Growing Self-Organizing Map Approach for Semantic Acquisition Modeling

Mengxue Cao*, Aijun Li*, Qiang Fang* and Bernd J. Kröger†
*Laboratory of Phonetics and Speech Sciences, Institute of Linguistics
Chinese Academy of Social Sciences, Beijing, China
Email: mengxuecao@outlook.com, liaj@cass.org.cn, fangqiang@cass.org.cn
†Neurophonetics Group, Department of Phoniatrics, Pediaudiology, and Communication Disorders, Medical School
RWTH Aachen University, Aachen, Germany
‡Cognitive Computation and Applications Lab., School of Computer Science and Technology
Tianjin University, Tianjin, China
Email: bkroeger@ukaachen.de

Abstract—Based on the incremental nature of knowledge learning, in this study a growing self-organizing neural network approach for modeling the acquisition process of semantic features is proposed. The Growing Self-Organizing Map (GSOM) algorithm is extended and applied to the problem of language acquisition. Based on that algorithm, experiments are conducted using Standard German children’s books corpus. A cyclic reinforcing and reviewing training procedure is introduced to model the teaching and learning process between children and their communication partners. Experimental results indicate that (1) GSOM has good ability to learn the semantic categories presented within the training data, that (2) clear semantic boundaries can be found in the network representation, and that (3) cyclic reinforcing and reviewing training leads to a detailed categorization of lexical items as well as to a detailed clustering, while keeping the already-learned clusters and already-developed network structure stable. Experiments show that our GSOM approach is a good method for modeling semantic learning during language acquisition.

Keywords—Neural network, growing self-organizing map, semantic feature map, language acquisition

I. INTRODUCTION

During language acquisition, infants face the task of learning various kinds of information and of organizing that information into linguistic categories. During that process, however, infants do not receive explicit language instructions, nor are they able to make inquiries about the structure that they are learning [1]. Instead, they must discover the linguistic categories of their native language from their interactions with communication partners. The task is further complicated by the fact that they do not know how many categories to discover along any particular input dimension [1]. The appropriate underlying learning mechanisms for infants, however, are still unclear. In this study we propose a feasible approach that it is capable to explain the acquisition process of semantic categories. That approach can contribute to the merging process of Information and Communication Technologies [3] and Cognitive Infocommunications [2] to some extent by modeling the teaching and learning process between infants and their communication partners.

Kohonen [4] proposed a kind of self-organizing neural network, known as the Self-Organizing Map (SOM), which has the ability to project high-dimensional data onto a two-dimensional feature map. Its high visualizing feature enables the analyst to overview the category structures of a data set. Later, Ritter and Kohonen applied the SOM algorithm onto semantic tasks [5]. Their research revealed that SOM has the ability to detect the “logical similarity” between words and group similar words into clusters.

In recent years, the topographic perceiving feature and the self-organizing ability of SOM have been applied increasingly for tasks on modeling the human acquisition process. In linguistic field, for example, Kröger et al. [6] modeled the acquisition of vowel and consonant categories using SOM. Zinszer et al. [7] developed a SOM model which models the first language lexical attrition phenomenon, and Warlaumont et al. [8] investigated the reinforcement effect on vocal motor learning based on SOM.

Although SOM is reasonable for modeling the topographic structure and the knowledge reorganization of a learning process, it does have limitations in modeling the incremental nature of knowledge growth. Due to the phenomenon of catastrophic interference [9], SOM has difficulties in adapting new knowledge into an existing trained network. In other words, the structure of SOM cannot be extended easily, thus cannot be directly used to model the knowledge learning process realistically.

Many researchers made an effort aiming to overcome that problem, such as the Growing Cell Structures (GCS) proposed by Fritzke [10] and the Neural Gas Algorithm (NGA) proposed by Martinetz and Shulten [11]. However, GCS has limitations in representing high dimensional data, while NGA is restricted by its limited map size. Those algorithms, therefore, are not suitable for modeling linguistic categories.

Exploring extendable SOM in data mining field, Alahakoon et al. [12] proposed an extendable version of SOM called the Growing Self-Organizing Map (GSOM), which let new nodes smoothly join the existing network and dynamically extend the size of the network. Its dynamic structure was proved to be very effective for knowledge discovery applications.

