Data Mining from Functional Brain Images

Mitsuru Kakimoto
Corporate Research & Development Center, Toshiba Corporation
1, Komukai Toshiba-cho, Saiwai-ku, Kawasaki 212-8582 Japan
mitsuru.kakimoto@toshiba.co.jp

Chie Morita
Corporate Research & Development Center, Toshiba Corporation
1, Komukai Toshiba-cho, Saiwai-ku, Kawasaki 212-8582 Japan
chie.morita@toshiba.co.jp

Hiroshi Tsukimoto
Corporate Research & Development Center, Toshiba Corporation
1, Komukai Toshiba-cho, Saiwai-ku, Kawasaki 212-8582 Japan
hiroshi.tsukimoto@toshiba.co.jp

ABSTRACT

Recent advances in functional brain imaging enable identification of active areas of a brain performing a certain function. Induction of logical formulas describing relations between brain areas and brain functions from functional brain images is a category of data mining. It is difficult, however, to apply conventional mining techniques to functional brain images due to several reasons, such as the difficulty of reducing images to symbolic data, possible existence of correlations between adjacent pixels in a image and the limited number of samples available from a single subject. Tsukimoto and Morita presented an algorithm for data mining from functional brain images and showed that the algorithm works well for artificial data. The algorithm consists of two steps. The first step is nonparametric regression. The second step is rule extraction from the linear formula obtained by the nonparametric regression. The authors have applied the algorithm to real f-MRI images. This paper reports that the algorithm works well for real f-MRI data and has led to the discovery of certain rules for a finger tapping action and a speech-related action.

Categories and Subject Descriptors

I.2.6 [**Learning**]: Concept learning; J.3 [**Life And Medical Sciences**]: Medical information systems

Keywords

Knowledge discovery, Functional brain images, Nonparametric regression, Rule extraction, Human brain mapping

1. INTRODUCTION

Conventional data mining techniques deal with symbolic and/or numerical data contained in tables in which the independence of rows is tacitly assumed. This simple structure makes the data easy to mine. As demand for knowledge discovery from real world data grows, however, methods for deriving knowledge from structured data (including time series, images and data embedded in graphical structures) are strongly desired.

Knowledge discovery from functional brain images is a candidate field for the application of such methods. As a result of the ongoing development of non-invasive measurement of brain function, detailed functional brain images can be obtained, from which the relations between brain areas and brain functions can be understood. These relations, however, could be complicated since several brain areas might be responsible for a brain function. Some of them are connected in series, and others are connected in parallel. Brain areas connected in series are described by "AND" and brain areas connected in parallel are described by "OR". Therefore, the relations between brain areas and brain functions are described by rules.

It is of crucial importance for researchers involved in mapping brain functions to find such rules from functional brain images. Although several statistical methods, such as the Statistical Parametric Map[2] and Independent Component Analysis[3], are being used in human brain mapping, these methods can only present some principal areas for the function and cannot discover rules. Furthermore, several factors prevent conventional data mining techniques from being applied to functional brain imaging. First, observed images consist of real values and it is not easy to reduce them to simple symbolic data such as "active" / "inactive". Second, it is expected that strong correlations exist between adjacent pixels in an image. Therefore, a mining scheme should take the structure of the image into account so as to improve its quality. Third, in a usual function brain imaging experiment, the number of samples obtained from a single subject is limited. This scarcity of samples makes it hard to obtain accurate rules.

Tsukimoto and Morita have presented a new algorithm capable of extracting rules from such structured data, which is now called the Logical Regression Analysis(LRA). They confirmed that the LRA works well for artificial functional brain image data[11]. The LRA consists of two steps. In the first step, a linear formula describing the relation between an image and a brain function is derived using regression

Proceedings of the International Workshop on Multimedia Data Mining (MDM/KDD'2000), in conjunction with ACM SIGKDD conference. Boston, USA, August 20, 2000
(O.R. Zaïane, S.J. Simoff, eds.)

analysis. In order to obtain a linear formula from relatively few samples, nonparametric regression is used. The subsequent step extracts rules from the linear formula obtained in the previous step.

