NARS: NTCIR-12 MedNLPDoc Baseline

Eiji Aramaki
NAIST
aramaki@is.naist.jp

Shoko Wakamiya
NAIST
wakamiya@is.naist.jp

ABSTRACT

NTCIR-12 MedNLPDoc is a shared task of ICD coding task, which is a multi-labeling task to a patient medical record. This paper describes the baseline system of the task. The system is based on the simple word match with a disease name dictionary without any use of training data. This report presents the results of the baseline system, and discusses the basic feasibility of this system.

Keywords

Medical records, electronic medical records (EMR), named entity recognition (NER), shared task and evaluation

Team Name: NARS

Subtask: Task 1 (Phenotyping task)

1. INTRODUCTION

Medical reports using electronic media are now replacing those of paper media. Correspondingly, the information processing techniques in medical fields have radically increased their importance. In such a situation, the NTCIR-12 MedNLPDoc task is organized. In this task, participants' systems infer disease names in ICD (International Codes for Diseases) from textual medical records. Due to this practical setting, task participants' systems could directly support an actual daily clinical service, also clinical studies in various areas.

The objective of our challenge is to provide a baseline system for NTCIR-12 MedNLPDoc. The system utilizes only a simple word match between an input and the ICD dictionary (mentioned Section 3.1).

2. TASK & MATERIALS

2.1 What is ICD Code

The International Classification of Diseases (ICD) is the standard diagnostic coding system used in many countries for epidemiology, health management and clinical purposes. ICD is used to monitor the incidence and prevalence of diseases and other health problems, proving a picture of the general health situation of countries and populations. The World Health Organization (WHO) within the United Nations System maintains ICD.

In the latest version of the ICD coding system, ICD-10, each ICD code consists of a single alphabet prefix and numbers, which represent a major classification. In addition, more detailed classification can be represented by several digits of additional numbers as suffix, up to six characters in total. Because the major categories are limited to 21 sections, the major categories include a set of similar diseases.

2.2 Task

A training data set of medical records was taken from "ICD Coding Training, Second Edition", written in Japanese for training Health Information Managers (HIMs) [6]. We challenged the Task 1 (**phenotyping task)**, in which we assigned ICD-10 codes to each given medical record and submitted a set of ICD-10 codes.

The inputted medical records are in form of .xml format, as shown in Figure 1 (a). The example of output (the assigned set of ICD-10 codes) is shown in Figure 1 (b).

(a)

```
<data id="68" sex="m" age="49">
<text>
２００４年１２月２～１６日，前回入院.
今回２回目の入院.
前回他院にてアメーバ肝膿瘍の手術予定であったが，術前の検査でＨＩＶ陽性であったため当院入院.
ＳＴＳ陽性.
今回上記の疾患について外来フォロー中であったが吐血で入院.
食道潰瘍が判明.
</text>
```

(b)

```
id=0    B029,B169,F329,J00,K720,K759,L029,R05,R060,R234,R509,R53,R609
id=1    C169,C189,C80,E14,R509,R529,R609,R81
id=2    C169,C80,E049,E785,R21,R229,Z00
id=3    I48,I619,R011,R471,R58,R730,Z00
id=4    E14,G819,I10,I219,I639,R002,R81
id=5    E14,G839,H919,I639,L029,N059,N12,R270,R42,R51,R81
…
```

Figure 1: A medical record (id=68) and a set of ICD-10 codes.

3. METHODS

The proposed method utilizes a simple word match between an input and the ICD dictionary (mentioned Section 3.1).

3.1 ICD Dictionary (MEDIS Standard Masters)

The ICD dictionary (called MEDIS dictionary) consists of pairs of standard disease names and ICD codes. The dictionary can be downloaded from the MEDIS website.

3.2 Algorithm

Given an input text, the system looks up all terms in the ICD dictionary. If the term appears in the text, the system outputs its corresponding codes. The term matching process does not use any extra resources (exact match).

4. Experiments

4.1 Evaluation

Test data set consists of 78 clinical texts, which have three different code sets as follows.

- SURE (S): a sure code set consists of codes that all coders (three persons) utilized.
- MAJOR (M): a major code set consists of codes that two or three coders utilized.
- POSSIBLE (P): a possible code set consists of codes that at least one coder utilized.

