A Kernel Based Neural Memory Concept and Representation of Procedural Memory and Emotion

Tetsuya Hoya
Laboratory for Advanced Brain Signal Processing,
BSI RIKEN, 2-1, Hirosawa, Wakoh-City, Saitama 351-0198, Japan
e-mail: hoya@bsp.brain.riken.go.jp

Abstract

This paper explores a general concept of two-stage dynamic memory by means of Gaussian kernels and associated processing mechanisms. The memory system can be viewed as the extension to the previous work of hierarchically arranged generalised regression neural network (HA-GRNN) and its evolutionary process, in which two psychological functions, attention and intuition, are interpreted. Within the proposed mechanism, both the functionality of procedural memory and emotion are newly modeled together with the aforementioned two psychological functions, in terms of the associated interactive processes between the short-term (STM) and long term memory (LTM).

1 Introduction

To elucidate the cognitive processes of humans and the relevant psychological functions is a challenging but inevitable subject for the development of artificial intelligence (AI). In psychology, the study of human learning and memory has long been established [1]. In the study, a great variety of models have been proposed, which suggests that the human memory system is a central part of the cognitive process.

In the artificial neural network field, multilayered perceptron neural networks (MLP-NNs) have played a significant role especially in the study of pattern recognition tasks [2]. However, it is now well-known that in practice the learning of the MLP-NN parameters by a backpropagation (BP) [3] type algorithm quite often suffers from becoming stuck in local minima and requiring long period of learning, both of which are good reason for detracting their utility in online processing. Such networks also need for training from scratch when new training data is used. MLP-NNs therefore have appeared as unsuitable candidates for elucidating the learning mechanism of the brain [4].

In the early 1990's, Specht rediscovered the effectiveness of kernel discriminant analysis [5] within the context of artificial neural networks. This led him to define the notion of a probabilistic neural network (PNN) [6]. Subsequently, Nadaraya-Watson kernel regression [7, 8] was reformulated as a generalised regression neural network (GRNN) [9]. In the neural network context, both PNNs and GRNNs are categorised into a family of radial basis function neural networks (RBF-NNs) [10] in which the hidden neurons are represented by Gaussian response functions (or, Gaussian kernels). It was reported that the RBF-NNs are also biologically appealing [11].

The advantage of PNNs and GRNNs is that they are essentially free from the 'baby-sitting' required for the MLP-NNs, i.e., the necessity to tune a number of network parameters to obtain good convergence rate or worry about any aforementioned numerical problems. By exploiting the property of PNNs and GRNNs, simple and quick incremental learning is possible due to their inherent memory-based architecture, whereby the network growing/shrinking is straightforwardly performed [12]. Moreover, in [13], it is reported that a PNN even exhibits a capability to accommodate new classes whilst maintaining a reasonable generalisation performance. These then give a substrate for modeling psychological functions [12], which leads to a framework for the development of brain-like computers, or, in a more true sense of, 'artificial intelligence'. On the basis of the remarks in [14], the aforementioned features of PNNs are considered to be crucial for the development of brain-like computers.

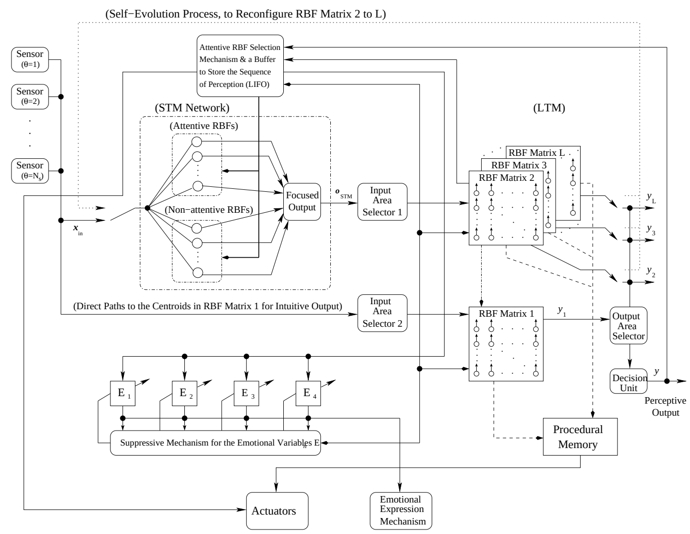

Figure 1: The new memory system.

