

Automatic Selection of Diagnosis Procedure Combination Codes Based on Partial Treatment Data Relative to the Number of Hospitalization Days

Kazuya Okamoto^{1,2*}, Toshio Uchiyama^{3#}, Tadamasa Takemura⁴, Naoto Kume², Tomohiro Kuroda^{1,4} and Hiroyuki Yoshihara²

¹ Division of Medical Information Technology and Administration Planning, Kyoto University Hospital, Kyoto, Japan

² Graduate School of Medicine, Kyoto University, Kyoto, Japan

³ Faculty of Business Administration and Information Science, Hokkaido Information University, Hokkaido, Japan

⁴ Graduate School of Applied Informatics, University of Hyogo, Kobe, Japan

Abstract

Objectives: The authors developed and evaluated a method of selecting accurate diagnosis procedure combination (DPC) codes based on standardized treatment information relative to the number of hospitalization days.

Methods: The authors used machine learning methods to generate DPC codes based on treatment data. The machine learning methods utilized were the Naïve Bayes method, the SVM method, and a combined method of the two methods. We prepared DPC code data and standardized treatment data corresponding to cases occurring in fiscal year 2008 at Kyoto University Hospital. To produce classification models, machine learning methods require a moderate amount of data corresponding to each DPC code; accordingly, we selected 166 DPC codes that were each related to at least 20 cases. The number of cases with these DPC codes was 10,123.

Results: DPC code selection was attempted using the Naïve Bayes method, the SVM method, and the combined method of the two; of these, the combined method yielded the best results, producing accurate DPC codes in 73.8% of cases. In addition, we were able to improve the precision of DPC code selection to 76.5% by utilizing partial treatment data gathered up to the 11th day of each hospitalization.

Conclusion: The present study confirmed the feasibility of automatic DPC code selection through machine learning methods based on treatment information. Our future work will include the construction of a system to select DPC codes automatically and the evaluation of this system to determine whether it can reduce doctors' workloads.

Keywords

DPC; Treatment information; Artificial intelligence; Machine learning

Correspondence to:

Kazuya Okamoto

Division of Medical Information Technology and Administration Planning, Kyoto University Hospital, Kyoto, Japan.

E-mail: kazuya@kuhp.kyoto-u.ac.jp

*This work was performed when T. Uchiyama was affiliated with NTT Corporation.

EJBI 2018; 14(1):45-51

Received: December 15, 2017

Accepted: January 29, 2018

Published: February 05, 2018

1 Introduction

The diagnosis procedure combination (DPC) is a diagnosis group classification created in imitation of the diagnosis-related groups (DRGs) used in the US. DRGs and DPC are classifications according to principal diagnosis, the presence of complications, operations and treatments, and other relevant criteria. DRGs and DPC have been used to analyze diseases and hospital managing statuses

[1, 2, 3, 4]. Thomas and Ashcraft have showed DRGs to be highly reliable to measure the severity of illness of patients [1]. Stang, Merrill and Kuss have calculated population-wide rates of hysterectomy across Germany on the basis of nationwide DRG statistics [2]. Lotter et al. have compared reimbursement of burns based on DRGs in four European countries [3]. Nakagawa, Yoshihara and Nakagawa have made indicators to evaluate the management efficiency and

medical activity in each DPC or DRG group [4]. In 2003, Japan introduced a payment system of medical fees based on DPC. Like the DRG-based payment system used in the US, the DPC-based payment system is a prospective payment system and is known as the DPC/PDPS (Per-Diem Payment System). Essentially, the medical fee associated with any particular hospitalization is determined based on the DPC assigned to that hospitalization. In hospitals that have adopted the DPC/PDPS system, doctors are requested to assign one DPC code to each case in order to calculate the medical fee associated with that case. In fiscal year 2011, the DPC/PDPS system was used in 1,449 hospitals and the beds in these 1,449 hospitals made up almost half of all hospital beds in Japan.

