ESANN 2011 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2011, i6doc.com publ., ISBN 978-2-87419-044-5.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100817300.

Abstract Category Learning

Atsushi Hashimoto and Haruo Hosoya

Department of Computer Science
The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan

Abstract. Motivated by a neurophysiological experiment on prefrontal cortex, we study a scheme for learning abstract categories. An abstract category represents a set of vectors that are identical to each other modulo substitution, e.g., 'ABAB', 'BABA', 'ACAC', etc. We employ a clustering-based unsupervised learning method for such abstract categories, in which the recognition step is reduced to the problem of maximal perfect weight matching. Our simulations using artificial inputs confirm that the scheme learns abstract categories robustly even with a certain level of noise in the inputs.

1 Introduction

In this paper, we study an unsupervised learning of *abstract categories*. An abstract category represents a set of input vectors that are identical to each other modulo substitution. For example, the inputs 'AABB' and 'BBAA' belong to the same abstract category, while 'AABB' and 'ABBB' do not. Our motivation for studying such learning comes from a neurophysiological experiment reported by Shima et al. [6]. According to this, neurons in monkey prefrontal cortex responded selectively to abstract categories of motor sequences. That is, the activities of some neurons depended not on the individual movement (such as "pull", "push", and "turn") or the specific sequence of movements (such as "pull-push-pull-push" and "pull-pull-push-push"), but on the abstract pattern like 'XXYY', 'XYXY', or 'XXXX' that described the sequence with a different movement substituted for each variable.

Note that naive approaches fail to solve this problem. For example, one might think that an existing clustering algorithm could be used if we ignore the identity of each component in the input vector but retain whether each component is equal to or different from the previous component. However, while this could properly group together 'AABB' and 'BBAA', this would wrongly group together 'ABBA' and 'ABBC', which belong to different abstract categories in our setting.

In this paper, we describe a novel learning scheme for abstract categories. In this scheme, analogously to typical clustering methods, we use a fixed number of templates to maintain representative abstract categories and repeat recognition and learning steps alternately. However, in the recognition step, we need to find a template that represents most appropriately the abstract category of a given input; we will show that this step can be solved as the problem of maximum weight perfect matching.

In order to investigate the performance of the proposed method, we have conducted several smallish simulations using artificial data. In these, we have

ESANN 2011 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2011, i6doc.com publ., ISBN 978-2-87419-044-5.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100817300.

confirmed that the method properly learned to represent their abstract categories from "concrete" input vectors, even when certain levels of noises were present in the inputs.

2 Abstract Category Learning

In this paper, an input data is an N-dimensional vector $(x_1, \cdots, x_N)$ of K-value variables ranging over the domain $\{e_1, \cdots, e_K\}$. For convenience, we represent the vector by the following $N \times K$ matrix.

$$D = \begin{pmatrix} d_{11} & \cdots & d_{1K} \\ \vdots & \ddots & \vdots \\ d_{N1} & \cdots & d_{NK} \end{pmatrix}, \text{ where } d_{nk} = \begin{cases} 1 & (x_n = e_k) \\ 0 & (otherwise) \end{cases}$$

As mentioned in the introduction, our purpose is to classify input vectors according to which elements in the input are the same and which are different. This means that the matrices representing inputs belonging to the same abstract category are in a column-wise permutation relation. For example, the input vectors (e_1, e_2, e_1) and (e_3, e_1, e_3) represented by the following D_1 and D_2

$$D_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad D_2 = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

are in the same abstract category since multiplying D_1 with a certain permutation matrix from the right equals D_2.

$$D_1 \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} = D_2$$

Formally, we say that a matrix DP is a *like-data* of an input data D when P is a $K \times K$ permutation matrix. We then define the abstract category of D as the set consisting of all like-data of D.

$$\mathrm{Abs}\,(D) = \{DP \mid P : \text{ permutation matrix of } K \times K\}$$

Our learning method uses a set $\mathcal{T}$ of $N \times K$ matrices (of values in $[0, 1]$) called *templates*; the size of the set $\mathcal{T}$ is fixed. We define the *squared abstract distance* between an input data D and a template T as follows

$$\mathrm{Dist}\,(D, T) = \min_{C \in \mathrm{Abs}(D)} \|C - T\|^2$$

where $\|\cdot\|$ is the Euclidean distance. Note that, when a template represents exactly the abstract category of an input, the squared abstract distance between the input and the template equals zero. Therefore the recognition problem is to

ESANN 2011 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2011, i6doc.com publ., ISBN 978-2-87419-044-5.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100817300.

find the template that has the minimal squared abstract distance with the given input D[1]

$$\operatorname*{argmin}_{T \in \mathcal{T}} \mathrm{Dist}\,(D, T)$$

In learning, we want to find a set of templates that expectedly have minimal distances from any inputs. Thus, the learning problem is to minimize the objective function $F(\mathcal{T})$ defined by

$$F(\mathcal{T}) = \mathrm{E}_D \left[\min_{T \in \mathcal{T}} \mathrm{Dist}\,(D, T) \right]$$

where $E_D[\cdot]$ is the expectation with respect to the input distribution. By applying the stochastic gradient method [1] to the above objective function, the following algorithm can be derived.

