Automatic Information Categorization through Concept Formation of Associative Memory Model

TOSHIKI KINDO, HIDEYUKI YOSHIDA and TAKEHIKO SHIDA
Information & Network Research Laboratory
Matsushita Research Institute Tokyo, Inc.
3-10-1, Higashimita, Tama-ku, Kawasaki, Kanagawa 215-8501
JAPAN

Abstract: - From the viewpoint of practical applications, an important ability of neural networks is to organize representations of the external world through learning. There is, however, a problem that the representations of the neural networks are too complicated to estimate the capability of the assistant system in practical use. A geometrical method to analyze the representation of an associative memory is introduced. As a result, a practical application of associative memory models - information categorization- is presented. In this application, the concept formation ability of associative memory is important.

Key-Words: - information filtering, categorization, associative memory, concept formation, geometrical method

1 Introduction
In this paper, we introduce a method to analyze associative memory models and its practical application, an information filtering and categorizing system.

From the viewpoint of practical applications, an important ability of neural networks is to organize representations of the external world through learning. The ability suggests that we are able to realize personal assistant systems in various areas by making use of neural networks that have learned our needs. There is, however, a problem that the representations of the neural networks are too complicated to estimate the capability of the assistant system in practical use.

We have given a geometrical method to analyze not only the representation of an associative memory model but also its dynamics[1]. As a result, we reveal the relation between the structure of the external stimuli and the representation of the model. Based on our analysis, we develop an adaptive personal information filtering and categorizing system[2,3,4].

In this paper, we introduce that understanding of the dynamics/representation of the associative memory model assists us to develop the practical information filtering/categorizing system with the model.

Our motivation of this application development is as follows. We have a happy image of the world connected through the internet. In this world, anyone gets any information through the internet from anywhere. Unfortunately it is not true. In the actual world, there is too much information, which includes SPAM, for anyone to get information. Another serious problem is that there are many dangerous contents. We try to solve the problems by making use of neural networks. Our target application of this research is to realize a personal information filtering/categorizing engine on internet.

In this development of this system, we the associative memory model because of following reasons: As its learning algorithm is similar to the case based reasoning of AI, the learning speed is faster than MLP etc.. We would like to claim that this system is realized without understanding of semantic meanings by making use of the associative memory model.

2 Associative Memory Model
The model is one of foundational models of neural networks. The model is one of the Recurrent neural networks. The associative memory model is a network consisting of N mutually interconnected neurons.

The weight of connection from the j-th neuron to the i-th neuron is denoted as $W_{ij}$, which is the $(ij)$ component of the weight matrix $W$. The present state of the model is represented by a column vector $x=(x_1, ..., x_N)^T$ whose i-th component $x_i$ is the state of the i-th neuron +1/-1. The typical learning algorithm is as follows:

$$W_{ij}=(1/N) \sum_j S_j - a_{ij}, \quad (1)$$
where \( N \) is the number of neurons included the associative memory model and \( a = (P/N) \).

And the recalling process is given by

\[
x_j(t+1) = \text{sgn}(W_{ij} x(t)).
\]

The following two figures show the dynamical behaviors of overlaps of the model. The horizontal axis of the figures indicates time and the vertical axis indicates the direction cosine between the state of the associative memory model and the stored vector. Figure 1 shows the behavior in a case that the memory ratio is smaller than the storage capacity. In this case, the associative memory model succeeds to recall the stored vector when the initial direction cosine is large. Figure 2 shows the behavior in a case that the memory ratio is larger than the storage capacity. In this case, the associative memory model does not succeed to recall the stored vector.

![Fig. 1, Dynamical behavior of direction cosine in a case, \( a<0.14 \).](image1)

![Fig. 2, Dynamical behavior of direction cosine in a case, \( a>0.14 \).](image2)

### 2.1 Geometrical Picture

We introduce the geometrical picture of the associative memory model. The norm of the state vector is constant. Accordingly, all of the state vectors represented as points on a \((N-1)\) dimensional sphere in the associative memory model.

The recalling process is divided into two steps, flow step and quantizing step:

flow step: \( x(t) \rightarrow u(t+1) \) (3)
qingentizing step: \( u(t+1) \rightarrow \text{sgn}(u(t+1)) \) (4)

where \( u(t+1) = (N)^{1/2} \frac{Wx(t)}{|Wx(t)|} \). In the recalling process, the flow step is important compared to the quantizing step because any state vector never changes through only the quantizing step. This obvious statement suggests that the flow step outlines the dynamics of the associative memory model.

