VECIMS 2003 – International Symposium on
Virtual Environments, Human-Computer Interfaces, and Measurement Systems
Lugano, Switzerland, 27-29 July 2003

User Type Identification by Mixing Weight Estimation of Mixture Models based on State Space Modeling

Norikazu Ikoma, Witold Pedrycz*, Keiko Baba, Takahiro Hyakudome,
Yosuke Matsumoto, Nagatomo Nakamura**, Hiroshi Maeda

Faculty of Engineering, Kyushu Institute of Technology, Fukuoka 804-8550, Japan
*Dept.of Electrical & Computer Engineering, University of Alberta, Edmonton, T6G 2H4, Canada
**Dept.of Economics, Sapporo Gakuin University, Sapporo, 069-8555, Japan

Abstract – ***An approach to adaptive user interface using mixture model and state space model is proposed. Mixture model is applied to response data of many users to extract user types in a preliminary experiment. Estimated components are regarded as "user types". Online identification of the type of a new user from his/her response series is done by state space model, where the weights of the components constitute the state vector. In the state space model, the system equation defines a time-smoothness of the weights and the observation equation consists of a mixture model allocated to the time-varying weights. State estimation is done by using particle filter. We propose to use the identification result of the new user to an adaptive user interface by showing an appropriate screen based on the estimated weights. Numerical simulation illustrates type identification result of new user. Real data analysis using key-typing performance with methods using both-hands, right(dominant)-hand, left(non-dominant)-hand, and one finger is also reported.***

I. INTRODUCTION

Recent developments of computers and related technologies allow us to make user interfaces more intelligent. Among many researches trying this, adaptation to user is very attractive to achieve the intelligent user interface [3], [9], [12], [17], [19]. To make the adaptation, knowing the current state of the user is fundamental. It is also an important task to know what type of users is now using the system [15], [16]. We focus on the latter task in this article.

When the potential users are pre-determined or it is possible to administrate the users, the most effective way to know the user information is to store its preference for all users. However, it requires user authentication and registration of the preference automatically or manually. Consequently, this approach is not suited for public system such as guidance system at information, web-site with anonymous access, and so on. It is also the case for a touch panel user interface on intelligent wheel chairs that we are now developing, which can adapt the size of button on the panel depending on user [1], [12], [13], [14].

Another approach to make user adaptation is to classify the users into several typical groups and using this information effectively [2], [15],[16]. We call the group a "type", and we adopt the more exact definition of a type as users group in which users have similar properties such as their response to the same situation. For systems in which it is difficult to make authentication and registration, adaptation based on types instead of individual users is an efficient way for user adaptive interface. Here, knowing the types of a user is an important task for the user adaptation. In some cases, an expert of user interface design can determine the types from his/her knowledge of the typical user [15],[16]. In other cases, the groups are a priori defined. But otherwise, we have to extract the type information hidden in huge data set of user response by some un-supervised learning algorithm, such as the EM (Expectation-Maximization) algorithm [5].

In this paper, we propose an approach to adaptive user interface using the information of user type by using mixture model and state space model. A mixture model is used to extract the user types from response data of many users in a preliminary experiment. We regard the components estimated by mixture model as a probability distribution of user type, and use them in the state space model with time-varying weight for each component. Identification of the type of a new user from his/her small observation series of response data is done by state space model. In the state space model, online estimation of the time-varying weight parameters is done by particle filtering [7]. Then we can know the degree of membership of the user to each component (type). The proposed model exploits two initial components as a prior in the context of Bayes estimation; 1) initial value as mixing weights of the mixture model, 2) smooth change of mixing weights represented by state space model.

II. MODEL

A. *Extraction of User Types by Mixture Model*

Let $\mathbf{U}$ be an input space (subspace of $\Re^m$) that corresponds to information presented to the user. Denote an input variable (vector in general) by $\mathbf{u} \in \mathbf{U}$. Let $\mathbf{Y}$ be an output space (subspace of $\Re^n$) that corresponds to the measurement of the user response, and define the output variable (vector in general) as $\mathbf{y} \in \mathbf{Y}$. Let $\mathbf{Z}$ be a space of external input (subspace of $\Re^l$) that corresponds to the condition given to the user, and define the external input variable (vector in general) as, $\mathbf{z} \in \mathbf{Z}$.

