Inference of ICD Codes by Rule-Based Method from Medical Record in NTCIR-12 MedNLPDoc

Masahito Sakishita
Faculty of Informatics, Shizuoka University
msakishita@kanolab.net

Yoshinobu Kano
Faculty of Informatics, Shizuoka University
kano@inf.shizuoka.ac.jp

ABSTRACT

We propose an effective method which automatically assigns proper ICD codes for diagnosis. Unfortunately, the number of electronic Japanese medical records available would not be sufficient for statistical machine learning methods to perform well. Therefore, we observed characteristics of medical records manually, writing rules to make effective methods by hand. Our system achieved the highest F-measure score among all participants in the most severe evaluation criteria. Through comparison with other approaches, we show that our approach could be a useful milestone for the future development of Japanese medical records processing.

Team Name

KIS

SubTask

MedNLPDoc

Keywords

ICD code, medical record, NLP

1. INTRODUCTION

Our ultimate goal is to the development of text processing field of medical informatics.

It is not easy for a human to derive an appropriate ICD code from a given medical record. There are two reasons why it is not easy. Firstly, the coding task requires knowledge of medical technical words in the medical record. Because there are a lot of technical terms in a medical record, the coding task is limited to professionals who learned to manage sufficient knowledge. Secondly, only doctors with actual clinical experiences could understand real intention of diagnosis. In other words, expert techniques and experiences are required if a non-professional wishes to guess the intention and for coding without examining an actual patient.

From features of medical records, we made five rule-base methods using four reference materials, ICD コーディングトレーニング(第 2 版) (ICD Coding Training(Second edition))[1], ICD10 国際疾病分類第 10 版（2003 年改訂）(International Classification of Diseases 10 edition (revision in 2003))[2], ICD10 対応標準病名マスター (Standard Disease Name Master compatible with ICD10)[3] and ライフサイエンス辞書プロジェクト (Life Science Dictionary Project)[4].

Our system achieved the best performance regarding the strict match score of this MedNLPDoc task.

2. METHOD

We suggest five methods that output appropriate ICD code given a medical record text. Our final system is a combination of these five methods. We describe our methods one by one below.

2.1. Decision of target sentence

We define a “sentence” as a line of a medical record marked off by the Japanese periodical symbol, “。”.

We suggest that there are two types of sentences in medical records: sentences that include diagnosis, and sentences that do not include diagnosis. The latter type of sentences may include disease names which are not related to diagnosis. For example, this sentence contains a diagnostic result: “検査の結果で慢性化膿性中耳炎と診断され、手術目的に入院となる。(As a result of medical check, diagnosed as *chronic suppurative otitis media*, and hospitalization is needed for an operation.)”, where diagnostic result is “慢性化膿性中耳炎（*chronic suppurative otitis media*)”. The next example does not contain diagnostic result: “難聴を主訴に受診する。 (The patient saw a doctor and his/her main complaint is *deafness*.)”. Describing main complaint of the patient is the objective of this sentence.

When a sentence contains diagnostic result, and when that sentence contains the name of disease, our system output the name of the corresponding ICD code. We describe details of our method below.

We extract sentences that contain a keyphrase to narrow candidate sentences down. For example, the previous example sentence with diagnostic result “検査の結果で慢性化膿性中耳炎と診断され、手術目的に入院となる。” has a keyphrase of “と診断され（be diagnosed)” with its diagnosis name of disease before the keyphrase. In addition to the keyphrase “と診断され”, we used keyphrases of “と診断（be diagnosed with)”, ”の診断（diagnosis of)”, ”との診断（diagnosis of)”, ”の症状（symptoms of)”, ”を発症し（develop) ”,”を認め (see)”,”が確認された（be confirmed)”,”をきたし（cause) ”,”フォロー（follow-up)”,”が原因と考えられた（be thought to be the cause)”,”と考えられた（be thought)”,”が考えられた (be thought)”,”を指摘（point)”,”も認め（see)”,”が認め (see)”,”合併（combine)”,”併発（accompany)”,”で入院（for an operation)”,”最終診断（final diagnosis)”,”確定診断 (established diagnosis)”,”*Dx*” and ”疑い（doubt)”. We chose these keyphrases by manually verifying medical records written in reference [1] and medical records of MedNLPDoc training data. If a sentence contains a negation, e.g. “認めない（not see)”, that sentence is discarded from the candidate sentences. In this paper, we call this candidate sentence selection method above as “SCS”.

