

## Prediction-based Threshold for Medication Alert

Yoshimasa Kawazoe<sup>a</sup>, Kengo Miyo<sup>a</sup>, Issei Kurahashi<sup>b</sup>, Ryota Sakurai<sup>a</sup>, Kazuhiko Ohe<sup>a,b</sup>

<sup>a</sup> Department of Planning, Information and Management, The University of Tokyo Hospital, Tokyo, Japan

<sup>b</sup> Department of Medical Informatics and Economics, School of Medicine, The University of Tokyo, Tokyo, Japan

### Abstract

*This study presents a prediction-based approach to determine thresholds for a medication alert in a computerized physician order entry. Traditional static thresholds can sometimes lead to physician's alert fatigue or overlook potentially excessive medication even if the doses are below the configured threshold. To address this problem, we applied a random forest algorithm to develop a prediction model for medication doses, and applied a boxplot to determine the thresholds based on the prediction results. An evaluation of the eight drugs most frequently causing alerts in our hospital showed that the performances of the prediction were high, except for two drugs. It was also found that using the thresholds based on the predictions would reduce the alerts to a half of those when using the static thresholds. Notably, some cases were detected only by the prediction thresholds. The significance of the thresholds should be discussed in terms of the trade-offs between gains and losses; however, our approach, which relies on physicians' collective experiences, has practical advantages.*

### Keywords:

Medication alert, prescription alert, machine learning, Random Forest, decision support system, CPOE.

### Introduction

#### Background

Computerized physician order entry (CPOE) systems for medication have been widely implemented in recent years. The percentage of large hospitals with 400 or more beds that use CPOE was 78.6% as of 2011 in Japan [1]. The use of CPOE is expected to be efficient for the health care practice and reduce a physician's time [2]. The CPOE system is also expected to contribute to the prevention of malpractice [3, 4], such as the prescription of excessive medication. In many of these systems, the thresholds for the upper-limits of medication dose are static based on drug notes, and are applied without consideration of the patient's condition. These static thresholds are generally useful; however, medication doses occasionally exceed the thresholds as a result of considering a patient's condition, which creates false alerts and leads to physician's alert fatigue [5, 6]. Moreover, when taking a patient's poor condition into account, the static threshold may overlook potentially excessive medication doses even if the doses are under the static thresholds. To address this problem, we have developed a case-based approach that uses the past medication record for setting the thresholds. A previous study revealed that the system reduced the number of false alerts compared to the static thresholds method; however, it did not function well with an

insufficient amount of case data, which remained a limitation [7].

### Purpose

This study presents a prediction-based approach to determine dose thresholds using past medication records. First, we apply Random Forest (RF), a type of machine-learning algorithm, to predict the medication dose. To show the characteristics of the RF prediction, we compare the performance with other algorithms, such as Bagging (BAG) and Classification and Regression Tree (CART). Second, we present a method to determine the dose upper-thresholds based on the results of the predictions, and compare them with the results using the static threshold. Finally, we discuss the benefit of the RF predictions and the significance of the prediction-based dose thresholds.

### Materials and Methods

#### Health Insurance Claims in Japan

To develop a prediction model for medication dose, we used health insurance claims in Japan as the data source. The health insurance claims are monthly statements of medical expenses and contain comprehensive information on medical procedures in inpatients and outpatients. In almost all of the hospitals and pharmacies in Japan, the claims for payment are electronically available. The major information contained in the claims are date of visits, demographics such as age and gender, medical procedures, current diseases, and total medical expenses for each month. In the medical procedures section, the details of the medications, such as what types of drugs were used, how many tablets were used daily, how many days long, regular use or only as needed, oral or injection, are described. In the current diseases section, the basis for medical procedures, e.g., patient's diseases or diagnoses are described for the purpose of insurance claims. The coding system used in the disease section, developed by Medical Information System Development Center (MEDIS-DC) Japan, is called the MEDIS standard disease master [8], and it contains approximately 23,000 disease codes.

#### Data Collection

The University of Tokyo Hospital (UTH) is an educational hospital with more than 1,200 beds and 750,000 visits annually. We collected the medication records of the inpatients and outpatients including ambulatory settings at UTH from January 2007 to December 2011; and multiple prescribed medication records for the same patients during different visits or on different days were included. We defined medication dose as the daily amount of tablets used regularly, excluding tablets used only as needed.