We assume that a type of users for a given input $\mathbf{u}$ and external input $\mathbf{z}$ is modelled by a probability density function

$$f_j(\mathbf{y}|\mathbf{u}, \mathbf{z}; \theta_j) \tag{1}$$

where j denotes a j-th type of the users, and θ_j is a parameter vector that specifies the density function. Eq.(1) will be referred to as a component.

For many users, where the type of each user is unknown, the law of measurement variable with given input $\mathbf{u}$ and external input $\mathbf{z}$ is modelled by the following mixture

$$f(\mathbf{y}|\mathbf{u}, \mathbf{z}; \Theta) = \sum_{j=1}^{p} w_j f_j(\mathbf{y}|\mathbf{u}, \mathbf{z}; \theta_j) \tag{2}$$

where

$$\Theta = \{\theta_1, \theta_2, \cdots, \theta_p\} \tag{3}$$

is a set of parameter vectors for all components. Let

$$\mathbf{w} = [w_1, w_2, \cdots, w_p]^T \tag{4}$$

denotes a mixing weight vector (which consists of weight variable w_j for each user type $j = 1, 2, \cdots, p$). $\mathbf{w}$ satisfies two conditions

$$\sum_{j=1}^{p} w_j = 1, \qquad w_j \geq 0 \;\; (j = 1, 2, \cdots, p) \tag{5}$$

For estimating the parameters in eq.(2), we need to collect data of response for many users in different situation by a preliminary experiment. Here, for many users with different types, we measure the response of the users by giving various conditions of input/external input, then we obtain a data set to estimate the mixture model of eq.(2).

In general, it is unknown what kind of user type should be prepared in advance, except the case where each type corresponds to each group that is defined objectively. In the latter case, we simply apply traditional estimation method such as maximum likelihood, or moment matching method to estimate each component distribution, and let the weights of components be uniform. For the former case, i.e., there is no information about types, we have to extract the types of users using un-supervised learning algorithm such as EM (Expectation-Maximization) algorithm [5]. We also assume the number of components $\hat{p}$ is determined by an appropriate method.

By estimating the mixture model from the data set, we have obtained the parameter estimates $\hat{\Theta} = \left\{\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_p\right\}$ and weight vector $\hat{\mathbf{w}} = \{\hat{w}_1, \hat{w}_2, \cdots, \hat{w}_p\}$. Now, we have obtained components denoted by eq.(1) together with $\hat{\theta}_j$ and weight $\hat{w}_j$ for $j = 1, 2, \cdots, p$, and then, we regard the components as the types of the users.

B. *Identification of User Type by State Space Model*

After extracting the types of users by mixture model, let us consider how to use this result to identify the type of new-coming user. Since the type of a new-coming user is unknown, it must be estimated by some method. For this purpose, we propose to estimate the mixing weight of mixture model based on measurement data of the new-coming user's response. We regard the estimated weight as membership values of the user for each type.

In this situation, many measurements for various input/external input patterns cannot be obtained for the new-coming user. It is because the new-coming user corresponds to a customer in real situation and we cannot ask many questions to him/her (otherwise, he/she will go away). Although many questions cannot be asked, we can obtain a few observations of input/output response of the user. Consequently, we must estimate the type of new-coming user from a small number of observations.

To cope with the situation of a small number of observations, we use a state space model and its Bayesian estimation to identify the type of a new-coming user. The identification of the user type is carried out by estimating the mixing weight vector, $\mathbf{w}$, which is now supposed to be time varying, and so it is rewritten as $\mathbf{w}_k$ with time index k (count of the measurement). The estimation task is to obtain conditional distribution of $\mathbf{w}_k$ given the measurement series and input/external-input series up to current time. It updates the estimate each time a new measurement comes.

B.1. State Space Model for User Type Identification

We consider that the mixing weight vector smoothly changes with respect to the time index k such that the following expression (system equation) holds

$$\mathbf{w}_k = \mathbf{w}_{k-1} + \mathbf{v}_k \tag{6}$$

where $\mathbf{v}_k$ is random vector of dimension p distributed according to a certain probability distribution, for example, uni-modal distribution with mode value equal to zero $\mathbf{0}$. $\mathbf{v}_k$

is called "system noise". The initial value of mixing vector, $\mathbf{w}_0$, is a random vector of uni-modal distribution with center be the vector estimated by mixture model, $\hat{\mathbf{w}}$. Note that this initial distribution and smoothness of eq.(6) define a prior distribution of $\mathbf{w}_k$ for $k = 1, 2, \cdots$ in a context of Bayes estimation.