This paper reports the application of the LRA to real f-MRI data obtained in the experiments of finger tapping and speech actions. Section 2 briefly outlines a scheme of the data mining from functional brain images. Section 3 gives exact formulation of nonparametric regression used in the analysis. Section 4 explains the rule extraction algorithm from a linear formula. Section 5 shows experimental results obtained by applying the LRA to real f-MRI images and discusses its meaning in brain science. Conclusions are presented in Section 6.

2. DATA MINING FROM FUNCTIONAL BRAIN IMAGES

Fig.1 is a schematic illustration of a 2-dimensional functional brain image with a circle representing a contour of a brain. In Fig.1, the image is divided into 6×6(=36) pixels. In an experiment using f-MRI, subjects are told to do some task and rest for a while repeatedly. If a pixel includes or intersects brain areas responsible for that task, activation of the area by the task results in enhanced value of the pixel. Detecting difference of the value of a pixel between the image taken while the subject is doing the task and one while he/she is resting thus makes it possible to identify areas responsible for the task.

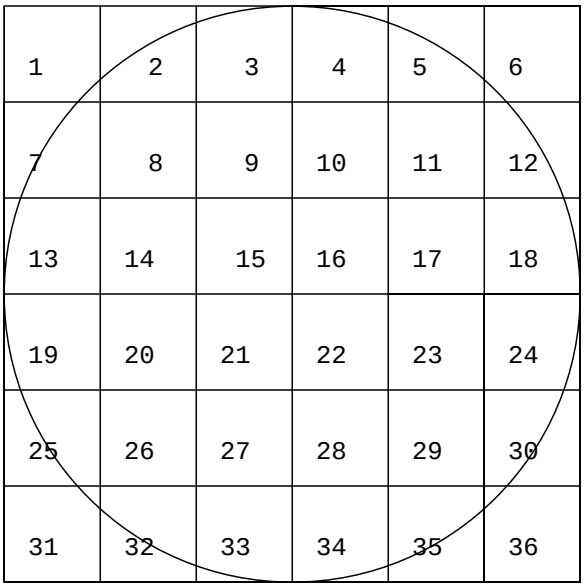


Figure 1: Brain image

Although values of pixels in a real f-MRI image are continuous, for simplification, we assume here that the values are binary, that is, on(active) or off(inactive). Also, we assume that there are seven samples. Table 1 shows the data. In Table 1, 'on' and 'off' mean that the pixel is active and inactive, respectively. Y in class shows that a subject is doing a certain task and N shows that he/she is resting.

Table 1: Data

sample	1	2	·	36	class
S1	on	off	·	off	Y
S2	on	on	·	off	N
S3	off	off	·	on	N
S4	off	on	·	on	Y
S5	on	off	·	off	N
S6	off	on	·	on	N
S7	off	off	·	on	Y

If an activity of a pixel is strongly correlated with the class value, the pixel is considered to be a part of an area responsible for the function. On the contrary, if a pixel is always inactive, it is considered to be a part of inhibitory area. Otherwise, i.e. a pixel's activity does not have any correlation with the class value, it is regarded as irrelevant to the function. Combining pixel values of responsible areas and negation of pixel values in inhibitory areas produces a logical formula that describes a rule governing the brain function. It is thus clear that rule extraction from functional brain images is formulated as a typical supervised inductive learning.

What makes the rule extraction difficult, however, is that variation of a pixel value correlated to brain function is so subtle that there is no clear-cut way to reduce observed numerical pixel values to simple 'on'/'off' symbols. Combining regression analysis and rule extraction, the LRA evaluates quantitative significance of each pixel respecting the brain function, which makes fast and rigorous rule extraction possible.

3. NONPARAMETRIC REGRESSION

As explained earlier, the LRA uses regression in its first step. In functional brain image analysis, each image has more than a thousand pixels, which mean that there are more than a thousand independent variables. The number of samples obtained in a single experiment is around one hundred, which is significantly small compared with a number of independent variables. It is therefore impossible to use conventional linear regression which requires a larger number of samples than the number of independent variables.

Another problem inherent in image analysis is that strong correlation is expected between adjacent pixels. This also applies in the analysis of functional brain images where contribution of each pixel to brain function cannot be regarded as truly independent.

As shown below, these two problems are resolved simultaneously in a framework of nonparametric regression. The next subsection explains the conventional 1-dimensional nonparametric regression[1]. Extension to 2-dimensions, which is used in the analysis of functional brain images, is described in the later subsection.

