Chem. Senses 36: 413–424, 2011

doi:10.1093/chemse/bjq147
Advance Access publication February 22, 2011

An Artificial Neural Network Approach for Glomerular Activity Pattern Prediction Using the Graph Kernel Method and the Gaussian Mixture Functions

Zu Soh[1], Toshio Tsuji[1], Noboru Takiguchi[2] and Hisao Ohtake[3]

[1]Department of System Cybernetics, Graduate School of Engineering, Hiroshima University, 1-4-1, Kagamiyama, Higashihiroshima, Hiroshima, 739-8527, Japan, [2]Division of Material Sciences, Graduate School of Natural Science and Technology, Kanazawa University, Kakuma, Kanazawa, 920-1192, Japan and [3]Department of Biotechnology, Graduate School of Engineering, Osaka University, 2-1, Shuita, Yamadaoka, 565-0871, Japan

Correspondence to be sent to: Zu Soh, Department of System Cybernetics, Graduate School of Engineering, Hiroshima University, 1-4-1, Kagamiyama, Higashihiroshima, Hiroshima, 739-8527, Japan. e-mail: sozu@bsys.hiroshima-u.ac.jp

Accepted November 27, 2010

Abstract

This paper proposes a neural network model for prediction of olfactory glomerular activity aimed at future application to the evaluation of odor qualities. The model's input is the structure of an odorant molecule expressed as a labeled graph, and it employs the graph kernel method to quantify structural similarities between odorants and the function of olfactory receptor neurons. An artificial neural network then converts odorant molecules into glomerular activity expressed in Gaussian mixture functions. The authors also propose a learning algorithm that allows adjustment of the parameters included in the model using a learning data set composed of pairs of odorants and measured glomerular activity patterns. We observed that the defined similarity between odorant structure has correlation of 0.3–0.9 with that of glomerular activity. Glomerular activity prediction simulation showed a certain level of prediction ability where the predicted glomerular activity patterns also correlate the measured ones with middle to high correlation in average for data sets containing 363 odorants.

Key words: odor qualities, affactory bulb, olfactory receptor, rats

Introduction

The fact that odors affect human memory and emotions (Manley 1993; Herz and Engen 1996) in addition to enriching our lives places increased importance on information related to odorants, especially in the fragrance, food, and beverage industries (Shahidi et al. 1986; Aznar et al. 2001; Limpawattana et al. 2001). These industries are eager to find new techniques for evaluating odors other than sensory evaluation, which carries problems related not only to consistency across human panels but also to unstable factors within such panels, including sensory fatigue and variations in health condition. The problems inherent in sensory evaluation can be solved only when prediction of odor qualities from odorant molecules becomes possible. In this regard, recent olfactory system analysis in the field of biological research is facing similar problems. By way of example, a combination of behavioral experimentation on rats and measurement of neural activity suggested a relationship between odor quality and neural activity on glomeruli distributed over the surface of the olfactory bulb (Youngentob et al. 2006). Although some lesion studies have suggested that odor perception is not simply related to activated glomeruli (Slotnick and Bisulco 2003), thereby leaving room for argument, another lesion study on mice confirmed that activity in some glomeruli evokes emotions of fear (Kobayakawa et al. 2007). Although this study suggests the possibility of predicting odor qualities from glomerular activity, the sheer scale of the whole odorant space (considered to consist of 400 000 kinds of odorant molecules; Mori 2003) precludes exhaustive measurement. Accordingly, the forecasting of glomerular activity from existing sample data is necessary to achieve the goal of odor quality prediction.

To the best of our knowledge, however, no techniques for glomerular activity prediction have so far been proposed. This paper presents a neural network model for glomerular activity prediction based on the structure of the olfactory system. Because glomeruli involve signals accumulated from receptor neurons (Mori 2003), the main issue in predicting glomerular activity is how the function of receptor neurons should be modeled. To solve this problem, the proposed model employs 2 engineered approaches—a graph kernel

method (Kashima et al. 2004) and an artificial neural network (ANN). The abilities of the model were validated using high-resolution images of glomerular activity patterns in rats (accessed through the Glomerular Activity Response Archive database web site at http://gara.bio.uci.edu). It should be noted that vertebrates share a common olfactory structure, meaning that the use of data sets from rats does not preclude the model's application to other animals.

Material and methods

Glomerular activity patterns

Glomeruli are round clusters of axon terminals accumulated from receptor neurons distributed over the surface of the olfactory bulb. Interestingly, axons from receptor neurons (which are distributed on the surface of the nasal chamber) expressing the same receptor proteins accumulate on the same glomeruli (Mori and Yoshihara 1995). Because each type of receptor protein binds with specific groups of odorants (Buck and Axel 1991; Skoufos et al. 2000), receptor neurons respond exclusively to particular odorants. As a result, the glomerular activity patterns evoked on them are also odor-specific (Mori and Yoshihara 1995; Uchida et al. 2000; Nagao et al. 2002; Mori 2003). In addition, some analysis data have suggested that activity patterns are related to odor qualities (Macrides and Chorover 1972; Hoshino et al. 1998; Youngentob et al. 2006; Kobayakawa et al. 2007). These patterns have been revealed only recently using the 2-deoxyglucose method, and a database on them is provided online (the Glomerular Activity Response Archive, available at http://gara.bio.uci.edu/; Johnson and Leon 2007). Activities are normalized to a z-score so that the activity strength becomes independent of odorant concentration. Further, activity patterns are shown on a 2D map expanded from the surface of the olfactory bulb (see Figure 1).

Using these measured activity patterns, the parameters of the proposed model were adjusted and prediction accuracy was evaluated. Because the Glomerular Activity Response Archive provides activity patterns only in image format, all patterns used in this study were quantified according to the color correspondence described on the web site. The patterns were then divided into a lattice with 1805 squares as shown in Figure 1, where the number of the squares is derived from the actual number of glomeruli (about 2000) found on the olfactory bulb of rats. The purpose of the proposed model is thus to predict the activity strength of each lattice from an input odorant.

Marginalized graph kernel

The marginalized graph kernel is a kernel definition for implementation of graph kernel method (Kashima et al. 2004) and was used for analysis of quantitative structure–activity relationships (Mahé et al. 2005). This kernel, which is employed for calculation of receptor neuron response, provides metrics of similarity between 2 labeled graphs (for odorant molecules in this case) by calculating the probability of correspondence for partial structures as shown in Figure 2. Labeled graph representations of odorant molecules $\tilde{G}_m(\tilde{V}_m, \tilde{E}_m)$and $G_q(V_q, E_q)$are shown in Figure 3, where the atoms and bonds, respectively, correspond to nodes $V \in \tilde{V}_m V_q$ and edges $E \in \tilde{E}_m E_q$. In addition, unique labels are assigned to nodes ($N1$, $N2$, …) and bonds ($E1$, $E2$, …) depending on the type of atom and bond. The following equation describes the

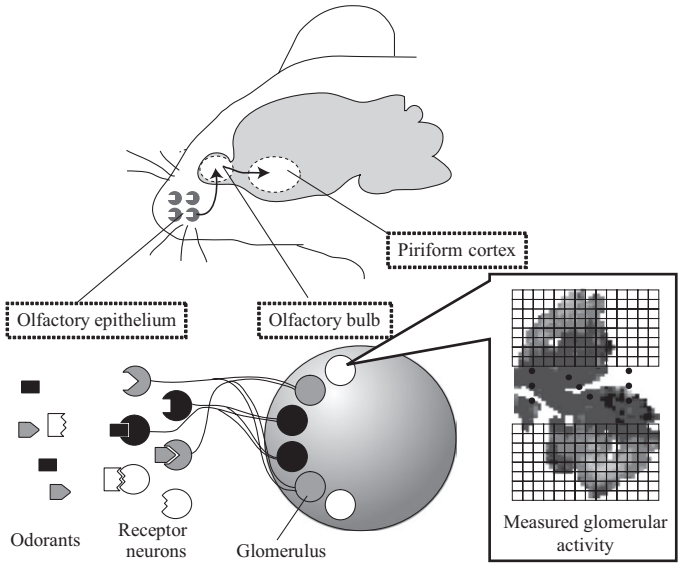

Figure 1 Schematic figure of receptor neurons and glomeruli.