Straightforward application of eq.(6) results in $\mathbf{w}_k$ not a satisfying the constraint eq.(5). To cope with this, we use logarithm of weight and normalization as follows. Let elements of $\mathbf{w}_k$ be denoted by $w_j(k)$, $j = 1, 2, \cdots, p$. Notation of element for system noise vector $\mathbf{v}_k$ is in similar way. By taking logarithm and exponential on eq.(6) for each element, we have

$$\tilde{w}_j(k) = \exp\left\{\log w_j(k-1) + \tilde{v}_j(k)\right\}, \quad (7)$$

and $w_j(k)$ is calculated by normalizing $\tilde{w}_j(k)$ as

$$w_j(k) = \tilde{w}_j(k) / \sum_{i=1}^{p} \tilde{w}_i(k). \quad (8)$$

Let $\mathbf{y}_k$ (measurement vector at time k) be according to the mixture model

$$\mathbf{y}_k \sim f(\cdot|\mathbf{u}_k, \mathbf{z}_k; \mathbf{w}_k, \hat{\Theta}) \quad (9)$$

which will be called an "observation equation".

A pair of system equation, originally eq.(6) but now eqs.(7) and (8), along with the observation equation eq.(9) is called "state space model". In this study, the mixing weights constitute the "state" of the model.

B.2. State Estimation

"State estimation" is a task to obtain the estimate of the state (here, weight vector) from the given observation series up to current time k

$$\mathbf{y}_{1:k} = \{\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_k\}. \quad (10)$$

Being more precise, the estimate means the calculation of the conditional distribution of $\mathbf{w}_k$ given observation series eq.(10), denoted by $p(\mathbf{w}_k|\mathbf{y}_{1:k})$. This distribution is called a "filter". The conditional distribution at time $k+1$, $p(\mathbf{w}_{k+1}|\mathbf{y}_{1:k})$, is called "one-step-ahead prediction" and at time before k such that $k - L$ $(L > 0)$, $p(\mathbf{w}_{k-L}|\mathbf{y}_{1:k})$, is called "smoothing". According to the aim of adapting the user interface depending on the estimate of mixture weight, the filtering estimate is suitable for our purpose.

In the state space model defined here, both the system equation, eq.(7), and the observation equation, (9), are nonlinear with respect to the state vector, $\mathbf{w}_k$. System noise distribution of system equation (7) can be non-Gaussian. Thus, nonlinear non-Gaussian method is required for state estimation. Here, we use Monte Carlo Filter (MCF) [10] for state estimation, which is a special instance of "particle filter" [7]. Several methods have been proposed [8]. The key idea of particle filter is to approximate the non-Gaussian distribution of the state by many particles in state space. It can be seen as a special case of sequential Monte Carlo Method [11].

From the estimated distribution $p(\mathbf{w}_k|\mathbf{y}_{1:k})$, we can use some characteristic values of the distribution as the estimation result of the weight vector $\mathbf{w}_k$. A practical example of this is the minimum mean square error estimate

$$\hat{\mathbf{w}}_k = \int \mathbf{w}_k p(\mathbf{w}_k|\mathbf{y}_{1:k}) d\mathbf{w}_k. \quad (11)$$

C. *Adaptation to User*

An outline of adaptation scheme by using the results of user type identification is shown here. We assume that strategies for each type are collected by appropriately designed inquiry prior to the adaptation. It is done for some typical users for each user type to present various input and external input patterns.

Collected strategies are written in either explicit or implicit form depending on the design of the inquiry. When an explicit form is used, we have an optimal strategy $\mathbf{u}_{[j]}$ of type j for all types, $j = 1, \cdots, p$. Then, together with the mixing weight vector estimated by state space model to these optimal strategies, we have a strategy for adaptation to the new-coming user, $\mathbf{u}_k$, as a weighted sum of optimal strategy such as

$$\mathbf{u}_k = \sum_{j=1}^{p} \hat{w}_j(k)\mathbf{u}_{[j]}. \quad (12)$$

On the other hand, when an implicit form is used, strategies are represented by performance indices such as $I_j(\mathbf{u}|\mathbf{z})$, $j = 1, 2, \cdots, p$. Here $\mathbf{u}^*_{[j]}$ which takes a maximum value of $I_j(\mathbf{u}|\mathbf{z})$ is the optimal input for j-th type. Together with mixing weight vector estimated by state space model, we have a strategy for adaptation to the new-coming user, in a performance index from of

$$I(\mathbf{u}|\mathbf{z}; \hat{\mathbf{w}}_k) = \sum_{j=1}^{p} \hat{w}_j(k) I_j(\mathbf{u}|\mathbf{z}). \quad (13)$$

Then, we will have the optimal input $\hat{\mathbf{u}}_k$ that achieves the maximum of eq.(13).