3.1 1-dimensional nonparametric regression

Nonparametric regression is defined as follows: Let y stand for a dependent variable and $t_j (j = 1, .., m)$ stand for independent variables. Then, the regression formula is as fol-

lows:

$$\hat{y} = \sum a_j t_j + e (j = 1, .., m),$$

where a_j are real numbers and e is a zero-mean random variable. When there are n measured values of y,

$$\hat{y}_i = \sum a_j t_{ij} + e_i (i = 1, .., n).$$

In usual linear regression, the coefficients a_j are defined so that residual error(i.e., difference between measured value of y and calculated value by the formula) is minimized, whereas in nonparametric regression, continuity of coefficients is also taken into account[Miwa, private communication]. Therefore, the evaluation value is now described as follows:

$$1/n \sum_{i=1}^{n} (y_i - \hat{y_i})^2 + \lambda \sum_{j=1}^{m} (a_{j+1} - a_j)^2$$

where $\hat{y}$ is an estimated value. The second term in the above formula is the difference of first order between the adjacent coefficients, that is, the continuity of the coefficients. λ is the coefficient of continuity. Consideration of two extreme cases facilitates understanding of the characteristics of the above evaluation value. When $\lambda \to 0$, the evaluation value consists of only the first term, that is, error, which means the usual regression. In this case, the effective number of coefficients is exactly the same as the actual number of coefficients. On the other hand, if λ is infinitely large, the evaluation value consists of only the second term, that is, continuity, which means that the error is ignored and a_j is a constant. This is equivalent to the case where there is a single coefficient. The effective number of coefficients is thus controlled by the value of λ. By determining the value of λ adaptively, a nonparametric regression scheme can handle the situation in which the number of samples available is smaller than the number of coefficients.

We determined the value of λ using the *leave-one-out method* cross validation[6], whose formulation can be described as follows. Let $\mathbf{X}$ stand for $n \times m$ matrix. Let t_{ij} be an element of $\mathbf{X}$. Let $\mathbf{y}$ stand for a vector consisting of y_i. $m \times m$ matrix $\mathbf{C}$ is as follows:

$$\mathbf{C} = \begin{pmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & 2 & -1 & \\ & & & & \cdots \end{pmatrix}$$

Off diagonal elements of $\mathbf{C}$ not written explicitly are exactly 0. Cross validation function CV is as follows:

$$CV = n \tilde{\mathbf{y}}^{\mathbf{t}} \tilde{\mathbf{y}}$$

$$\tilde{\mathbf{y}} = \mathbf{Diag}(\mathbf{I} - \mathbf{A})^{-1} (\mathbf{I} - \mathbf{A}) \mathbf{y}$$

$$\mathbf{A} = \mathbf{X} (\mathbf{X}^{\mathbf{t}} \mathbf{X} + (\mathbf{n} - 1) \lambda \mathbf{C})^{-1} \mathbf{X}^{\mathbf{t}},$$

where $\mathbf{DiagA}$ is a diagonal matrix whose diagonal components are $\mathbf{A}$'s diagonal components. Then the coefficients $\hat{\mathbf{a}} = (a_1, \cdots, a_m)^t$ are obtained by

$$\hat{\mathbf{a}} = (\mathbf{X}^{\mathbf{t}} \mathbf{X} + \mathbf{n} \lambda_{\mathbf{o}} \mathbf{C})^{-1} \mathbf{X}^{\mathbf{t}} \mathbf{y},$$

where λ_o is the λ that minimizes the cross validation function CV.

3.2 2-dimensional nonparametric regression

The nonparametric regression scheme explained in the previous subsection is extended to the 2-dimensional case and applied to f-MRI data. In 2-dimensional data, there are four adjacent measured values [1], whereas in 1-dimensional data, there are only two. Hence, the evaluation value for the continuity of coefficients a_j is modified so that continuity with adjacent four pixels is taken into account. For example, pixel 8 in Fig.1 has four adjacent pixels (2, 7, 9 and 14), and the evaluation value is as follows:

$$(a_8 - a_2)^2 + (a_8 - a_7)^2 + (a_9 - a_8)^2 + (a_{14} - a_8)^2.$$

4. RULE EXTRACTION

4.1 Rule extraction in the discrete domain

In this subsection, a method for rule extraction in the discrete domain is explained. The main idea is to find a Boolean function which is nearest to a given linear formula in the Boolean function space.

