International Journal of Engineering & Technology, 7 (3.33) (2018) 223-226

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Building the knowledge base for non-combinable codes according to the Korean Standard Classification of Diseases

Mijung Kim[1]*

[1]Faculty, Department of Health Administration, Kwangju Women's University
**Corresponding author E-mail: mjkim@kwu.ac.kr*

Abstract

The purpose of this study is to develop a knowledge base for non-combinable combinatorial codes to improve the accuracy of disease classification. We defined the rules related to non-combinable codes according to the list of code pairs proposed by the HIRA and the KCD-7 classification rules. A knowledge base was created according to defined rules and verified. To validate the knowledge base, inpatients who were billed for diabetes mellitus in December 2016 were selected as the subject of the study. As a result, the number of combinatorial codes proposed by the HIRA was 1,195, but the number of code pairs generated in the knowledge base was 25,439. Non-combinable codes by confirming with an indication of the HIRA have discovered 1,391 cases. As a result of verification with the code pair of the proposed knowledge base, 100 combinations were found. Non-combinable codes by confirming with an indication of the HIRA have discovered 1,391 cases. As a result of verification with the code pair of the proposed knowledge base, 3,525 combinations were found. It is meaningful that a convenient authoring tool that can automatically catch combinatorial codes was developed to build a knowledge base.

Keywords: *Coding rules; Insurance claims code; Knowledgebase*

1. Introduction

The release of public data and its utilization by private sectors are being accelerated world widely. In South Korea, to guarantee the right to access the open data by people and to contribute to the enhancement of quality of life and the development of the national economy through the private sector's utilization of open data, the data held and managed by public institutions are released. Among them, the representative public data in the healthcare field is the health insurance claims data of Health Insurance Review & Assessment Service (HIRA). HIRA is the only national institution to review and assess the health insurance. This data is a beneficial one to understand the state of national health, and the use of medical services. Since insurance claim is reviewed based on the final diagnosis of patient, the diagnosis code is the most important but despite the same patient, as different codes are assigned according to the skills, department, knowledge level of those who give the diagnosis codes to each patient, the question on the accuracy of diagnosis code has been raised continuously [1-2]. The use of inappropriate codes causes a cut in insurance, resulting in the hospital's financial disadvantage, and also lowers the credibility of data when using it [3]. Therefore, to improve the accuracy of diagnosis code in the claim, Health Insurance Review & Assessment Service developed and has been monitoring the item of code unable to be used in combination entry rate indicator from 2016 and expanding it step by step. The code unable to be used in combination means the code that cannot be used together based on the complete code. We named it as non-combinable code for convenience. Non-combinable code entry rate indicates the ratio of non-combined code together.

The diagnosis code used when claiming the health insurance is from the KCD-7, the seventh revision of Korean standard Classification of Disease. KCD-7 is based on ICD-10, the tenth revision of International Standard Classification of Diseases, with the extension in accordance with domestic reality. Taegeuk mark is affixed after the code, but when classifying the code, Taegeuk mark is not used. In case that two codes are assigned to one diagnosis name, only one crucial code between them is selected and designated as a principal diagnosis code. Although for the statistical purpose, mainly the code of 3 digits, that is, the sub-class, is used, when claiming the insurance, the complete code is used.

The concordance rate between the diagnosis code provided by the doctor and the code reclassified by the medical record officer was 81%, and the concordance rate of other diagnosis was only 47.5% [4]. Although the knowledge-based model which induced the doctor to provide the accurate diagnosis was developed reflecting the domestic reality that when the doctor selects the diagnosis name, the disease classification codes are encoded and stored [5], it was designed as the function to support the entry of diagnosis name by doctor so that the diagnosis name at detailed level can be entered not the comprehensive diagnosis name. In some researches, to solve the diagnosis name classification error and the omission of the code in the obstetric patient, the rule-based diagnosis code classification supporting system was developed by defining the obstetric disease classification guidelines as rule, but it approached with the concept that suggests the additional code according to the obstetric patient information such as the number of births, number of the stillborn births, delivery method, others. [6]. The disease classification that classifies the disease name prepared mainly in a declarative sentence by the natural language processing technology and the machine learning was suggested [7], but it is not suitable for out domestic medical environment that the doctor selects the codified disease name and stores the disease classification code.

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this study, we built the knowledge base with defined the rules for non-combinable codes to improve the accuracy of classification of diseases.

The rest of this paper is organized as follow: section 2 describes the material and methods for building the knowledge base. Section 3 describes the result and evaluation of the knowledge base.