Table 1 - Statistical summary for the collected medications

Drug	Tablet Dose (mg)	Age			Gender M:F	Dose (Tablets)			Total Records	Alert (%)
		Mean	SD	Range		Mean	SD	Range		
Ursodeoxycholic acid	100	64.0	13.4	4-97	1:0.9	5.5	1.76	1-15	73,918	11.5
Carvedilol	2.5	65.1	16.2	0-95	1:0.4	1.9	1.21	0.02-9	48,379	9.6
Sennoside	12	62.1	16.8	6-100	1:1.1	1.9	0.89	0.5-10	126,483	5.2
Loxoprofen	60	55.0	18.1	3-99	1:1.4	2.0	0.95	0.5-6	170,775	4.4
Brotizolam	0.25	60.1	16.4	3-96	1:1	1.1	0.33	0.25-3	148,812	4.3
Nifedipine	20	68.0	12.0	9-97	1:1.1	1.7	0.84	0.5-6	55,245	2.5
Famotidine	20	61.8	16.2	4-97	1:0.9	1.4	0.49	0.5-3	130,650	2.3
Pravastatin	10	68.1	11.8	12-99	1:1.3	1.1	0.30	0.5-4	31,063	1.9

The records also include patient demographics such as gender, age, and disease names presented at the time of the medication. All of these data were not extracted from actual claims, but from the hospital information system database; however, the data include the same information as described in the claims. We selected the records for the eight drugs most frequently causing alert in January 2011 in our hospital. A statistical summary of the collected medications is shown in Table 1.

## Random Forest

### General Procedure

Random Forest is a machine-learning algorithm for both classification and regression problems and developed by Breiman and Cutler [9]. It has become popular and appears to be very powerful in many different applications such as gene selection in microarray data, face recognition, etc. [10]. Random forest belongs to a family of ensemble learning algorithms that is a committee of weak learners. In RF, CART, a family of the decision tree, is used as a weak learner. For the classification problems, RF outputs the majority of class votes across all trees. On the other hand, for regression problems, RF outputs the mean of the votes. We present a brief description of the regression RF procedure, which takes the following steps:

**Step 1:** Making a bootstrap sample B1, B2... BN consisting of N subjects from the original training data. In each sample, not all features are used, but a small number of features are randomly selected. Two-thirds of the data are used for training a tree and one-third are used for testing the tree, which are called Out-Of-Bag (OOB) data.

**Step 2:** Growing CARTs for the training data in the samples B1, B2... BN. In a CART, the split variable in a node of the tree is decided to increase the difference of Gini-Indexes between a parent and its two child nodes. Once the variable is decided and the node is split into two, the same process is applied to each child node. In RF, no pruning is used and the tree is grown to its fullest possible extent, which is distinct from the standard CART. The OOB data are not used for training the tree, but are used to estimate the prediction error and then to evaluate the variable importance.

**Step 3:** Random forest outputs the mean of the votes across all trees based on the OOB estimation.

There are two parameters that should be given in RF. The first parameter is the number of features that are randomly selected from all features. Although it could be determined by heuristic reasoning, it is recommended that one-third of all features be used in the regression RF. The second parameter is the number of trees included in RF. In theory, the stability of the results increases with the number of trees grown. In practice, we can use

the number of trees when the decrease in error rate becomes saturated, because the OOB-based error estimation gradually decreases as the number of trees increases. According to our preliminary experiments, we used 200 trees in a forest in this study.

### Variable Importance in RF

There are two measurements of variable importance available in RF. The first measurement is the permutation importance that is computed by comparing the prediction error of the original data with that of the OOB data, in which the variables are randomly permuted. The other measurement is the Gini importance that is computed by adding up the reduction of the Gini-Index for each node over all trees in the forest. In this study, we used the permutation importance to show which feature is useful for predicting the medication dose and discussed whether these features are consistent with clinical interpretation.

## CART and BAG

Generally, RF performs better than both CART and BAG. We compared RF with CART and BAG to confirm the generality of its algorithm. The classification and regression tree [11] is a well-known decision tree model, which is used as a weak learner in RF. Unlike RF, pruning is usually used in CART to avoid over fitting to training data. However, in this study, we used CART without pruning in the same way as RF to clarify the contribution of the ensemble in RF. Bagging is another well-known tree-based algorithm [12], which also belongs to a family of ensemble learning that create CARTs. The difference between RF and BAG is that RF uses randomly selected sample features to grow the trees, whereas BAG uses all features. The number of trees used in BAG should also be given. Here, we used 200 trees for developing BAG, which was the same setting in RF.