Selecting a DPC code for a case virtually never directly improves medical care, and the task regularly takes up time that could otherwise be devoted to patient care. Consequently, in some divisions of Kyoto University Hospital, chief physicians or assistant chief physicians select the DPC codes for all hospitalizations, taking the burdensome selection process upon them to save doctors' time. Additionally, Suzuki et al. have pointed out the difficulty of accurately selecting only one DPC code for each patient because many patients have more than one disease [5]. For this reason, Suzuki et al. examined the capability of machine learning methods to select suitable DPC codes automatically based on the contents of discharge summaries [5]. Similarly, Okamoto et al. have improved the precision of automatic DPC code selection based on discharge summaries by testing several machine learning methods and combined methods of the machine learning methods [6]. They showed that DPC codes can be accurately assigned to cases automatically. However, the use of data from discharge summaries for this purpose will not help reduce doctors' workloads because the discharge summary is usually written after the patient's discharge and the selection of a DPC code is required for the patient to be discharged, meaning that a DPC code must be selected before the discharge summary is written.

Therefore, in order to reduce doctors' workloads, the authors attempt to generate a means of accurate automatic DPC code selection based on treatment information which is stored before patient discharge.

2 Objectives

In this paper, the authors develop a method of accurate automatic DPC code selection based on standardized treatment information and evaluate the developed method. In the course of hospitalization, the patient undergoes a series of treatments, and thus, treatments are different depending on the phases of the hospitalization. Moreover, doctors are requested to decide a DPC code in an early date of hospitalization. Therefore, we evaluate how precisely DPC codes are selected based on partial treatment data relative to the number of hospitalization days.

3 Materials and Methods

3.1 DPC Codes

Hospitals adopting the DPC/PDPS system are required to submit DPC codes for each admission case to the Japanese government, which is a part of data set called File Format 1. DPC codes consist of 14 digits. The first six digits of a DPC code indicate the disease name while the remaining eight digits indicate disease severity, complications, the age of the patient, the purpose of admission, operations required, treatments needed, and other variables. For instance, in the DPC code 060030xx01x0xx, the first six digits (060030) indicate that the disease was malignant tumors of the small bowel and the remaining eight digits (xx01x0xx) show that the patient underwent small bowel resection. In fiscal year 2008, the number of DPC codes used in the DPC/PDPS system was 1,572 and the number of diseases indicated by the 1,572 DPC codes was 452. In fiscal year 2011, the number of DPC codes used was 1,880 and the number of diseases indicated by the 1,880 DPC codes was 474.

3.2 Treatment Data

Hospitals adopting the DPC/PDPS system must submit data on medical treatments to the Japanese government. The data on medical treatments are standardized and sorted into three types of files known as File D, File E and File F. Files E and F include complete treatment information; File E includes medical billing information according to the medical treatment fee system; File F includes more detailed information according to the daily medical records which corresponds to the minimum unit of the master file of medical materials or examinations, etc. File D, in contrast, includes only treatment information which is not included in terms of payment in the DPC/PDPS system. For instance, microscopic examination of pathological tissue is a treatment which is included in terms of payment in DPC/PDPS system, and so File D does not include the information of this; Pathologic diagnosis expense is a treatment which is not included in terms of payment in DPC/PDPS system, and so File D includes the information of this.

3.3 Machine Learning Methods

In this study, the authors use machine learning methods known as supervised learning, which learn a classified training set and then classify a sample set. These methods thus require a classified training set.

We describe the Naïve Bayes method [7] and the SVM (support vector machine) method [8], both of which are useful machine learning methods. Several previous studies have shown that the precision of automatic classification can

increase when machine learning methods are combined [9, 10, 11, 12]. In particular, Uchiyama et al. have shown that a combined method of machine learning methods using corresponding levels of confidence calculated by logistic regression analysis is effective [12]. Okamoto et al. have likewise demonstrated that the combined methods are effective means of selecting DPC codes based on discharge summaries [6]. Hence we explain not only the Naïve Bayes method and the SVM method but also their combined method of machine learning methods.

Naïve bayes method

The Naïve Bayes method measures frequencies of elements in a training set to obtain a classification model [7]. In this paper, we utilize a multinomial model, which takes into consideration occurrence frequencies of elements in data. For data in a sample set, the multinomial model calculates the likelihood of each class and classifies data into the class whose likelihood is highest in all classes. Note that, when we count the number of elements, we smooth the data using the Good-Turing method to prevent zero probability [13].