2 The Two-Stage Dynamic Memory Concept

Fig. 1 shows the two-stage dynamic memory system based upon the RBF (kernel) networks. As in the figure, the STM consists of 1) a single modified RBF-NN with an attentive selection mechanism, 2) a buffer to store the sequence of perception given as the interactive processing within the Gaussian kernel matrices in the LTM, 3) four emotional variables connected to the emotional expression mechanism of AI, and 4) a suppressive mechanism for the four emotional variables. In contrast, the LTM consists of 1) input/output area selectors, 2) the Gaussian kernel (RBF) matrices, and

3) memory arrays representing the procedural memory.

2.1 The STM Network

The STM network of the proposed memory system has a simple two-layered structure in itself as that of the HA-GRNN [12]. As in Fig. 1, within the STM network, the RBFs can be divided into two parts; 1) the attentive RBFs which tend to generate relatively higher activations for a particular set of incoming vectors and 2) the rest. The main role is thus to temporarily buffer the incoming sensory vectors and eventually transfer (some of) them to the LTM. The STM network output $\boldsymbol{o}_{STM}$ is given as a vector rather than a scalar:

$$\boldsymbol{o}_{STM} = [\boldsymbol{o}'_{STM}; \theta]^T \quad (1)$$

where $\boldsymbol{o}'_{STM} = [o_1, o_2, \ldots, o_L]^T$ is the feature vector transferred from the STM network and θ denotes the sensory input number (i.e. $\theta = 1, 2, \ldots, N_s$) for which the feature vector was obtained (e.g., $\theta = 1$: microphone, 2: CCD camera, ..., with varying L). The learning of the STM network is then summarised as follows:

Step 1: (At an initial stage) if the number of the centroids is less than M_{STM}, add an RBF with its centroid vector $\boldsymbol{c}_i = \boldsymbol{x}$ in the STM. Then, set $\boldsymbol{o}'_{STM} = \boldsymbol{x}$.

Step 2: Otherwise,
1) If the activation of the least activated centroid (h_j, say) $h_j < th_{STM}$ (where th_{STM} is a given threshold), replace it with a new one with $\boldsymbol{c}_j = \boldsymbol{x}$ and set $\boldsymbol{o}'_{STM} = \boldsymbol{x}$.
2) Otherwise,

$$\boldsymbol{o}'_{STM} = \lambda \boldsymbol{c}_k + (1 - \lambda)\boldsymbol{x} \quad (2)$$

where $\boldsymbol{c}_k$ is the centroid vector of the most activated centroid (k-th, say) h_k and λ is a *smoothing* factor ($0 \leq \lambda \leq 1$).

In Step 2 above, a smoothing factor λ is introduced in order to regulate how fast the attentive focus of the STM network is changed by the newly incoming sensory vector which has not appeared before. This can then be regarded as 'selective attention' to a particular event/object. For instance, if the factor is small, $\boldsymbol{o}'_{STM}$ becomes more like $\boldsymbol{x}$, then this shows a sign of 'carelessness'. In contrast, if the factor is large, the STM network becomes 'sticky' to particular patterns. The above learning scheme can then be considered as a process similar to last in first out (LIFO) stack.

