kyoto: Kyoto University Baseline at the NTCIR-11 MedNLP-2 Task

Tetsuaki Nakamura
Kyoto University

Kay Kubo
Osaka University

Shuko Shikata
Kyoto University

Mai Miyabe
Kyoto University

Eiji Aramaki
Kyoto University

mednlp.office@gmail.com

ABSTRACT
Since more electronic records are now used at medical scenes, the importance of technical development for analyzing such electronically provided information has been increasing significantly. This *NTCIR-11 MedNLP-2 Task* is designed to meet this situation. This task is a shared task that evaluates natural language processing technologies especially on Japanese medical texts. The task has three subtasks: (1) the Extraction task, which is to recognize complaints and diagnoses in medical texts; (2) the Normalization task, which is the ICD-coding task for complaint and diagnosis in the texts; (3) free task. This paper is the report on our results. For the Extraction task, we used a standard named entity recognition technique that is based on conditional random fields. For the normalization task, we used the string similarity between the input term and the MEDIS ICD-10 dictionary. For the free task, we proposed to design a glossary of medical terms for patients. The experimental results in the Extraction task showed reasonably high performance (precision: 77.10%, recall: 17.74%, F-measure: 58.97). However, the results in the Normalization task showed low performance (precision: 33.69%, recall: 33.69%, F-measure: 33.69). Finally, we show an example of the glossary described above as the result of the free task.

Keywords
Natural language processing (NLP), medical informatics, machine learning, conditional random field (CRF)

Team Name
Kyoto University Design School

Subtasks
(1) Task 1 (Extraction task)

(2) Task 2 (Normalization task)

(3) Task 3 (Free task)

1. INTRODUCTION
This paper describes the system that we used for NTCIR11 MedNLP-2. The objective of our challenge is to provide a baseline system for NTCIR-11 MedNLP-2. To do so, we only utilized open software tools and resources, which were provided free of charge. Our methodology also relied on a standard method, which was well known in each field. Due to the simplicity of the method, our system is easy to implement and is easy to re-build for everyone.

In the Task 1, we utilized a standard sequential labeling algorithm, conditional random fields based on word based IOB representation.

In the Task 2, we utilized the edit-distance based similarity between the input term and the MEDIS ICD-10 dictionary.

The experimental results in the Extraction task showed reasonably high performance (precision: 77.10%, recall: 17.74%, F-measure: 58.97). However, the results in the Normalization task showed low performance (precision: 33.69%, recall: 33.69%, F-measure: 33.69).

In this paper, we are showing an example of the glossary described above as the result of our free task.

The rest of this paper is organized as follows: (1) we describe resources and tools of our system in Section 2; (2) the methods in our system are explained in Section 3; (3) our results are described in Section 4; (4) our results and conclusions are discussed in Section 5 and 6.

2. MATERIALS
The corpus for the NTCIR-11 MedNLP contains two types of data: (1) the Dummy Patients' Medical Reports, and (2) the Questions from the actual past state examinations. The question sentences and graphics are eliminated. In this corpus, elements representing time were tagged with <t> </t> (t-tag), and elements representing symptom and diagnosis were tagged with <c> </c> (c-tag). Symptom and Diagnosis (c-tags) also included the ICD-10 codes as their

attributes. Some elements were also marked with modalities (positive, negation, and so on).

3. METHODS

3.1 Task 1: Extraction of Complaint and Diagnosis Task

The named entity recognition is based on conditional random field (CRF). CRF is widely used in various NLP tasks, such as part-of-speech tagging [1], named entity recognition [2], information extraction [4], parsing [5], and so on. Especially, it is often used as a term identifier. In this study, we used CRF++[1] distributed as a tool of CRF.

To utilize CRF, we convert the texts into word based IOB representation with morphological features as follows:

Word (Lexicon)	1-2 gram
Parts of speech	1-3 gram
Inflected forms of word	1-3 gram
Script Types (Hiragana, katakana, Chinese characters, or combinations)	1-3 gram

In order to obtain these features, we used MeCab[2], one of the major Japanese morphological analyzers.

