Application of MedLEE to Process Medical Text Reports in Taiwan

Mei-Yi Lin[a], Jen-Hsiang Chuang[a*], Der-Ming Liou[a], Chen-Huan Chen[b]

[a] Institute of Health Informatics and Decision Making, National Yang-Ming University, Taipei, Taiwan

[b] Faculty of Medicine, School of Medicine, National Yang-Ming University, Taipei, Taiwan

[*] The corresponding author

chuang@ym.edu.tw

Abstract

Computerized clinical information is extensively stored as free-text reports. However, information stored in text reports is not computable and it is difficult to access these data for clinical, teaching, or research purpose. Medical language processing (MLP) offers a good opportunity to extend the use of the electronic medical record (EMR) by converting narrative text into coded data. The purpose of this study was to apply the MedLEE program, which was an MLP system developed at Columbia University, to automate extraction of clinical information from medical reports produced at Taipei Veterans General Hospital (VGH). In this study, we used MedLEE to develop a system to determine the patients with congestive heart failure (CHF) from discharge summaries, chest x-ray reports and nuclear medicine reports. We used 50 cases to train the system for adding terms to the lexicon and developing the inference rules. Then we used another independent set of 300 patients to test the system to compare with the administrative data (ICD-9 codes). The recall, specificity, and precision for the automated abstraction system/ICD-9 codes in the testing set were 0.74/0.52, 0.99/0.99, and 0.97/0.97, respectively. We conclude that although the MedLEE program was developed in USA, it is feasible to apply it to process and analyze medical text reports produced in Taiwan with good performance.

Keywords: medical language processing, congestive heart failure

Introduction

With the coming of information revolution, the computerized hospital information systems have been extensively used. Several large clinical databases are increasingly established. Clinical databases provide the large amount of data which are generated from the routine process of care. They are unstructured texts which are usually stored as narrative reports. Although clinical databases provide a rich and convenient source of clinical data, most of them are difficult to use for decision support or data analysis.

Natural language processing in the medical domain offers a good opportunity to extend the use of the electronic medical record (EMR) by converting narrative text into coded data[1]. Many medical language systems have been established, such as LSP system[2], Canonical Phrase Identification System (CAPIS)[3], SymText[4], MedLEE[5;6], etc. However, few NLPs in the medical domain have been generally implemented in Taiwan[7]. Traditionally, researchers use the available administrative data (i.e., ICD-9 codes) to obtain the clinical information and estimate the trend of the diseases. However, ICD-9 codes tend to underestimate to the prevalence of the diseases compared to medical reports. So, we hope to apply the MLP technique to enhance the performance of extracting medical reports in this thesis.

At Columbia Presbyterian Medical Center, an MLP system in the clinical domain, called MedLEE (Medical Language Extraction and Encoding System), has been developed and tested in several studies[8-13]. Therefore, we would like to apply the MedLEE program to automatically extract the clinical information from the medical reports produced in Taiwan. Furthermore, congestive heart failure (CHF) is a major problem in public health. Nearly 5 million Americans are living with heart failure, and 550,000 new cases are diagnosed each year in USA[14]. In Taiwan, the research also revealed that patients with progressive chronic disease (congestive heart failure, chronic pulmonary disease, cancer) had a higher unplanned readmission rate than those with chronic impairment following an acute episode (stroke, traumatic brain injury, hip fracture)[15].

Due to the reasons described above, we hope to utilize the computer to automatically detect whether a patient has CHF to efficiently manage such patients. At the same time, it may enhance the quality of medical care. The purpose of this study was to apply the MedLEE program to develop an automated abstraction system for detecting CHF from medical text reports produced in Taiwan.

Methods

In our study, the data sources included discharge summaries, chest x-ray reports and nuclear medicine reports. After we excluded the patients without ICD-9 codes in their administrative data, we randomly selected 50 patients as a training set and 300 patients as a testing set from the reports during a three-month period (discharged between July 1, 2003 and September 30, 2003) at Taipei Veterans General Hospital (VGH).

Research Process

The research process overview is shown in Figure 1.