Let (f_i) be the values of a linear formula. Let $(g_i)(g_i = 0$ or $1)$ be the values of Boolean functions. The basic method is as follows:

$$g_i = \begin{cases} 1 (f_i \geq 0.5), \\ 0 (f_i < 0.5). \end{cases}$$

This method minimizes Euclidean distance.

Generally, let $g(x_1, ..., x_n)$ stand for a Boolean function, and let $g_i (i = 1, ..., 2^n)$ stand for values of a Boolean function and then the Boolean function is represented by the following formula:

$$g(x_1, ..., x_n) = \sum_{i=1}^{2^n} g_i a_i,$$

where $\sum$ is disjunction, and a_i is the atom corresponding to g_i, that is,

$$a_i = \prod_{j=1}^{n} e(x_j) \ (i = 1, ..., 2^n),$$

where

$$e(x_j) = \begin{cases} \overline{x_j} (e_j = 0), \\ x_j (e_j = 1), \end{cases}$$

where $\prod$ stands for conjunction, $\overline{x}$ stands for the negation of x, and e_j is the substitution for x_j, that is, $e_j = 0$ or 1. The above formula can be easily verified.

Fig. 2 shows a case of two variables, x and y. Crosses stand for the values of a linear formula and circles stand for the values of a Boolean function. Values on the horizontal axis, $00, 01, 10$ and 11, stand for the domains. For example, 00 stands for $x = 0, y = 0$.

In this case, the values of the Boolean function $g(x, y)$ are as follows:

$$g(0, 0) = 1, g(0, 1) = 1, g(1, 0) = 0, g(1, 1) = 0.$$

[1] One might think of a neighbor consisting of eight surrounding pixels. We did not pursue this possibility.

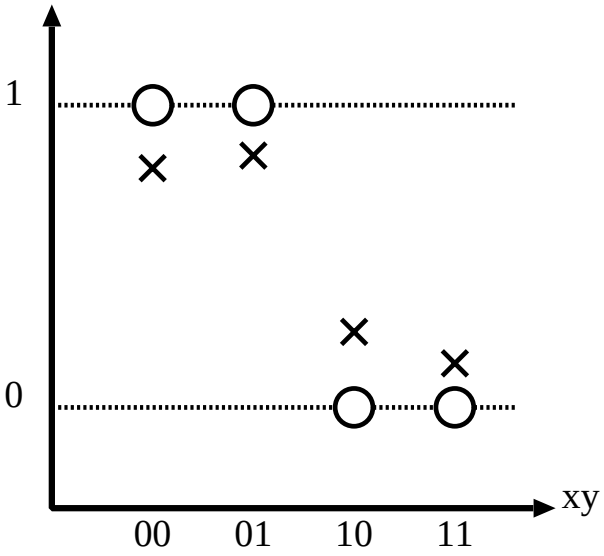


Figure 2: Approximation

Therefore, in the case of Fig. 2, the Boolean function is as follows:

$$\begin{aligned} & g(x,y) \\ = \; & g(0,0)\bar{x}\bar{y} + g(0,1)\bar{x}y + g(1,0)x\bar{y} + g(1,1)xy \\ = \; & 1\bar{x}\bar{y} + 1\bar{x}y + 0x\bar{y} + 0xy \\ = \; & \bar{x}\bar{y} + \bar{x}y \\ = \; & \bar{x}. \end{aligned}$$

4.2 The fast polynomial time algorithm

A naive implementation of the rule extraction above requires computational time which grows exponentially with the number of independent variables. Tsukimoto presented the polynomial time algorithm[9], [10], the outline of which is now described.

Let a linear formula be as follows:

$$f = p_1 x_1 + \ldots + p_n x_n + p_{n+1},$$

The Boolean function which approximates f is obtained by the following steps.

1. Check if

   $$x_{i_1} \cdot \cdot x_{i_k} \overline{x}_{i_{k+1}} \cdot \cdot \overline{x}_{i_l}$$

   exists in the Boolean function after the approximation by the following formula:

   $$p_{n+1} + \sum_{i_1}^{i_k} p_j + \sum_{1 \leq j \leq n, j \neq i_1, \ldots, i_l, p_j \leq 0} p_j \geq 0.5$$

   .