2.2 The Input/Output Area Selectors

In the LTM, the incoming STM network output $\boldsymbol{o}_{STM}$ is firstly transferred to the input area selector as in Fig. 1. The role of the input area selector is to collect the input nodes of all the RBF units in the RBF matrices that belong to the particular area corresponding to the sensory input specified by the number θ and then forward the STM network output $\boldsymbol{o}_{STM}$. In contrast, the output area selector will gather all the output values of the RBF units, within the area for a particular pattern classification task, and form a decision unit (by following the 'winner-takes-all' strategy) that will tell us the final classification result. The right part of Fig. 2 illustrates an example of this; there are two modal columns depending upon the types of stimuli; one for auditory and the other for visual. Then, the respective columns are subdivided further into several areas corresponding to the specific

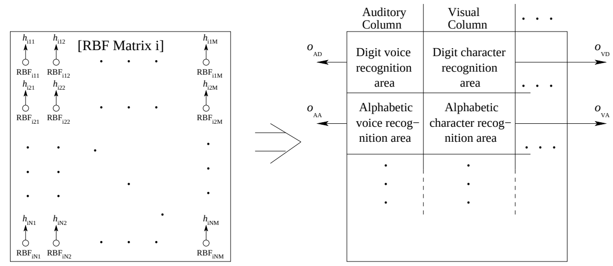

Figure 2: The i-th RBF matrix and its modality-dependent areas.

cognitions/perceptions. In the right part of Fig. 2, it is considered that there are four distinct areas, i.e., digit character/voice recognition (visual/auditory) and alphabetic character/voice recognition areas (visual/auditory), with the respective classification results, o_{VD}, o_{AD}, o_{VA}, and o_{AA}. In each area, a PNN/GRNN (with the assembly of the RBF units and a decision unit) so formed thus represents the dynamic memory for the corresponding pattern classification. In Fig. 1, the final classification result y from the RBF matrices is then given as the largest value among all the weighted outputs of the particular area within the RBF Matrix (1 to L):

$$y \;=\; \max(y_1, y_2, \ldots, y_L), \quad (3)$$

where $y_i = v_i \cdot o_i$, o_i is the output from the particular area in RBF Matrix i, and the weights $v_1 >> v_2 > v_3 > \ldots > v_L$. Note that the weight value v_1 for RBF Matrix 1 is much larger than the others. This discrimination indicates the generation of the 'intuitive output'. In practice, the intuitive outputs can be exploited to obtain e.g., faster classification results [12].

2.3 A Layer of RBF Matrices

As in Fig. 1, a layer of RBF matrices represents the LTM of the proposed new memory. Each RBF matrix (from 1 to L) consists of the RBF units, and multiple neural networks can coexist within the RBF matrices by way of the addressing pointers to other RBF units. As in the LTM networks of HA-GRNN, the hierarchical structure of the RBF matrices is constructed based upon the 'significance' or 'attractiveness' of information represented by the activation of centroids. In Fig. 1, RBF Matrix 1 indicates a collection of RBF units which can generate intuitive outputs as LTM Net 1 within the HA-GRNN [12], as there is no buffering process by the STM. In contrast, RBF matrices (2-L) represent the normal LTM. (Note that in practical implementation RBF Matrix 1 is not necessarily apart from other matrices and thus it does not mean there is a special agency for 'intuition'.) Although this layered LTM principle is inherited from that of HA-GRNN, the formation of each LTM layer (i.e., RBF matrix) here is based simply upon an assembly of RBF units, whilst the formation within the HA-GRNN is limited by a single GRNN. Hence, the neural representation is quite versatile and not restricted to the conventional fixed representations, such as layered networks or two-dimensional neuronal map (as in Kohonen's self-organising maps), which leads to a more brain-like memory representation.

2.4 The RBF Unit

In the kernel matrices, as shown in Fig. 3, each RBF unit can be seen as a memory element and is composed of

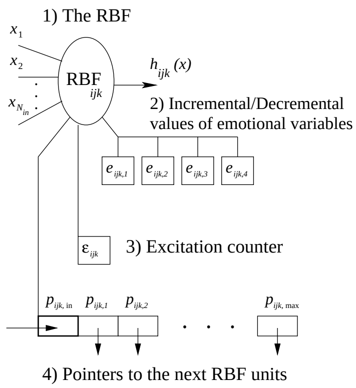

Figure 3: A Gaussian kernel (RBF) unit.