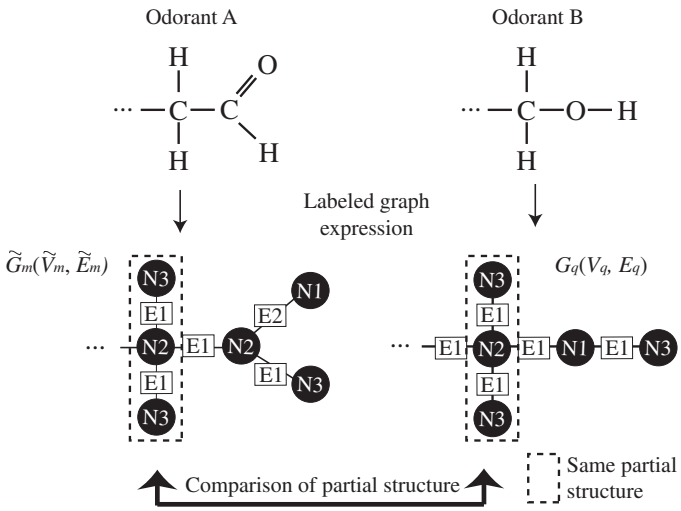

Figure 2 Labeled graph expression of odorant molecules for the marginalized graph kernel. For marginalized graph kernel calculation, the odorants are expressed using a labeled graph in which, for example, $N1$ and $N2$ correspond to oxygen, carbon and hydrogen, respectively, whereas $E1$ and $E2$ correspond to the bond type. Calculation is carried out to find the total correspondence of partial structures, which are shown as sequences of labeled nodes and edges.

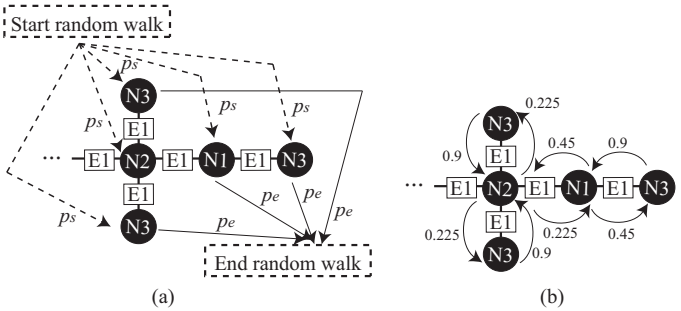

Figure 3 Probability of path occurrence. The probabilities of path matching represent a measure of the similarity between graphs. The occurrence of a path is defined by the probability obtained from a random walk algorithm. **(a)** The starting and ending probability. Assuming that the random walk can uniformly start at any node, the occurrence for a particular starting node is defined as p_s = 1/(total number of nodes). In addition, assuming the random walk can stop at any node with the same probability, the occurrence for ending nodes is a constant value p_e. **(b)** The transition to successive nodes with a probability of (1 - p_e), where p_e = 0.1 (see eq. 4).

marginalized graph kernel used for calculating the similarity between 2 graphs (Kashima et al. 2004):

$$K(\tilde{G}_m, G_q) = \sum_{h_m \in V_M} \sum_{h_m \in V_M} p(h_m) p(h_q) K_L(b(h_m), b(h_q)), \quad (1)$$

$$K_L(b(h_m), b(h_q)) = \begin{cases} 1, & b(h_m) = b(h_q), \\ 0, & b(h_m) \neq b(h_q), \end{cases} \quad (2)$$

where h_m and h_q are the paths of the graphs consisting of sequences of nodes and edges, $p(h_m)$ and $p(h_q)$ are the probabilities of occurrence for paths h_m and h_q, and $K_L(b(h_m), b(h_q))$ is a function that calculates the correspondence of the labels of h_m and h_q. Only when the paths selected from the 2 graphs share exactly the same structure, $(b(h_m) = b(h_q))$ are the probabilities of this correspondence $(p(h_m))p(h_q))$ added up.

The probabilities $p(h_m)$ and $p(h_q)$ are defined using the concept of the random walk algorithm shown in Figure 3. Considering a path h starting from node V_1, then passing through nodes $V_2, V_3, \ldots, V_{R-1}$ and ending at node V_R, the following equation gives the probability $p(h)$ for the occurrence of path h:

$$p(h) = \left\{ p_s \prod_{r=1}^{R} p(V_r | Vr-1) \right\} p_e, \quad (3)$$

which is a multiplication of the starting probability p_s from node V_1, the sequence of transition probabilities $p(V_r|V_{r-1})$ $(r = 1, 2, \ldots, R)$ and the ending probability p_e at node V_R. The random walk algorithm selects the starting node V_1 using the uniform probability $(P_s = 1/N_{\max})$ (shown in eq. 3 and Figure 3a), where $N_{\max}$ is the total number of nodes in the graph. The transition probability $p(V_r|V_{r-1})$ from node V_{r-1} to the next node V_r is then given depending on the number of N_{neighbor} neighboring nodes described by the following equation (see Figure 3b):

$$p(V_r|V_{r-1}) = (1 - p_e)\frac{1}{N_{\text{neighbor}}}, \quad (4)$$

where p_e is the constant probability when the random walk ends at node V_{r-1}.

This marginalized graph kernel is now finding application in the field of chemical informatics (Ralaivola et al. 2005) because it enables the definition of structural similarities between arbitrary chemical compounds. The next section explains the application of the proposed model to the prediction of neural activity on glomeruli.

Proposed model

In order to construct a prediction model for glomerular activity, 2 problems must be solved: the first is the large variety of possible odorants, which increases the difficulty of parameter adjustment, and the second is the complexity of the binding mechanism between receptor neurons and receptor proteins. As a solution to the first problem, the model takes the form of a neural network structure. The learning ability of the neural network enables automatic parameter adjustment from a sample data set. Figure 4 shows the structure of the proposed model, which is composed of a receptor layer, 2 hidden layers, a Gaussian layer and an activity pattern layer. The second problem is solved in the receptor layer using the marginalized graph kernel (Kashima et al. 2004) described in the last section. The model's structure is outlined below.

Model structure

Receptor layer

Receptor neurons are activated by a wide variety of odorants depending on the type of receptor protein expressed on their surface (Buck and Axel 1991). Because the binding between receptor proteins and odorants involves a number of molecular properties, each receptor neuron has a highly complex receptive range (Araneda et al. 2000). In this study, we focused on the structure of odorants as a determinant evoking neural activity because although it has now been found that stereochemical theory (Amoore 1963) is oversimplified, several studies have reported the importance of molecular structure considerations such as functional group and molecular length (Araneda et al. 2000; Haddad et al. 2008). This motivated us to employ the graph kernel method, which can quantify similarities between molecular structures, to approximate the function of receptor neurons.