After SCS, morphological analysis is performed by the Kuromoji morphological analyzer. We used a custom dictionary for Kuromoji where Wikipedia entry words and disease names are registered. Disease names are taken from ICD10 対応標準病名マスター (Standard Disease Name Master compatible with ICD10) [3]. We changed the weight of words in the dictionary in order to make disease names of the dictionary appear preferentially.

When the disease name is included in the morphological analysis result, we derive corresponding ICD code in the table of ICD10 対応標準病名マスター[3].

2.2. Dealing with English words

There are many English words used as technical terms in the Japanese medical records. These English words are written in alphabets. Because these English words are often professional which are not registered in the morphological analyzer’s dictionary, we cannot deal with it directly.

We used the Life Science Dictionary [4] to translate English words into Japanese words. In this technique, we only use dictionary entries which exactly matched with the English words in the medical record.

2.3. Dealing with paraphrase words

There are many inconsistent spelling variations appear in the medical records. We deal with this problem by our technique below. For example, ” *Alzheimer’s disease*” has variations like “アルツハイマー型認知症（*Alzheimer dementia*)”, “アルツハイマー（*Alzheimer*)”, "アルツハイマー病（*Alzheimer disease*)" and "*Alzheimer*". We make “アルツハイマー型認知症” as a normalized word for these variations in order to assign an appropriate ICD code. We use the redirection relations of Wikipedia to make such normalizations, i.e. redirected words correspond to normalized words.

2.4. Dealing with disease names including various body parts

Our technique described in section 2.1, descriptions like “○○の癌” or “○○と××に損傷” will only output corresponding ICD codes of “癌（*cancer*)” or “損傷（*damage*)”, ignoring ”○○” and ” ××”. However, these ignored words could include information required to output appropriate ICD codes.

We decided to focus on “悪性新生物 (*malignant neoplasm*)”and”損傷(*damage*)” in our technique. Our system outputs ICD codes from combination of words as follows.

	the same meaning words
悪性新生物	癌,悪性新生物,悪性腫瘍,転移
損傷	骨折,捻挫,ストレイン,脱臼,断裂,表在損傷,挫傷,打撲,皮下血腫,皮下出血,内出血,皮下異物,擦過創,虫刺傷,切創,挫創,割創,裂創,開放創,貫通創,咬創,刺創,損傷,挫滅損傷,挫滅創,切断

Table 1. the same meaning words of “悪性新生物 (*malignant neoplasm*)” and “損傷 (*damage*)”

We define rules to detect ICD codes using combination of words expressed various parts of body, and the words in Table 1.We manually made a list of body parts using the document [2].

If a sentence contains both a word of the body parts and a word listed in Table 1, our system outputs a corresponding ICD code. For example, in a case of “C00.0 口唇の悪性新生物，外側上唇 (*malignant neoplasm* of lips, out upper lips)” in ICD10 国際疾病分類第 10 版（2003 年改訂） (International Classification of Diseases 10 edition (revision in 2003)) [2], our system outputs C000 if a candidate sentence contains “唇 (lip)” and a word that indicates “悪性新生物” as listed in Table 1.

While we only check sentences selected by our technique described in section 2.1 in the case of “損傷 (*damage*)”, we used the whole medical record for the case of “悪性新生物 (*malignant neoplasm*)”. This is because there are special keyphrases used for “悪性新生物 (*malignant neoplasm*)”. For example, ”１年前に子宮癌の手術を受けている。(The patient had an operation last year to abate her *uterine cancer*)”.

