compared to current one.

In PATON model, the response generation behavior is realized by the attention generating system that activate internal / output action by recognition of word input (Fig.2). The “what” action is realized by a combination of recognition on non-verbal input and recollection to speech production area (Fig.3 (a)). The effect of attribute designating words, like “color” or “shape”, is realized by long lasting attention to corresponding attribute areas (Fig.3 (b)). The problem is acquisition of those attention vectors from the conversations.

3-2. Probabilistic search by reinforcement learning

For the acquisition of attention vector, we assume that infants have instinct to feel pleasure when they success communication through conversation. Then, the attention search task becomes the reinforcement learning with immediate reward. For the search method of vector, we adopted random probabilistic search as most primitive one that requires minimum apriori knowledge on the target field.

[1]We give same interpretation on incomplete action of pet shaped robot

To be published in Proceedings of ICONIP2000(International Conference On Neural Information Processing), 2000 Nov. 14-18, Korea

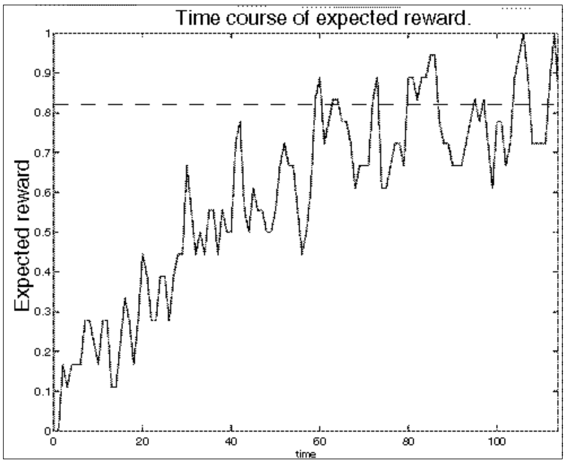


Fig.4 Learning curve of “what” acquisition task. Moving average of the reward by conversation success along learning time increases up to theoretical limit by random search.

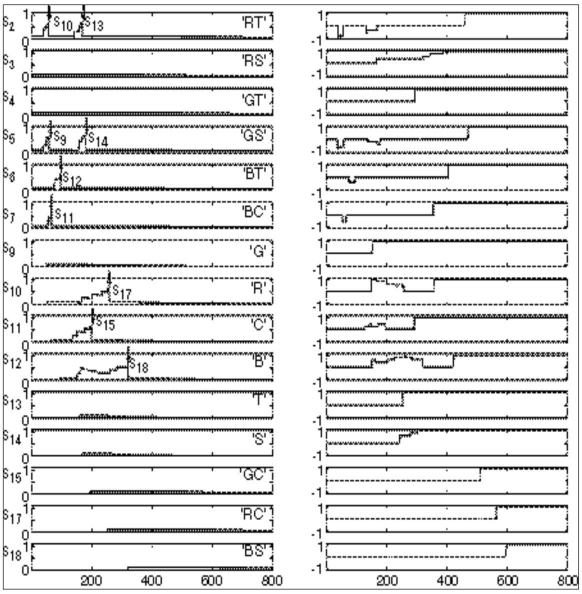


Fig.5 Concept acquisition process by the syntax constrained two words sentence. The syntax designates first word to indicates attribute, and the second to action. The search space became narrow, and words are acquired within rather small trial. For the initial knowledge, some words S2,…,S8 are assumed to be known in advance. By the two word sentence learning, new attribute designating word “color”, “shape”, and their instance words such as “green”(G) or “Triangle”(T) are acquired. Left half of the figure shows change of instability index for each word along learning time, and right half shows correct answering rate of left corresponding word. The arrows indicate timing of cell division by the instability index increase.