III. SIMULATION

To illustrate how the proposed approach works, we have conducted numerical simulations as follows. Let the dimension of input space $\mathbf{U}$ be $m = 3$, and the dimension

Table 1. Simulation parameters of Gaussian mixture model.

j	w_j	$\mathbf{c}_j$	Σ_j
1	0.3	$\begin{bmatrix} 3 & 3 & 3 \\ 3 & 2 & 1 \end{bmatrix}$	$\begin{bmatrix} 1.0 & 0.1 \\ 0.1 & 1.0 \end{bmatrix}$
2	0.4	$\begin{bmatrix} -3 & -3 & -3 \\ 3 & 4 & 5 \end{bmatrix}$	$\begin{bmatrix} 0.8 & 0.3 \\ 0.3 & 0.8 \end{bmatrix}$
3	0.2	$\begin{bmatrix} -3 & -3 & -3 \\ -3 & -2 & -1 \end{bmatrix}$	$\begin{bmatrix} 0.8 & -0.5 \\ -0.5 & 0.8 \end{bmatrix}$
4	0.1	$\begin{bmatrix} 3 & 3 & 3 \\ -3 & -4 & -5 \end{bmatrix}$	$\begin{bmatrix} 0.3 & -0.2 \\ -0.2 & 0.3 \end{bmatrix}$

of output space $\mathbf{Y}$ be $n = 2$. We assume that there is no external input for convenience. Three input patterns, such that $\mathbf{u}_1 = [1, 0, 0]$, $\mathbf{u}_2 = [0, 1, 0]$, and $\mathbf{u}_3 = [0, 0, 1]$ are used as input $\mathbf{u}$ here. Four types are assumed here, and they are denoted by Gaussian component $N(\mathbf{C}_j \mathbf{u}^T, \Sigma_j)$ for $j = 1, 2, 3, 4$. Values of components' parameters, $\mathbf{C}_j$ and Σ_j for $j = 1, 2, 3, 4$, and weights of mixture model appeared in eq.(2) are shown in Table 1.

Corresponding to a preliminary experiment to extract the user types, we have firstly generated a data set from the Gaussian mixture. Generating the random input among $\mathbf{u}_1$, $\mathbf{u}_2$, and $\mathbf{u}_3$ with equal probabilities and fed it into the Gaussian mixture model, then we obtain the output data plotted in Figure 1, here the number of samples is 1,000. Secondly, for this data set, we have applied EM algorithm to obtain parameter estimation of the mixture model. We have assumed the number of components is known in this experiment.

Using the estimated components by EM algorithm, we have proceeded an experiment of the user type estimation as follows. Firstly, we have generated data that correspond to the new-user's response by giving random input into Gaussian components, where time index k varies from 1 to 200 and 1st component was used for $k = 1, \cdots, 50$, 2nd component was used for $k = 51, \cdots, 100$, and 3rd and 4th components are used for $k = 101, \cdots, 150$ and $k = 151, \cdots, 200$, respectively. The data, $\mathbf{y}_k$ for $k = 1, \cdots, 200$, are plotted in Figure 2.

The proposed method for user type estimation have been applied to the data in Figure 2. Initial condition of mixing weights are independently according to Gaussian distribution with mean as estimated by EM algorithm and variance 0.1 (given value). User type estimation results are shown in Figure 3. For each period, in which certain component was used for the data generation, weight value of correct component is raising gradually, and other weights are going down as the passage of time k.