The ICD codes define “C” as “悪性新生物 (*malignant neoplasm*)”, and “S” as “損傷 (*damage*)”. Our system covered almost all the ICD codes containing “C” and “S“, including various body parts.

We removed words listed Table 1 from the dictionary used in technique 2.1. These words e.g. “癌 (*cancer*)：C80” are sometimes used to refer specific concepts e.g. “肺癌 (*lung cancer*)：C349”. In this case, we don’t need “癌：C80” but only “肺癌：C349”.

2.5. Inferring ICD codes from XML tags

We suggest another technique that outputs ICD codes using information in XML tags. The dataset of MedNLPDoc contains XML tags, for example:.

<data id="27" sex="FEMALE" age="67">
<text type="既往歴">
</text>
<text type="現病歴">
</text>
<text type="手術">
</text>

where 既往歴, 現病歴, 手術 means anamnesis, clinical history of present illness, and operation respectively.

We focused on two tags: “<text type＝”既往歴”>”and “<text type＝”家族歴”>” because there are categories of ICD codes directly correspond to these two types. Therefore, if there is a tag “<text type＝”既往歴”>”or “<text type＝”家族歴”>” in a given medical record, our system outputs an ICD code by discovering a clue from words inside these tags. For example,

<text type="既往歴">
３７歳、子宮がん手術。
６０歳、高血圧、内服治療中。
</text>

In this case, sentences ”３７歳、子宮癌手術。 (Age 37, an operation to abate *uterine cancer.*)” and ”６０歳、高血圧、内服治療中。 (Age 60, *high blood pressure*, under treatment by *oral administration*)” are the potential clues. After extracting these clues, we apply the same method described in 2.4.

3. EXPERIMENT and RESULT

We conducted two types of experiments. We submitted our result to the MedNLPDoc task and its evaluation was returned.

However, the test dataset used for evaluation is not provided. We conducted another experiment using the training data to show the effectiveness of the methods we suggest, using the tenique2.1 as a baseline to compare with.

Table 2 shows the result of the experiment using the training data. In Table 2, "perfect match" means the number of codes perfectly matched with the correct ICD codes. "3_digits match" means the number of output codes which three digits (a top alphabet in ICD codes and next two numbers) are matched. Total number of correct answers is 772. We compared a couple of different combinations of our sub-methods in Table 2, each described in section 2.1, 2.2, 2.3, 2.4, and 2.5.

	perfect match	3_digits match	Estimated number	P (perfect)	R (perfect)	F (perfect)	P (3_digits)	R (3_digits)	F (3_digits)
2.1	101	161	424	23.82	13.08	16.89	37.97	20.85	26.92
2.1+2.2	110	176	450	24.44	14.25	18.00	39.11	22.80	28.81
2.1+2.2+2.3	116	185	505	22.97	15.03	18.17	36.63	23.96	28.97
2.1+2.2+2.3+2.4	135	232	575	23.48	17.49	20.04	40.35	30.05	34.45
2.1+2.2+2.3+2.4+2.5	145	245	597	24.29	18.78	21.18	41.04	31.74	35.79

Table 2. Differences between combinations of methods in Precision (P), Recall (R) and F-measure (F)

Because the F-measure becomes better when more methods are stacked, each individual method can be regarded as effective. When the method 2.4 is added, the growth of F-measure is the largest. Regarding *malignant neoplasms* and *damage*, we can write coding rules easier by hand because corresponding ICD descriptions explicitly discriminates "body part and *damage*", "body part and the *cancer*", etc. Additionally, *malignant neoplasms* and *damage* are frequently appeared in the training data, which made the contribution larger.

When the method 2.4 is added, the growth of F-measure is the largest. Regarding *malignant neoplasms* and *damage*, we can write coding rules easier by hand because corresponding ICD descriptions explicitly discriminates "body part and *damage*", "body part and the *cancer*", etc. Additionally, *malignant neoplasms* and *damage* are frequently appeared in the training data, which made the contribution larger.