In this paper, that Growing Self-Organizing Map algorithm is adapted for language acquisition modeling.
II. GROWING SELF-ORGANIZING MAP

A. Structure of Growing Self-Organizing Map

Compared with traditional SOM, the structure of Growing Self-Organizing Map (GSOM) is simpler. Instead of having a size-predetermined rectangular map, the network of GSOM does not have a fixed size or shape. Starting from 4 initial nodes, new nodes can grow at boundary nodes and smoothly join the existing network (see Fig. 1b). Thus, the network can be dynamically expanded to any direction outwards depending on the new growing nodes (see Fig. 1a).

Two factors, the accumulative error (AE) and growth threshold (GT), are introduced into GSOM. The error value is calculated by the Euclidean distance between an input vector and the weight vector of the best matching unit (BMU). Thus, each BMU node has an error value as an additional characteristic parameter, and its value is accumulated throughout the training process. When the AE value of a BMU exceeds the value of GT, that corresponding Voronoi region [13] is said to be underrepresented, and new nodes are then introduced to the network. A high GT value will result in a less-spread map, while a low GT value will produce a well-spread map [12].

B. Basic Training Process

Training data (i.e. A set of N training items consisting of feature vectors) are treated as input tokens for the network training process. The basic training process of Growing Self-Organizing Map contains 2 phases:

1) Initializing Phase: At the beginning, 4 neurons are initialized. Their weight vectors are assigned randomly within interval 0 and 1, and their AE values are initialized to 0. Instead of what is stated in [12], in our approach, GT is set arbitrary to fit our experimental requirements.

After initialization, all starting nodes are boundary nodes and thus are free to grow in any direction outwards. That results in great flexibility concerning network growth.

2) Growing Phase:

a) : Input tokens are presented to the network one by one sequentially. Each token will be trained for several times before next token enters. That is consistent with the gradual learning process in practice: parents often teach their children just one word at a time, and repeat it for several times.

b) : In GSOM, the weight update and network reorganization processes are performed locally. Therefore, learning rate and neighborhood size are initialized to their initial value with each new input token. The learning rate is defined as a function of the total number of nodes in the network by (1), where $\alpha$ is the reduction factor of learning rate with $0 < \alpha < 1$; $\varphi(n) = 1 - R/n(t)$ is a function of the total number of nodes in current network; $R$ is a constant set to 3.8 as in [12]; and $n(t)$ is the number of nodes in the network at time $t$.

$$LR(t + 1) = \alpha \times \varphi(n) \times LR(t)$$

(1)

c) : The distances between weight vectors and training vectors are calculated using Euclidean distance measure, and the best matching unit (BMU) with minimum distance is detected within the network for the current training token.

d) : If the AE value of this BMU exceeds GT and this BMU is a boundary node, then new nodes grow at all free direct neighboring positions. Then weight vectors of new nodes are initialized with regard to the weight vectors of this BMU and its neighbors (that process is called weight distribution).

e) : If the AE value of this BMU does not exceed GT, then weight update is applied to this BMU and its neighbors within the neighborhood. Gaussian distribution is chosen as a part of the neighborhood function that can be represented by $h(t) = \beta \times \exp(-d_{ix}^2/2\sigma_i^2)$, where $\beta$ is the reduction factor of neighborhood size with $0 < \beta < 1$; $d_{ix}$ represents the distance between weight vector $i$ in BMU and training vector $x$; $\sigma_i$ represents the current neighborhood size. The weight update function can be expressed by (2). The Euclidean distance between this BMU vector and the training vector is accumulated as the AE value of this BMU.

$$\omega_i(t + 1) = \omega_i(t) + LR(t) \times h(t) \times (x(t) - \omega_i(t)), i \in N$$

(2)

f) : If the AE value of this BMU exceeds GT but this BMU is not a boundary node, then error distribution is performed. Then, the error value of this BMU is reduced to GT/2, and the error values of its immediate neighbors are increased by $\gamma GT$ ($0 < \gamma < 1$).

g) : Then several iterations are done for the current training token (go through steps c) to f) several times). The learning rate and neighborhood size decrease at each iteration. The iteration process stops when neighborhood size reduces to unity.

h) : Then, the next training token is processed by repeating steps b) to g) until all training tokens are presented.

C. Checking Process

The checking process does not change the network. It is performed to check whether the trained network has learned a good representation of categories represented within the training data, by identifying the winner positions in the trained network for each token. That can be considered as a calibration phase if known data are used [12]. The closeness of each token to each neuron in the network is measured by Euclidean distance.