2. Design the knowledge base

This study was carried out in 3 steps. Firstly, we analysed the non-combinable codes. The list of non-combinable codes announced by HIRA and the coding rules of KCD-7 were carefully examined. Secondly, we defined the rules for the non-combinable cases. Thirdly, we built these rules as a knowledge base. Lastly, we evaluated the knowledge base.

2.1. Analysis non-combinable codes

2.1.1. Monitoring items of HIRA

Acquisition of knowledge on the diagnosis code unable to be used in combination was made through KCD-7, disease classification guidelines, and the knowledge of disease classification expert [8-9]. The indicator of the code unable to be used in combination published by Health Insurance Review & Assessment Service in 2018 was total 16 indicators and 3,796 pairs of code were designated as non-combinable codes as in Table 1. The number of items related to diabetes mellitus is the largest at 1,195. So indicator of diabetes mellitus was selected for the subject.

Table 1: 16 indicators and 3,796 pairs of code by HIRA

No	Indicator	Number of code pairs
1	Diabetes mellitus	1,195
2	Musculoskeletal system	1,109
3	Symptoms	99
4	Gastrointestinal diseases	10
5	Genitourinary diseases	38
6	Infectious diseases	81
7	Conflicting terms	57
8	Late effects	137
9	Complications	85
10	Injury	34
11	The perinatal period	110
12	Congenital diseases	207
13	Pregnancy status	101
14	Mental illness	7
15	health status and contact with health services	49
16	Cause of onset	477
Total		3,796

The formula of Non-combinable code entry rate is as in (1).

$$\text{NCC entry rate} = \frac{\text{Number of occurrences (A and B code)}}{\text{Total number of claims (A or B code)}} \times 100 \quad (1)$$

Diabetes mellitus is categorized by codes from E10.00 to E14.9 in the Korean Classification of Diseases. When insurance claims are charged, remove the punctuation mark of the code generally. Since the code was defined based on the complete code, although the number is high, if it is described as a rule, it can be simplified. Focus on diabetes mellitus.

2.1.2. Coding rules of KCD-7

According to KCD-7 coding rules, more codes should be included in the code unable to be used in combination. In this study, out of the KCD-7 coding rules, the codes corresponded to the index of non-combinable code by HIRA were analyzed. The general coding rules are: first, once the diagnosis is confirmed, no code associated with the symptoms of that disease and higher than the test value shall not be assigned separately; second, specific disease classification code shall be assigned as much as possible, and the comprehensive code shall not be assigned together; third, the disease classification code corresponded to "include" of specific disease classification code shall not be assigned together; fourth, the disease name can be expressed two or more, and each disease classification code can be assigned, but in case that they are combined into one and can be classified as another disease classification code, they should be classified as latter one only.

2.2. Defining rules for knowledge base

Based on the main code, additional codes that cannot be used with it were written in the table. For the specific knowledge, the rule should be defined based on the complete code, and some knowledge could be defined as a rule in the sub-class level. In other words, for some codes, the non-combinable codes could be defined by the sub-class of the codes that have the same numbers in the initial three digits. Some others were defined by the group of codes that have the same numbers in the initial four digits. The rest were defined by the basis of complete codes. For example, in the case of diabetes, the types of diabetes were not to be overlapped, and the case with complication was not to be combined with the case without complication. When the diabetes mellitus is selected as the main code, ‘glycosuria’ and ‘elevated blood glucose level’ code are not to be entered.

In the results of comparing the index of code unable to be used in combination published by Health Insurance Review & Assessment Service and the code unable to be used in combination according to KCD-7 coding rules, since in case of those made by Health Insurance Review & Assessment Service, they suggested the code based on the complete code, multiple pair of codes are omitted although actually, there are more codes unable to be used in combination, which is deemed to be because the codes unable to be used in combination suggested by Health Insurance Review & Assessment Service were prepared based on the actual cases of insurance claim.

2.2.1. How to create rules

There are three ways to create rules. The first is how to define narrow rules based on the complete code directly. The second is to use an asterisk mark to determine the likelihood search when the three or four unit classification is the same. Third, if the code of a particular position is a fixed number, it can be specified by an exclamation point, and the number to be entered at that position can be listed.

There are some commands and marks for creating rules grammar. We can use an asterisk mark, an exclamation mark and a wave mark. Using an asterisk, it searches for all codes that begin with the character before the asterisk (*). An exclamation mark (!) searches for all codes that contain specially restricted characters in place. Characters that can be specifically entered are enclosed using square brackets ([]). Write an exclamation point followed by a square bracket, and commas separate the numbers in square brackets. A wave mark can be used in square brackets. “SET” means that the code listed after the word represents reference codes. "BAN" means that codes listed after the word cannot be used with codes listed before it.