The receptor layer of the proposed model consists of M receptor units (see Figure 4). Assuming that a representative odorant evokes a particular receptor unit the most, the response to the odorant is defined by the similarity between an input odorant and the representative odorant. The receptor layer consists of M receptor units represented by

marginalized graph kernel as described in the last section. The output of the receptor unit U_m^1 is normalized to the range of [0, 1] using the following equation:

$$U_m^1 = \frac{K(\tilde{G}_m, G_q)}{\sqrt{K(\tilde{G}_m, \tilde{G}_m)K(G_q, G_m)}}, \quad (5)$$

where $K(\cdot,\cdot)$ is the marginalized graph kernel defined in equation (1), and $\tilde{G}_m$ and G_q are the representative odorant and input odorant, respectively, in the labeled graph expression. In addition, each atom was labeled using its atom name together with the Morgan index value (Morgan 1965; Mahé et al. 2004) as shown in Figure 5. This setting prevents the random walk algorithm from generating the same path for different structures. Otherwise, paths extracted from a ring structure and a straight chain structure, for example, would be identical if the same labels were used. The Morgan index assigns a number to each atom using iteration calculation.

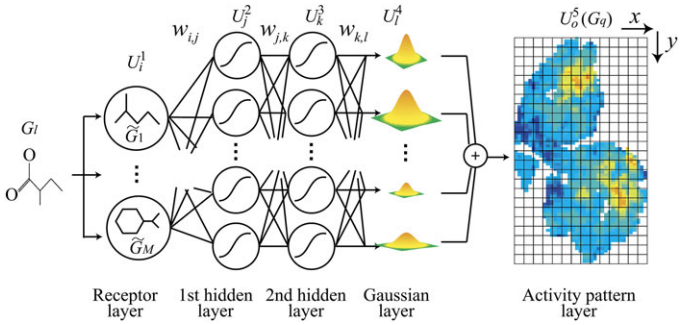

Figure 4 Structure of the proposed model. The proposed model consists of a receptor layer, two hidden layers, a Gaussian layer and an activity pattern layer. The odorant structures are represented by labeled graphs. Each unit in the receptor layer has a representative odorant that activates the unit the most. The output of the receptor layer is determined by the structural similarity between the representative odorant and the input odorant. The output of the receptor layer is converted into peak values of Gaussian units in the Gaussian layer through the hidden layers. Each Gaussian function is allocated to different coordinates on the activity pattern, and the glomerular activity pattern is represented by the sum of the Gaussian functions. This figure appears in color in the online version of *Chemical Senses*.

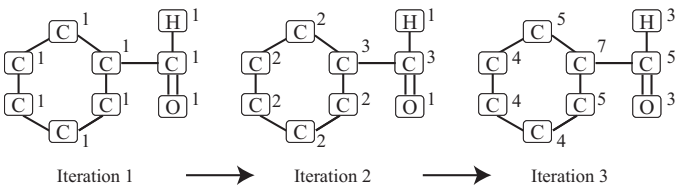

Figure 5 Label settings with the Morgan index. The Morgan index is introduced to differentiate odorant structures by calculation using a simple iterative algorithm. The left side shows the initial state in which the integer 1 is assigned to each node. The center of the figure shows the state of Iteration 2, where the value assigned to each node is the sum of the values assigned to the neighboring nodes in Iteration 1. The right part of the figure shows the state of Iteration 3 with the same process repeated.

For the first iteration, an index value of 1 is assigned to all nodes. For subsequent iterations, the index of each atom is calculated by summation of the numbers assigned to its neighboring nodes. In this study, we set the number of iteration to 5.

Figure 6 shows the relationship between odorant structures and the output of a receptor unit. The calculation is performed using equations (1–5). In the figure, geranyl acetate (shown on the left) is set as a representative odorant, and the output to each odorant is shown in the lower row. This figure illustrates how output falls with larger structural differences between odorants.

Hidden layers

Hidden layers are classic feed-forward neural network model elements that convert the input from the receptor layer into the strength of the activity at the connected region on the glomerular layer. Each of the 2 hidden layers consists of MN sigmoidal function units. Here, N represents the number of Gaussian functions assumed to approximate the major glomerular activity evoked by the representative odorant. The outputs of the hidden layers are, respectively, given by the following equations:

$$U_j^2 = \frac{1}{1+\exp\left\{-a\left(\sum_m w_{m,j} U_m^I - \theta\right)\right\}}, \quad (6)$$

$$U_k^3 = \frac{1}{1+\exp\left\{-a\left(\sum_j w_{j,k} U_j^2 - \theta\right)\right\}}, \quad (7)$$

where a and θ are the gain and threshold constant of the sigmoidal function, respectively.

Gaussian and activity pattern layers

The outputs of the hidden layers are input to the Gaussian layer through connective weights $w_{k,l}$ and generate the glomerular activity strength according to the following equation:

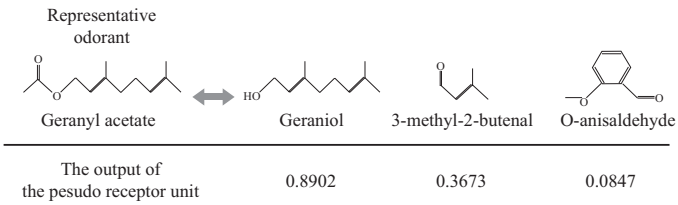

Figure 6 Output U_m^1 calculated using marginalized graph kernel. The upper row shows odorants, and the lower row shows their calculated graph kernel values in relation to the representative odorant (geranyl acetate). Calculation is carried out using equations (1–5), and the number of iterations for the Morgan index was set to 5. These values indicate the similarity between odorants and are defined as the output of the receptor layer.

$$U^4_{l,(x,y)} = \sum_k w_{k,l} U^3_k \exp\left\{-\alpha(x_{c,l}-x)^2 - \beta(y_{c,l}-y)^2\right\}, \quad (8)$$

where $(x_{c,l}, y_{c,l})$ denotes the center coordinates on the activity pattern layer to which a Gaussian unit is connected, and the parameters α and β control the width of the Gaussian curve. The activity pattern layer consists of 1805 linear function units each allocated to a set of coordinates corresponding (x, y) to the lattice squares shown in Figure 1. The units add up the input from the Gaussian layer according to the following equation:

$$U^5_{0,(x,y)}(G_q) = \sum_1 U^4_{l,(x,y)}. \quad (9)$$

Consequently, the glomerular layer allows the prediction of glomerular activity $U^5_{0,(x,y)}(G_q)$ for an input odorant G_q.