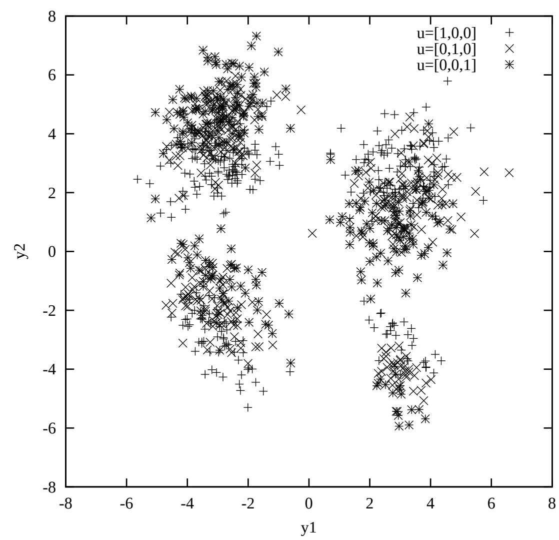

Fig. 1. Simulated data for mixture estimation.

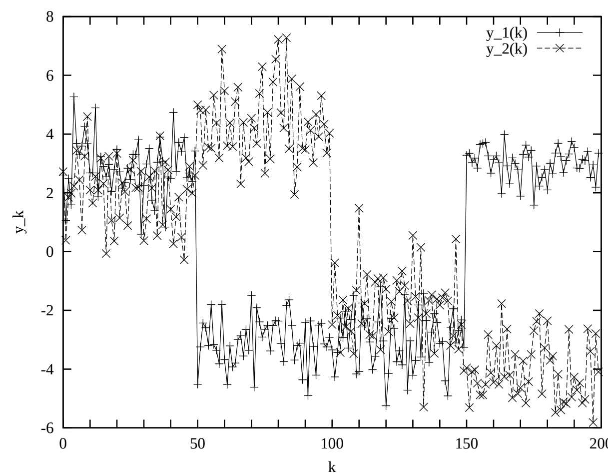

Fig. 2. User response data for type estimation.

IV. EXPERIMENT

As a real world example of the proposed method, user type estimation based on key typing performance and input methods has been examined. Performance data have been collected from 13 test subjects who are every day PC user. The task of test subject is to enter a paragraph of English text which is randomly selected from newspaper. Four kinds of input method are assumed here, which are (1)using both hands (possible to blind touch), (2)using right (dominant) hand only, (3)using left (non-dominant) hand only, and (4)using one finger (of non-dominant hand) only.

From raw data, which is a series of key typing time

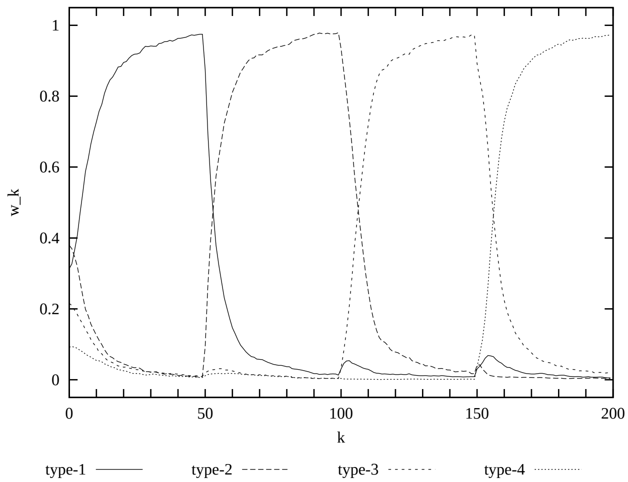

Fig. 3. Estimated weight for simulation.

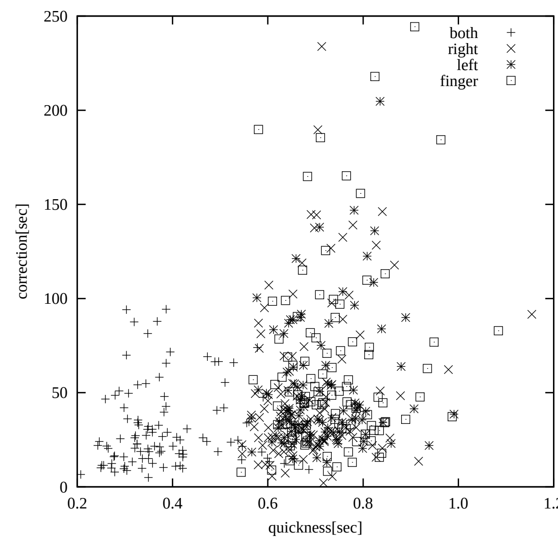

Fig. 4. Key-typing data (Quickness-Correction).