## Experiment Settings

### Data Arrangement

To improve the efficiency of the analysis, we arranged the collected data by reducing the features. There were over 22 million records for the current diseases, and approximately 13,000 types of diseases were observed. The frequency distribution of the diseases was highly skewed with a long tail on the right. The most frequently observed disease was hypertension, followed by diabetes. The top 1,500 types of diseases accounted for 90% of the frequency. Based on this finding, we used the top 1,500 types of diseases as features for developing the prediction model. Thus, our dataset included 1,503 features for a medication record, i.e., two for gender and age, 1,500 for current diseases, and one for the dose of medication. Among these features, the dose and the age are presented as real values, and the gender and the current diseases

are presented as binary values. In addition, we reordered the features in order of frequency for every drug, which was measured by how many times a certain disease was presented at the time a certain drug was prescribed. According to this reordering, a disease that is frequently presented for a drug is highly ranked, while a disease that is not frequently presented for a drug is ranked lower. For example, the most highly ranked disease was hepatocellular carcinoma followed by liver cirrhosis for Ursodeoxycholic acid; whereas, for Carvedilol, chronic heart failure and dilated cardiomyopathy were the two most highly ranked diseases.

**Build and Evaluate Prediction Models**

To build the prediction model for each algorithm, experiments were conducted as follows. For each drug, 30,000 medications were randomly selected and then further divided into two datasets: a training dataset containing 20,000 medications and a test dataset containing 10,000 medications. Performance of the prediction was measured as the mean of the correlation coefficient (n = 10 for RF and CART, n = 5 for BAG) between the predicted doses and the actual doses using the test dataset. To determine how many features are appropriate to use, the performance was calculated by gradually increasing the number of the features from 10 to 1,500. The features were already ranked in order of frequency as described above. For this analysis, we used statistical software R version 2.15.1 on a Linux OS.

**Determine Threshold: Drawing Boxplot for RF Votes**

To determine the thresholds based on the prediction model, we applied a boxplot for RF votes. The boxplot is a common and robust way to define a threshold for outlying data. The boxplot's data item has five main values: Low, Q1, Median, Q3, and High. The difference between Q1 and Q3 is called the interquartile range or IQR. The High value is determined by a minimum of 1.5 times the IQR above Q3. The Low value is determined similarly. Thus, when RF votes are given for a medication, we can calculate the high value of the boxplot and use it as an upper-threshold for the medication dose. For evaluation, we compared the static thresholds described in drug notes to the predicted thresholds. We used 20,000 medications for building RF and 10,000 medications for the evaluation, which are the same datasets used for building the prediction model.

**Results**

**Performances of the Prediction Models**

Figure 1 shows the performance of the prediction among the three algorithms. The horizontal axis shows the number of features used, and the vertical axis shows the mean of the correlation coefficient. The results showed that regardless of the type of drug, RF performed better than the other algorithms. In all algorithms, the performances reached their highest point when the number of features used was around 200. Specifically, RF showed 0.80 to 0.95 in its highest point, except for two drugs. The relatively low performances of 0.52 and 0.69 were observed for Loxoprofen and Brotizolam respectively, and these trends were also observed in the other algorithms. In addition, RF maintained its performance even after reaching the highest point; whereas, in BAG and CART, the performance decreased as the number of features increased.

**Evaluation of the Prediction Thresholds**

Figure 2 illustrates the result for 'Carvedilol 2.5 mg Tab' by showing the relationships between the actual doses and the pre-

dicted thresholds and the static threshold. Among 10,000 medications, 602 cases were detected in total. . Among them, 57 cases were detected by both thresholds, 228 cases were detected only by the predicted thresholds, and 317 cases were detected only by the static threshold.