SVM

A support vector machine (SVM) [8] finds a hyper plane which can effectively separate a training set described in a vector space into two classes. Thus SVMs are basically binary classifiers. Here, we use linear SVMs, which are the simplest SVMs. Linear SVMs are known to be effective when the number of vector elements is large, as it often is in document classification. When SVMs are applied to multiclass classification problems, they must be combined and a class must be selected based on proper criteria. In this study, we use the one-versus-the-rest classification [14], the most popular SVM technique.

Combined method

In this study, we also utilize a combined method of machine learning methods that was proposed by Uchiyama et al. [12]. First, the combined method eliminates levels of confidence of machine learning methods through logistic regression analysis. Confidence is an expression of the correctness of each machine learning method. When the combined method eliminates corresponding levels of confidence, it uses *classification scores* of the most probable class and the second most probable class. Classification scores are indicators based on which machine learning methods classify data, for example, likelihood in the Naïve Bayes method and distances from hyperplanes in SVMs. A level of confidence P is calculated as follows:

$$P = \frac{\exp Z}{1 + \exp Z} \quad (3)$$

$$Z = \beta_0 + \beta_1 L_1 + \beta_2 L_2 \quad (4)$$

where $\beta_0, \beta_1, \beta_2$ are model parameters and L_1, L_2 are classification scores of the most probable class and the second most probable class. The combined method calculates the values of model parameters $\beta_0, \beta_1, \beta_2$ by using logistic regression analysis corresponding to every machine learning method and every class in a class set.

Next, the combined method classifies sample data through the following steps: i) It classifies the sample data using more than one machine learning method. ii) It calculates corresponding levels of confidence based on classification scores of the most probable class and the second most probable class corresponding to every machine learning method. iii) It classifies the sample data into the class selected by the machine learning method with the highest level of confidence.

3.4 Experiments

The authors used machine learning methods to accurately select DPC codes based on treatment data. Namely, the DPC code data were output variables of the machine learning methods and all treatment data included in Files D, E and F were input variables of the machine learning methods.

We prepared data corresponding to cases occurring in fiscal year 2008 at Kyoto University Hospital. In the cases, every patient was admitted after April 1st 2008 and was discharged before March 31st 2009. The number of cases submitted to the DPC/PDPS system was 14,460 and the number of DPC codes related to these cases was 1,133. When generating classification models, machine learning methods require a moderate amount of data corresponding to each class, i.e., each DPC code, and thus, we only chose DPC codes that were each assigned to at least 20 cases. The number of chosen DPC codes was 166 and the number of cases assigned to these DPC codes was 10,123. Table 1 shows the top-10 list of DPC codes in terms of the number of cases.

The machine learning methods we used in experiments were the Naïve Bayes method, the SVM method and the combined method of the two methods (introduced in 3.3.3). Note that we implemented the Naïve Bayes method, used SVM-Light [15] utilizing the SVM method, and created the combined method. In the Naïve Bayes method, elements and the occurrence frequencies of the elements correspond to treatment types and the number of treatments, respectively. In the SVM method, likewise, elements of vectors and the values of the elements correspond to treatment types and the number of treatments, respectively.

Experiment 1

We evaluated the feasibility of accurate automatic selection of DPC codes by applying the Naïve Bayes method, the SVM method and the combined method of the two methods to the

Table 1: Target DPC codes and the number of the cases (top 10).

DPC code	DPC name	# of cases
020200xx970xxx	Macula and posterior pole degeneration, presence of an operation.	352
060050xx97100x	Cancer of liver and intrahepatic bile duct, presence of another operation, presence of operation and treatment 1.	291
050050xx9910xx	Angina and chronic ischemic heart disease, presence of operation and treatment 1.	242
020110xx97x0x0	Cataract and disease of the lens, presence of an operation, one eye.	215
100070xxxxxx0x	Type 2 diabetes except for diabetic ketoacidosis.	201
070560xx99x0xx	Autoimmune disease with generalized organ dysfunction.	198
020110xx97x0x1	Cataract and disease of the lens, presence of an operation, both eyes.	194
120020xx99x41x	Cancer of cervix and uterine body, presence of operation and treatment 2,4, presence of a complication.	180
030250xx991xxx	Sleep apnea, presence of operation and treatment 1.	179
120010xx99x31x	Cancer of ovary and uterine adnexa, presence of operation and treatment 2,3, presence of a complication.	174