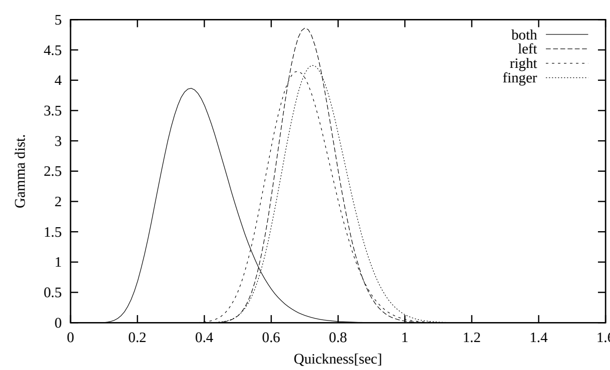

Fig. 5. Estimated distribution for Quickness.

and ASCII code of the key, three kinds of characteristic value are calculated. They are (1)Quickness Q by average time of key interval in a task, (2)Accuracy A by error rate between entered text and original text based on longest common sequence(LCS), and (3)Correction degree C by average interval time of use of the Backspace key in a task.

Eight cases are examined for each input method, for each test subject, under experimental design. Thus the number of data is $13[\text{person}] \times 8[\text{case}] = 104$ for each input method. Scatter plot of the data between Quickness and Correction degree is shown in Figure 4.

Estimation of component distributions have been proceeded as follows. Here we have assumed that each characteristics (i.e.,Q,A,C) are mutually independent for convenience, based on the fact that correlations are low. Then, we have applied a probabilistic model for each characteristic values, and we have chosen a Gamma distribution based on the likelihood value for all characteristics. Estimated probability densities are shown in Figure 5 for Quickness. Looking at the figure, method using both hands is well separated from other methods. So we consider that there are two types, i.e., both hands and single hand. We have used the component of input method finger as the type of single hand, as representative.

User type estimation has been proceeded as follows. As the performance data of new-coming user, we randomly select cases for each input method from the original data, which are used in the estimation above, and make a sequence of the cases. The length of the sequence is 160, and the sequence consists of four period of equal length (i.e., 40) corresponds to each input method with order of finger, both-hands, right-hand, and left-hand. Estimation results are shown in Figure 6. Looking at the figure, we can see that the correct type has been estimated with high value of weight for each period of length 40, i.e., single hand type for $k = 1, \cdots, 40$ sampled from finger method data, both hand type for $k = 41, \cdots, 80$ sampled from both-hands method data, single hand type for $k = 81, \cdots, 120$ sampled from right-hand method data, and single hand type for $k = 121, \cdots, 160$ sampled from left-hand method data.

V. CONCLUSION

We have proposed an approach to adaptive user interface in which we assume the existence of user types represented by mixture model and estimate the type of new-coming user using state space model. The state space model consists of a system equation with time-varying mixing weight and mixture model as observation equation with the time-varying mixing weights.

There are related researches to our approach as follows. Justification by using Dirichlet distribution and ap-

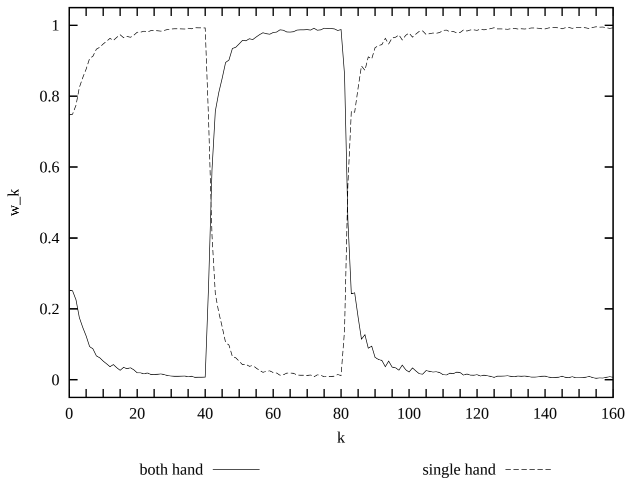

Fig. 6. Estimated weight for key-typing data.

proximate solution have been proposed in [18], use of Markov Chain Monte Carlo for more general situation including the problem setting of [18] has been proposed in [6], and the use of particle filter for further more general situation has been proposed in [4]. In future works, investigation for the relation to these researches and use of these ideas are interesting.