2. Connect the terms existing after the approximation by logical disjunction to make a DNF formula.

3. Execute the above procedures up to a certain (usually two or three) order.

4.3 Extension to the continuous domain

Continuous domains can be normalized to [0,1] domains by some normalization method. So we assume here that the values lie in [0,1] domains without loss of generality. First, we have to present a system of qualitative expressions corresponding to Boolean functions, in the [0,1] domain. The expression system is generated by direct proportion, reverse proportion, conjunction and disjunction. The direct proportion is $y = x$. The inverse proportion is $y = 1 - x$, which is a little different from the conventional one ($y = -x$), because $y = 1 - x$ is the natural extension of the negation in Boolean functions. The conjunction and disjunction will be also obtained by a natural extension. The functions generated by direct proportion, reverse proportion, conjunction and disjunction are called continuous Boolean functions, because they satisfy the axioms of Boolean algebra. For details, refer to [8]. In the domain [0,1], linear formulas are approximated by continuous Boolean functions. The method for deriving such an expression is exactly the same as the one in the discrete domain[7], [12].

5. EXPERIMENTS

Two f-MRI experiments are conducted, each with a single subject. Data obtained consist of 32 series of cross-sectional brain images, i.e. slices(Fig. 3). Nonparametric regression is performed on each series of images and average residual error of the regression is calculated. Small error value of a slice is deemed to be an evidence of the existence of signals correlated to the task on the slice. Rule extraction is performed on these slices to identify brain areas related to the task. Detailed results of each experiment are given in the following subsections.

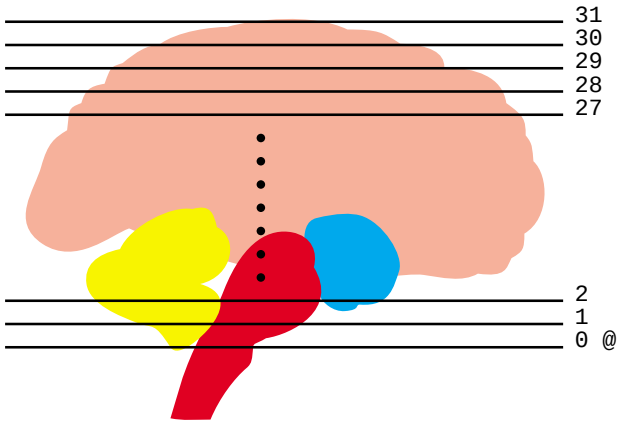


Figure 3: Slices

5.1 Finger tapping

In a finger tapping task experiment, the subject is asked to tap four fingers with thumb using his right hand.

The experimental conditions are summarized as follows:

magnetic field intensity:	1.0 Tesla
number of pixels:	64×64
number of slices:	32
subject:	male(36 years old)
number of task samples:	30
number of rest samples:	33

Table 2 shows the error of nonparametric regression for each slice. Note that slices are numbered from bottom to top, i.e., slice 0 is lower part of the brain and slice 31 is located at the top of the brain.

Table 2: Error (Finger tapping)

slice	err.	slice	err.	slice	err.	slice	err.
0	0.73	8	1.00	16	0.11	24	0.96
1	0.02	9	0.83	17	0.96	25	0.10
2	0.38	10	0.45	18	0.40	26	0.54
3	0.58	11	1.00	19	0.71	27	0.53
4	0.09	12	0.10	20	0.93	28	0.71
5	0.01	13	0.90	21	0.09	29	0.58
6	0.62	14	0.75	22	0.39	30	0.71
7	0.37	15	0.09	23	0.47	31	0.89

Slices with small errors (0.1 or less) are listed as follows:

slice	5	1	4	21	15	25	12
error	0.01	0.02	0.09	0.09	0.09	0.10	0.10

The errors of No.8 and No.11 slices are nearly 1.0, which means that these slices have no relations to finger tapping.

Figure 4 - Figure 10 are graphical representations of the rules for slices of No. 1, 4, 5, 12, 15, 21 and 25, respectively. In the figures, white areas show activation areas related to finger tapping and dark gray areas show inhibition areas. Note that these figures illustrate cross sections seen from below, i.e., left side of figures correspond to the right side of subject's head and vice versa. Note also that upper side and lower side of figures correspond to the front and back of subject's head, respectively.