1) Gaussian kernel (or RBF) itself, 2) an excitation counter ϵ_{ijk}, 3) incremental/decremental values e_{ijk} to update the emotional variables E_n $(n = 1, 2, 3, 4)$ within the STM, and 4) addressing pointers $p_{ijk,m}$ $(m = 1, 2, ..., \max)$ which enable the RBF to *link* with other RBF units specified by the addresses.

The activation (output) of the kernel unit at the j-th row and k-th column of the i-th RBF matrix is given as

$$h_{ijk}(\boldsymbol{x}) = \exp(-\frac{\|\boldsymbol{x} - \boldsymbol{c}_{ijk}\|_2^2}{\sigma^2}) \qquad (4)$$

where $\| \ldots \|_2^2$ denotes the L_2 norm, $\boldsymbol{c}_{ijk}$ is the centroid vector, and σ is the radius. In another point of view, the Gaussian response function in (4) can be also translated as a measurement of similarity, since as the input vector $\boldsymbol{x}$ becomes closer to the centroid vector $\boldsymbol{c}_{ijk}$, the output $h_{ijk}(\boldsymbol{x})$ is increased.

Then, by exploiting the addressing pointers, a single RBF unit can be also seen as a multiple-input-multiple-output (MIMO) system and thereby the 'memory-chain' concept is developed. To illustrate the memory-chain concept, an example using four RBF units is given in Fig. 4 (note that in Fig. 4 only the RBF and the addressing pointers are depicted for convenience). In the figure, suppose that for a certain

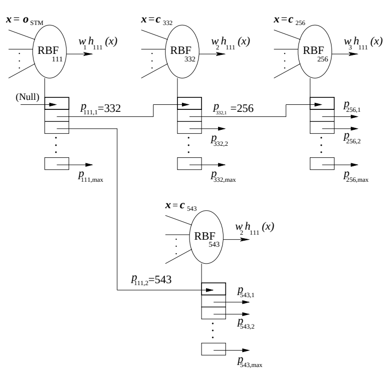

Figure 4: An example of a memory-chain represented by the RBF units.

modality RBF_{111} was most strongly activated by the STM output vector $\boldsymbol{o}_{STM}$ among all the RBF units within a particular area of RBF Matrix (1 to L) and that the strong activation of RBF_{111} is transferred to the remaining three RBF units in the chain, i.e., RBF_{332}, RBF_{256}, and RBF_{543}, via the addressing pointers. Then, the four associated RBF units become active for the subsequent perceptive processes. This memory-chain concept can be extended to, for instance, form a network for digit voice classification task, i.e., a star-shaped network consisting of ten distinct clusters of RBFs (i.e., each consists of single-layered sub-networks) describing the pattern space from /ZERO/ to /NINE/), each having a set of RBF units for the corresponding digit represents a network for the digit voice classification task. Then, the actual pattern classification can be performed as: 1) forward the incoming feature vector to all the RBF units within the neural network represented by the star-shaped network for each cluster (or, sub-network), obtain the sum of output values, (in this case, we will have ten different values) and 3) applying the decision unit to the star-shaped graph (and followed by the 'winner-takes-all' strategy), find the maximum amongst the ten sum. values and the class number which corresponds to the digit.

Similar to the memory-chain concept above, Minsky developed a concept called as 'Knowledge-line' (K-line) [15] in which each node in a semantic network is linked to other networks (agencies) via the K-lines. However, unlike K-lines, the proposed memory-chain concept also takes account for the generalisation capability in each node (RBF), which is advantageous in practical sense, i.e., the memory space required can be small compared to that of K-lines.