Fig. 1: The initial structure of GSOM. (a) The network can be expanded to any direction at the beginning. (b) New nodes can grow the network at boundary nodes. [12]
D. Reinforcing and Reviewing Training

During language learning process, children cannot learn the whole knowledge (i.e. all semantic categories represented by the training tokens) at once, so that imperfections in clustering are inevitable. During semantic acquisition, those errors could represent a current incapability of distinguishing words with different meanings. Reflected in the network, that fact is represented by those neurons which represent many words after performing the checking process. This is comparable to the following natural learning situation: when parents teach their children, and if they find that their children are always confused by some words, parents will repeat those words and reinforce the differences between those words, in order to help their children get them distinguished. During that reinforcement process, some learned words will also show up in the communication between parents and children.

When growing phase is completed and checking is performed, the trained network may end up with some “unsolved” edge nodes, which means some nodes may represent the characteristics of many tokens comprising different semantic categories. In order to resolve those “high-density” nodes, a series of additional reinforcing and reviewing training can be performed.

During the reinforcing phase, the trained network from growing phase is the starting network. The training process is similar to that in the growing phase. Training data are consisted of those words represented by “high-density” neurons in the checking results. The initial learning rate is increased to give more weights to the input token, and the GT value is decreased to stimulate the network growth at “high-density” nodes.

During the reviewing phase, the trained network from current reinforcing phase is the starting network. The training process is similar to that in the growing phase. Compared with reinforcing phase, more tokens from the training set are used in the reviewing process to simulate the reoccurring of learned words. The initial learning rate and the GT value are set to the same as those in the growing phase.

A reinforcing phase followed by a reviewing phase together forms a combined training process of reinforcing and reviewing training. According to our experimental requirements, this combined training process repeats in a cyclic (or iterative) way for several times.

III. EXPERIMENT AND RESULTS

A. Standard German Children’s Books Corpus

In this study, the Standard German children’s books corpus [14] was used as the basis for generating our training set. The Standard German children’s books corpus comprises transcriptions of 40 books targeted to children from age 1 to 6. In total, 6513 sentences and 70512 words are transcribed in this corpus. Morphologically distinct forms of the same word are counted as separate words (e.g. “Blume” meaning flower and “Blumen” meaning flowers, are treated as two different words). The corpus therefore consists of 8217 different words, which is assumed to approximately represent a 6-years-old child’s mental lexicon. In this study, only nouns were used as training data, and only the first 332 most frequent nouns were chosen. The top ten frequent nouns are listed in TABLE I.

<table>
<thead>
<tr>
<th>Frequency counts in corpus</th>
<th>Frequency counts in training data</th>
<th>German words</th>
<th>English translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>392</td>
<td>78</td>
<td>Mama</td>
<td>mom</td>
</tr>
<tr>
<td>278</td>
<td>55</td>
<td>Bar</td>
<td>bear</td>
</tr>
<tr>
<td>235</td>
<td>47</td>
<td>Papa</td>
<td>dad</td>
</tr>
<tr>
<td>217</td>
<td>43</td>
<td>Mom</td>
<td>moon</td>
</tr>
<tr>
<td>190</td>
<td>38</td>
<td>Kinder</td>
<td>children</td>
</tr>
<tr>
<td>147</td>
<td>29</td>
<td>Katze</td>
<td>cat</td>
</tr>
<tr>
<td>145</td>
<td>29</td>
<td>Frau</td>
<td>wife</td>
</tr>
<tr>
<td>106</td>
<td>21</td>
<td>Bett</td>
<td>bed</td>
</tr>
<tr>
<td>105</td>
<td>21</td>
<td>Maidchen</td>
<td>girl</td>
</tr>
<tr>
<td>104</td>
<td>20</td>
<td>Wasser</td>
<td>water</td>
</tr>
</tbody>
</table>

B. Training Data Set

Two native speakers of Standard German (undergraduate students from RWTH Aachen University) developed a list of semantic features for the corpus by a simple brain storming procedure. In total, 1715 features were developed. To reduce dimensions of the training vectors, we dropped features which only occur once, while keeping all words distinguishable by their semantic features. Finally, 724 features were kept (the top ten features are listed in TABLE II). Therefore, each word in our training data set is represented by 724 feature vectors. Binary coding is used for the representation of each word. Thus, for each word, among its 724 vectors, “1” is used to mark its own features and “0” is used to mark features not belong to this word.