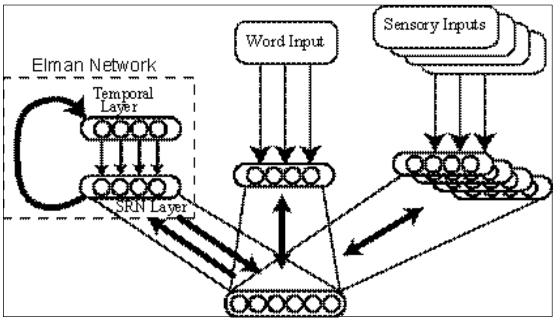


Fig.6 Addition of SRN on PATON attributes. The SRN layer is independent of external input, and has its internal delayed feedback. Right side overlapped areas are conventional sensory attributes. Word sequence that satisfy syntax in Fig.7 is given, and SRN and its related connection learns to predict next incoming word. The connection between symbol later and SRN is asymmetric.

For this task, we used PATON model that has four attribute areas of word input, color input, shape input and word output. Each of input area is given preprocessed input pattern, and the word output area make response to outer world when it is given a recollection pattern from the symbol layer. We also assumed recognition of word input “what” and some boosting up initial memory, such as Red Triangle (RT) or Green Square (GS) and so on, to enable initial conversation.

Given an input sensory signal and word “what”, the system generates attention vector that has learnable probability distribution, and PATON model makes action by the attention. When PATON has succeeded to recollect proper word output pattern to the corresponding attribute area as the result of the given vector, reward is given and the attention system increases the probability to produce the attention vector. Fig.4 shows moving average of the given reward along learning time.

Just same learning strategy applies to two word sentence of “Color what” and “Shape what”. Here, we assume a syntax that the first word designate attribute, and second word indicates action with context of the first word. In the initial stage of learning, PATON cannot recognize words “color” and “shape”, and does not know their meaning. Beginning from the state, PATON learns to recognize those words and seeks for the attention vectors that satisfy task requirement at the same time. The syntax works as constraint for search space and the reinforcement learning becomes

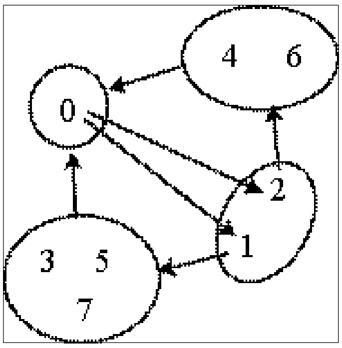


Fig.7 Simplified syntax for learning.
(0) indicates common first word. Third word groups (3,5) or (4,6) are decided by the second word (1,2). In the experiment, new word (7) is added to the group (3,5) after the initial syntax is learned, and the convergence process afterward is observed.

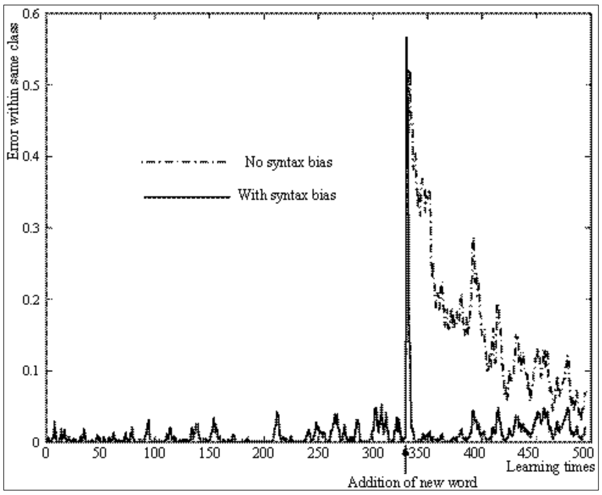


Fig.8 New word learning process by syntax extended PATON. New word (7) is added to group (3,5) at the timing of lower arrow. Dashed line afterward shows variance of next word prediction strength from SRN layer by back propagation learning rule. The solid line after the addition is the variance by new learning rule.

possible within limited time.

As the result of learning, the meaning of “color” and “shape” is represented as the internal action that designates attention to corresponding attribute area. And more, concrete color concept of “Red”, “Blue” etc. and shape concept, “Triangle”, “Square” and so on, are acquired at the same time (Fig.5).

To form concept from event memory, we need to cut off unnecessary attribute. In our model, we realized concept learning by evaluating stability of reward when one attribute is used for the recognition and action generation. When unnecessary attribute is used for the action generation, the result has high probability of error and it leads to larger reward instability.