2.5 Representation of Procedural Memory

If the AI has skilled a learned sequence of particular actions/movements, then, in the context of memory representation, some information about the perceptive sequence contained in the memory-chain of RBF units will become solid and be transformed in a different form of representation. For clarity, this transformation is represented by the area for procedural memory shown in the right corner in Fig. 1. In such representation, all the RBF units except the first will have much smaller variance , e.g., in (4), the variance asymptotically approaches a certain small value. The procedural memory then contains only the centroid vectors transferred from the RBF units in the chain plus a special link connected to the first RBF unit within the normal LTM (i.e., the RBF matrices) for receiving the STM output $\boldsymbol{o}_{STM}$ (or, filtered sensory input vector) and executing the sequence of real actions/movements. This indicates that the experienced actions can be performed without being conscious [1]. In other words, during the execution, the sequence described by the procedural memory-chain is not monitored by the STM at all. For instance, let us turn back to the example of the memory-chain in Fig. 4. Provided that here the memory-chain with the three RBF units, RBF_{111}, RBF_{332}, and RBF_{256}, reside in a part of the motor area and the sequence describes a kinetic action, it can be formulated that the values of their activations directly controls the actuators of robots. For instance, the values stored within the centroid vectors could be used as, e.g., the target positions for PID controllers.

3 Interpretation of Emotion

In Hobson's argument [16], the psychological function 'emotion' is considered as one of the fundamental components describing consciousness. In the philosophical context, it is considered/implied that short-term memory plays a key role for describing the functionality of consciousness (e.g., [17]). On the other, an ethological model for entertainment robots (such as SDR-3X/AIBO) with emotional expressions has appeared in the literature [18]. The modeling, however, seems to resort to rather static symbolic mechanisms and it is thus considered that its extension to more reconfigurable and flexible brain-like representation is questionable.

Inspired/motivated by these studies, in this paper, the STM with four emotional states is considered as in Fig. 1. In the figure, the four emotional variables E_n ($n = 1, 2, 3, 4$) representing i) pleasure, ii) anger, iii) sadness, and iv) amenity, respectively. Accordingly, in Fig. 3, every RBF unit RBF_{ijk} comes also with four auxiliary variables $e_{ijk,n}$ ($n = 1, 2, 3, 4$) used for updating the corresponding emotional variables within the STM.

As in Fig. 1, a buffer is used to temporarily store the locations/activations of subsequently excited RBF units during the successively occurred perceptions. For instance, the following scenario is considered in the context of pattern classification: if the AI performs a series of perceptions, 1) to look at a picture of his father (face image recognition) passed by a few years ago, 2) to remember its father's voice (voice identification), and 3) to remember his pen (object image recognition), then it is naturally considered that the three pattern classification tasks, 1) face image recognition, 2) voice identification, and 3) object image recognition, are subsequently involved. Now, let us assume that within the RBF matrices there is already formed the memory representation (or, a neural network) for each pattern classification task and suppose that RBF_{111} was firstly activated when the feature vector extracted from the image of the father's face was received from the STM. Then, suppose that the two RBF units RBF_{332} and RBF_{256} in the memory-chain shown in Fig. 4 were successively activated, which respectively match the feature vector of his utterance and the image of his pen. Then, the three perceptive processes involve the respective updates of the emotional variables E_n ($n = 1, 2, 3, 4$), e.g., by a simple summation operation:

$$E_n = E_n + e_{111,n} + e_{332,n} + e_{256,n}$$

Note that here the interpretation of the memory-chain in Fig. 4 is totally different from the example of procedural memory in Section 2.5; since, except the first (i.e., RBF_{111}), the RBFs so activated correspond to the consequent of 'recalling' the previously stored events/objects. Namely, for the recalling, it is postulated that the activation of $\mathrm{RBF}_{332}/\mathrm{RBF}_{256}$ is led by not the actual sensory input (i.e., the image of its father) but the linkage due to the RBF pointers. In such a case, the input to those two RBFs is given simply as their respective centroid vectors (as illustrated in Fig. 4). Then, since the activation was caused by the RBF linkage, the actual output values h'_{332} and h'_{256} so obtained may be set to the decaying value of $h_{111}(\boldsymbol{x})$:

$$\begin{aligned} h'_{332}(\boldsymbol{x}) &= w_2\, h_{332}(\boldsymbol{c}_{332})h_{111}(\boldsymbol{x}) = w_2\, h_{111}(\boldsymbol{x}) \quad (5)\\ h'_{256}(\boldsymbol{x}) &= w_3\, h_{111}(\boldsymbol{x}) \end{aligned}$$

where $w_1 (= 1, \text{say}) > w_2 > w_3$. This may represent 'memory-fading' within the association of memory.