Interpretations of these rules in terms of brain physiology are given as follows:
Movements of the non-dominant hand usually induce neural activity in motor and sensory areas in both hemispheres, and cause higher activity in the dominant hemisphere (contralateral to the non-dominant hand). In this case, the subject was left-handed and he moved his right hand (non-dominant). Higher activity was observed in the right (dominant) motor and sensory areas than in those on the left as shown in Figs. 7 - 9.

- Fig.6, 4, 5(slice No. 5, 1, 4)
  Higher activity was observed in the right cerebellum. The result agrees with the fact that the neural activity in the cerebellum ipsilateral to the moving hand is higher than in the cerebellum contralateral to the moving hand.

- Fig.7(slice No.12)
  Activity in the right (dominant) motor-sensory area.

- Fig. 8(slice No.15)
  Neural activity was clearly observed in motor, sensory, and supplementary motor areas in the right (dominant) hemisphere, and diffusive activity in the left motor-sensory area.

- Fig.9(slice No.21)
  Activity in the right (dominant) premotor area related to motor programming and pattern generation.

- Fig.10(slice No.25)
  Activity in the right (dominant) premotor and supplementary areas.

5.2 Shiritori

The next experimental task is *shiritori*, which is a well-known Japanese word game for two or more players. Each player utters a word that starts with the same syllable as the last syllable of a word uttered by a previous player. An example is shown below.

toki (time) → kimono (wear) → nomimono (drink) → nonki (optimism) → ...

This process is repeated until someone fails to come up with a word. In our *shiritori* experiment, the subject is presented a single Japanese character at the beginning of each task period, then he begins playing *shiritori* by himself starting with the character. While doing the task, he does not actually utter words but speaks silently.

The experimental conditions are as follows:

magnetic field intensity:	1.0 Tesla
number of pixels:	64×64
number of slices:	32
subject:	male(45 years old)
number of task samples:	40
number of rest samples:	40

Table 3 shows the errors of nonparametric regression. The result of nonparametric regression is worse than that of finger tapping. In *shiritori*, no slice has an error of 0.1 or less and the least error is 0.11. That means that *shirotori* is complicated and therefore is related to several areas such as speech area, vision area, auditory area, motor area, and so on.

Slices with small errors (0.2 or less) are as follows:

slice	7	13	6	16	3
error	0.11	0.15	0.15	0.17	0.19

There are strong correlations between *shiritori* and slices No.7, 13, 6, 16 and 3, since errors in the slices are small.

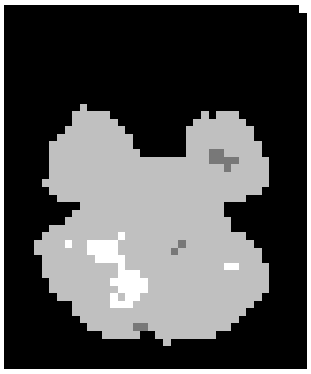


Figure 4: Rule (Finger 1)

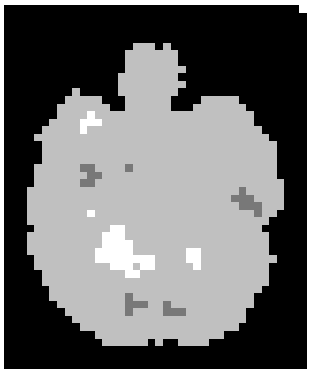


Figure 5: Rule (Finger 4)

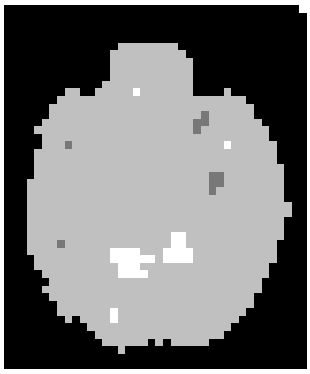


Figure 6: Rule (Finger 5)

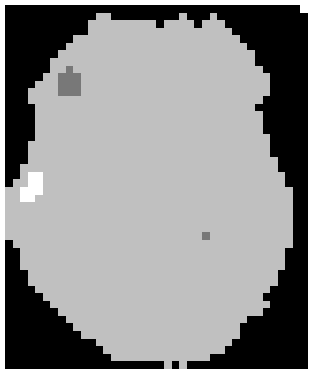


Figure 7: Rule (Finger 12)