chosen cases. The evaluation was performed via 10-fold cross-validation. We measured the precision of the 14 digits of the automatically selected DPC codes and the precision of their first six digits, which indicate the disease name. Furthermore, because an automatic DPC code selection system that proposes several candidate DPC codes could be effective, we had the methods generate five possible 14-digit DPC codes and measured their ability to produce one correct DPC code out of the five. We also had the methods generate five possible candidates of the first six digits of the DPC codes and measured their ability to produce one correct six-digit sequence. These four measurements, namely, the precision of the 14 digits of a single automatically selected DPC code, the precision of the first six digits of a single automatically selected DPC code, the precision of one of five suggested 14-digit DPC codes, and the precision of one of five suggested first-six-digit sequences, are referred to as *the 14-digit top-1 precision*, *the 6-digit top-1 precision*, *the 14-digit top-5 precision* and *the 6-digit top-5 precision*, respectively. However, the combined method cannot be used to select additional DPC codes, because a combined method of machine learning methods used here selects a single one DPC code by selecting one machine learning method. Hence, the combined method selects the remaining four of the five possible DPC codes according to the Naïve Bayes method.

Experiment 2

We evaluated how precisely DPC codes are selected based on partial treatment data relative to the number of hospitalization days. Each machine learning method makes a classification model by using treatment data up to the x -th day of hospitalization and classifies a sample set consisting of treatment data up to the x -th day of hospitalization by utilizing the classification model. In this experiment, we evaluated the precision achieved using the Naïve

Table 2: The results of automatic DPC code selections based on treatment data using the Naïve Bayes method, the SVM method and the combined method of the Naïve Bayes method and the SVM.

Machine Learning Methods	Precision (%)		Prec. of top 5 (%)	
	14 digits	6 digits	14 digits	14 digits
Naïve Bayes method	63.2	71.5	87.6	90.6
SVM	72.1	75.7	86.4	89.8
Combined method	73.8	78.7	89.9	93.2

Bayes method, the SVM method and the combined method of the two methods when $x = 1, \dots, 20$. As in experiment 1, we measured the 14-digit top-1 precision, the 6-digit top-1 precision, the 14-digit top-5 precision and the 6-digit top-5 precision. Note that, in order to choose five candidates, the combined method selects the remaining four DPC codes according to the Naïve Bayes method, as in experiment 1.

4 Results

The results of experiment 1 are shown in Table 2. The second and third columns show the 14-digit top-1 precision and the 6-digit top-1 precision, respectively. The fourth and fifth columns show the 14-digit top-5 precision and the 6-digit top-5 precision, respectively.

Figures 1-4 show the 14-digit top-1 precision, the 6-digit top-1 precision, the 14-digit top-5 precision and the 6-digit top-5 precision, respectively, in experiment 2. In these figures, the Naïve Bayes method and the combined method of the Naïve Bayes method and the SVM method are abbreviated as NB and NB+SVM, respectively. Moreover, NB+SVM (all)

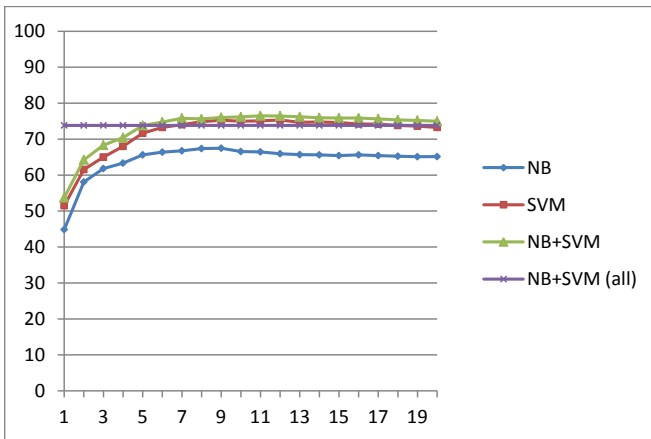


Figure 1: The precision of the 14 digits of automatically selected DPC codes relative to the number of hospitalization days from which data was used (the 14-digit top-1 precision).