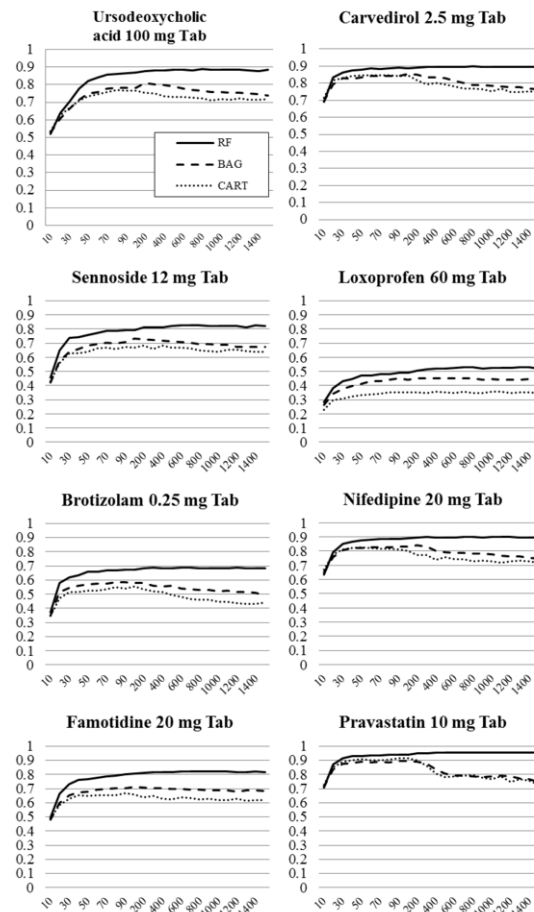


Figure 1- The performances of the prediction among the three algorithms

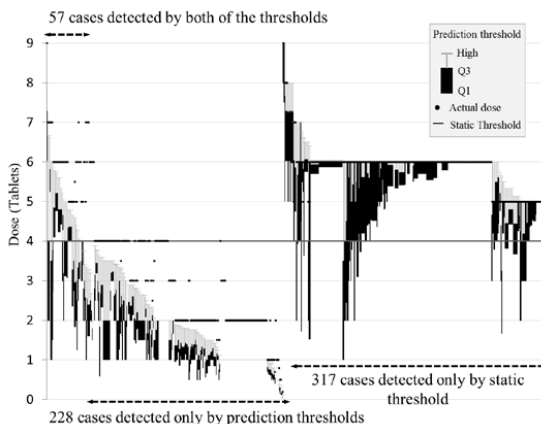


Figure 2 - The relationships between the actual doses and two types of thresholds for Carvedilol Tablet

This result means that when we apply the predicted thresholds for medication alert, 317 medications are considered to be appropriate even if the doses are above the static threshold; however, 228 medications are considered to be inappropriate even if the doses are below the static threshold. The summary of all results is shown in Table 2. The results indicate that when we apply the predicted thresholds for medication alert, there will be a reduction in four drugs, whereas the opposite will happen for the other four drugs. In total, the predicted thresholds will reduce the alerts by a half of those using the static threshold.

Table 2 - The evaluation results of thresholds

	Number of the Cases Detected by Thresholds				Proportions (Prediction/Static)
	All Cases	by Both Thresholds	Only by Prediction	Only by Static	
Ursodeoxycholic acid	949	113	17	819	0.02
Carvedilol	602	57	228	317	0.72
Senoside	327	29	195	103	1.89
Loxoprofen	505	15	465	25	18.60
Brotizolam	1,167	300	5	862	0.01
Nifedipine	627	37	171	419	0.41
Famotidine	183	8	171	4	42.75
Pravastatin	53	0	41	12	3.42
TOTAL	4,413	559	1,293	2,561	0.50

Discussions

Dose Prediction from Health Insurance Claims

Drug safety has been a notable issue for clinical practice. Recently, machine-learning techniques have been applied for predicting dose of warfarin [13, 14]. In those studies, detailed clinical data, such as biochemical examination and genetic variation were used as features to predict the narrow therapeutic dose of the drug. Unlike those studies, we used patient demographics and current diseases, which are commonly available from health insurance claims, to develop the prediction model that enables us to apply a wide range of drugs to an alert system. Although current diseases in the health insurance claims are described not only for accurate diagnoses but also for insurance claims, the performance of RF using the claims showed high value. This indicated that it is useful to predict drug dose based on diseases described in the insurance claims even if they are not intended for only accurate diagnoses.