Figure 1 shows the overall process which consists of the medical language processing system (MedLEE) and a set of inference rules to process the structured output generated by MedLEE for abstracting clinical information from free-text reports. We could expand the lexicon and develop the inference rules by the process of parsing the training set of medical records. Afterward, a set of inference rules were developed to abstract the clinical information that were relevant to congestive heart failure. Finally, the administrative data (ICD-9 codes) were used to compare with the automated abstraction system in the aspect of finding the patients with CHF.

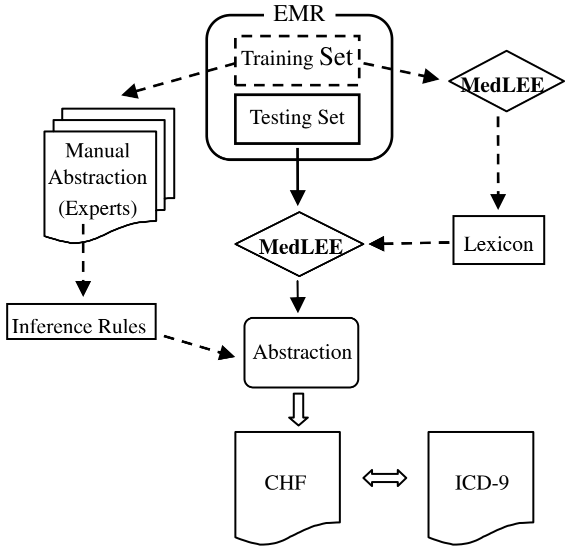

Figure 1 Research Process Overview

Automated Abstraction System

The system architecture overview is shown in Figure 2. We applied MedLEE to build the automated abstraction system to process medical reports produced in Taiwan. It needed a series of steps from pre-processing to post-processing. The system mainly consisted of an MLP system (MedLEE) and a set of inference rules which processed the structured output generated by the MLP system. In the first step which built the automated abstraction system, we must make the action called “pre-parsing”. Pre-parsing would transform the original reports into the suitable format that was acceptable by MedLEE. The structured data generated by MedLEE were the XML format. In order to facilitate to automatically abstract information, we transformed the XML output to Oracle database and stored the inference rules as the stored procedure.

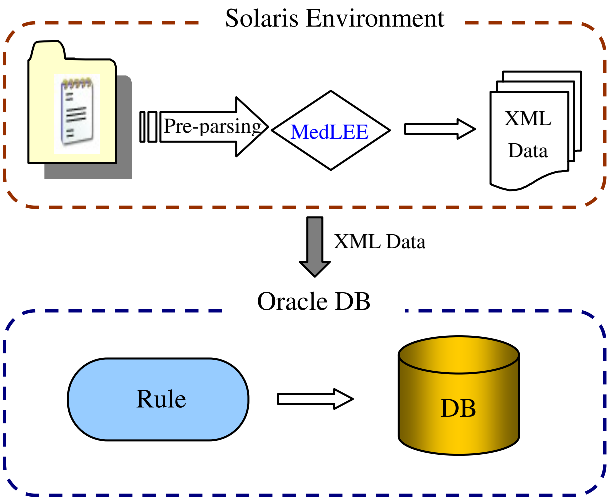

Figure 2 System Architecture

The process of developing an automated abstraction system for a new domain still required expanding the lexicon and constructing another set inference rules to abstract clinical information of our interest.

Although MedLEE had the rich terms in lexicon, adding new terms to the lexicon can help researches find the patterns of interest in the face of a new domain. When the terms were not included in the lexicon, there were two approaches to expand the lexicon: adding new terms into the lexicon or writing a program to pre-tag these terms before being parsed by MedLEE. As we know, expanding the lexicon was an important process. In order to implement it, we must adequately realize the MedLEE output that would be introduced in next section.