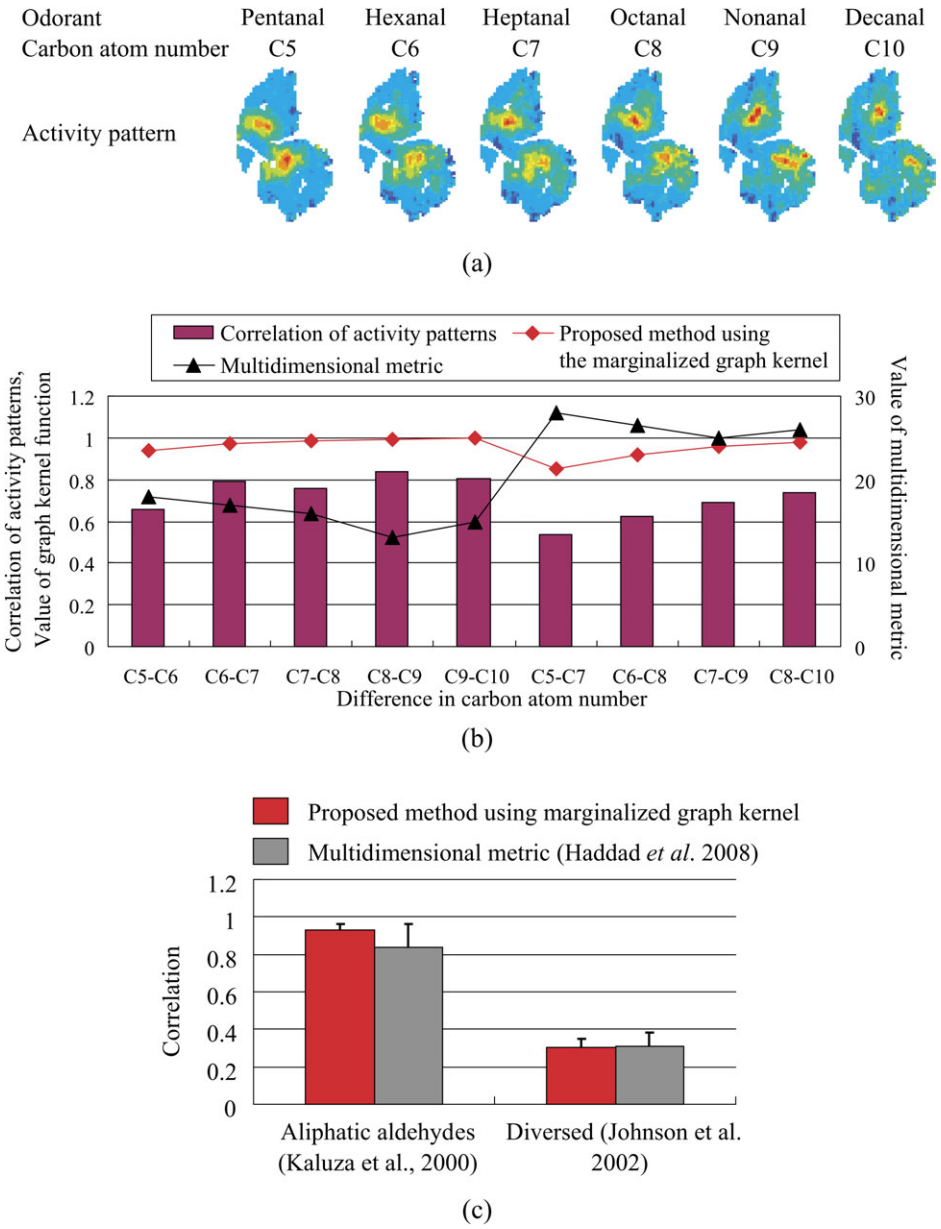

Figure 7 Comparison between the proposed model using marginalized graph kernel and the multidimensional metric. **(a)** The data set of pairs of odorants and activity patterns (from the Glomerular Activity Response Archive) used for comparison. **(b)** The correlation of activity patterns with the proposed method using marginalized graph kernel and multidimensional metric. The multidimensional metric shows the reversed trend because it is a metric of dissimilarity, whereas others deal with similarity. **(c)** The average correlation of activity patterns with the proposed method and multidimensional metric. Horizontal axis shows each odorant dataset. Error bars reflect correlation confidence interval at 0.05. The both metric have equivalent correlation value. This figure appears in color in the online version of *Chemical Senses*.

Although the number of linear units is determined based on the order of the actual number of glomeruli, it should be noted that individual units do not correspond to actual glomeruli; we do not intend to predict the activity of each glomerulus.

Learning algorithm

To enable automatic adjustment of the parameters included in the model so that it can estimate the corresponding glomerular activity from the input odorant, a 2-step learning algorithm is proposed. Before learning, an arbitrary odorant data set that contains L odorants and corresponding glomerular activities is divided into learning data set 1, learning data set 2, and a validation data set.

Step 1

Using learning data set 1, step 1 allocates the Gaussian functions at the proper coordinates on the glomerular layer and adjusts the width of the Gaussian curve. Each glomerular activity pattern is approximated using different N Gaussian functions. This process is carried out using the orthogonal least squares learning algorithm developed by Chen et al. (1991) generally used for function approximation of radial base function networks.

Step 2

Step 2 adjusts the connective weights between each layer $(w_{m,j}, w_{j,k}, w_{k,l})$ using learning data sets 1 and 2. In this step, odorants in learning data set 1 are first assigned to receptor units as representative odorants. The odorants in data sets 1 and 2 are then input to the model so that the hidden layer produces peak values for the Gaussian functions according to equation (8). As a result, the output of the glomerular layer approximates the measured glomerular activity pattern. Finally, a steepest-gradient method called RPROP (Riedmiller and Braun 1993) is employed to adjust the connective weights through minimization of the sum of the mean squared errors (MSEs) between the outputs of the glomerular layer and the measured glomerular activity with the following equation:

$$E(G_q) = \frac{1}{1805} \sum_{x,y}^{X,Y} \left(U_{T,(x,y)}(G_q) - U^5_{(x,y)}(G_q) \right)^2, \qquad (10)$$

where $U_{T,(x,y)}(G_q)$ is the measured activity at coordinates (x,y) for input odorant G_q, and 1805 is the total number of linear units on the glomerular layer.

Results

Comparison between the marginalized graph kernel used in the proposed model and the multidimensional metric

This section validates the ability of the marginalized graph kernel used in the proposed model by comparing it with the multidimensional metric proposed by Haddad et al. (2008). This metric uses the Euclidian distance between 1664 odorant descriptors to represent the difference between odorants. Haddad et al. reported that the metric better correlates with the neural response reported in previous studies (Kaluza and Breer 2000; Johnson et al. 2002) than traditional metrics such as the carbon atom number. For comparison, we employed the same strategy and odorant data set.

Before outlining the analysis results, let us introduce activity correlation r_A referring to the similarity of glomerular activities defined by the correlation between them and the graph kernel value referring to the structural similarity of odorants calculated using the marginalized graph kernel. In addition, we propose the kernel-activity correlation r_k referring to the correlation between the activity correlation r_A and the graph kernel value.