If the knowledge base is built with the codes unable to be used in combination based on the KCD-7 sub-classification unit, the number of rules can be reduced, and more error cases can be found. In the case of diabetes mellitus, there exist 13 broad rules and 30,915 non-combinable code pairs. The broad rules for the diabetes mellitus are described in table.2.

Table 2: Broad rules for non-combinable code pairs related to diabetes mellitus

Type	Main code	Non-combinable codes
Complications	E10*~E14*	E10.9, E11.9, E12.9, E14.9
Exclusion codes	E10*~E14*	E74![0~9]
Abnormal findings	E10*~E14*	R81, R73![0,9]
Symptom	E10.1! ~E14.1![0,1,2,8]	E87.9
DM type	E10*	E11*, E12*, E13*, E14
	E11*	E10*, E12*, E13*, E14
	E12*	E10*, E11*, E13*, E14
	E13*	E10*, E11*, E12*, E14
	E14*	E10*, E11*, E12*, E13
Principle diagnosis selection error	E10.9	E10*
	E11.9	E11*
	E12.9	E12*
	E13.9	E13*
	E14.9	E14*

However, there are considerations when creating detailed rules by general rules. A broad rule may produce the same code pair and the same detail rules can also be made. For example, a diabetes code with complications cannot be accompanied by a complication-free diabetes code, and a rule that diabetes codes with different diabetes types cannot be duplicated can result in overlapping combination codes. The authoring tool has the function of eliminating redundant code pairs and excluding the same code pairs. Figure 1 shows the authoring tools to define a rule.

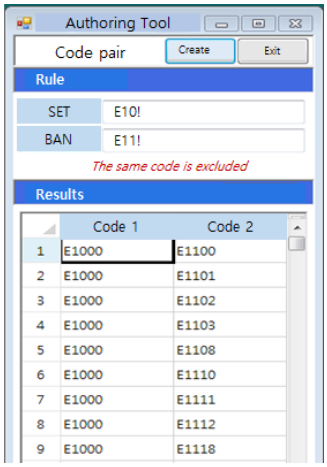

Fig. 1: The authoring tools to define a rule

Rules related to diabetes mellitus were generated using the authoring tool. The number of the specific rule is 30,170 except for pairs of codes that overlap and the same code pairs. It is necessary the ability to filter duplicates in the authoring tool.

2.2.2. The phase of construction of the knowledge base

The phase of construction of the knowledge base is as follow. First, define the rules using the authoring tool according to the grammar. The generated rules are retrieved from the KCD-7 code master table under the syntax by using the instance generator, and the retrieved codes and the combinable codes are mapped as 1: 1 and stored in the knowledge base. Rules created in the authoring tool can be viewed and modified at any time by the user. Table 2 shows an example according to the phase of generating code pairs that cannot be used together.

Table 2: The phase of rules generated by the authoring tool

Phase	Sample
1) Authoring broad rules	SET E10.1! ~ E14.1! [0,1,2,8] BAN E87.9
2) Interpretation	E87.9 cannot be used with E10.10 or E10.11 or E10.12 or E10.18, E11.10 or E11.11 or E11.12 or E11.18, E12.10 or E12.11 or E12.12 or E12.18, E13.10 or E13.11 or E13.12 or E13.18, E14.10 or E14.11 or E14.12 or E14.18
3) Generating Non-combinable code pairs	(E10.10, E87.9)(E10.11, E87.9)(E10.12, E87.9) …… (E14.11, E87.9)(E14.12, E87.9)(E14.18, E87.9)

2.3. Proposed system

The knowledge base was built with non-combinable codes that should not be entered additionally. Figure 2 is an overview of the knowledge-based diagnosis coding system. Non-combinable code pairs were generated automatically according to authored rules.