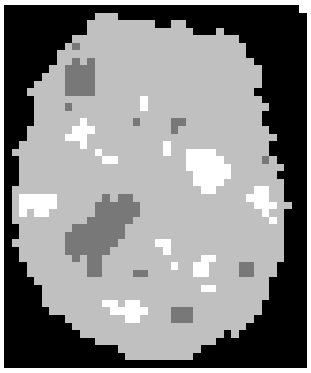


Figure 8: Rule (Finger 15)

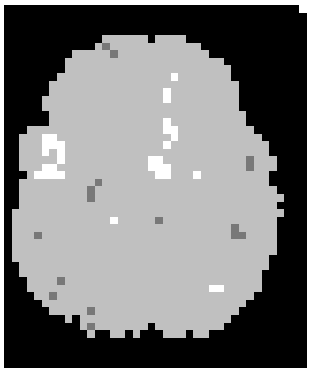


Figure 9: Rule (Finger 21)

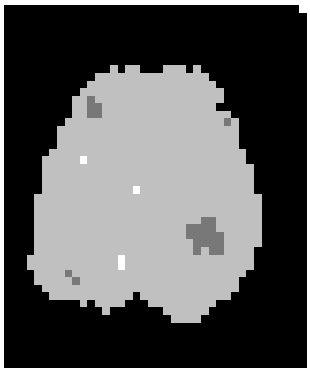


Figure 10: Rule (Finger 25)

Table 3: Error (Shiritori)

slice	err.	slice	err.	slice	err.	slice	err.
0	0.92	8	0.88	16	0.17	24	0.21
1	0.89	9	0.85	17	0.62	25	0.87
2	0.87	10	0.67	18	0.44	26	0.86
3	0.19	11	0.61	19	0.49	27	0.31
4	0.80	12	0.84	20	0.71	28	0.58
5	0.66	13	0.15	21	0.70	29	0.68
6	0.15	14	0.76	22	0.84	30	0.91
7	0.11	15	0.71	23	0.85	31	0.38

Figure 11 - Figure 15 show rules for slices No. 3, 6, 7, 13 and 16, respectively.

Some of the rules presented are interpreted as follows:

- Fig. 13(slice No.7)
  Activation of the left prefrontal area.
  The left prefrontal area was activated related to working memory required for mental word generation .
- Fig. 11(slice No.3)
  Activation of the right cerebellum.
  This showed that the cerebellum was related to some cognitive functions in addition to motor functioning. The cerebellum predominantly connects with the contralateral frontal cortex, and the right cerebellum was activated associated with the left prefrontal area.

Physiological meanings of other rules observed during the task are to be studied in future.

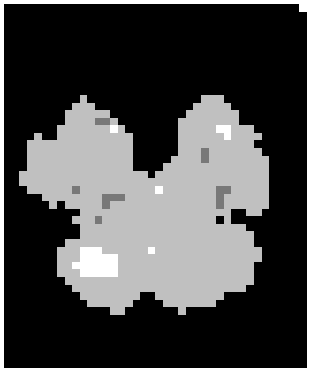
6. CONCLUSIONS

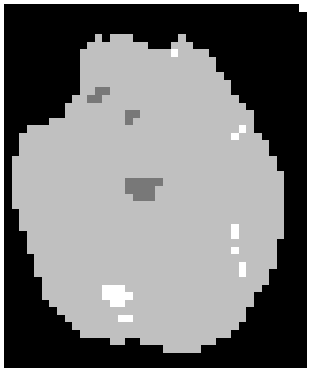
We have applied the Logical Regression Analysis to real f-MRI images obtained by experiments of finger tapping and speech actions, i.e., *shiritori* tasks. It is confirmed that the nonparametric regression extended for functional brain image analysis, which consists of the first step of the LRA, can successfully identify slices of a brain relevant to the tasks. Rule extractions, the second step of the LRA, performed on these relevant slices induced rules which are reasonably interpreted in terms of brain physiology.