First, comparison was implemented with a relatively small data set that shares the same structure (except the carbon atom number) consisting of aliphatic aldehydes with 5–10 carbons (C5–C10). The corresponding neural activity was ascertained from the Glomerular Activity Response Archive (Johnson and Leon 2007) rather than adopting that of olfactory neurons (Kaluza and Breer 2000) used in the study of Haddad et al. (2008) because our objective was to predict glomerular activity. As a result, the graph kernel value was seen to have a high level of correspondence to glomerular activity patterns, with a correlation of $r_k = 0.93$ ($p = 6.4 \times 10^{-7}$) as shown in Figure 7, whereas the same value was –0.69 ($p = 0.03$) for the multidimensional metric. It should be noted that the sign of the correlation coefficient for the

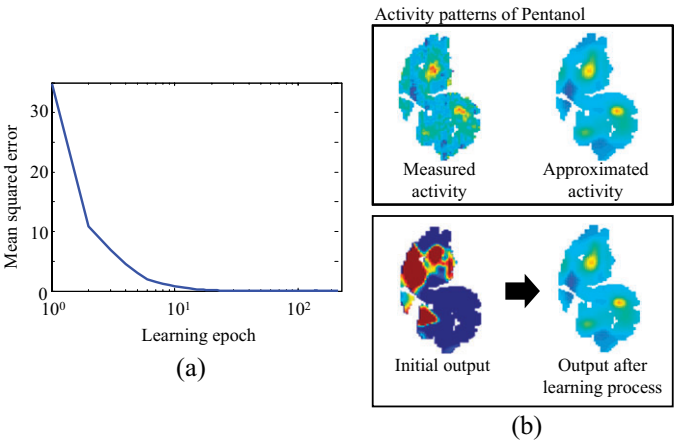

Figure 8 Learning results for the data set obtained from Johnson et al. (2002). **(a)** The learning process successfully reduced the error between the outputs of the model and the measured glomerular activities. The upper side of **(b)** shows the approximated activity using the same number of Gaussian functions as the model, which facilitates comparison and evaluation between the outputs of the model and the measured activities. The lower side of (b) shows the learning ability of the model using pentanol as an example and compares the initial outputs of the model before and after the learning process. This figure appears in color in the online version of *Chemical Senses*.

multidimensional metric is reversed because it measures dissimilarities between odorants, whereas the other one measures similarities. Although it appears that the multidimensional metric underperformed the marginalized graph kernel in this case, in the original paper of Haddad et al., the multidimensional metric was compared with a different neural activity data set (percentages of discriminating olfactory neurons; Kaluza and Breer 2000), and the authors reported a correlation of $r = 0.84$ ($p = 0.004$). This result suggests that the marginalized graph kernel can equally account for neural activity as the multidimensional metric does in the case of odorants from a single chemical family.

We then used a more diverse odorant set (Johnson et al. 2002) for comparison. This data set contains 54 odorants including carboxylic acids, ketones, alcohols and so on. The activity patterns adopted for this comparison were the same as those used in Haddad's study. The kernel-activity correlation obtained in this way was 0.30 ($p = 3.5 \times 10^{-13}$) and was about 0.30 ($p < 1 \times 10^{-10}$) for the multidimensional metric. Therefore, an equivalent correlation values were observed.

From the results described above, we considered the possibility of estimating the glomerular activity space using multiple marginalized graph kernel axes. Although, the marginalized graph kernel principally cannot describe odorant properties besides structural similarity, this simplification brings an advantage over the multidimensional metric in that it requires only one algorithm for calculation while offering a nearly equivalent level of performance. It should be mentioned that Haddad et al. also proposed an optimized multidimensional metric and obtained a correlation of about $r = 0.55$ ($p < 1 \times 10^{-10}$) regarding the data set (Johnson et al. 2002). In our model, a similar optimization technique was applied using the neural networks as well as predicting glomerular activity.

Learning and prediction abilities of the model

This section reports on verification of the learning algorithm and the prediction ability of the proposed model. The simulation described here was implemented using Matlab numerical computing software. To test the model, we questioned its abilities in 2 areas: (i) its capacity to predict the glomerular activity measured in the previous study and (ii) prediction result accuracy when a data set suitable for the proposed model is provided.

Firstly, the model was tested employing a data set ($Q = 54$ odorants) listed in a previous study by Johnson et al. (2002)

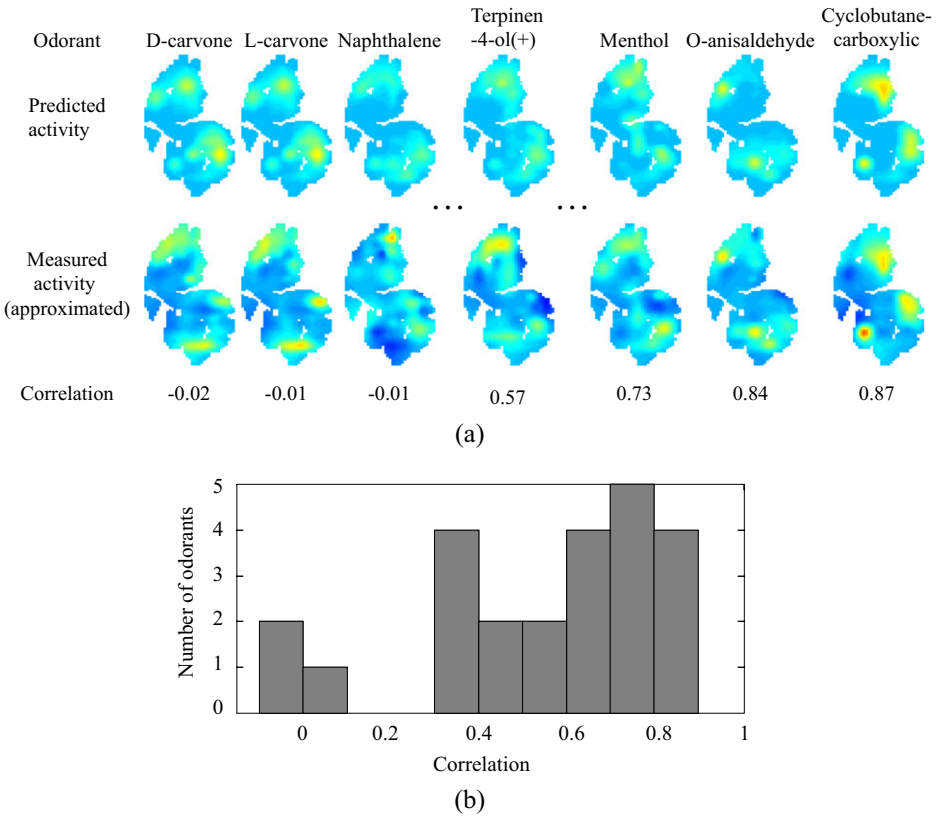

Figure 9 Prediction results for the data set obtained from Johnson et al. (2002). **(a)** Example of prediction results. The 3 best and worst predicted glomerular activity patterns are shown in the upper column, and the measured activity patterns are shown in the lower column. In addition, a predicted result with an activity correlation of about $r_A = 0.5$ is shown in the center for reference. **(b)** Histogram showing prediction accuracy. The horizontal axis is the activity correlation between predicted and measured activity patterns, and the vertical axis is the number of odorants in each bin of activity correlation. This figure appears in color in the online version of *Chemical Senses*.