3.2 Task 2: Normalization Task

In order to give ICD-10 codes on complaint and diagnosis, we calculated string similarity between the target words and all the names of diseases in the MEDIS Standard Masters (*Hyoujyun-Byoumei* Master)[3], a dictionary of disease (symptom and diagnosis) names. Each disease name in the MEDIS has an ICD-10 codes.

To calculate the string similarity, we utilized an edit-distance similarity. When two or more terms had the same score, a code was randomly applied among these.

Note that this method does not use the training corpus (unsupervised system).

3.3 Task 3: Free Task

At medical scene, patients often do not know the meaning of the medical terms used by doctors. Patients complain the difficulty of such professional terms, and it is a clamant task to arrange the medical terms into plainer vocabularies that are easier to understand. In 2009, the National Institute for Japanese Language has tried to simplify some medical terms [3]. However, the total number of terms that have been simplified has been only 57, and there are plenty more to work on for improvement.

We, therefore, propose a glossary of medical terms exclusively for patients. In order to produce the glossary, we firstly extracted the ICD-10 codes (appeared twice or more) from the NTCIR-11 corpus. As a result, we extracted 316 medical terms that should be simplified. We, then, selected 98 terms as the entry of the glossary, and have added specific but simplified explanations to each term. The remaining 218 terms are indexed as the related terms of the 98 entries.

4. RESULT

4.1 Task 1

In order to evaluate our system, we calculated precision, recall and f-measure. The results are shown in Table 1.

Table 1. Task 1 Results.

		Precision	Recall	$F_{\beta=1}$
NER (Total)		87.72	62.48	72.98
Modality	Positive	77.65	56.92	65.69
	Family	100.00	28.57	44.44
	Negation	73.73	47.58	57.84
	Suspicion	70.00	12.73	21.54

4.2 Task 2

Task 2 is consisted of two subtasks: (1) to give ICD attributes to the test data of Task 1, and (2) to give ICD attributes to the gold standard data provided after Task 1 submission.

To evaluate the task 2 result, we calculated precision, recall and f-measure. The results are shown in Table 2.

Note that gold standard data values are the same because the system gave ICD codes to all entities.

Table 2. Task 2 Results.

	Precision	Recall	$F_{\beta=1}$
For the test data of Task 1	31.94 %	19.96 %	24.57
For the gold standard data	33.69 %	33.69 %	33.69