3-3. Acceleration by syntax use

To try modeling of word acquisition acceleration by syntax acquisition, we extended PATON model to include SRN as one of attribute areas. In our model, the role of syntax in word acquisition is the estimation of new word class or meaning when the word appears in specific position of a sentence. Compared to the probabilistic search in last section, the syntax information immediately restrict meaning or role of new coming word, and system can assign the meaning to new word quickly. The extended SRN attribute layer has a recurrent type internal context layer, and does not has direct connection to/from external world (Fig.6). We call it syntax PATON.

In the learning task, word pattern sequence that obeys simplified syntax in Fig.7 is given to the word input layer of syntax PATON. When incoming word is new, the symbol layer quickly learns to recognize it by by competitive learning. The SRN learns to predict next coming word by back propagation rule. In the initial stage of learning, number of word in the sentence is limited. After the learning progress, SRN became to be able to predict next word. Then, new word is added to the sentence, and SRN re-learns to predict including the added word. The learning speed is rather slow with conventional BP learning rule, and we evaluate effect of new supplemental learning rule that make use of the syntax information.

In the learning rule, the system learns connection from SRN layer to symbol layer with larger coefficient when unexpected word is detected in a specific syntax position of input sentence. At the moment of unexpected word detection, the context layer of SRN is representing specific syntax situation, and the excitation pattern is used to predict next word through the connection from SRN layer to the symbol layer. So, the learning of connection from the SRN exciting neuron to the symbol layer new word representing neuron quickly assigns syntax information to the word.

As the PATON computation is continuous in time, the learning occurs when PATON dynamics converges to a stable state after input word is presented. In the syntax learning phase, word sequence set (0-1-3), (0-2-4), (0-1-5), (0-2-6) is given 2000 times counting each sequence presentation as one time. In the additional learning phase after that, a sequence (0-1-7) is added and SRN learns whole sequence set 1000 times. Fig.8 shows change of prediction strength variance to word (3,5) before the addition, and to word (3,5,7) after the addition. With conventional BP learning, the prediction strength does not converge for a while after addition. But it quickly converges to low value successing previous syntax knowledge when we use new learning rule.

4. Discussion

As most of readers has noticed already, our recognition and response generation mechanism are not restricted to language. There is a possibility that our method is applicable to other symbolic looking process such as problem solving or navigation. At the primary stage of learning, there is a possibility that language understanding has same basic mechanism with non language processing.

5. Conclusion

We explained memory model PATON as a neural network model of infant primary word acquisition process, and proposed the syntax PATON to explain acceleration of the acquisition process. Original PATON is a model to explain symbolization and manipulation of sensory information. In this paper, we extended PATON to encode pattern sequence rule and activate learning rule depending on the sequential situation. Though it is primitive, PATON became a model that can represent word meaning and syntax at the same time. As the result, we have succeeded to realize acceleration of word acquisition by supplemental learning of syntax in addition to original probabilistic learning strategy. Next problem of our model is evaluation. To evaluate this model, we need to confirm existence of similar process in infant and animal concept learning behavior.

References

[1] Elman J.L.: Distributed representations, simple recurrent networks, and grammatical structure, Machine Learning, 7, 195-225, 1991

[2] Elman J.L.: Learning and development in neural networks: the importance of starting small, Cognition, 48, 71-99, 1993

[3] Omori T., Mochizuki A., Mizutani K., Nishizaki M.: Emergence of Symbolic Behavior from Brain Like Memory with Dynamic Attention, Neural Networks, Vol.12, No.7-8, p1157-1172, 1999

[4] Omori T., Nishizaki M.: Representation and Learning of Word Meaning Using Memory Model with Dynamic Attention, Second International Conference on Cognitive Science 99, 189-194, 1999

[5] Omori T., Nishizaki M.: Incremental Knowledge Acquisition Architecture that is driven by the Emergence of Inquiry Conversation, Proc. of IEEE System Man & Cybernetics 99,pp219,1999

[6] Mizutani K., Omori T.:On-line Map Formation and Path Planning for Mobile Robot by Associative Memory with Controllable Attention,Proc. of IJCNN'99, 1999

[7] Imai, M.: Constraint on word learning constraints. Japanese Psychological Research, 41(1) 5-20, 1999.