ACKNOWLEDGEMENT

This study was carried out under the ISM Cooperative Research Program (2002-ISM·CPR-1005). Authors also thank to Dr.A.Doucet, Signal Processing lab. in CUED for useful comments.

REFERENCES

[1] K.Baba, T.Hyakudome, N.Ikoma, and H.Maeda: On the evaluation of the User-Adaptive Human Interface for Intelligent Wheelchair by using Virtual Reality, *Proc. of 3rd Annual Conference of SOFT Kyushu Chapter*, pp.57-60 (2001),in Japanese.

[2] K.Baba, T.Hyakudome, Y.Matsumoto, N.Ikoma, and H.Maeda: Estimation of user type by state space model for user adaptive interface, *Proc. of 18th Fuzzy System Symposium*, (2002),in Japanese.

[3] S.K.Chang, M.F.Costabile and S.Levialdi: Modeling users in an adaptive visual interface for database systems, *Journal of Visual Languages and Computing*, 4, pp.143-159 (1993).

[4] R.Chen and J.S.Liu: Predictive Updating Methods with Application to Bayesian Classification *Journal of the Royal Statistical Society*, Ser.B, Vol.58, No.2., pp.97-415 (1996).

[5] A.Dempster, N.Laird, and D.Rubin: Maximum likelihood estimation from incomplete data via EM Algorithm, *Journal of Royal Statistical Society*, B, vol.39, pp.1-38 (1977).

[6] J.Diebolt and C.P.Robert: Estimation of Finite Mixture Distributions through Bayesian Sampling *Journal of the Royal Statistical Society*, Ser.B, Vol.56, No.2., pp.363-375(1994).

[7] A.Doucet, J.F.G.de Freitas, and N.J.Gordon (eds): *Sequential Monte Carlo Methods in Practice*, New York, Springer (2001).

[8] N.J.Gordon, D.J.Salmond, and A.F.M.Smith: Novel approach to nonlinear/non-Gaussian Bayesian State Estimation, *IEE Proceedings-F*, **140**, No.2, 107-113, (1993).

[9] E.Kantorowitz and O.Sudarsky: The adaptive user interface, Communications of the ACM, vol.32, No.11, pp.1352-1358 (1989).

[10] G.Kitagawa: Monte Carlo filter and smoother for non-Gaussian nonlinear state space models, *Journal of Computational and Graphical Statistics* **5**,No.1,1-25, (1996).

[11] J.S.Liu and R.Chen: Sequential Monte Carlo methods for dynamic systems, *Journal of the American Statistical Association*, **93**, 1032-1044, (1998).

[12] H.Maeda, K.Haro, and N.Ikoma: An intelligent database interface adapting itself to the degree of user's skill, *Proc. of the 8th International conference on Fuzzy Systems(FUZZ-IEEE'99)*, Vol.2, pp.1153-1158 (1999).

[13] J.Morita, K.Baba, N.Ikoma, and H.Maeda: A Development of Human Interface for Intelligent Wheelchair Adapting to User's Characteristic, *Proc. of 17th Fuzzy System Symposium*, pp.253-256 (2001),in Japanese.

[14] J.Morita, K.Baba, N.Ikoma, and H.Maeda: On a Method of Level Judgement in User Adaptive Interface, *Proc. of 13th Annual Meeting of Biomedical Fuzzy System Assoc.*, pp.136-139 (2000),in Japanese.

[15] R.Parasuraman, T.B.Sheridan and C.D.Wickens: A model for types and levels of human interaction with automation, *IEEE Trans. On Systems, Man, and Cybernetics*, Part.A, Vol.30, No.3, pp.286-297 (2000).

[16] F.Patern and C.Mancini: Effective levels of adaptation to different types of users in interactive museum systems, *Journal of the American Society for Information Science*, 51(1), pp.5-13 (2000).

[17] F.D.Rosis, S.Pizzutilo, and B.D.Carolis: Formal description and evaluation of user-adapted interfaces, *Int. J. Human-Computer Studies*, 49, pp.95-120 (1998).

[18] A.F.M. Smith and U.E.Makov: Quasi-Bayes Sequential Procedure for Mixtures, *Journal of the Royal Statistical Society*, Ser.B, Vol.40, No.1., pp.106-112 (1978).

[19] G.Viano et.al.: Adaptive user interface for process control based on multi-agent approach, *Proc. of the Working Conference on Advanced Visual Interface*, pp.201-204 (2000).