A Method to Detect Errors in Electronic Discharge Summaries Based on Named Entity Recognition

D.S. Yuan
EMR and Intelligent Expert System Engineering Research Centre
Ministry of Education
Innovation Joint Research Centerfor Cyber-Physical-Society System
College of Biomedical Engineering and Instrument Science
Zhejiang University
Hangzhou, China

Y. Tian
EMR and Intelligent Expert System Engineering Research Centre, Ministry of Education
Innovation Joint Research Centerfor Cyber-Physical-Society System
College of Biomedical Engineering and Instrument Science
Zhejiang University
Hangzhou, China

T.S. Zhou
EMR and Intelligent Expert System Engineering Research Centre
Ministry of Education
Innovation Joint Research Centerfor Cyber-Physical-Society System
College of Biomedical Engineering and Instrument Science
Zhejiang University
Hangzhou, China

J.S. Li
EMR and Intelligent Expert System Engineering Research Centre
Ministry of Education
Innovation Joint Research Centerfor Cyber-Physical-Society System
College of Biomedical Engineering and Instrument Science
Zhejiang University
Hangzhou, China

***Abstract*—The hospital discharge summaryis an essential document, containing clinical and administrative information necessary for the continuity of care after the patients are discharged from hospital. The utilization of electronic discharge summaries has grown in popularity. However, many transcription errors and spelling mistakes exit, potentially reducing the medical quality of patient care. To solve this problem, this paper presents a novel approach to detect these errors automatically by using Named entity recognition (NER). The NER model was trained by 450 discharge summaries and rich features set was used to improve the recall and precision. Experiment on the independent test set validated the good performance of NER. The follow-up error detection using the trained NER discovered that the mistakes and ambiguous information that frequently occurred in discharge summaries.**

Keywords-named entity recognition; natural language processing; electronic discharge summary; error detectio; conditional random field

I. Introduction

The hospital discharge summary, containing clinical and administrative information necessary for the continuity of care after the patients are discharged from hospital, is an essential document for general practitioners (GPs) to communicate the condition of the corresponding patient.

Traditionally, discharge summaries are paper-based with deficiencies long being recognized as limited interoperability, clerical error, etc. With the development and widespread use of clinical information systems, the utilization of electronic discharge summaries is on the increase. Although being considered to have many advantages over paper-based ancestors, the use of electronic discharge summaries in clinical practice is still limited, and studies have shown that a higher number of errors and/or omissions in the electronic discharge summaries than in the handwritten ones[1].

Traditional data entry in electronic discharge summaries is carried out by manually typing data into computer by medical staffs, inevitably resulting in, manual transcription errors or spelling mistakes which will make the discharge summaries inadequate, potentially leading to serious harm to patients[2]. For instance, errors associated with manual transcription of medications from medication charts to discharge summaries could result in drug misuse or drug overdose, which could be fatal to discharged patients.

To Date, methods proposed to reduce discharge summary errors include automatic prevention with advanced information technology and manual correction by hospital clinical staff[3]. The former involves integrating clinical decision support modules into the original hospital information systems, which requires a huge workload in China, considering the heterogeneity between the systems. Besides, it can’t guarantee the correction of discharge summaries and lacks measures for detection. Similarly, the manual intervention for correction will cost a lot of time and money and the result is often not satisfying. Hence, an efficient automatic detection to reduce discharge summary errors is needed for the safe handover of care to the primary care provider.

A new method to detect errors in Chinese discharge summary is proposed in this paper. The method will warn the medical staff of potential errors in this document. Considering that discharge summaries are narrative and the complexity of Chinese text processing, we combines machine learning-based named entity recognition (NER) method with clinical terms dictionary and structured data related to the discharge summary, to ensure the high detection and accuracy. In this paper, we will take detecting mistakes about discharge medications and diagnosis results as an example to introduce how this method works.

II. MATERIALS AND METHODS

A. *NER in Electronic Discharge Summaries*

Before the detection of discharge summaries, it's supposed to extract the information of prime importance from them. Due to their unstructured characteristic, Natural language processing (NLP) is necessary. NER in clinical text, which is used to identify clinically relevant entities such as diseases and drugs, can help us find all the clinical entities to be detected.