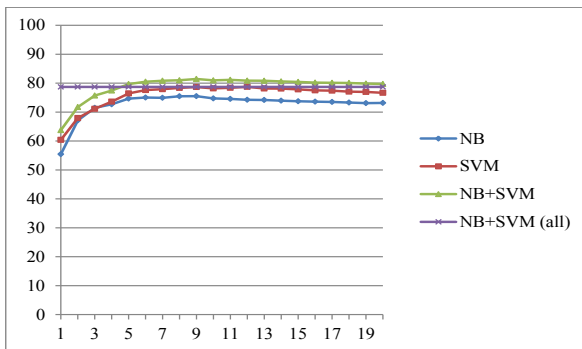


Figure 2: The precision of the first six digits of automatically selected DPC codes relative to the number of hospitalization days from which data was used (the 6-digit top-1 precision).

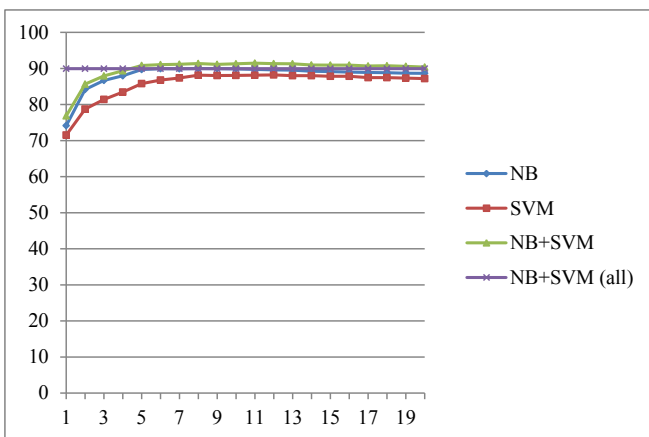


Figure 3: The precision of one of five suggested 14-digit DPC codes relative to the number of hospitalization days from which data was used (the 14-digit top-5 precision).

indicates the combined method which makes a classification model by using the treatment data of entire hospitalization as a training set and selects DPC codes by using a sample set consisting of the treatment data of entire hospitalization. The results of NB+SVM (all) are shown for comparison.

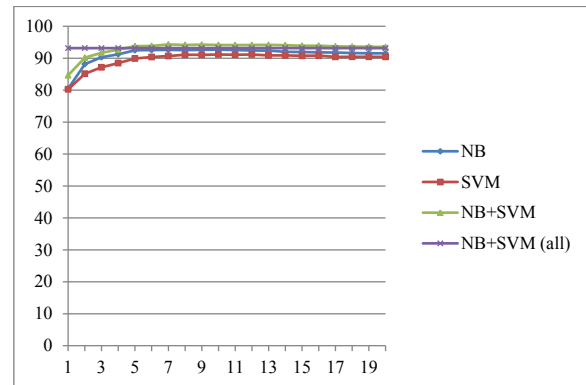


Figure 4: The precision of one of five suggested first-six-digit sequences relative to the number of hospitalization days from which data was used (the 6-digit top-5 precision).

5 Discussion

In experiment 1, the combined method of the Naïve Bayes method and the SVM method yielded results that were more precise than those of either the Naïve Bayes method or the SVM method. The 14-digit top-5 precision and the 6-digit top-5 precision achieved using the combined method were almost 90% and over 90%, respectively; these levels of precision demonstrate that the combined method is useful. In experiment 2, all four precision values were maximized as the number of hospitalization days from which data were used approached 10, showing that DPC code selection based on partial treatment data relative to the number of hospitalization days is ideal and that characteristic treatments are tend to be administered at the beginning of hospitalization. In actuality, characteristic treatments, such as operations and expensive medicines, are administered at the beginning of hospitalization and a DPC code seems to be determined based on the characteristic treatments.