To model the semantic acquisition process, the frequency of word occurrence was taken into account when building the training set. The words were presented proportional to the frequency of occurrence of a word in the corpus (see the second column of TABLE I).

The 332 words arranged in sequence by descending order of their frequency counts constitute the basis of the training set. Based on that, repeated words were randomly inserted to the 332-word list regarding to the frequency counts of each word. To keep the frequency order, we exclusively allowed higher frequent words to be inserted after lower frequent words. For example, the first rank is “Mama” and then “Bär”. After “Bär”, “Mama” has 50% possibility to be inserted. Then the next rank is “Papa”. After “Papa”, “Mama” and “Bär” each has 50% possibility to be inserted. The insertion of certain words would be stopped when words reached their frequency counts. In total, 1929 word tokens comprised our training data set.

C. Experiment Procedure and Results

As introduced in Section II-B and Section II-D, basic growing training and reinforcing and reviewing training experiments were conducted respectively. The experiment process was divided into 21 training steps. Step 1 represents the basic growing training. The following 20 steps represent the steps in the cyclic reinforcing and reviewing training (from step 2, even numbers represent reinforcing training steps and odd numbers represent reviewing training steps). In total, one basic growing training and 10 cycles (20 steps) of reinforcing and reviewing training were performed in this experiment. The 1929-word

\begin{table}[h]
\centering
\caption{Top ten frequent nouns in the 332 word data set}
\begin{tabular}{|c|c|c|c|}
\hline
Frequency counts in corpus & Frequency counts in training data & German words & English translation \\
\hline
392 & 78 & Mama & mom \\
278 & 55 & Bar & bear \\
235 & 47 & Papa & dad \\
217 & 43 & Mom & moon \\
190 & 38 & Kinder & children \\
147 & 29 & Katze & cat \\
145 & 29 & Frau & wife \\
106 & 21 & Bett & bed \\
105 & 21 & Maidchen & girl \\
104 & 20 & Wasser & water \\
\hline
\end{tabular}
\end{table}
TABLE II: Top ten semantic features in the 332 word data set

<table>
<thead>
<tr>
<th>Frequency</th>
<th>German features</th>
<th>English translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>81</td>
<td>ist ein Gegenstand</td>
<td>is an object</td>
</tr>
<tr>
<td>63</td>
<td>hat zwei Augen</td>
<td>has two eyes</td>
</tr>
<tr>
<td>60</td>
<td>es gibt verschiedene Arten</td>
<td>there are different types</td>
</tr>
<tr>
<td>56</td>
<td>hat einen Kopf</td>
<td>has a head</td>
</tr>
<tr>
<td>47</td>
<td>hat eine Nase</td>
<td>has a nose</td>
</tr>
<tr>
<td>46</td>
<td>hat zwei Beine</td>
<td>has two legs</td>
</tr>
<tr>
<td>42</td>
<td>hat eine Haut</td>
<td>has a skin</td>
</tr>
<tr>
<td>37</td>
<td>ist aus Kunststoff</td>
<td>is made of plastic</td>
</tr>
<tr>
<td>37</td>
<td>ist aus Metall</td>
<td>is made of metal</td>
</tr>
<tr>
<td>36</td>
<td>ist ein Mensch</td>
<td>is a human</td>
</tr>
</tbody>
</table>

training set was used as the main training data, and the 332-word list (see Section III-B) was used as the testing set.

1) Basic growing training: First, a growing training was performed based on the training set of 1929 training tokens with 724 feature vectors in each. After training, a checking process was done by inputting the 332-word list with their feature vectors to the trained network, and winner neurons were found. The trained network structure and the checking result are shown in Fig. 2.

From Fig. 2, general clusters of semantic categories can be found, such as “Persons”, “Animals”, “Clothes”, “Transportation”, “Housewares”, “Snacks” and “Body parts”. That proves GSOM has the ability to learn the semantic categories in the training data and build semantic clusters even with such high dimensional training vectors. A separation between “Living” and “Dead” can be seen that “Living” things are located on the top-left and down-right part, while “Dead” things are located through the middle part.

In the network, some neurons represent more than one word. Taking a closer look at the position of those neurons, we find most of them are located at the edge areas of the network. With proper reinforcing and reviewing training, we expected those neurons at network edge areas could stimulate the network to grow further in that region in order to resolve those words represented by a “high-density” neuron.