4 Conclusion

In this paper, a novel two-stage dynamic memory concept has been proposed. The memory system exploits the properties of kernel regression neural networks. Based upon the memory concept, the three psychological functions, attention, intuition, and emotion, have been modeled in terms of the interactive processes between the STM and LTM. In the LTM, a layer of RBF matrices and the associated mechanisms have newly been proposed as an extension of the LTM networks in the HA-GRNN [12]. Then, the representation of procedural memory and interpretation of emotion by an artificial model have been attempted. This paper has focused only upon presenting a framework for modeling psychology-oriented brain functions. Due to the limit of the space, many other aspects and features (e.g., the actual learning procedures) are left unexplained and thus a complete picture has not been given in this paper. The whole picture will be presented elsewhere. Future work is directed towards the development of the AI motivated by the proposed memory concept and justification of the effectiveness by observing and analysing the behaviour.

References

[1] J. A. Anderson, *Learning and Memory: an Integrated Approach*, 2nd Ed., New York: John Wiley, 1995.

[2] C. M. Bishop, *Neural Networks for Pattern Recognition*, Oxford, U,K.: Oxford Univ. Press, 1996.

[3] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," in D. E. Rumelhart and J. L. McClelland (Eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, vol. 1, chapter 8, Cambridge, MA: MIT Press, 1986.

[4] A. Roy, "Artificial neural networks - a science in trouble," *SIGKDD Explorations*, 1(2), pp. 33-38, 2000.

[5] D. J. Hand, *Kernel Discriminant Analysis*, Research Studies Press, 1984.

[6] D. F. Specht, "Probabilistic neural networks," *Neural Networks*, 3, pp. 109-118, 1990.

[7] E. A. Nadaraya, "On estimating regression," *Theory Probab. Applic.*, 10, pp. 186-190, 1964.

[8] G. S. Watson, "Smooth regression analysis," *Sankhy, Series A*, 26, pp. 359-372, 1964.

[9] D. F. Specht, "A general regression neural network," *IEEE Trans. Neural Networks*, 2-6, pp. 568-576, 1991.

[10] P. D. Wasserman, *Advanced Methods in Neural Computing*, in chapter 8, "Radial basis-function networks" pp.147-176, New York: Van Nostrand Reinhold, 1993.

[11] T. Vetter, A. Hurlbert, and T. Poggio, "View-based models of 3D object recognition: invariance to imaging transformations," *Cerebral Cortex*, vol. 3, pp. 261-269, 1995.

[12] T. Hoya, "Notion of intuition and attention modeled by hierarchically arranged generalized regression neural networks," subject to revision for the *IEEE Trans. Systems, Man, and Cybernetics - Part B: Cybernetics.*

[13] T. Hoya, "On the capability of accommodating new classes within probabilistic neural networks," subject to revision for the *IEEE Trans. Neural Networks.*

[14] G. Matsumoto, Y. Shigematsu, and M. Ichikawa, "The brain as a computer," in *Proc. Int. Conf. Brain Processes, Theories, and Models*, Cambridge, MA: MIT Press, 1995.

[15] M. Minsky, *The Society of Mind*, Simon & Schuster: New York, 1985.

[16] J. A. Hobson, *Consciousness*, W. H. Freeman and Company: New York, Japanese translation; Tuttle-Mori Agency Inc.: Tokyo, 1999.

[17] D. C. Dennett, (1991). *Consciousness Explained*, Little Brown: Boston, 1991.

[18] R. C. Arkin, M. Fujita, T. Takagi, and R. Hasegawa, "Ethological modeling and architecture for an entertainment robot," *Proc. Int. Conf. Robotics and Automation*, pp.453-458, Seoul Korea, 2001.