### 2.2 D function

Here we introduce a useful function to analyze the flow step. We refer the function to D function. It is given by this equation,

\[
D(x) = \frac{1}{N}(Wx,Wx),
\]

where \( x \) is a state of the associative memory model.

The D function is a positive definite function. It depends on only the state vector of the associative memory model.

The D function has two important properties. The first one is that D function monotonically increases along the streamline on the sphere. The second one is that if \( S \) are random vectors, \( D(S) \sim (1+a) \). The second property shows that the stored vectors are distributed around the stored band on the sphere. The stored band is given by this equation,

\[
D(x) = (1+a).
\]

The stored band separates the sphere into two parts. We refer the two parts to upper side and lower side. Following figure shows that the dynamical behavior of the D function in the recalling processes.
Fig. 3, Dynamical behavior of D function

Figure 3 shows that the values of the D function increase monotonically in almost recalling steps. This figure shows that there is the store vector at the middle of the flow. The next figure whose horizontal axis indicate the inner product, (S ,x), overlap and whose vertical axis indicates the D value shows the fact clearly. Figure 4 shows that several exceptional recalling steps, at which the D function decreases, are observed in the cases that the stored vectors are recalled. This phenomenon suggests that the quantizing step is dominant compared to the flow step near the stored vector. But there is no fixed point of the flow near the stored vector. Thus the recalling process is summarized by the schematic figure(Fig.5).

The recalling process starts a point in the upper side of the flow and the point moves along the streamline. If the point closes to the stored vector, the point is attracted to the stored vector by the quantizing step. Otherwise, the point is carried down the lower side. This simple picture of the associative memory dynamics suggests that if we control the flow, the performance of the model is improved. It is true. We have given several improvements in reference 1.

2.3 Concept Formation Ability

Now, we consider a following problem. When we embed only the patterns of the level 1 whose correlations are shown in Fig.6 into the associative memory model, where are the patterns of the level 2 and of the level 3 distributed on the sphere? We can get its answer with the D function easily.

This table shows the values of the D function of each level where C1 (C2) indicates the correlation between the pattern of the level 1 and that of the level 2 (the pattern of the level 2 and that of the level 3). The italic numbers show the maximum values in the same row.

<table>
<thead>
<tr>
<th>C1</th>
<th>C2</th>
<th>Level1</th>
<th>Level2</th>
<th>Level3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3</td>
<td>0.3</td>
<td>1.2</td>
<td>0.8</td>
<td>0.4</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5</td>
<td>1.9</td>
<td>2.1</td>
<td>1.8</td>
</tr>
<tr>
<td>0.7</td>
<td>0.7</td>
<td>6.0</td>
<td>8.0</td>
<td>9.3</td>
</tr>
</tbody>
</table>

When we remember that the D function increases in the recalling process monotonically, this table shows that in order to be organized higher concepts, the correlation between the stored patterns have to be strong. The following figure supports this statement. When the correlation is strong, there are two points to which the streamlines converge. The points
correspond the higher concepts. Otherwise, there are no similar points.

We summarize this section as follows: the associative memory model organizes concept patterns when stored patterns have strong correlation. This suggests that the model works as a good information categorizing tool because the higher concepts represent the stored vectors.

3 Information Categorization
From here, we present an application, personal information filtering and categorizing system with the associative memory model.

3.1 Personal Profiling with Associative Memory

Our idea is that personal information filter system is a system to classify items in 2 groups, interesting and dislike. In this context, each word has just two meaning, interesting and dislike. In this context, semantics of each word is not important. Two concepts (interesting & dislike) are important.

If we are able to neglect semantic meaning of words, it is easy that we apply the concept formation ability of the associative memory model to the system as a personal profiling module. This schematic figure shows the block diagram of the personal profiling module with the associative memory model. Input items with key words are transformed to vectors with this dictionary. The associative memory receives the vectors and organizes concept vectors corresponding to user’s interests.

In order to make a practical system, we need an algorithm organizing a small dictionary and also we need not only to identify the two concepts, (interest and dislike), but also to estimate relations between words without semantic analysis.