Random Forest Predictions

The results of the prediction model revealed that regardless of the types of algorithm or drugs, the performances reached their highest points when the number of features used was around 200. These results indicate a sufficient number of the features used to build the prediction model. In contrast to RF, the performance of BAG and CART decreased as the number of features increased. This is due to over fitting to training data, which generally occurs when a model has too many features relative to the number of samples. Although RF is a collection of CART, ensemble-based learning helps to avoid the over fitting and re-

duces the variation of prediction errors. Bagging is also an ensemble of CART; however, its performance was also decreased, similar to CART. The reason is assumed to be the random sampling of the features, which provides variety for the trees and improves the generalization of the prediction [9].

Understanding Clinical Reasons

The quantification of the variable importance is useful for interpreting data and understanding the underlying phenomenon. The variable importance measures a feature’s contribution to the reduction of the prediction errors in the forest. Figure 3 shows the top 10 features in a certain experiment with 200 features. There are several interesting features relevant to the clinical understandings. First, for patient demographics, age was the most important factor among all other features in almost all drugs, and gender was also important but not as much as age. This indicates that age and gender are primal and convenient markers to estimate the tolerance of drug dose.

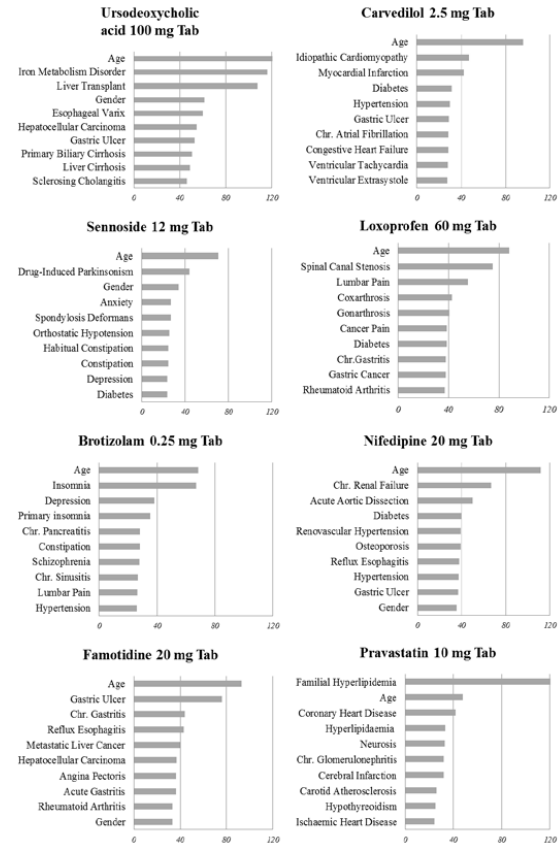


Figure 3 - Variable importance: top 10 among 200 features in a certain experiment

Second, there were some clinical reasons that explain how the features could affect the dose. For example, ‘familial hyperlipidemia’ in ‘pravastatin 10 mg Tab’ was the most important feature compared to the others. Clinically speaking, familial hyperlipidemia is a genetic disorder that causes greatly elevated levels of cholesterol, and thus the patients with familial hyperlipidemia need a higher dose of an anti-cholesteremic agent than patients with other types of hyperlipidemia. Thus, it is reason-

ble to suppose that the familial hyperlipidemia is a good feature for predicting the required dose of pravastatin.

### Significance of the Prediction Thresholds

Merely discussing the increase or decrease in the number of alerts is not enough. We should also discuss the trade-offs between gains and losses; in other words, the sensitivity and specificity of alert system [6]. However, knowing the correct answers for medication alerts is difficult because we could not identify the outcomes responsible for those medications. In light of this, our approach has some practical advantages. The approach relies on data from past medications and physicians' collective experiences, which are assured as appropriate for a patient's condition. As a result, the developed prediction model was capable of predicting the patient-specific medication dose. As shown above, the performance of prediction in RF is high in most of the drugs. Therefore, it is reasonable to suppose that the thresholds based on those predictions have a certain level of evidence.

### Availability of the Prediction-based Approach

One of the advantages of our approach is that it can detect inappropriate medications even if the doses are below the static threshold. Our results show that the medications can be detected only by the prediction thresholds in 1,293 cases. Although it may lead to an increase in false alerts when we apply them, it is more significant to prevent iatrogenic injury caused by excessive medication. Furthermore, our approach is also capable of suppressing the alerts even if the doses are above the static threshold. These advantages are helpful to improve the system by leaving room for interpretation either way. Moreover, in term of visualization, the display of boxplot-based threshold and the variable importance in RF will help physicians to understand the reason for the alert.