Horiguchi et al. proposed several rules by which to determine the DPC codes for cases to which disease names had already been assigned, and showed that 95.8% of cases were assigned proper DPC codes according to their rules [16]. In particular, they pointed out that when the first six digits of a DPC code are given, the remaining eight digits can be automatically selected with high precision. However, since they did not clearly divide their data into a training set and a sample set, their study did not have sufficient cross-validation. In the present study, the 14-digit top-1 precision of the combined method was 73.8% and the 6-digit top-1 precision of the combined method was 78.7%. The results confirmed that when the first six digits of a DPC code are given, the remaining eight digits can be automatically selected with high precision of 93.8%.

Nishi and Uozawa selected characteristic treatment data from Files E and F and constructed a decision tree using this data which was able to correctly determine the first six digits

of the DPC code in 74.5% of cases [17]. Since they selected the treatment data which they considered characteristic, however, the data were not standardized. In the present study, in experiment 1, the combined method yielded the first six digits of the appropriate DPC code with a higher level of precision: 78.7%. In experiment 2, the precision increased to 81.5% when data up to the 9th day of hospitalization were incorporated. Although it is difficult to compare the results of the present study with those of Nishi and Uozawa because of difference data sets used in each study, the precision of our results was higher in spite of the smaller size of training data and the larger number of possible DPC codes. The results achieved in the present study appear to be superior to those achieved in the study by Nishi and Uozawa.

In experiments 1 and 2, when the combined method selected five candidates, it preferred the Naïve Bayes method. If the combined method were to prefer the SVM method, the precision could increase. This is not likely, however, given that, in the other experiments where the combined method preferred the SVM method, the 14-digit top-5 precision and the 6-digit top-5 precision were 87.7% and 91.1%, respectively, which were lower than the levels achieved when the combined method preferred the Naïve Bayes method. Additionally, in DPC code selections using partial treatment data relative to the number of hospitalization days, the precision achieved when the combined method preferred the Naïve Bayes method was always higher than the precision achieved when the combined method preferred the SVM method.

In experiment 2, the machine learning methods made classification models by using partial treatment data from up to the x -th day of hospitalization to classify a sample set consisting of partial treatment data from up to the x -th day by utilizing the classification models. We compared this technique with the machine learning methods using the treatment data from the entire hospitalization as a training set. As x increased, the precision of the alternative methods approached asymptotically to the precision of the machine learning methods trained on treatment data from the entire hospitalization to select DPC codes using a sample set consisting of treatment data from the entire hospitalization. Also, the precision of the alternative methods was always less than the precision of the machine learning methods used in experiment 2 when $x = 1, \dots, 20$. Thus we confirmed that a patient with a particular diagnosis does tend to undergo a series of treatments and that machine learning methods can learn effectively when trained on partial treatment data relative to the number of hospitalization days.

We consulted chief physicians in some divisions about the implications of automated coding in practical terms. According to them, they take a few minutes to select each DPC code and feel the processes laborious especially when they must change DPC codes. Though their comments cannot always reflect practical situations, we obtained suggestions about effective use of an automated coding system. If automatic selection of DPC codes by machine learning methods is to be implemented to reduce doctors' workloads, one problem will have to be resolved. Machine learning methods need a moderate amount of data

corresponding to each class. In the present study, we selected 166 DPC codes; in contrast, the number of DPC codes used in the DPC/PDPS system in fiscal year 2008 was 1,572. However, the 166 DPC codes used in the present study were assigned to 70.0% of the actual cases submitted to the DPC/PDPS system. To confirm the viability of automatic DPC code selection using machine learning methods, we consulted a doctor, who pointed out that, even if machine learning methods could be used to identify only partial DPC codes for only 70.0% of cases that alone would significantly reduce doctors' workloads. Another drawback is that the DPC codes used in the DPC/PDPS system are revised every two years. Nevertheless, it is not difficult to translate the codes of an old DPC code system into a new version.

6 Conclusion

The authors evaluated automatic DPC code selection by machine learning methods based on treatment information for the purpose of reducing doctors' workloads. In 73.8% of cases, accurate DPC codes were selected when the combined method of the Naïve Bayes method and the SVM method was used. In addition, we were able to improve the precision of DPC code selection to 76.5% by incorporating partial treatment information from up to the 11th day of hospitalization. Our future work will include the construction of a system that selects DPC codes automatically and the evaluation of this system to determine whether it can reduce doctors' workloads.