7. ACKNOWLEDGMENTS

We would like to thank Dr. T. Kobayakawa of the National Institute of Bioscience Human-Technology for instruction and planning of f-MRI experiments. This research is partly supported by Grant-in-Aid for Scientific Research on Priority Areas "Discovery Science" from the Ministry of Education, Science and Culture, Japan(grant 10143106).

8. ADDITIONAL AUTHORS





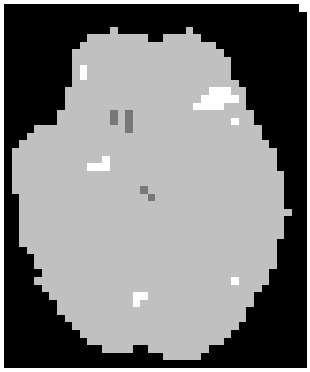


Figure 11: rule (shiritori 3) Figure 12: rule (shiritori 6) Figure 13: rule (shiritori 7)

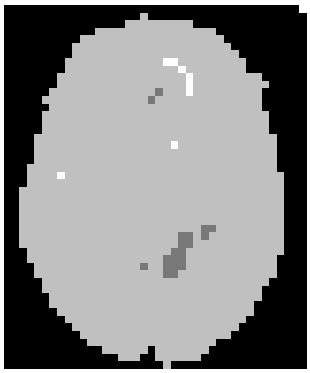


Figure 14: rule (shiritori 13)

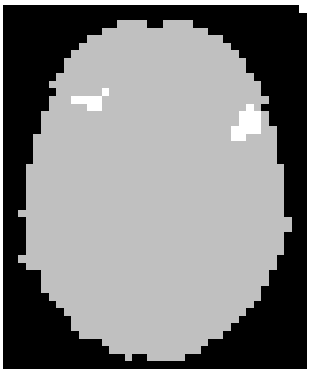


Figure 15: rule (shiritori 16)

Additional authors: Yoshiaki Kikuchi (Tokyo Metropolitan University of Health Sciences, Higashi-ogu 7-2-10, Arakawa-ku, Tokyo,116-8551 Japan
email: ykikuchi@post.metro-hs.ac.jp)

9. REFERENCES

[1] Eubank, R.L.: Spline Smoothing and Nonparametric Regression, Marcel Dekker, New York, 1988.

[2] Friston et al.: SPM course notes, 1997. http://www.fil.ion.bpmf.ac.uk/spm

[3] McKeown, M., Makeig, S., Brown, G., Jung, T.-P., Kindermann, S., Lee, T.-W., and Sejnowski, T.J. : Spatially independent activity patterns in functional magnetic resonance imaging data during the stroop color-naming task. *Proceedings of the National Academy of Sciences, 95*, pp.803-810, February 1998. http://www.cnl.salk.edu/~tewon/ica_cnl.html

[4] Morita,C. and Tsukimoto,H.: Knowledge discovery from numerical data, *Knowledge-based Systems*, Vol.10, No.7, pp.413-419, 1998.

[5] Posner, M.I., Raichle,M.E.: Images of Mind, W H Freeman & Co, 1997.

[6] Stone, M.: Cross-validatory choice and assessment of statistical prediction (with discussion), *Journal of the Royal Statistical Society*, Series B, 36, pp.111-147, 1974.

[7] Tsukimoto, H.: The discovery of logical propositions in numerical data. *AAAI'94 Workshop on Knowledge Discovery in Databases*, pp.205-216, 1994.

[8] Tsukimoto, H.: On continuously valued logical functions satisfying all axioms of classical logic. *Systems and Computers in Japan*, Vol.25, No.12. pp.33-41, 1994.

[9] Tsukimoto,H. and Morita,C.: Efficient algorithms for inductive learning-An application of multi-linear functions to inductive learning, *Machine Intelligence 14*, pp.427-449, Oxford University Press, 1995.

[10] Tsukimoto,H., Morita,C., Shimogori,N.: An Inductive Learning Algorithm Based on Regression Analysis, *Systems and Computers in Japan*, Vol.28, No.3. pp.62-70, 1997.

[11] Tsukimoto,H. and Morita,C.: The Discovery of Rules from Brain Images, *Discovery Science, Proceedings of the First International Conference DS'98*, pp.198-209, 1998.

[12] Tsukimoto,H: Extracting Rules from Trained Neural Networks, *IEEE Transactions on Neural Networks*, Vol.11, No.2, pp.377-389, 2000.