NER includes traditional rule-based approaches and more advanced machine learning-based approaches in recent years. Rule-based approaches cost great amounts of time and effort, and is prone to mistakes. Besides, rule-based approaches completely depend on linguistic style and writing habit specific of the author, which often results in bad portability. Hence, in this paper, the machine learning-based approaches are applied to implement the NER.

B. *Datasets and Annotation*

Four months (August 2013 to November 2013) of discharge summaries were collected from the electronic health record (EHR) database of some hospitals in Zhejiang, China. After excluding some incomplete notes, we randomly selected 600 discharge summaries for this study, of which 450 notes are for training and others are for testing. All patient health information (PHI) was manually removed before annotation. Four types of clinical entity—disease, test, medication and procedure—were annotated manually according to a predefined guideline. We used the common "BIO" tags, in which "B" represents the beginning of an entity, "I" represents the content of an entity and "O" represents the outside of an entity. An example of annotated entities labeled with "BIO" tags is shown in Figure 1.

Sentence	BIO Tags
入院后告病危，完善血气分析、PCT、痰培养、血常规、血生化等辅助检查。(After admission with critical ill, blood gas analysis, PCT, sputum culture, blood routine, blood biochemical and other auxiliary examinations were conducted.)	入院/O/后/O/告/O/病危/O，/O完善/O血/B-test气/I-test分析/I-test、/O PCT/B-test、/O痰/B-test培养/I-test、/O血常规/B-test、/O血/B-test生化/I-test等/O辅助/O检查/O。/O

FIGURE I. AN EXAMPLE OF ANNOTATED ENTITIES LABELED WITH "BIO" TAGS.

C. *Machine Learning-Based Algorithm*

NER problems can be considered as either a classification problem, such as Support Vector Machine (SVM), or a sequence-labeling problem, such as Hidden Markov Model (HMM), Maximum Entropy Markov Model (MEMM) and Condition Random Field (CRF)[4]. Research has shown that the sequence-labeling algorithms outperform the classification ones. Among these sequence-labeling algorithms, CRF is the most widely used algorithm for NER, which achieves global normalization and performs well even with a small-scale training corpus. In this study, we used CRF algorithm to implement NER.

When using CRF, we regard every sentence as an observation sequence. Given an sequence of observed words of length n :$X = (x_1, x_2, \cdots\cdots, x_n)$, let S be a set of states in a finite state machine, each corresponding to a label $l \in L$ (e.g. disease, medication, etc.). Let $Y = (y_1, y_2, \cdots\cdots, y_n)$ be the sequence of states in S that correspond to the labels assigned to words in the input sequence X. CRF uses an undirected graph to model the conditional distribution of status Y conditioned on observations X, and the conditional probability $p(y|x)$can be represented by a first-order Markov chain in the following form:

$$P(y|x) = \frac{1}{Z(x)} \exp\left(\sum_i \sum_k \lambda_k f_k(y_{i-1}, y_i, x) + \sum_i \sum_k \mu_k g_k(y_i, x)\right) \quad (1)$$

where$f_k(y_{i-1}, y_i, x)$ and $g_k(y_i, x)$ are both one of k functions that describes a feature, in which λ_k and μ_k are their weight factors. Besides, $Z(x)$ is a normalization factor of all state sequences:

$$Z(x) = \sum_y \exp\left(\sum_i \sum_k \lambda_k f_k(y_{i-1}, y_i, x) + \sum_i \sum_k \mu_k g_k(y_i, x)\right) \quad (2)$$

After being trained by annotated corpus, a weight value can be learned for each such feature function to maximize the conditional log likelihood of labeled sequences in a training set $D = \{(x, l)_1, (x, l)_2, \cdots\cdots, (x, l)_n\}$ [5]. The maximum log-likelihood estimation function can be written as:

$$L(D) = -\mathrm{Log}(\prod_{k=1}^{N} p(y^k | x^k, w) \ . \quad (3)$$

In this study, we used Stanford Named Entity Recognition[6] to optimize equation (3).