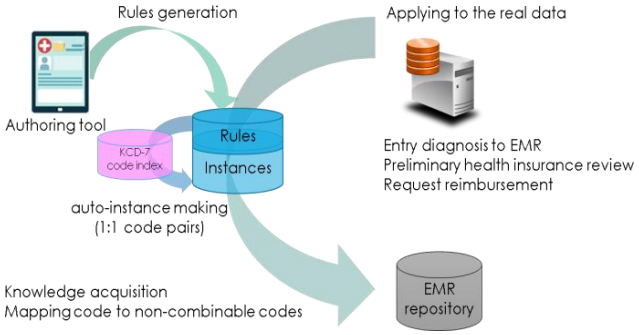

Fig. 2: The overview of the knowledge-based diagnosis coding system

The complete code represents the most detailed code. Since the detailed rules based on the complete code are too many, the detailed rules are designed to be automatically generated according to comprehensive rules when they can be handled according to them. The patient's diagnosis is entered by the doctor into the electronic medical record. The doctor must fill out the hospital discharge summary and select the final diagnosis when the patient is discharged. However, there are cases where codes such as diagnostic impression, abnormal test findings, and symptoms before the diagnosis is confirmed are written in the discharge summary. The diagnostic code entered there is used for insurance claims, medical statistics, and issuing a medical certificate. Each division has been performing a time-consuming double task separately that requires the doctor to enter the combinatorial code entered. Verifying disease codes entered into the knowledge base when the physician writes the summary of discharge, combination codes are filtered. When the main code is entered, combinatorial codes that are paired with the corresponding main code are listed in the instance, and an error message is displayed when the listed code is input as an additional code. If an error check has not been performed at this stage, the insurance claim department can verify the diagnosis name of the patient to be charged at the pre-examination or insurance claim stage to the knowledge base.

3. Results and Discussion

The actual claim data of the Health Insurance Review and Assessment Service were used to evaluate the validity of the proposed system. The subjects were hospitalized patients who were diagnosed with diabetes and had over one other diagnosis during December 2016. A total of 64,287 patients were diagnosed with diabetes mellitus. Non-combinable codes by confirming with an indication of the HIRA have discovered 1,391 cases (1,338 patients). As a result of verification with the code pair of the proposed knowledge base, 3,525 (3,184 patients) combinations were

found. More errors were caught using the proposed knowledge base than HIRA's index. The error detection rate was 2.4 times higher. This result was evaluated just whether the code can not be combined by the principal diagnosis purely. And input error is represented by code pair because HIRA announced it in code pairs. It does not have to be this way. Rules can be generated by other types.

4. Conclusion

Disease codes written on medical records must be assigned according to principles of the standard classification disease correctly. As a way to do this, it is to make sure that non-combinable codes for each code cannot be used together. We have constructed a non-combinable code knowledge base around diabetic patients and verified their validity.

It is meaningful that a convenient authoring tool that can automatically catch combinatorial codes was developed to build a knowledge base. This authoring tool can be used to generate knowledge, and it is easy to make detailed rules by only defining general rules. Since the made rules can be corrected regardless of system or type of programs, they can reuse again. The suggested knowledge base was built independently of a specific program or system, so it has good scalability and portability. When using the proposed knowledge base, more code input errors can be caught and prevented. If this knowledge base is applied to the preliminary review system, the risk to claim the insurance with wrong codes can decrease. Moreover, if it is loaded to electronic medical record systems, the accuracy of diagnosis code collection for statistics can be guaranteed as well as the selection of diagnosis name selected by doctors. Since the number of non-combinable codes is considerable, processing speed remains a problem.

Acknowledgement

This research was supported in part by Research Funds of Kwangju Women's University in 2018(No.KWUI18-037).

References

[1] Lee JH, Shim MS, “The Accuracy of the ICD-10 Code for Trauma Patients Visiting on Emergency Department and the Error in the ICISS”, *Journal of Trauma and Injury*, Vol. 22, No. 1, (2009), pp.108-115.

[2] Larkey LS, Croft WB, “Technical Report - Automatic assignment of icd9 codes to discharge summaries”, the *University of Massachusetts at Amherst, Amherst, MA.*, Ph.D. thesis, 1995

[3] Pakhomov SV, Buntrock JD, Chute CG., “Automating the assignment of diagnosis codes to patient encounters using example-based and machine learning techniques”, *Journal of the American Medical Informatics Association*, Vol. 13, No. 5, pp.516-525, 2006

[4] Bae SO, Kang KW, Boo YK, Lee Y, Cheo HS, Choi HY, “A Study on the Difference in Disease Coding of Doctors, Medical Insurance Review Nurses and Medical Record Administrators based on Coding Simulation”, *Journal of Health Informatics and Statistics*, Vol. 40, No. 3, (2015), pp.161-174.

[5] Kim MJ, “A Knowledge-based Model for Efficient Classification of Diseases”, *Jeju National University*, Korea, Doctoral thesis, 2017

[6] Kim MJ, Kim HC, “Building Rule-based support system for disease code classification”, *The e-Business Studies*, Vol. 15, No. 4, (2014), pp.61-81.

[7] Serguei V.S. Pakhomov, JAMES D. Buntrock, Christopher G. Chute, “Automating the Assignment of Diagnosis Codes to Patient Encounters Using Example-based and Machine Learning Techniques”, *Journal of the American Medical Informatics Association*, Vol. 13, No. 5, (2006), pp.516-525

[8] Statistics Korea Korean standard classification of disease 7th revision, Korean medical record association, 2015

[9] Statistics Korea, *Classification of disease manual*, Statistics Korea,(2015), pp: 3-35