1. For an input D, solve the recognition problem described above; let U be the found template (recognition step).
2. Update the template U by

$$U \leftarrow U + \alpha \left\{ \operatorname*{argmin}_{C \in \mathrm{Abs}(D)} \|C - U\|^2 - U \right\}$$

   where α is a learning rate (learning step).
3. Repeat 1–2.

Details of the derivation of the algorithm are omitted here for lack of space.

In our experience, the algorithm tends to fall into a local optimum, depending on the initial values of templates. However, we found that we can often avoid such local optimum by incorporating a standard Kohonen-style neighborhood learning [4]; our simulations shown later used this.

2.1 Recognition by maximal perfect weight matching

In the recognition step, we need to calculate $\mathrm{Dist}\,(D, T)$ for a template T, for which we need to find the permutation P that gives the minimal squared error between DP and T. Since a brute-force search from $K!$ possible permutations would be prohibitively slow, we need a more efficient method. Indeed, such a method can easily be found with a simple observation. First, note that the squared abstract distance, which is the objective function for the problem of finding the desired permutation P, is rewritable as follows.

$$\begin{aligned} \|DP - T\|^2 &= \mathrm{Tr}\left[(DP - T)^T (DP - T) \right] \\ &= \mathrm{Tr}\left[P^T D^T DP - 2T^T DP + T^T T \right] \\ &= \|D\|^2 - 2\mathrm{Tr}\left[T^T DP \right] + \|T\|^2 \end{aligned}$$

[1] The recognition problem here is in fact equivalent to the parameterized matching [2].

ESANN 2011 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2011, i6doc.com publ., ISBN 978-2-87419-044-5.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100817300.

Therefore, for fixed D and T, the minimization of the squared abstract distance is equivalent to the maximization of the trace in the second term. Maximizing this trace can in fact be solved as the maximum weight perfect matching problem. Indeed, let $d_1, \cdots, d_K$ and $t_1, \cdots, t_K$ be the column vectors of the matrices D and T; construct a bipartite graph such that each vertex on one side corresponds to a vector d_i and each vertex on the other side corresponds to a vector t_j; let the weight of an edge between d_i and t_j be their inner product. Then, the maximum weight perfect matching problem is to find the bijection f maximizing $\sum_i d_i \cdot t_f(j)$, which is nothing but maximizing $\mathrm{Tr}\left[T^T DP\right]$.

A number of efficient algorithms for solving this problem are known; among others, Munkres's algorithm [3] has time complexity $O\left(K^3\right)$.

3 Simulation

We conducted two simulations to confirm the performance of our learning scheme. The first task was inspired by the neurophysiological study [6] mentioned in the introduction. An input vector was generated by one of the patterns 'XXXX', 'XXYY', or 'XYXY' with each variable replaced by a different symbol chosen from 'A', 'B' or 'C'. We used 10 templates for learning that were initialized with random values. After 1500 cycles of learning, eight of the templates represented an abstract category of inputs. Fig. 1a shows the templates before learning and 1b shows those after. In order to see how well a learned template recognized an actual input, we defined the matching index as the sum of the weights divided by N, where the weights are those obtained after solving the recognition problem by matching (Section 2.1). Note that the matching index equals 1 if the template exactly represents the abstract category of the input. Fig. 1c and d show the matching indices for all templates (indicated by the line types) and all data (indicated by the radial items), before learning (c) and those after (d). The result shows that some templates indeed acquired selective responsiveness to one abstract category, reproducing a qualitatively similar result to the neurophysiological experiment. Moreover, although some data ('CBCB', 'ACAC', 'AABB' and 'CCAA') were not used for inputs during learning, the templates were able to respond to them properly after learning, thus succeeding in generalization.

In the second task, we used a slightly larger dimension of dataset in order to examine the robustness against noise. An input data was an image of 10×10 pixels whose values can be one of three colors. Each input image was generated by dividing it into several subregions and painting a color on each subregion. Fig. 2a shows examples of input images (the colors are displayed by textures), where each row shows two images with the same region division but different colorings. The intention of this task was to categorize the inputs in such a way that the subregions with the same texture were regarded as belonging to the same "object." For example, the bottom left image in Fig. 2a might represent a small rectangle block stacked on top of a larger rectangle block whereas the bottom right image might represent a single block with a hole in the middle. Furthermore, in this task, we added noise to the input images by replacing a

ESANN 2011 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2011, i6doc.com publ., ISBN 978-2-87419-044-5.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100817300.