D. *Feature Set*

In addition to the algorithm, finding an appropriate feature set also matters. In our study, four types of feature were used: (1) single character; (2) word; (3) part-of-speech (POS) tags; (4) orthography. The single character means individual Chinese characters, punctuation, foreign letters and numbers, which constitute the clinical document. The major characteristic of Chinese narrative text is that it's formed with continuous Chinese characters without any space. Because of this, we also chose word as another feature and we used the Stanford Word Segmenter[7]to implement segmentation. While processing, POS tags were generated as well by the software. Choosing POS tags as the machine learning feature has a great effect on the recognition of Chinese entities' boundary. For example, the verbal word "diagnose" or "conduct" usually indicates the left boundary of clinical entity. Feature orthography was used to recognize the entities consisting of English letters, such as some laboratory tests.

E. Error Detection

After applying NER on the discharge summary, we can evaluate the note by error detection. The prerequisite for composing a quality discharge summary is the provision of an adequate summary of the patient's hospital stay and details of any aftercare required. Hence, during the error detection, we try to find the missing or ambiguous information and mistakes. The procedures of error detection are shown in Fig. 2. When all entities are stored, the first step is to confirm whether these kinds of entities cover all the information that the note is supposed to convey. Then we detect entities in different ways according to their category. The category such as disease, is compared to the dictionary that consists of terms in ICD-10(The International Classification of Disease, 10th Revision). The comparison will tell whether the entity conforms to the specification. For example, when a doctor is writing the discharge summary of a patient diagnosed with type 1 diabetes mellitus, it won't convey the accurate information about the patient's health condition if he just types the word "diabetes". Besides, the category such as medication, is compared to the structured data related to the discharge summary like laboratory test report. In this way, we can find clearly whether there are any omissions or mistakes in this kind of entities.

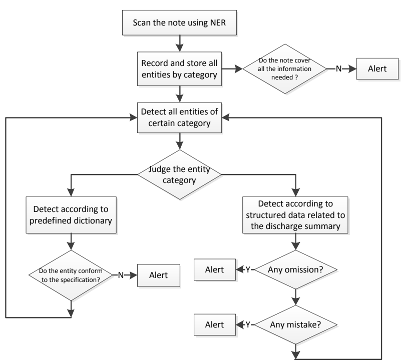

FIGURE II. ERROR DETECTION WORKFLOW

III. RESULTS

There were 27019 sentences and 361779 characters in the selected 600 discharge summaries. The proportion of each type of entities in 600 notes was 14.20% for medications, 63.69%for tests, 15.40% for procedures, and 6.71% for diseases. We divided these notes into two subsets randomly, one-quarter for testing and three-quarters for training.

After training the CRF-based NER classifier with the training set, a NER model was serialized for recognizing clinical entities in the discharge summary. The graphical user interface (GUI) that displays the NER processing on a discharge summary, is shown in Fig. 3. An existing text file can be opened as input file, which was segmented. Then we loaded the trained NER model as NER classifier and ran it for identification. As a result, the entities in the notes were highlighted with different colors according to different labels, and each type of entities was collected for statistics.

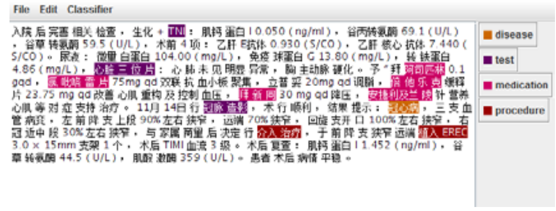

FIGURE III. THE GRAPHICAL USER INTERFACE (GUI) OF NER

Table 1 shows the performance of the CRF-based classifier on testing dataset. The numbers in the table represent F-measures followed by corresponding recall and precision values in parentheses. The overall F-measure is 88.32%,which is good enough to meet the requirement for the follow-up error detection. For each type of entity, precision is always higher than recall. The results for Tests and diseases are the best, of which F-measures are both above 90%,while the result for procedures is the worst.

TABLE I. PERFORMANCE OF CRF-BASED NER

Entity	**F-measure(recall / precision)**
Overall	88.32(89.01 / 87.53)
Medications	86.59(87.17 / 85.42)
Tests	90.92(91.27 / 90.15)
Procedures	77.39(81.97 / 74.03)
Diseases	91.13(91.29 / 90.11)

After we confirmed the good performance of CRF-based NER, we randomly selected 500 discharge summaries from the dataset which were used for NER model training and testing, of which 9 were not used because their discharge status are dead , leaving 491 for study. In this experiment, we detected mistakes about discharge medications and diagnosis results. The result of automatic error detection on these 491 discharge summaries is shown in Table 2. By error type, column 2 presents the numbers of discharge summaries with the particular error and column 3 shows the numbers of the certain error existing in all 491 notes. Medication omission is the most common error. The average number of each type of error per discharge summary is also presented in Table 2. The highest average number of errors per discharge summary is related to medication omission.