Table 1 Odorant list

1-Decanol	Hexanal	2,3,5-Trimethylpyrazine
1-Heptanol	Dodecane	2,3-Dimethylpyrazine
1-Hexanol	Tridecane	2,5-Dimethylhexane
1-Nonanol	Tetradecane	2,5-Dimethylpyrazine
1-Octanol	Hexadecane	2,6-Dimethylpyrazine
1-Pentanol	1,3,5-Triisopropylbenzene	2-Decanone
Decanal	1,7-Octadiene	2-Methylheptane
Nonanal	1,7-Octadiyne	2-Nonanone
Octanal	2,3,4-Trimethylpentane	2-Octynoic acid
Heptanal	2,3,5,6-Tetramethylpyrazine	2-Undecanone
5-Methylfurfural	Isoamyl butyrate	*trans*, *trans*-2,4-decadienal
9-Decen-1-ol	Menthyl isovalerate	*trans*-2-tridecenal
Adoxal	Methyl heptanoate	Undecane
α-Angelica lactone	Methyl octanoate	
Amyl acetate	Methyl-*trans*-2-octenoate	
Decane	Nonane	
Dodecanal	Norbornane	
Ethyl caprylate	Pentyl propionate	
γ-Undecalactone	Phenyl propionate	
Hexyl acetate	Phytol	

as used for analysis of the marginalized graph kernel in the last section. Simulation was performed under the following conditions:

1. Data sets—$Q = 54$ odorants were randomly divided into 3 data sets: learning data set 1 contained 15 odorants assigned to receptor units as representative odorants, learning data set 2 contained 15 odorants, and the other 24 odorants fell into the validation data set.
2. Neural network structure—the unit numbers for each layer were set as follows: $M = 15$ units for the receptor layer (corresponding to the number of representative odorants) and $J = K = L = MN = 75$ units for the hidden layers and the Gaussian layer.
3. Learning algorithm parameters—learning step 1, in which every glomerular activity was approximated using different $N = 5$ Gaussian functions, was stopped when the rate of decrease in the MSE became lower than 1×10^{-6}, or when the number of learning epochs exceeded 30 000. Learning step 2 was stopped when the MSE (see eq. 9) fell below the learning goal of $E = 1 \times 10^{-3}$ or when its rate of decrease become lower than 1×10^{-6}.
4. Evaluation—The prediction target of the measured activity pattern was smoothed using the same number of Gaussian functions as those included in the model. Prediction accuracy was evaluated from the correlation with these smoothed activity patterns. This evaluation process allowed us to focus on the learning and prediction abilities of the model while avoiding errors caused by comparison between smoothed and unsmoothed activity patterns.

Figure 8 shows the results of the learning simulation. Although the learning goal was not met, the MSE decreased to a level very close to it (Figure 8a). Figure 8b shows an example of the glomerular activity indicated by the model before and after learning and indicates an improvement in output. After this learning process, the proposed model could produce learned glomerular activity with an activity correlation of $r_A > 0.99$ for all odorants in learning data sets 1 and 2. This result confirmed its learning ability.

Prediction simulation was then carried out. Figure 9a shows the 3 best and worst prediction results and Figure 9b is a histogram showing the prediction performance for the odorants in the validation data set. Simulation was repeated 10 times with randomly shuffled data sets. The results showed that $63.2 \pm 10.7\%$ of predicted activity patterns correlated with measured ones with an activity correlation of $r_A > 0.5$ (note the example of the activity pattern with $r_A > 0.5$ in the center of

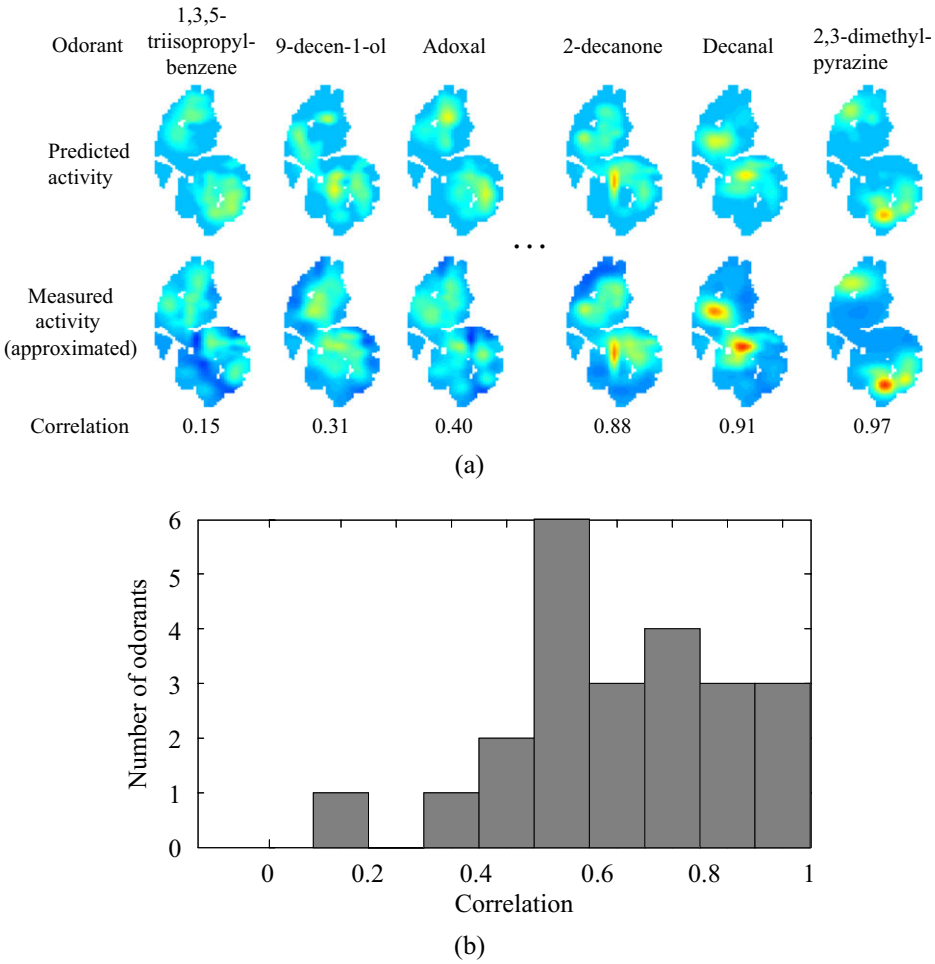

Figure 10 Prediction results for the data set containing odorants with a high kernel-activity correlation. **(a)** The predicted and measured activity patterns when both the learning data sets and the validation data set contain odorants with a kernel-activity correlation of $r_k > 0.5$. **(b)** The accuracy of prediction through a histogram of the correlation between predicted and measured activity patterns. This figure appears in color in the online version of *Chemical Senses*.

Figure 9a). In addition, the average correlation coefficient was 0.55 ± 0.26. These results show a certain level of prediction ability for the proposed model.

Prediction performance was then tested in a case where all odorants in the data set had a higher kernel-activity correlation. A data set containing 53 odorants with a kernel-activity correlation of $r_k > 0.5$ was extracted from the 363 odorants in the database. The odorant list is shown in Table 1. Using the same process as that outlined for the first simulation, learning and prediction were performed. Figure 10 shows an example of the best and worst prediction results along with a histogram of prediction accuracy. The results indicated that, despite some prediction difficulty for a few odorants, most activity patterns were successfully predicted. Simulation was repeated 10 times with random selection of the learning data set and the validation set. It was observed that about 79.9 ± 5.5% of predicted glomerular activities showed an activity correlation of $r_A > 0.5$ to the measured ones, and average correlation was 0.67 ± 0.21. These results confirmed the prediction ability of the model when a suitable data set is provided.

Further, Figure 11 summarizes the prediction accuracy of simulation performed on the 2 data sets. These results suggest that the model predicts significantly better for the data set containing odorants with a high kernel-activity correlation.

Discussion

This paper proposed a glomerular activity prediction method using a combination of the graph kernel method and neural network. Here, we discuss about contribution of the both method on prediction ability.