When the method 2.3 is added, the growth of F-measure is the smallest. There are paraphrases of the terminology which are not in Wikipedia. Paraphrases are less in train data in the first place.

4. Comparison with other teams

Figure 3 illustrates comparison between our system's result and others teams' results in the "SURE" evaluation metric of the MedNLPDoc task.

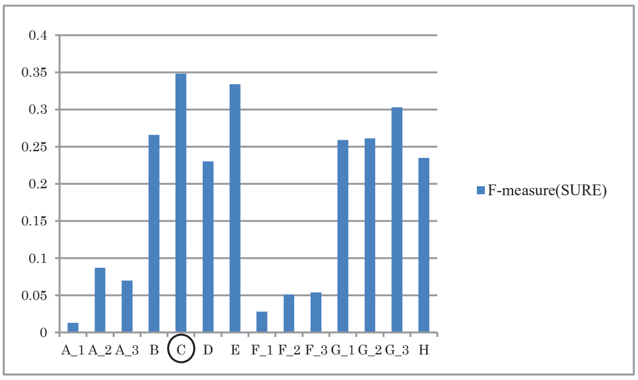

Figure 3. Comparison with other teams in F-measure (SURE), where C indicates our result.

Our team is shown as C. Our system performed the highest score among all participants. The MedNLPDoc dataset was created by three human annotators, all of them annotated on the same documents individually. Evaluation of the MedNLPDoc was performed by comparing participants' system output ICD codes with the correct answer. The "Sure" metric is an evaluation using gold standard data of ICD codes which all of three annotators agreed to annotate. There are two more evaluation metrics, "Major" for ICD codes of more than two annotators agreed, "Possible" for more than one of the annotators agreed. Because the inter-annotator discrepancy is quite low in this dataset, the "Sure" metric is considered as most reliable. Therefore, F-measure (SURE) is the most severe and reliable total evaluation criteria.

5. FUTURE WORK

There should be two decision criterion required to achieve the ultimate goal of this ICD codes assignment study. The first decision is whether symptoms are explicitly described or not in medical records. This decision would have almost been achieved by our approach except for *cancers*. About *cancers*, our system can output ICD codes of *cancers* by our technique described in 2.4 without SCS. If there is a word standing for cancer in a sentence, our system infers which type of *cancer* is the result of diagnosis, then outputs a corresponding ICD code. However, it is uncertain whether there is a *cancer* in the patient's body even if there is a word standing for *cancer* in medical records. We wish to design our system which can detect a *cancer* actually exist or not in future.

The second decision is whether we should output ICD codes or not when we find out symptom or name of disease in medical records. Let us consider *cough* for example. The *cough* often appears in medical records. In order for the code of the *cough* to be assigned, we need to know how strong an effect of the *cough* gives to a patient. We could notice this by deriving relationship of the *cough* and main diagnosis.

If we could properly define the criterion for these two decisions, we can output more accurate ICD codes. Then we can recognize a bad effect to a patient's body by these decisions that could contribute to the real clinical works.

6. CONCLUSION

Medical records contain some features like inclusion of diagnosis names, paraphrases, etc. From such features, we made five rule-based methods that output ICD codes accurately. We discussed contributions of each method in the section of experiment and result. Our system performed best among participants. However, it is still difficult to output ICD codes perfectly. In order to make better ICD coding in future, it will be required to analyze relationship between a patient's symptom and his/her disease.

7. REFERENCES

[1] 鳥羽克子･診療情報管理ネットワーク（編集）『ICD コーディングトレーニング(第２版)』 医学書院

[2] ICD10 国際疾病分類第 10 版（2003 年改訂） http://www.dis.h.u-tokyo.ac.jp/byomei/icd10/

[3] ICD10 対応標準病名マスター http://www2.medis.or.jp/stdcd/byomei/index.html

[4] ライフサイエンス辞書プロジェクト http://lsd-project.jp/ja/download/