2) Cyclic reinforcing and reviewing training: Based on the current trained network, a cyclic reinforcing and reviewing training was conducted. Each cycle contains two phases, a reinforcing phase and a reviewing phase. The trained network in each step was used as the starting network of next step.

For the reinforcing phase, in each current trained network, by examining the checking result, if a neuron represents more than 4 words, or the average Euclidean distance between the neuron and its represented words is bigger than 2.5, those corresponding words are then taken as the further training data. For simplicity, no frequency influence was considered in this phase. For example, 218 words were chosen for the first reinforcing training step. The trained network from the growing training was used as the starting network of the first reinforcing training step. The initial learning rate was increased to give more weights to the input token simulating the reinforcement from parents, and the GT value was decreased to stimulate the network growth at those “high-density” node regions.

For the reviewing phase, the starting network was the trained network from current reinforcing phase. The 332-word list comprised the training set. The initial learning rate and the the GT value were set to the same values as those of growing training in Section III-C1.

After each training step, the same 332-word checking (as already used for the growing training) was performed. The trained network structure and the checking result of the first reinforcing training phase are shown in Fig. 3. From Fig. 3, we can clearly see a growth at the down-left part of the network. Among those neurons, lots of “high-density” neurons have been solved, and only a few still exist. In addition, more words are represented across the retrained network. Some local reorganization can be noticed during this reinforcement training procedure. However, the learned clusters are clearly kept and no catastrophic interference is caused by the reinforcing training. In addition, some new clusters such as “Time”, “Plants”, “Places”, “Numbers”, “Names” and “Minds” are formed at the expanded parts of the network. Those results indicate that the reinforcing training can help the network distinguish “unsolved” words and build more detailed clusters while keeping the learned network structure stable.

The trained network structure and the checking result after 10 training cycles are shown in Fig. 4. From Fig. 4, we can see that the down-left part of the network in Fig. 3 has developed into a network with comparable size and shape of the upper part. Although very few “high-density” nodes still exist, the whole network are now very well developed. Although some reorganizations happen, the learned clusters and network structure in the network in Fig. 3 are well kept and not destroyed by the cyclic reinforcing and reviewing training.
The trained network structure and checking result after the first reinforcing training, and then the network keeps expanding itself, but the increasing rate gradually decreases. The curve in Fig. 6 declines with the training process after the first reinforcing training, and then gradually becomes stable. That reveals the process that the network occupies another nearby neuron region after the first reinforcing training and gradually forms into a compact network. The final network ends up with a reasonable size of 934 nodes, and a good neuron representation resolution of 279 words (84% of all words are resolved).

B. Semantic Representation

During the training process, the network gets gradually better semantic representation of words. Fig. 7 shows the maximum number of words represented by a single neuron. A generally declining trend can be found with the development of the cyclic reinforcing and reviewing training. Although some fluctuations can be noticed during the middle and late stages of the training process, the average words represented by a neuron in the network decreases continuously as shown in Fig. 8. That means the cyclic reinforcing and reviewing training can help the network resolve “high-density” nodes, thus can help children to disambiguate the meaning of words in the case of clustered words at one node.
training. The reinforcing training steps simulate a situation process was modeled by a cyclic reinforcing and reviewing their stable regions in the network.

the network structure can keep stable throughout the growing the knowledge learning process. The weight distribution passes factors, the network can produce a vivid biologic picture of algorithm based on the accumulative error and growth threshold learning process in practice. By introducing the GSOM al-

next, the training process of GSOM simulates the gradual in inputs one at a time and training it for iterations till the model the incremental nature of knowledge growth. By reading
to learn the semantic features in the training data and build semantic clusters even with such high dimensional training vectors. From the checking results, clear semantic boundaries can be found in the network neuron representation. Thus, cyclic reinforcing and reviewing training is proved to be able to help the network distinguish “unsolved” words and build more detailed clusters while keeping the learned clusters and network structure stable.

Although the GSOM is a highly abstract neurocomputation-mental model, it could be carefully interpreted as biologically plausible to a specific degree, because it incorporates important neurofunctional principles like self-organization, associative learning, Hebbian learning, adaptation, and neural plasticity.

ACKNOWLEDGMENT
This research was funded by the NSFC Key Project with No. 61233009, and the CASS Innovation Program.

REFERENCES