The simulation results showed that prediction accuracy depends on the kernel-activity correlation of the odorants included in data sets. To clarify this dependency, relationship between prediction accuracy and kernel-activity correlation for each prediction target odorant is tested.

Figure 12 shows the relationship of the data set obtained from Johnson et al. (2002). As shown in Figure 12, a middle correlation of 0.33 ($p = 2 \times 10^{-7}$) between them was found, meaning that when the prediction target is an odorant whose

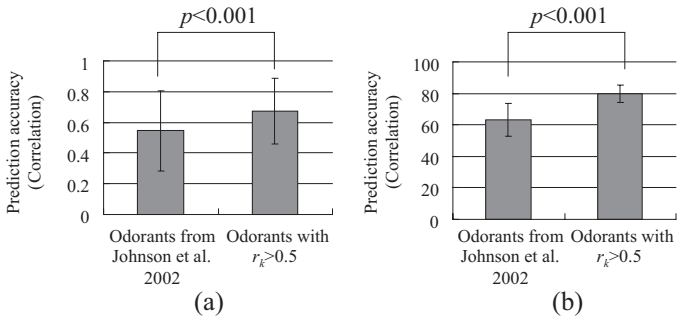

Figure 11 Summary of prediction accuracy. The both figures shows performance of the proposed model regarding a data sets obtained from Johnson et al. (2002) and another data set containing the odorants with a high kernel-activity correlation. **(a)** The average activity correlation between the measured and predicted activity patterns. **(b)** Average percentage when the predicted activity pattern correlates with the measured ones at $r_A > 0.5$. A *t*-test at a significance level 0.001 confirmed a superior level of prediction ability for the data set containing odorants with a high kernel-activity correlation.

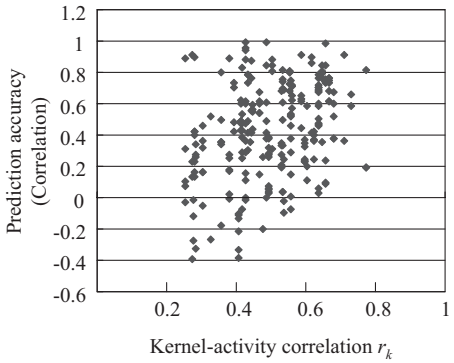

Figure 12 Scatter plot of prediction accuracy and its kernel-activity correlation. Prediction accuracy is defined as the correlation between predicted and measured activity patterns. The correlation between prediction accuracy and the kernel-activity correlation was $r = 0.33$ ($p = 2.0 \times 10^{-7}$). Accordingly, when the prediction target has a higher kernel-activity correlation, better prediction results can be expected.

structural similarity to other odorants correlates to the similarity of activity patterns, prediction accuracy tends to be high. This result also suggests that prediction accuracy can be improved by improving the calculation algorithm of the graph kernel method.

We then tested advantages of using a neural network over more straightforward and simpler method. The model's performance was compared with a simple prediction method in which the activity of the learning data set odorant most closely related to the prediction target (i.e., the one with the highest graph kernel value) is found, and its activity pattern is taken as the prediction result. This evaluation protocol was implemented to validate the ability of the neural network that merges multiple activity patterns in learning data set 1 to predict that of the target odorant. Simulation was implemented using 100 pairs of odorants and glomerular activity patterns

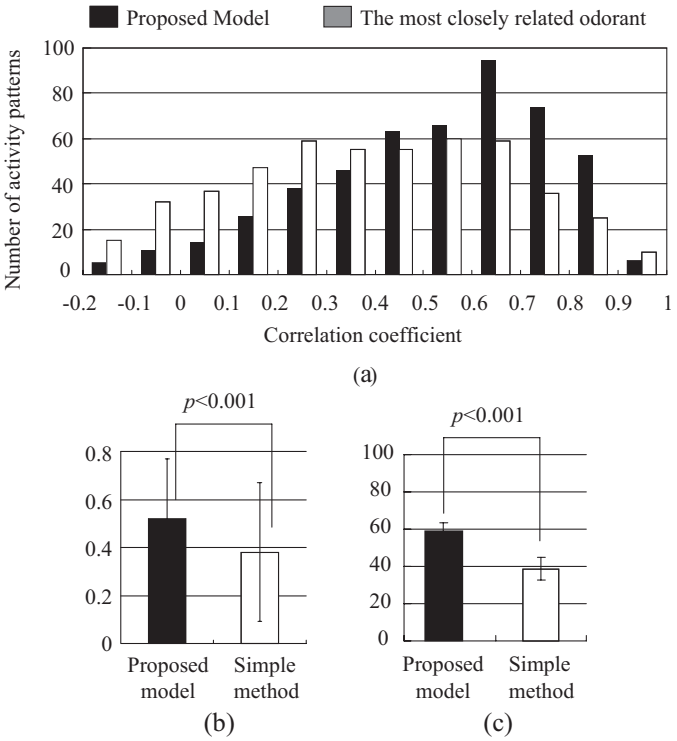

Figure 13 Performance comparison between the proposed model and a simple method. **(a)** Prediction accuracy between the proposed model (black bars) and a simple method (white bars) that predicts activity patterns by taking that of the odorant with the highest graph kernel value (i.e., the most closely related odorant). **(b)** The average correlation between measured and predicted activity patterns. **(c)** The average percentage when the predicted activity pattern correlates with the measured ones at $r_A > 0.5$. As a result of *T*-test at significance level 0.001, we confirmed that the proposed model performed better prediction.

were randomly selected as a data set for simulation out of the 363 provided in the Glomerular Activity Response Archive (Johnson and Leon 2007). In the selected data set, learning data sets 1 and 2 each consisted of 25 random odorants. The other 50 odorants were used for validation.

The configurations for the simulation were set to the same as that in the simulation in the last section. The histogram in Figure 13 shows the distribution of prediction accuracy for 500 prediction targets (50 odorants/trial × 10 trials). Comparing the prediction results of the model (red bars) and those of the simple method (gray bars), it is observed that the proposed model predicted more activity patterns correlating with the measured ones with $r_A > 0.5$ and fewer patterns with $r_A < 0.5$ than the simple method. The percentage of activity pattern predictions with $r_A > 0.5$ was then compared using a *t*-test at significance level 0.001 as shown in Figure 13b. These results confirmed that the proposed model had a significantly better level of prediction ability (60.1 ± 4.4%) than the simple method (38.8 ± 6.0%). The model also predicted significantly better average correlation between predicted and measured activity patterns (0.53 ± 0.25) than the simple method (0.38 ± 0.28) as shown in Figure 13c.

Advantages of using neural network thus were confirmed from these comparison results.

Conclusion and future work

The purpose of this study was to predict glomerular activity patterns. The proposed method employed a combination of a neural network model and a graph kernel method.

The marginalized graph kernel, which is one of the definitions for implementation of graph kernel method, serves as a simple way to define a metric of structural similarity between 2 molecules (odorants) (Kashima et al. 2004). As previous studies reported that the structure of an odorant is an important determinant for neural response (Araneda et al. 2000; Johnson and Leon 2007), the marginalized graph kernel was applied for the first stage of glomerular activity prediction. Its ability was verified in comparison to the multidimensional metric proposed by Haddad et al., which accounts for neural response with mid-to-high correlation (Haddad et al. 2008). The marginalized graph kernel has a critical advantage in that it requires only a single algorithm for calculation, in contrast to the multidimensional metric, which requires 1664 kinds of odorant properties and thus the same number of calculation algorithms.