Overview of the MLP Output

MedLEE is composed of several different modules where each module processes and transforms the text reports to the structured output. The structured output consists of a series of primary findings (i.e., **finding**, **problem**, **procedure**, **med**, **device**, **bodymeas**, etc), along with its corresponding modifiers (i.e., **bodyloc**, **certainty**, **status**, **quantity**, **degree**, **change**, etc). Besides, the output is the XML format. Extensible Markup Language (XML) is a simple, very flexible text format. The following Figure 3 shows an example of a simplified version of structured output that MedLEE parsed the sentence "increased infiltration in bilateral lung fields".

```
<problem v = "infiltrate">
   <bodyloc v = "lung"><region v = "bilateral">
</region></bodyloc>
   <change v = "increase"></change>
</problem>
```

Figure 3 Sample Output by MesdLEE

Constructing the Inference Rules

Constructing the inference rules is a complicated task. Before constructing the inference rules, we must understand the organization of the original text reports and how the MLP system represents the information. Negated findings are a worth noticing problem when constructing the rules. Before constructing the rules, we should understand the influence of negated finding and further point out the location. The following is a simplest logic of inference rules that detect CHF:

if problem is in ("congestive heart failure";" congestive cardiac failure")

and certainty-modifier is not in ("no"; "rule out"; "borderline certainty")

and status-modifier is not in ("resolved")

then

CHF=1;

Evaluation

In our study, we measured reliability and validity. Reliability can be measured as the percentage of agreement among the experts who generated the gold standard. Reliability is a measure of gold standard quality. And the kappa statistic was computed. As regards the validity, we calculated recall (sensitivity), precision (positive predictive value), and specificity with their respective 95 percent confidence intervals (95%CI) for the inference on congestive heart failure.

Results

The Kappa value of determining CHF from two raters who constitute the gold standard is 0.87. Recall, precision, and specificity of inferring CHF for each subject are presented with their respective 95 percent confidence intervals in Table 1.

Table 1 Performance Measure of the CHF Inference for All Subjects

	MD1	MD2	System	ICD-9
Inference – Congestive heart failure (N=103)				
Recall	0.97 (0.94-1.00)	0.87 (0.81-0.93)	0.74 (0.66-0.83)	0.52 (0.42-0.62)
Precision	1	0.99 (0.97-1.00)	0.97 (0.94-1.00)	0.97 (0.94-1.00)
Specificity	1	0.99 (0.97-1.00)	0.99 (0.97-1.00)	0.99 (0.97-1.00)

For the disease inference in Congestive heart failure, the MD1 had recall of 97 percent, precision of 100 percent and specificity of 100 percent. The MD2 had recall of 87 percent, precision of 99 percent and specificity of 99 percent. The automated system had recall of 74 percent, precision of 97 percent and specificity of 99 percent. The administrative data had a lower performance with 52 percent recall, but higher precision and specificity. The performance of the system and the ICD-9 code was lower than the experts. However the system is not statistically different from one of the experts.

Discussion

Because it is difficult to define CHF, the diagnosis of CHF is a challenging job. With regard to extracting CHF from medical text reports, it is more difficult than extracting other diseases. We studied the ability of the automated abstraction system by a rule-based algorithm to identify the patients with CHF. The system is superior to applying the administrative data (ICD-9 code), but some issues were still worthy to discuss. We grouped an analysis of the errors that influenced the validity into four categories: inadequacies of the report format, MedLEE parsing errors, inference rules errors, and errors in the gold standard.

- The inadequacy of the report format means that a word is separated into two strings in the end of each line. In such situation, the MLP system cannot parse the term correctly (Figure . Quality of EMR could reduce the performance of the automated abstraction system.

> The 77 year-old female patient is a case of primary hypothyroidi
> sm and polycystic kidney disease with ESRD diagnosed about 10 ye
> ars ago and under regular HD on W2,4,6. She had been admitted to
> hospital due to gross hematuria and CHF with acute pulmonary ed
> ema. Ejection fraction (92/03) showed LVEF/RVEF: 35%/34%.

Figure 4 the report format causing lower performance

- Parsing errors mean that MedLEE cannot process the texts and detect the conditions correctly. Spelling error and the informal syntax caused the parsing errors. In the medical reports, spelling error was the serious problem. For example, the word “cardiomegaly” had four incorrect spellings which were discovered in the reports: “cardiomegally”, “cadiomegaly”, “cardimegaly”, and “cardiomealy”. In addition, the informal syntax, such as “*Ejection fraction (92/03) showed LVEF/RVEF: 35%/34%*”, could not be parsed by the system. It might lose the valuable information.
- Inference-rule errors resulting from the rules are unable to cover all of the relevant terms. Because the case number of the training set might not be enough, some terms were uncovered in advance. In future studies, we will apply the refined inference rules to improve the system performance for studying CHF. For example, “impaired LV function” and “LV dysfunction” were missed from the rules for inferring CHF.
- Another kind of errors results from the gold standard itself. In some cases, the automated abstraction

system adhered to the criteria, but the inferring results were different from the gold standard. The errors were due to human error in manual abstraction, so gold standard lost some findings that the system regarded them as the presence. The kind of error would cause the lower precision.