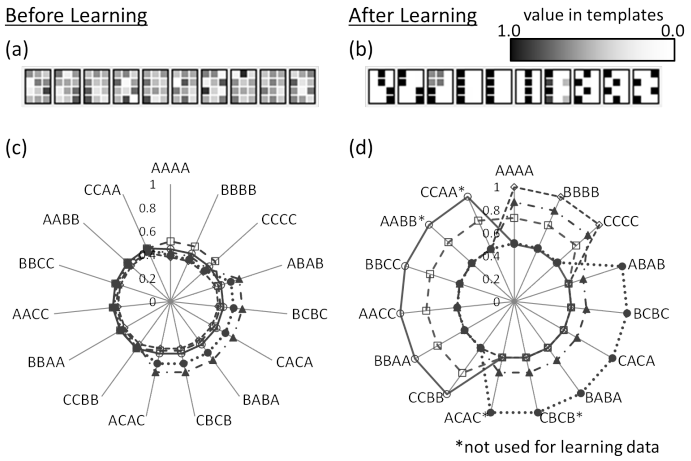

Fig. 1: The result of a simulation of abstract category learning with the task inspired by a neurophysiological study.

certain percentage (noise level) of pixels with random colors and examined the effect of the noise on the performance of learning. Fig. 2b shows the templates after learning when adding no noise to the inputs (left) and when adding noise to 50% of the pixels (right). We can see that the resulting templates were quite similar in both cases. Then, we tested the recognition ability of the learned templates with respect to the noise level. For this, we measured the rate of the correct regonition of noise-added inputs (i.e., the noise-added input is matched with the same template as the noiseless input). Fig. 2c and d show that, as learning progressed, the correct rates became higher, both when the noise level was 10% and when 50% during learning, though the latter case yielded a lower correct rate. Further, we measured the distance between the templates that were matched with the noise-added input and the noiseless one (where the distance is defined as the Euclidean distance between the closest like-data in the abstract categories represented by the two templates). As the figures show, as learning progressed, the template distances became close to zero, though they tended to be slightly higher when the noise level during learning was 50%. Taken together, these results show that our scheme is robust against noise at least in this smallish task.

4 Conclusion

In this paper, we have proposed a learning scheme for abstract categories. The proposed scheme does not categorize inputs according to "geometric" similarities, and is thus different from the usual clustering-based approaches. Such categorization is known to be performed in the prefrontal cortex and we have

ESANN 2011 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2011, i6doc.com publ., ISBN 978-2-87419-044-5.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100817300.

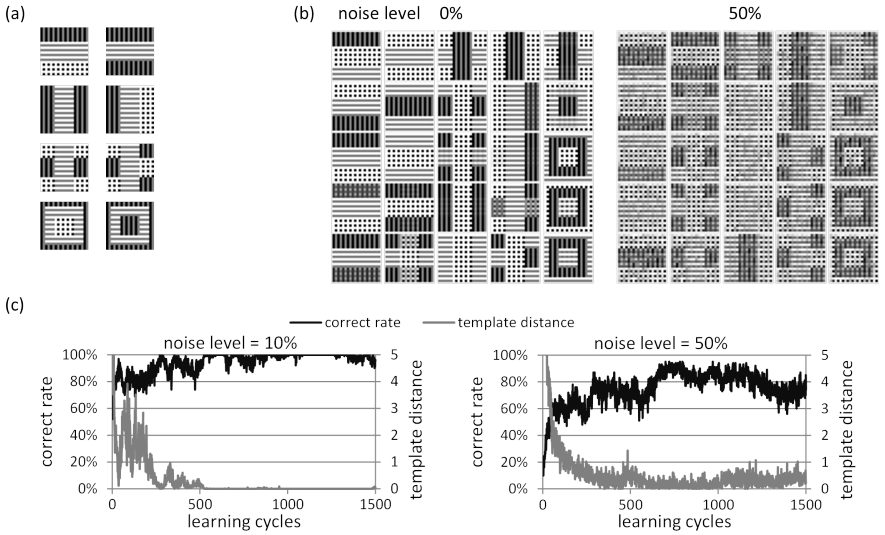

Fig. 2: Results of image pattern categorization independent of textures.

shown that our scheme can yield a result similar to a neurophysiological experiment [6]. A psychophysical phenomenon known as perceptual categorization [5] seem related to abstract categorization and we speculate that our scheme or some extension could reproduce this. Also, in order to gain an insight into how the brain realizes such cognitive functionality, we plan to investigate a neural network model implementing abstract category learning.

References

[1] S. Amari. A theory of adaptive pattern classifiers. *IEEE Transactions on Electronic Computers*, pages 299–307, 1967.

[2] B.S. Baker. A program for identifying duplicated code. *Computing Science and Statistics*, pages 49–49, 1993.

[3] F. Bourgeois and J.C. Lassalle. An extension of the Munkres algorithm for the assignment problem to rectangular matrices. *Communications of the ACM*, 14(12):802–804, 1971.

[4] T. Kohonen. *Self-Organizing Maps*. Springer, 2001.

[5] S. Li, D. Ostwald, M. Giese, and Z. Kourtzi. Flexible coding for categorical decisions in the human brain. *Journal of Neuroscience*, 27(45):12321, 2007.

[6] K. Shima, M. Isoda, H. Mushiake, and J. Tanji. Categorization of behavioural sequences in the prefrontal cortex. *Nature*, 445(7125):315–318, 2006.