The neural network model is intended to convert graph kernel values into activity patterns. The neural network has the important feature of allowing automatic parameter adjustment using a variety of sample odorant data sets (learning data sets); otherwise, it would be impossible to approximate the complex relationships between multiple graph kernel values and activity patterns. The learning simulation confirmed that the model successfully converted learned odorants into corresponding activity patterns. We then repeated learning and prediction simulation using different data sets and observed a certain level of prediction ability in the model. We also observed a significantly higher level of prediction ability compared with a simple method in which the activity pattern is predicted using that of the data set odorant most closely related to the prediction target (i.e., the one with the highest graph kernel value).

Although these simulation results confirmed a certain level of predictability for glomerular activity, prediction accuracy needs to be improved. We consider that the following 2 tasks should be addressed to improve the model. First, the graph kernel method should be improved as enantiomers would be classified as identical even though they evoke distinct activity patterns (Rubin and Katz 2001). Prediction ability can therefore be improved by expanding the algorithm to allow evaluation of enantiomers. Another task is to increase the number of input dimensions for improvement. As previous studies have suggested that neural activity is not determined simply by odorant structures (Soucy et al. 2009), improvement of the model will necessarily be limited unless other odorant properties are added as inputs.

Due to the enormous variety of odorants in existence, we believe the proposed approach is necessary for future investigation into the features of glomerular response and that it can serve as a simulation tool to help reveal the mechanisms behind the olfactory system.

Funding

This study was supported in part by a Grant-in-Aid for Scientific Research [20 9124] from the Research Fellowships of the Japan Society for the Promotion of Science for Young Scientists.

Acknowledgements

The authors would like to acknowledge associate professor Hisashi Kashima, the University of Tokyo, for providing a sample program of the marginalized graph kernel method.

References

Amoore J. 1963. Stereochemical theory of olfaction. Nature. 198:271–272.

Araneda RC, Kini AD, Firestein S. 2000. The molecular receptive range of an odorant receptor. Nat Neurosci. 3:1248–1255.

Aznar M, Lopez R, Cacho JF, Ferreira V. 2001. Identification and quantification of impact odorants of aged red wines from Rioja. GC-Olfactometry, quantitative GC-MS, and odor evaluation of HPLC fractions. J. Agric Food Chem. 49:2924–2929.

Buck L, Axel R. 1991. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. Cell. 65:175–187.

Chen S, Cowan CFN, Grant PM. 1991. Orthogonal least squares learning algorithm for radial basis function networks. IEEE Trans Neural Netw. 2:302–309.

Haddad R, Khan R, Takahashi YK, Mori K, Harel D, Soble N. 2008. A metric for odorant comparison. Nat Methods. 5:425–429.

Herz R, Engen T. 1996. Odor memory review and analysis. Psychon Bull Rev. 3:300–313.

Hoshino O, Kashimori Y, Kambara T. 1998. An olfactory recognition model based on spatio-temporal encoding of odor quality in the olfactory bulb. Biol Cybern. 79:109–120.

Johnson BA, Ho SL, Xu Z, Yihan JS, Yip S, Hingco EE, Leon M. 2002. Functional mapping of the rat olfactory bulb using diverse odorants reveals modular responses to functional groups and hydrocarbon structural features. J Comp Neurol. 449:180–194.

Johnson BA, Leon M. 2007. Chemotopic odorant coding in a mammalian olfactory system. J Comp Neurol. 503:1–34.

Kaluza JF, Breer H. 2000. Responsiveness of olfactory neurons to distinct aliphatic aldehydes. J Exp Biol. 203:927–933.

Kashima H, Tsuda K, Inokuchi A. 2004. Kernels for graphs in kernel methods. In: Bernhard S, Tsuda K, Jean-Philippe V, editors. Computational biology. Cambridge, MA: MIT Press. p. 166–181.

Kobayakawa K, Kobayakawa R, Matsumoto H. 2007. Innate versus learned odour processing in the mouse olfactory bulb. Nature. 450:503–508.

Limpawattana M, Yang S, Yang DS, Kays SJ, Shewfelt RL. 2001. Relating sensory descriptors to volatile components in flavor of specialty rice types. J Food Sci. 73:S456–S460.

Macrides F, Chorover SL. 1972. Olfactory bulb units: activity correlated with inhalation cycles and odor quality. Science. 175:84–87.

Mahé P, Ueda N, Akutsu T, Perret JL, Vert JP. 2004. Extensions of marginalized graph kernels. In: Brodley CE, editor. Proceedings of the Twenty-First International Conference on Machine Learning. New York: ACM. p. 552–559.

Mahé P, Ueda N, Akutsu T, Perret JL, Vert JP. 2005. Graph kernels for molecular structure-activity relationship analysis with support vector machines. J Chem Inf Model. 45:939–951.

Manley CH. 1993. Psychophysiological effect of odor. Crit Rev Food Sci Nutr. 33:57–62.

Morgan H. 1965. The generation of unique machine description for chemical structures - a technique developed at chemical abstracts service. Journal of Chemical Documentation. 5:107–113.

Mori K. 2003. Grouping of odorant receptors: odour maps in the mammalian olfactory bulb. Biochem Soc Trans. 31:134–136.

Mori K, Yoshihara Y. 1995. Molecular recognition and olfactory processing in the mammalian olfactory system. Prog Neurobiol. 45:585–619.

Nagao H, Yamaguchi M, Takahashi Y, Mori K. 2002. Grouping and representation of odorant receptors in domains of the olfactory bulb sensory map. Microsc Res Tech. 58:168–175.

Ralaivola L, Swamidassa SJ, Saigoa B, Baldia P. 2005. Graph kernels for chemical informatics. Neural Netw. 18:1093–1110.

Riedmiller M, Braun H. 1993. A direct adaptive method for faster backpropagation learning: the RPROP algorithm. Proc IEEE Int Conf Neural Netw. 1:586–591.

Rubin BD, Katz LC. 2001. Spatial coding of enantiomers in the rat olfactory bulb. Nat Neurosci. 4:355–356.

Shahidi F, Rubin LJ, D'Souza LA. 1986. Meat flavor volatiles: a review of the composition, techniques of analysis, and sensory evaluation. Crit Rev Food Sci Nutr. 24:141–243.

Skoufos E, Marenco L, Nadkarni PM, Miller PL, Shepherd GM. 2000. Olfactory receptor database: a sensory chemoreceptor resource. Nucleic Acids Res. 28:341–343.

Slotnick B, Bisulco S. 2003. Detection and discrimination of carvone enantiomers in rats with olfactory bulb lesions. Neuroscience. 121:451–457.

Soucy ER, Albeanu DF, Fantana AL, Murthy VN, Meister M. 2009. Precision and diversity in an odor map on the olfactory bulb. Nat Neurosci. 12:210–220.

Uchida N, Takahashi YK, Tanifuji M, Mori K. 2000. Odor maps in the mammalian olfactory bulb: domain organization and odorant structural features. Nat Neurosci. 3:1035–1043.

Youngentob SL, Johnson BA, Leon M, Sheehe PR, Kent PF. 2006. Predicting odorant quality perceptions from multidimensional scaling of olfactory bulb glomerular activity patterns. Behav Neurosci. 120: 1337–1345.