In addition, we detected a big difference concerning the usage of the term “R/O”. In general, the abbreviation “R/O” means “rule out” in medical reports. It is a negative meaning. For the reason, our inference rules regarded “R/O” as the negative meaning. But at the department of cardiology in Taipei Veterans General Hospital, the term “R/O” is not the absolutely negative meaning. The abbreviation has the doubtful meaning. Such difference in the expression of “R/O” would reduce the performance of the automated abstraction system.

Conclusions

Although the MedLEE program was developed in USA, it is feasible to apply it to process and analyze medical text reports produced in Taiwan. The MLP system extended the use of the EMR and further obtained more valuable information. In the future, the MLP system will be extensively applied to other domains.

Acknowledgements

This study was supported in part by grant NSC 92-2320-B-010-058 from National Science Council. We would like to thank Dr. Carol Friedman for providing MedLEE program and adding new terms to the lexicon. We would like to thank Dr. Po-Hsun Huang and Dr. Hao-Min Cheng reading the reports.

Reference

[1] Spyns P. Natural language processing in medicine: an overview. Methods Inf Med 1996; 35(4-5):285-301.

[2] Sager N, Lyman M, Bucknall C, Nhan N, Tick LJ. Natural language processing and the representation of clinical data. J Am Med Inform Assoc 1994; 1(2):142-160.

[3] Lin R, Lenert L, Middleton B, Shiffman S. A free-text processing system to capture physical findings: Canonical Phrase Identification System (CAPIS). Proc Annu Symp Comput Appl Med Care 1991;843-847.

[4] Haug PJ, Koehler S, Lau LM, Wang P, Rocha R, Huff SM. Experience with a mixed semantic/syntactic parser. Proc Annu Symp Comput Appl Med Care 1995;284-288.

[5] Friedman C, Alderson PO, Austin JH, Cimino JJ, Johnson SB. A general natural-language text processor for clinical radiology. J Am Med Inform Assoc 1994; 1(2):161-174.

[6] Friedman C, Hripcsak G, Shagina L, Liu H. Representing information in patient reports using natural language processing and the extensible markup language. J Am Med Inform Assoc 1999; 6(1):76-87.

[7] Hsiang-yu Yuan, Jou-wei Lin, Jau-min Wong. Extracting clinical information from colonscopy finding reports. Medical Informatics Symposium in Taiwan 2002.

[8] Chuang JH, Friedman C, Hripcsak G. A comparison of the Charlson comorbidities derived from medical language processing and administrative data. Proc AMIA Symp 2002;160-164.

[9] Extracting Information on Pneumonia in Infants Using Natural Language Processing of Radiology Reports.: 2003.

[10] Chuang JH. Automated Abstraction of Medical Records for Assessing Patient Outcomes. New York : Columbia University, 2003.

[11] Hripcsak G, Friedman C, Alderson PO, DuMouchel W, Johnson SB, Clayton PD. Unlocking clinical data from narrative reports: a study of natural language processing. Ann Intern Med 1995; 122(9):681-688.

[12] Friedman C, Knirsch C, Shagina L, Hripcsak G. Automating a severity score guideline for community-acquired pneumonia employing medical language processing of discharge summaries. Proc AMIA Symp 1999;256-260.

[13] Hripcsak G, Kuperman GJ, Friedman C, Heitjan DF. A reliability study for evaluating information extraction from radiology reports. J Am Med Inform Assoc 1999; 6(2):143-150.

[14] American Heart Association. 2004 2004.

[15] Dai YT, Wu SC, Weng R. Unplanned hospital readmission and its predictors in patients with chronic conditions. J Formos Med Assoc 2002; 101(11):779-785.