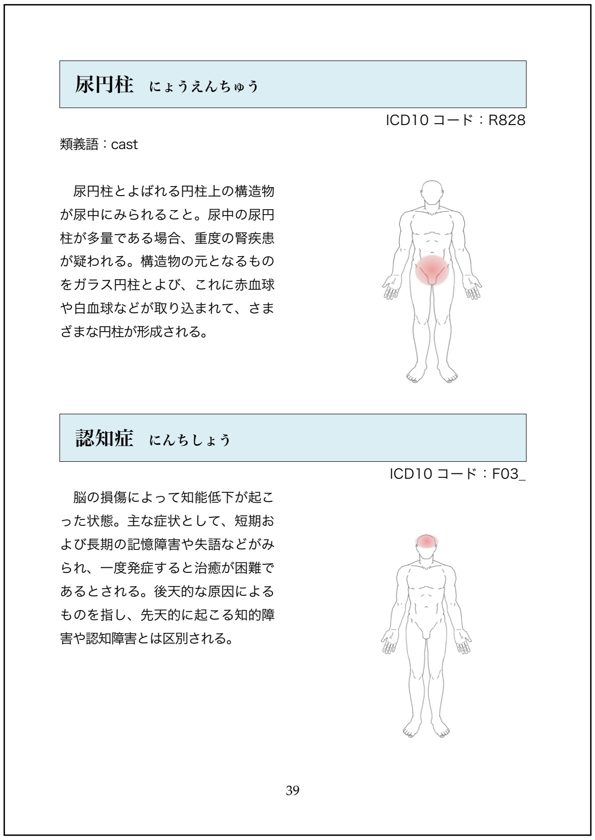
4.3 Task 3

An example of the glossary we made is shown in Fig.1. As described in chapter 3.3, this glossary is to promote the patients to fully comprehend doctors' words to know more precisely about their disease. For this reason, we paid much attention using as plain Japanese vocabularies as possible. We also appended a human body figure to each explanation. The each high-lightened circle on the figure shows the related body part of the entry. This glossary will be distributed on the web site (http://mednlp.jp).

[1] http://crfpp.googlecode.com/svn/trunk/doc/index.html

[2] http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html

[3] http://www2.medis.or.jp/stdcd/byomei/index.html



尿円柱　にょうえんちゅう

ICD10 コード：R828

類義語：cast

尿円柱とよばれる円柱上の構造物が尿中にみられること。尿中の尿円柱が多量である場合、重度の腎疾患が疑われる。構造物の元となるものをガラス円柱とよび、これに赤血球や白血球などが取り込まれて、さまざまな円柱が形成される。

認知症　にんちしょう

ICD10 コード：F03_

脳の損傷によって知能低下が起こった状態。主な症状として、短期および長期の記憶障害や失語などがみられ、一度発症すると治癒が困難であるとされる。後天的な原因によるものを指し、先天的に起こる知的障害や認知障害とは区別される。

39

Figure 1. An Example of the Glossary

5. DISCUSSION

In task 1, the result in NER evaluation is 72.98 ($F_{\beta=1}$). Considering the baseline does not utilize the extra resource, this value is almost the result of pure machine learning. This score is a benchmark for contribution of extra rehouses.

In task 2, the performance is quite low, which indicates the gap between MEDIS terminology and medical daily expression. Because our approach did not rely on any other machine learning techniques, this result could be the benchmark for machine leaning approaches.

In task 3, we believe various applications could be build using this resource.

6. CONCLUSION

We developed a baseline system to extract complaints and diagnoses from medical texts and a system to give ICD-10 codes on complaint and diagnosis in the texts. We also developed a glossary of medical terms for patients. The proposed system consists of open software and tools that enable rapid implementation. We believe that this result and system might be a benchmark to other systems.

7. ACKNOWLEDGMENTS

This study was supported in part by JST PRESTO. T.N. designed and performed experiments, analyzed data and wrote the paper (Task 1 and 2); K.K. designed and wrote the paper (Task 3); S.S. analyzed data; E.A. and M.M made our baseline system; E.A. supervised the project.

8. REFERENCES

[1] R. Dearden. Structured prioritised sweeping. In *Proceedings of the Eighteenth International Conference on Machine Learning, ICML '01*, pages 82–89, San Francisco, CA, USA, 2001. Morgan Kaufmann Publishers Inc.

[2] A. McCallum and W. Li. Early results for named entity recognition with conditional random fields, feature induction and web-enhanced lexicons. In *Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003 - Volume 4, CONLL '03*, pages 188–191, Stroudsburg, PA, USA, 2003. Association for Computational Linguistics.

[3] National Institute for Japanese and Linguistics, A proposal for the simplification of hospital language (in Japanese: Byouin no kotoba wo wakariyasuku suru teian), http://www.ninjal.ac.jp/byoin/

[4] D. Pinto, A. McCallum, X. Lee, and W. Croft. Table extraction using conditional random fields. In *Proceedings of the 26th ACM SIGIR*, pages 235–242, 2003.

[5] F. Sha and F. Pereira. Shallow parsing with conditional random fields. In *Proceedings of the Human Language Technology conference and the North American chapter of the Association for Computational Linguistics (HLT-NAACL2003) Workshop on Building and Using Parallel Texts: Data Driven Machine Translation and Beyond*, pages 213–220, 2003.