7 Acknowledgement

This work was supported by KAKENHI (15K00466).

8 Author Contribution

K. Okamoto, T. Uchiyama, T. Takemura and H. Yoshihara have contributed to the conception and design of research; K. Okamoto, T. Takemura, N. Kume and T. Kuroda collected and managed the research data; K. Okamoto and T. Uchiyama performed experiments; K. Okamoto, T. Uchiyama, and T. Takemura performed the interpretation the data and results. K. Okamoto and T. Uchiyama has drafted the article; K. Okamoto, T. Uchiyama and T. Kuroda revised the article. All authors have given approval of the final version to be published.

9 Conflict of Interest

The authors declare no conflicts of interest associated with this manuscript.

References

1. Thomas JW, Ashcraft ML. Measuring Severity of Illness: a Comparison of Interrater Reliability among Severity Methodologies. *Inquiry*. 1989; 26: 483-492.

2. Stang A, Merrill RM, Kuss O. Hysterectomy in Germany: a DRG-based nationwide analysis, 2005-2006. *Dtsch Arztebl Int.* 2011; 108: 508-514.
3. Lotter O, Jaminet P, Amr A, Chiarello P, Schaller HE, Rahmanian-Schwarz A. Reimbursement of burns by DRG in four European countries: An analysis. *Burns.* 2011; 37: 1109-1116.
4. Nakagawa Y, Yoshihara H, Nakagawa Y. New Indicators Based on Personnel Cost for Management Efficiency in a Hospital. *J Med Syst.* 2011; 35: 625-637.
5. Suzuki T, Yokoi H, Fujita S, Takabayashi K. Automatic DPC Code Selection from Electronic Medical Records: Text Mining Trial of Discharge Summary. *Methods Inf Med.* 2008; 47: 541-548.
6. Okamoto K, Uchiyama T, Takemura T, Adachi T, Kume N, Kuroda T, et al. Automatic DPC Code Selection from Discharge Summaries Using Several Machine Learning Methods. *Trans Japanese Soc Med Bio Eng.* 2011; 49: 40-47 (Japanese).
7. McCallum A, Nigam K. Comparison of Event Models for Naïve Bayes Text Classification. *Proceedings of AAAI-98 Workshop on Learning for Text Categorization*; 1998. p. 41-48.
8. Boser BE, Guyon IM, Vapnik VN. A Training Algorithm for Optimal Margin Classifiers. *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory*; 1992. p. 144-152.
9. Schapire RE, Singer Y. Improved Boosting Algorithm Using Confidence-rated Predictions. *Machine Learning.* 1999; 37: 297-336.
10. Schapire RE, Singer Y. BoosTexter: A Boosting-based System for Text Categorization. *Machine Learning.* 2000; 39: 135-168.
11. Kudo T, Matsumoto Y. A Boosting Algorithm for Classification of Semi-Structured Text. *Trans IPS Japan.* 2004; 45: 2146-2156 (Japanese).
12. Uchiyama T, Bessho K, Uchiyama T, Oku M. A Multi-classifier Combination Using Confidence Estimation. *Proceedings of the 84th Knowledge Base Systems*; 2009. p. 33-38 (Japanese).
13. Chen SF, Goodman J. An Empirical Study of Smoothing Techniques for Language Modeling. *Technical Report at Harvard University.* 1998; TR-10-98.
14. Hamel LH. *Knowledge Discovery with Support Vector Machines.* Massachusetts: Wiley-InterScience; 2009.
15. Joachims T. Making large-Scale SVM Learning Practical. In: Schölkoph B, Burges CJC, Smola AJ, editors. *Advances in Kernel Methods: Support Vector Learning.* Massachusetts: MIT-Press; 1999.
16. Horiguchi H, Yasunaga H, Hashimoto H, Ishikawa KB, Kuwabara K, Anan M, et al. Development of the Standard DPC Coding Logic. *J Med Inf* 2008; 28: 73-82 (Japanese).
17. Nishi T, Uozawa M. The DPC Disease Decision Model from Medical Treatment Data by Data Mining. In: *Proceedings of the 29th Joint Conference on Medical Informatics*; 2009: p. 56-459 (Japanese).