The outputs of the baseline system were evaluated through these three types of gold standard data for each code set above. Note that there is a relationship of S⊂M⊂P (SURE is a subset of MAJOR, MAJOR is a subset of POSSIBLE).

Performance of the coding task was assessed using the F-measure (β=1), precision, and recall [8]. Precision is the percentage of correct codes found by the baseline system. Recall is the percentage of codes presents in the corpus that were found by the system. F-measure is the harmonic mean of precision and recall. We employed two matching levels as follows:

- Exact match.
- Rough: Partial match in the first three letter in the code (A00, C16, etc.).

In total, we have three gold standard data sets and two matching methods. Therefore, six types of precision, recall, and F-measure are calculated. For example, SURE and EXACT match results consists of the following three metrics:

- Precision $^{EXACT}_{sure} = |S \cap R| / |R|$
- Recall $^{EXACT}_{sure} = |S \cap R| / |S|$
- F-measure = 2 ・ Precision$^{EXACT}_{sure}$ ・ Recall$^{EXACT}_{sure}$/ (Precision$^{EXACT}_{sure}$ + Recall$^{EXACT}_{sure}$)

4.2 Result

Table 1 presents our results of exact match, evaluated by the methods above mentioned.

Table 1: Results

	Precision	Recall	F
SURE	0.173	0.388	0.235
MAJOR	0.314	0.408	0.354
POSSIBLE	0.370	0.265	0.309

5. CONCLUSION

As shown in Table 1, the overall performance is low (F-measure 0.235-0.354), indicating the difficulty of this task. From the practical viewpoints, we want two types of systems; (1) the high precision system, which suggests only reliable ICD codes, or (2) the high recall system, which presents all possible ICD cords.

Considering that the assignment of ICD-10 codes at actual daily clinical services often varies between each coder (or is sometimes dependent on policies of each medical institution), a system should be a tool that offers clinical coders several possible and optional codes with high accuracy (the latter system).

Unfortunately, the current baseline could not control the performance between the precision and the recall. That is one of the task to be solved in this task.

6. TOOLS

We also present a Windows-based implantation of the baseline system, "ICD viewer - Code-Kun," as shown in Figure 2. That is available at the website (https://sites.google.com/site/mednlpdoc/).

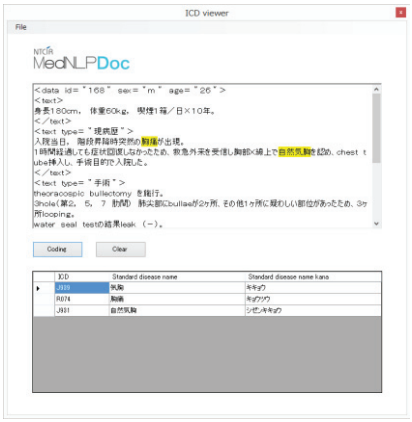

Figure 2: Baseline system named "ICD viewer - Code-Kun"

REFERENCES

[1] Chapman, W.W., Nadkarni, P.M., Hirschman, L., D'Avolio, L.W., Savova, G.K., and Uzuner, O. 2011. Overcoming barriers to NLP for clinical text: the role of shared tasks and the need for additional creative solutions. *J Am Med Inform Assoc*, 18, 540–543.

[2] Ozlem, U. 2008. Second i2b2 workshop on natural language processing challenges for clinical records, in *AMIA Annual Symposium proceedings*. 1252-1253.

[3] Voorhees, E.M. and Hersh, W. 2012. Overview of the TREC 2012 Medical Records Track. in The Twentieth Text REtrieval Conference.

[4] ShARe/CLEF eHealth Evaluation Lab. 2013 [cited 2014/06/04; Available from: https://sites.google.com/site/shareclefehealth/.

[5] Morita, M., Kano, Y., Ohkuma, T., Miyabe M., and Aramaki, E. 2013. Overview of the NTCIR-10 MedNLP task, In *Proceedings of NTCIR-10*.

[6] 鳥羽 克子, ICD コーディングトレーニング, (編集), 診療情報管理東京ネットワーク (編集), 医学書院

[7] Japanese Society of Internal Medicine. 2014. [cited 2014 2014/06/04]; Available from: http://www.naika.or.jp/.

[8] van Rijsbergen, C. J. 1975. *Information Retrieval*. Buttersworth, London.