TABLE II. THE RESULT OF ERROR DETECTION IN DISCHARGE SUMMARIES BY ERROR TYPE

Error Type	**Number of discharge summaries with the particular error N= 491 N (%)**	**Number of errors**	**Average number of errors per discharge summary with error**
Medication omitted	33(6.72)	49	0.61
Medication different	26(5.30)	33	0.41
Additional medication listed	12(2.44)	16	0.20
Diagnosis omitted	3(0.61)	3	0.04
Diagnosis ambiguous	26(5.30)	31	0.39

After we confirmed the good performance of CRF-based NER, we randomly selected 500 discharge summaries from the dataset which were used for NER model training and testing, of which 9 were not used because their discharge status are dead , leaving 491 for study. In this experiment, we detected mistakes about discharge medications and diagnosis results. The result of automatic error detection on these 491 discharge summaries is shown in Table 2. By error type, column 2 presents the numbers of discharge summaries with the particular error and column 3 shows the numbers of the certain error existing in all 491 notes. Medication omission is the most common error. The average number of each type of error per discharge summary is also presented in Table 2. The highest average number of errors per discharge summary is related to medication omission.

IV. DISCUSSION AND CONCLUSION

In this paper, we present a novel method to detect potential errors in discharge summaries automatically. The application of CRF-based NER successfully handled the complexity of Chinese text processing. With our algorithm and features set, F-measure has reached 88.32% after being trained with only a small-scale corpus. However, the recall and precision for procedures is unsatisfactory, mainly because the average character length of this kind of entities is too long to be recognized as a whole by the context window. Besides, word segmentation is one of the most important issues for Chinese text processing. In future study, domain knowledge dictionaries can be applied to word segmentation.

The result of error detection indicates that the writing of discharge summaries still lacks standardization. Medication omissions are common errors due to the transcription mistakes by typing. In addition, other errors in discharge summaries, such as ambiguous diagnosis also affect the accurate communication of patient’s health condition. Therefore, error detection in discharge summaries is essential for an adequate summary of the patient’s hospital stay and details of required aftercare. The method we proposed can improve the quality of discharge summaries efficiently.

ACKNOWLEDGMENTS

This work was supported by the China National Natural Science Foundation (Grant No. 61173127), National High-Tech R&D Program (No. 2013AA041201), and the Fundamental Research Funds for the Central Universities.

REFERENCES

[1] Callen JL, Alderton M, McIntosh J (2008) Evaluation of electronic discharge summaries: A comparison of documentation in electronic and handwritten discharge summaries. International Journal Of Medical Informatics 77 (9):613-620. doi:10.1016/j.ijmedinf.2007.12.002.

[2] McMillan TE, Allan W, Black PN (2006) Accuracy of information on medicines in hospital discharge summaries. Internal Medicine Journal 36 (4):221-225. doi:10.1111/j.1445-5994.2006.01028.x.

[3] Callen J, McIntosh J, Li J (2010) Accuracy of medication documentation in hospital discharge summaries: A retrospective analysis of medication transcription errors in manual and electronic discharge summaries. International Journal Of Medical Informatics 79 (1):58-64. doi:10.1016/j.ijmedinf.2009.09.002.

[4] Lei JB, Tang BZ, Lu XQ, Gao KH, Jiang M, Xu H (2014) A comprehensive study of named entity recognition in Chinese clinical text. Journal Of the American Medical Informatics Association 21 (5):808-814. doi:10.1136/amiajnl-2013-002381.

[5] Settles B (2004) Biomedical named entity recognition using conditional random fields and rich feature sets. Paper presented at the Proceedings of the International Joint Workshop on Natural Language Processing in Biomedicine and its Applications, Geneva, Switzerland.

[6] Stanford Named Entity Recognition, http://nlp.stanford.edu/software/CRF-NER.shtml.

[7] Stanford Word Segmenter, http://nlp.stanford.edu/software/segmenter.shtml.