In terms of adaptivity, our approach can be easily adopted in other hospitals because the health insurance claims that we used as the source for the prediction model are commonly available. Although there are differences in patient demographics, disease frequencies, and drug doses among the hospitals, our approach can take these differences into account to determine thresholds.

### Limitations and Future Works

Although we used 30,000 cases for building and testing the prediction model, we did not determine how many cases are required to build the prediction model. Thus, our approach may not apply to minor drugs that do not have a sufficient number of past records. When dealing with minor drugs, health insurance claims should be gathered from other hospitals. In addition, it is necessary to evaluate physicians' compliance with those alerts in the clinical setting while clarifying the outcomes responsible for those alerts.

### Conclusions

In this study, we presented a prediction-based approach to determine thresholds for the medication alert. The evaluation of eight selected drugs showed that the performance of the prediction using RF was high except for two drugs. The thresholds based on the prediction model reduced the alerts by a half of those when using the static thresholds. Although the significance of the thresholds should be assessed in more details, our approach, which relies on physicians' collective experiences, had some practical advantages. In future work, we

will evaluate physicians' compliance with those alerts in a clinical setting.

### Acknowledgment

The authors would like to thank their colleagues at Department of Planning, Information and Management, the University of Tokyo Hospital, Japan.

### References

- [1] Ministry of Health, Labour and Welfare. Survey of Medical Institutions. Available from: <http://www.e-stat.go.jp/SG1/estat/eStatTopPortal.do>. Accessed: Mar 22, 2013. (in Japanese)
- [2] Kuperman GJ, and Gibson RF. Computer physician order entry: benefits, costs, and issues. *Ann Intern Med.* 2003;139(1):31-9.
- [3] Kaushal R, Shojania KG, Bates DW. Effects of computerized physician order entry and clinical decision support systems on medication safety. A systematic review. *Arch Intern Med.* 2003;163:1409-16.
- [4] Judge J, Field TS, DeFlorio M, Laprino J, Auger J, Rochon P, Bates DW, and Gurwitz JH. Prescribers' responses to alerts during medication ordering in the long term care setting. *J Am Med Inform Assoc.* 2006;13(4):385-90.
- [5] Glassman PA, Simon B, Belperio P, Lanto A. Improving recognition of drug interactions. Benefits and barriers to using automated drug alerts. *Med Care.* 2002;40:1161-71.
- [6] van der Sijs H, Aarts J, Vulto A, Berg M. Overriding of Drug Safety Alerts in Computerized Physician Order Entry. *J Am Med Inform Assoc.* 2006;13:138-147.
- [7] K.Miyo, Y.S.Nittami, Y.Kitagawa, and K.Ohe. Development of Case-based Medication Alerting and Recommender System: A New Approach to Prevention for Medication Error. *Stud Health Technol Inform.* 2007;129(Pt 2):871-4.
- [8] MEDIS-DC. Available from: <http://www.medis.or.jp>. Accessed: Mar 22, 2013. (in Japanese)
- [9] Breiman L. Random forests. *Mach Learn.* 2001;45(1): 5-32.
- [10]Uriarte RD, Andres SA. Gene selection and classification of microarray data using random forest. *BMC Bioinformatics* 2006;7:3.
- [11]Breiman L, Friedman J.H, Olshen R.A, Stone C.J. *Classification & Regression Trees.* Chapman & Hall 1984.
- [12]Breiman L. Bagging predictors. *Mach Learn.* 1996;26(2):123-140.
- [13]Hu YH, Wu F, Lo CL, Tai CT. Predicting warfarin dosage from clinical data: A supervised learning approach. *Artif Intell Med.* 2012;56(1):27-34.
- [14]Cosgun E, Limdi NA, Duarte CW. High-dimensional pharmacogenetic prediction of a continuous trait using machine learning techniques with application to warfarin dose prediction in African Americans. *Bioinformatics* 2011;27(10):1384-9.

### Address for correspondence

Yoshimasa Kawazoe, M.D., Ph.D.  
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655, Japan  
E-mail: kawazoe-tky@umin.ac.jp