The first problem is “What keywords reflect our interest?” Here we introduce two probabilities:

\[ q(w_i) : \text{probability that an item with } w_i \text{ is interesting}, \]
\[ p : \text{probability that an item is interesting}. \]
Now such a keyword, $w_i$, satisfies $q(w_i) > p$ or $q(w_i) < p$ and it carry much information of user’s interests. Keywords with much information of our interests are useful. As so, Kallback divergence is a good quantity to select keywords.

$$KD(w_i) = q_i \log(q_i/p) + (1 - q_i) \log((1 - q_i)/(1 - p)), \quad (6)$$

where $q_i = q(w_i)$. Accordingly, our self-organized dictionary accumulates key words with large Kallback divergence.

Next we solve the second problem. The autocorrelation associative model organizes concept patterns. But we have no macro quantity to identify each concept pattern. To avoid this problem, we use a duplex associative memory model proposed by Kakeya and Kindo[5].

The duplex associative memory model has been presented as a high stored capacity model in order to demonstrate the validity of the geometrical analyzing method. The learning algorithm of the duplex associative memory model has two modes:

plus mode

$$W_{ij}^+ = (1/N)(S_i S_j - ä_{ij}), \quad (7)$$

and minus mode,

$$W_{ij}^- = (1/N)(S_i S_j - ä_{ij}). \quad (8)$$

The interesting items are embedded into the model by the plus mode. And the dislike items are embedded into the model by the minus mode. The duplex associative memory model organizes two basins in the pattern space. The centers of the basins correspond to the two concept patterns, “interest” and “dislike”. And the deviation between the input vector and the concept patterns reflects the strength of interest or dislike. Accordingly, the strength of the interest is estimated by

$$(I, W I) = \begin{cases} 
  \text{positive, if the item is interesting}, \\
  \text{negative, if the item isn’t interesting},
\end{cases}$$

where $I$ is a vector corresponding to an item. Thus the system sorting items by this value works a good personal information filtering system.

This is the foundational idea of our information filter. In our current system, we come up with our idea to obtain better estimation of the interests.

### 3.2 Decomposition of Personal Profile

Next, we present our information categorizing system without natural language processing.

Please suppose a case where we classify items into 3 groups. For this purpose, we need to obtain 3 sets of key words corresponding to the groups. In order to obtain the sets, we pick up a key word and decide the best group including the key word. In this decision, we have to estimate the semantic relations between words. An important problem solved by us is “how can we estimate it without natural language processing?”

Our idea to solve the problem is that each semantic relation includes its meaning and its strength. The weight matrix of the model dose not represent the semantic meaning but the weight matrix represents the strength of semantic relations. Furthermore the strength of the semantic relation has enough information to decide the best group including the key word.

When a user has several interests, the duplex associative memory model organizes several concept patterns through the use-system interaction. The geometrical analysis suggests that the model organizes such a representation. Each group corresponds to each interest of the user.

Now it is easy to obtain the sets of key words to classify items into categories. It is enough to cut the weak connections between nodes. It equals that the small components of the weight matrix are replaced by zero. The key words included the same group are given by the connected nodes with non-zero components of the weight matrix.

### 3.3 Current System of Information categorization

Figure 10 shows our current information categorizing system. Items sent from a contents provider are stored in item store. A user turn on the browser. The browser loads the items through the INSOP filter which rearranges the items in order the strength of the interest, which is estimated with Personal Profile. The user receives rearranged items in order of his/her interest. When the user inputs his/her interest, “interest” or “dislike”, to the browser, the browser sends the input INSOP adapter. The INSOP adapter modifies the Personal Profile. When the user clicks a bottom indicating a organized category on the browser, the INSOP filter calls the
Decomposer and receives a decomposed personal profile (Figure 11). After that, the INSOP filter send rearranged item with the decomposed personal profile. Furthermore the decomposed personal profiles are visualized by IViS(Figure 12).

![Block diagram of current information categorizing system](image)

**Fig.10** Block diagram of current information categorizing system.

![Input window of Browser](image)

**Fig.11** Input window of Browser.

![Category window](image)

**Fig.12** Category window. There are two categories

![Rearranged items with the decomposed personal profile](image)

**Fig.13** Rearranged items with the decomposed personal profile.

## 4 Conclusion

We present a practical application of associative memory models in information technologies - information categorization. In this application, the concept formation ability of associative memory is important. A part of this work is supported by Telecommunication Advancement Organization of Japan.

### References:


