journal homepage: www.intl.elsevierhealth.com/journals/ijmi

Multilingual chief complaint classification for syndromic surveillance: An experiment with Chinese chief complaints

Hsin-Min Lu[a,*], Hsinchun Chen[a], Daniel Zeng[a,d], Chwan-Chuen King[b], Fuh-Yuan Shih[c], Tsung-Shu Wu[b], Jin-Yi Hsiao[b]

[a] Management Information Systems Department, Eller College of Management, University of Arizona, 1130 East Helen Street, McClelland Hall 430, Tucson, Arizona 85721, USA
[b] Graduate Institute of Epidemiology, National Taiwan University, Taipei, Taiwan
[c] Department of Emergency Medicine, National Taiwan University Hospital, No. 7, Chung-Shan South Road, Taipei 100, Taiwan
[d] Institute of Automation, Chinese Academy of Sciences, Zhongguancun East Road #95, Beijing, China

ARTICLE INFO

Article history:
Received 3 January 2008
Received in revised form
18 August 2008
Accepted 19 August 2008

Keywords:
Communicable disease control
Medical records
Multilingual chief complaint classification
Syndromic surveillance
Statistical pattern extraction

ABSTRACT

Purpose: Syndromic surveillance is aimed at early detection of disease outbreaks. An important data source for syndromic surveillance is free-text chief complaints (CCs), which may be recorded in different languages. For automated syndromic surveillance, CCs must be classified into predefined syndromic categories to facilitate subsequent data aggregation and analysis. Despite the fact that syndromic surveillance is largely an international effort, existing CC classification systems do not provide adequate support for processing CCs recorded in non-English languages. This paper reports a multilingual CC classification effort, focusing on CCs recorded in Chinese.

Methods: We propose a novel Chinese CC classification system leveraging a Chinese-English translation module and an existing English CC classification approach. A set of 470 Chinese key phrases was extracted from about one million Chinese CC records using statistical methods. Based on the extracted key phrases, the system translates Chinese text into English and classifies the translated CCs to syndromic categories using an existing English CC classification system.

Results: Compared to alternative approaches using a bilingual dictionary and a general-purpose machine translation system, our approach performs significantly better in terms of positive predictive value (PPV or precision), sensitivity (recall), specificity, and F measure (the harmonic mean of PPV and sensitivity), based on a computational experiment using real-world CC records.

Conclusions: Our design provides satisfactory performance in classifying Chinese CCs into syndromic categories for public health surveillance. The overall design of our system also points out a potentially fruitful direction for multilingual CC systems that need to handle languages beyond English and Chinese.

[*] *Corresponding author*. Tel.: +1 520 621 2165.
E-mail addresses: hmlu@email.arizona.edu (H.-M. Lu), hchen@eller.arizona.edu (H. Chen), zeng@eller.arizona.edu (D. Zeng), cc_king99@hotmail.com (C.-C. King), wcsg@msn.com (T.-S. Wu), pingeshity@yahoo.com.tw (J.-Y. Hsiao).

1386-5056/$ – see front matter © 2008 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.ijmedinf.2008.08.004

1. Introduction

Modern transportation shortens the time needed for a person to travel from one side of the globe to the other. At the same time, it also shortens the time needed for a disease to spread. A case in point is the severe acute respiratory syndrome (SARS) episode which started in the Guangdong Province, China in November, 2002 and spread to Toronto, Vancouver, Ulaan Bator, Manila, Singapore, Hanoi, and Taiwan by March, 2003. The disease was finally brought under control and the whole episode ended in July, 2003. There were a total of 8096 known cases, and about 35% were outside mainland China (cf. http://www.who.int/csr/sars/en/).

The SARS experience indicates that an effective plan for infectious disease detection and prevention, in which syndromic surveillance may play an important role, should be considered on a global scale [1,2]. However, only a few countries have adopted formal syndromic surveillance systems. The U.S. public health system has significant experience in developing and adopting syndromic surveillance systems. However, leveraging such experience in international contexts is proven to be difficult. Multilingual data present a major barrier, as different languages are used by medical and public health practitioners in different parts of the world. This is particularly true for a major data source used by many syndromic surveillance systems: emergency department (ED) triage free-text chief complaints (CCs).

ED triage free-text CCs are short free-text phrases entered by triage practitioners describing reasons for patients' ED visits. ED CCs are a popular data source because of their timeliness and availability [3–6]. However, medical practitioners in other countries do not always use English when recording patients' CCs [7]. As a result, existing CC classification systems designed for English CCs cannot be directly applied in these countries as an important component of the overall syndromic surveillance strategy.

For automatic syndromic surveillance, free-text CC records need to be classified into predefined syndromic categories. This paper reports a study examining the importance of Chinese CCs as a data source for syndromic surveillance and aims to develop a Chinese CC syndromic classification approach. This research was motivated to answer the following research questions:

(a) How useful Chinese CCs are for syndromic surveillance and
(b) Whether an effective cross-lingual approach can be developed leveraging existing English CC classification methods.

CCs from EDs in Taiwan were collected and analyzed in our research. Medical practitioners in Taiwan are trained to record CCs in English. However, it is a common practice to record CCs in both Chinese and English. Furthermore, some hospitals record CCs only in Chinese. We systematically investigated the role and validity of Chinese CCs in the syndromic surveillance context. We then developed a system to classify Chinese CCs based on an automated mechanism to map Chinese CCs to English CCs.

The remainder of this paper is organized as follows. Section 2 provides the background for existing CC classification and cross-lingual information retrieval methods. The next section presents research opportunities and objectives of our research. Section 4 describes our findings regarding the importance of Chinese CCs. Sections 5 and 6 discuss system design of the Chinese CC classification system and experiments to study system performance. Section 7 concludes our discussion.

2. Research background

This section reviews existing CC classification research for both English and non-English CCs. Cross-lingual information retrieval and Chinese key phrase extraction and text segmentation are also reviewed as it provides technical foundation for this research.

2.1. English chief complaint classification methods

There are three main approaches for automated CC syndrome classification: supervised learning, rule-based classification, and ontology-enhanced classification. The supervised learning methods require CC records to be labeled with syndromes before being used for model training. Naive Bayesian [8–10] and Bayesian network [4] models are two examples of the supervised learning methods studied. One prerequisite of supervised learning methods is collecting a sufficient amount of training records, which is usually costly and time-consuming. Another major disadvantage of supervised learning methods is the lack of flexibility. New syndromic definitions may be required by public health practitioners as new events may indicate new surveillance focuses. However, it is often difficult to produce new training data for new syndromic definitions.

Rule-based classification methods do not require labeled training data. Such methods typically have two stages. In the first stage, CC records are cleaned up and transformed to an intermediate representation called "symptom groups" by either a symptom grouping table (SGT) lookup or keyword matching. In the second stage, a set of rules is used to map the intermediate symptom groups to final syndromic categories. For instance, the EARS system (http://www.bt.cdc.gov/surveillance/ears/) uses 42 rules for such mappings.

A major advantage of rule-based classification methods is their simplicity. The syndrome classification rules and intermediate SGTs can be constructed using a top-down approach. The "white box" nature of these methods makes system maintenance and fine-tuning easy for system designers and users. In addition, these methods are flexible. Adding new syndromic categories or changing syndromic definitions can be achieved relatively easily by switching the inference rules. The SGTs can typically be shared across hospitals.

A major problem with rule-based classification methods is that they cannot handle symptoms that are not included in the SGTs. For example, a rule-based system may have a SGT containing the symptoms "abdominal pain" and "stomach ache." This system, however, will not be able to handle "epigastric pain" even though "epigastric pain" is closely related to "abdominal pain."

The BioPortal CC classifier [11,12] is designed to address this vocabulary problem using an ontology-enhanced approach. The semantic relations in the Unified Medical Language System (UMLS), a medical ontology, are used to increase the performance of a rule-based chief complaint classification system. At the core of this approach is the UMLS-based weighted semantic similarity score (WSSS) grouping method that is capable of automatically assigning symptoms previously un-encountered to appropriate symptom groups.

In most chief complaint classifier studies, the performance of chief complaint classification methods is measured by sensitivity, specificity, positive predictive value (PPV), F measure, and F2 measure [4,5,8,13,14]. The F measure is a weighted harmonic mean of PPV and sensitivity. In the context of syndromic surveillance, sensitivity is often considered more important than precision and specificity [4]. The F2 measure gives sensitivity twice as much weight as precision and thus can reflect this emphasis on sensitivity.

In our previous study dealing with English CC records, we showed that the ontology-enhanced approach can achieve a higher level of sensitivity, F measure, and F2 measure when compared to a rule-based system that had the same symptom grouping table and syndrome rules [11].

2.2. Non-English chief complaint classification methods

Little research has focused on non-English CC classifications. One straightforward extension is adding non-English keywords into existing English CC classification systems. For instance, this approach has been applied to process Spanish CCs in EARS [15]. However, for other languages (such as Oriental languages), it would be difficult to incorporate them in an English-based system.

It is also possible to use International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9) codes instead of free-text CCs to classify ED records. ICD-9 codes are standardized, widely used, and can be more accurate than CCs in terms of reflecting true patient illness. Wu et al. used ICD-9 codes attached to ED records to classify Chinese CCs into eight syndromic categories [16,17]. However, as ICD-9 codes are primarily used for billing purposes, they are not always informative for syndromic surveillance [18,19]. As such, free-text CCs remain one of the most important data sources for syndromic surveillance [20].

2.3. Major cross-lingual information retrieval approaches

Existing English chief complaint classification methods can be leveraged in a multilingual context by incorporating cross-lingual information retrieval (CLIR) methods. Cross-lingual information retrieval (CLIR) uses a query in one language to retrieve documents in different languages [21]. Chinese CCs can be treated as documents in the target language, and an English CC classifier can be considered as a system performing query in English, the source language. There are two basic strategies in CLIR. The first strategy is translating documents in the target language to the source language (the language of original query) and performs information retrieval in the source language. The other strategy is translating queries in the source language to the target language and performs information retrieval in the target language [22].

Three major translation approaches are commonly used in CLIR research: machine translation-based approach, corpus-based approach, and dictionary-based approach. The machine translation-based approach [23,24] uses existing machine translation techniques to provide automatic text translation. Machine translation packages can be integrated into existing information systems. However, machine translation packages are often hard to customize. Moreover, in the context of syndromic surveillance, free-text CCs consist of mostly short phrases or incomplete sentences, which lack the contextual and grammatical structural necessary for machine translation.

The corpus-based approach [20,25–27] analyzes large document collections (parallel or comparable corpora) to construct a statistical translation model. It has the potential to translate emerging terminologies. However, parallel corpuses are usually very hard to obtain. Existing parallel multilingual corpuses are typically small and cover only a small numbers of subjects.

Dictionary-based approach [28–30] uses bilingual dictionaries to translate text. Bilingual dictionary are relatively easy to obtain due to recent significant lexicon development efforts; thus this method can often be implemented more easily. However, multiple definitions of a word may cause translation ambiguity (i.e., word sense ambiguity). Moreover, commonly seen medical and symptom-related terminologies are often absent in the multilingual dictionary collection.

2.4. Chinese key phrase extraction and text segmentation

Chinese sentences are written without word/phrase boundaries explicitly delimited. This creates significant problems for Chinese-based information retrieval and text processing. For example, the precision of an information retrieval system can drop significantly if a query is not processed at the word level [31]. As such, how to recognize words in written Chinese has been an important research topic. Note that in Chinese, words and phrases are used interchangeably as they refer to a complete and standalone lexicon pattern that contains more than one Chinese character and has independent meanings.

Chinese key phrase extraction and Chinese text segmentation are two related major research questions. Chinese key phrase extraction studies the problem of extracting important key phrases from a corpus. Chinese text segmentation, on the other hand, focuses on the problem of separating words in a given sentence. These two problems are not completely independent. A text segmentation system can benefit from a good key phrase list and a key phrase extraction system can benefit from good text segmentation results. The major difference between these two problems is that Chinese key phrase extraction usually does not assume the existence of a training dataset. However, it is common to formulate Chinese text segmentation as a supervised learning problem.

2.4.1. *Chinese key phrase extraction*

Similar to the task of constructing multi-word phrases in English, one way to construct the key phrase list is by running through a part-of-speech (POS) tagger and combining characters based on the tagging results. However, because of the lack of word boundaries in Chinese, a Chinese POS tagger needs to either have word segmented before POS tagging or perform word segmentation and POS tagging simultaneously [32]. Note that under the context of syndromic surveillance, there are few training corpora available to implement this approach.

Another popular Chinese key phrase extraction method relies on statistical evidence that reflects collocations or co-occurrences among Chinese characters. Pointwise Mutual information [33], a statistical metric used to measure the strength of association between two adjacent characters, is often the basis for such research. The method was used to extract words with two characters [34] or more [35]. A recent research used this approach to extract significant topics from a text collection of Chinese book and article titles [36].

An alternative approach that uses extended mutual information to measure the strength of co-occurrence among lexicon pattern of two or more characters was proposed by Chien [37]. All lexicon patterns were checked with respect to the extended mutual information measure and key phrases were extracted without length limitation. This approach often requires more computing resources as a larger pattern candidate space needs to be explored.

2.4.2. *Chinese text segmentation*

Existing Chinese text segmentation methods can be broadly classified into two categories: dictionary-based and statistical-based methods. We briefly summarize these methods below.

Dictionary-based approach is the simplest approach to segment Chinese text [38,39]. When a large-enough collection of phrases is available, this method can provide reasonable performance using straightforward implementation such as maximum forward match or maximum backward match. However, dictionary-based method has an obvious problem of identifying new words [40]. Thus if there is no suitable dictionary for text collections from a particular field, this method could perform poorly.

Similar to the problem of Chinese key phrase extraction, the collocation information such as *n*-gram can also be used to perform text segmentation. The compression-based method uses an adaptive language model originally designed for text compression and formulate the text segmentation problem as a hidden Markov model to insert spaces between characters [31]. Specifically, the Prediction by Partial Matching (PPM) compression scheme [41] was studied. This approach learns *n*-gram from a segmented training dataset. Given a sentence in testing dataset, the segmentation with highest compression is chosen. Experimental results showed good performance when training and testing dataset were from the same corpus. However, performance was significantly worse when training and testing dataset were from different corpus.

One way to alleviate the problem of mismatched training and testing dataset is to make use of a large-enough corpus. The web mining-based segmentation algorithm makes use of the *n*-gram collected by submitting corresponding queries to search engines such as Google and Yahoo [42]. After adjusting for the length of words, the combination of words with highest adjusted frequency are chosen as the segmentation result. Experiments showed that this segmentation algorithm outperformed existing state-of-art segmentation methods and were robust to text collections from different geographical areas [42].

3. Research opportunities and objectives

Our review of existing CC classification methods reveals several research opportunities. First, little research has investigated the role of non-English CCs in syndromic surveillance systems. Second, current syndromic surveillance research provides limited support for non-English CC processing.

Based on these observations, our research is aimed at: (a) gaining an empirical understanding of the importance of Chinese CCs in syndromic surveillance and (b) developing a Chinese CC classification system which leverages existing English-based CC classification research. The objective of our research is to bridge the technical gaps existing in the current multilingual CC classification research and develop practical automatic syndromic classification approaches that can handle both English and Chinese CCs. In Section 4, we summarize an empirical study motivated to gain knowledge about the importance of Chinese CC for syndromic surveillance. In Section 5, a multilingual CC classification system is described in detail.

4. An empirical study: The importance of Chinese chief complaints

In our multilingual CC research, we conducted an empirical study to investigate the prevalence and usefulness of Chinese CCs in the syndromic surveillance context based on a large dataset collected from a number of hospitals in Taiwan.

Our working definition of Chinese CCs is any CC records containing Chinese characters. Specialized punctuation marks, which belong to standard-compliant computerized Chinese character sets, are also considered as Chinese characters. In order to validate Chinese CCs as an input to syndromic surveillance systems, we developed a computer program to calculate the prevalence of Chinese CCs and selected random samples from our dataset for further analysis to better understand their importance. This section reports on the data collection effort, followed by a discussion of our experimental design and findings.

4.1. *The Chinese chief complaint dataset*

The Chinese CC dataset used in our study consisted of 939,024 chief complaint records from 116 hospitals in Taiwan. About 98% of these records had admission times from January 1, 2004 to January 26, 2005. We collected CCs from 10 medical centers, 39 regional hospitals, and 67 district hospitals. The collection covered about 60% of hospitals that had emergency departments in Taiwan.

4.2. *Data analysis design*

Manual evaluation of the nearly one million records in our Chinese CC dataset would be impractical. Our experimental investigation followed a two-step design. In the first step, a computer program was designed to distinguish whether a CC record contained Chinese characters. The prevalence of Chinese CCs was then calculated from the output of the program.

Since the focus of this study was to understand the importance of Chinese CCs for syndromic surveillance, in the second step we focused on the hospitals that had more than 10% of CC records containing Chinese characters. For each hospital meeting this threshold, a random sample of 30 Chinese CC records was drawn for manual review. In total, 20 hospitals met this condition and were reviewed. The 600 records from these 20 hospitals were then merged in a random order.

A coder read through all 600 records and classified the Chinese text in the records into four major categories: symptom-related, name entity, Chinese punctuation, and others. Two examples for the CC records belonging to the first “symptom-related” category were “今早開始腹痛 ; 剛吃藥後始雙眼腫，現呼吸不適，心悸 (verbatim translation: abdominal pain began this morning; eyes swollen after taking medication, shortness of breath, palpitations)” and “昨天開始腹瀉 (verbatim translation: diarrhea started yesterday).” From time to time, triage nurses might find that it was hard and inconvenient to translate names of places, people, restaurants, among others, and as a result, keep them in Chinese while still describing symptoms in English. For example, in the CC record “Diarrhea SINCE THIS MORNING. Group poisoning. Having dinner at 威爾康 restaurant,” the restaurant name was kept in Chinese while everything else was in English. This set of CC records was classified as “name entity.” The third category, Chinese punctuation, consisted of CCs with English phrases and Chinese punctuation marks. For example, the record “FEVER SINCE YESTERDAY, COUGH FOR 3–4 DAYS-THROAT INJECTED, LUNG:BS CLEAR” consisted of English expressions only. However, the nurse used the comma symbol available from the Chinese character set “,” instead of the comma symbol “,” commonly used in English sentences. The sentence might appear just like a normal English sentence in some systems (depending on the font used and language setting of the operation system). However, the underlying encoding was very different. The Chinese comma symbol took two bytes to store while the standard comma symbol took one byte only. This might be caused by the default input language setting of the workstations used by some hospitals. Finally, CCs that do not belong to any of these three categories were coded as others.

4.3. *Empirical findings*

Table 1 summarizes the prevalence of Chinese CCs. The overall prevalence of Chinese CCs in the entire Taiwan CC dataset is about 25%. Among the three types of hospitals covered by this dataset, medical centers have the highest prevalence rate of 52%, followed by district hospitals (19%), and regional hospitals (16%). The hospital with the highest prevalence at the medical center level is the MK Hospital (anonymized), which has 100% of its CC records in Chinese. The hospital with the second highest prevalence is the TDUMC Hospital (anonymized) with a prevalence of 18%.

Table 1 – Chinese chief complaint prevalence in Taiwan hospitals

	# Records	# Hospitals	% Chinese CCs
Medical Center	222,893	10	52%
Regional Hospital	484,123	39	16%
District Hospital	232,008	67	19%
Total	939,024	116	25%

It should be noted that the prevalence of Chinese CCs varies from zero to one hundred percent in our sample. In fact, 58% of hospitals have prevalence lower than 10%; 30% of hospitals are between 10% and 90%; and 12% of hospitals are higher than 90%. Strong between-hospital variation suggests that factors unique to each hospital may have strong influence on Chinese CC prevalence. Assuming Chinese CCs appear evenly across hospitals in different regions is thus not reasonable. Discarding Chinese CCs from further processing may potentially bias subsequent disease outbreak detection ability.

Table 2 summarizes the results of the analysis performed in the second step of our study. Twenty hospitals have Chinese CC prevalence higher than 10%. The second row of Table 2 reports the percentages of each of the four target categories, averaged across all 20 hospitals. The third row reports similar percentages averaged across all hospitals but weighted by the total number of Chinese CCs from each hospital. These results demonstrate that more than half (53.8%) of the Chinese CC records contain symptom-related information. About 14.63% of Chinese CCs are related to Chinese punctuations. Only about 7.36% of Chinese CCs are related to Chinese name entities.

5. A Chinese chief complaint classification approach

The empirical study reported above indicates the importance of Chinese CCs as a data source for syndromic surveillance. This section reports our work on designing and evaluating a CC classification system that can process both Chinese and English CCs.

It is possible to develop a Chinese CC classification approach from scratch. However, there are significant language processing issues and few comprehensive medical ontologies in languages other than English. Existing Chinese medical terminologies are only related to translations of medicine and disease names, none are designed for syndromic surveillance. Since there are many effective CC classification methods already developed for English CCs, we chose to leverage these methods. The language difference can be bridged by cross-lingual text processing techniques.

There are two major challenges hindering our effort to process Chinese CCs. The first is the lack of a Chinese key phrases list containing common medical phrases appearing in Chinese CCs. The second is the lack of a Chinese-English translation mechanism for important medical phrases rel-

Table 2 – Categories of Chinese chief complaints

Category	Symptom-related	Name entity	Chinese punctuation	Other
Simple average*	40.79%	13.97%	20.32%	24.92%
Weighted average**	53.80%	7.36%	14.63%	14.63%

* Equally weighed for all hospitals.

** Weighed by the number of Chinese CC records at each hospital.

evant to syndromic surveillance. Although some related general discussions and technical solutions have been discussed in the field of cross-lingual information retrieval, to the best of our knowledge, there are no readily available solutions that can be directly applied to process Chinese CCs.

Motivated to address these two challenges, we have designed a two-step method that consists of (a) statistical Chinese key phrase extraction based on the concept of mutual information and (b) symptom phrase translation. After translating Chinese CCs to English, the BioPortal CC classification system, which was developed in our prior research for English CCs [11], is then used to process translated CCs. In this section, we first discuss the techniques used to process Chinese CCs in order to translate them into English. The following subsections then present the detailed procedure that is used to classify Chinese CCs into syndrome categories.

5.1. *Chinese chief complaint preprocessing*

The goal of Chinese CC preprocessing is to translate Chinese CCs gathered from the field in such a way that the existing, well-tested English CC classifier can be reused. The set of all Chinese CCs (939,024 records in total) was processed using a statistical pattern extraction method based on the concept of mutual information to construct a key phrase list for syndromic surveillance. This key phrase list was then used to perform Chinese word segmentation and Chinese-English translation.

Note that translating Chinese CCs to English is different from typical translation tasks. The Chinese expressions in CCs are in most cases short phrases as opposed to complete sentences. Moreover, not every word or phrase is informative for syndromic surveillance purposes. As a result, the goal of Chinese CC preprocessing is not to provide verbatim translation of Chinese expressions in CC records. Instead, only information that is useful and relevant to syndromic surveillance should be extracted from the original Chinese CCs.

5.1.1. *Statistical Chinese key phrase extraction using extended mutual information*

Following the method proposed by Chien [37], we define the Extended Mutual Information (EMI) of a phrase [11,37,43] as:

$$\mathrm{EMI} = \frac{f(c)}{f(a) + f(b) - f(c)} \qquad (1)$$

where $f(c)$ represents the frequency of the pattern c; $c = c_1, c_2, \ldots, c_n$ is the pattern of interest (e.g., 上吐下瀉 ; vomiting and diarrhea); $a = c_1, c_2, \ldots, c_{n-1}$ and $b = c_2, c_3, \ldots, c_n$ are longest left and right subpatterns of c, i.e., a = “上吐下” (a partial word without meaning) and b = “吐下瀉” (a partial word without meaning). Based on this measure, EMI will be substantially higher than other random patterns if c is by itself a phrase and its subpatterns a and b appear in the text only because of c. For instance, c = “上吐下瀉” may appear in the text 9 times. Its subpatterns a = “上吐下” and b = “吐下瀉” appear in the text only because they are the subpatterns of c. In this case, we have $\mathrm{EMI} = 9/(9+9-9) = 1$. Intuitively, stronger co-occurrence indicates a higher chance of being a meaningful phrase. A EMI score of 1 indicates that c should be considered as a complete phrase.

Searching the whole candidate pattern space requires considerable computing power. Fortunately, each Chinese CC record can be treated as a separate document and punctuation marks such as comma and period can be used to further divide the text string. The maximum length of lexicon patterns is thus greatly reduced. As suggested by previous research [37,43], we construct a PAT tree [44] from divided text strings and stored the frequency of the semi-infinite strings in corresponding nodes. The PAT tree then could be used to provide an efficient structure of computation. Given a lexicon pattern, the frequency of its subpatterns could be easily retrieved by walking up and down the tree. The EMI measure was calculated solely from the information stored in the PAT tree. Lexicon patterns with EMI higher than a pre-specified threshold were considered as the candidate terms in the Chinese key phrase list.

5.1.2. *Key phrase list construction and translation*

To construct a high quality key phrase list, we used a low threshold to filter the output from the EMI method and manually reviewed 2533 candidate phrases. All candidate phrases contained at least two Chinese characters. These candidates were sorted in ascending order by phrase length (number of Chinese characters). One of the authors went through the candidates and removed them if (a) the candidate was not a meaningful phrase or (b) the candidate did not contain information relevant to syndromic surveillance or (c) the meaning of the candidate can be caught by the combination of shorter phrases that had been included. Table 3 provides a few examples of candidate phrases that were reviewed during the process. It took us about 4 h to extract four hundred and fifteen symptom-related key phrases from the 2533 candidate phrases.

We expanded the key phrase list using a general-purpose Chinese-English dictionary of about 220,000 entries (http://www.mandarintools.com/cedict.html). For each candidate Chinese phrase from the Chinese-English dictionary, we

Table 3 – Intermediate results of Chinese key phrase list construction

Candidate	Included (Yes/No)	Comment
自殺 (suicide)	Yes	
臉部 (face)	Yes	
吸不 (partial phrase, no meaning)	No	Not a phrase
鄰居 (neighbor)	No	Unimportant information
治療 (treatment)	Yes	
被割傷 (trauma)	Yes	
被打現 (partial phrase, no meaning)	No	Not a phrase
被狗咬 (bitten by a dog)	Yes	
被汽車 (partial word, no meaning)	No	Not a phrase

included it in our key phrase list if it appeared in our Chinese CC dataset for more than 5 times. Fifty-five additional key phrases were identified. The final symptom key phrase list contains 470 Chinese key phrases.

Three physicians in Taiwan were recruited to translate the extracted Chinese key phrases into English. We provided the physicians with a file listing the Chinese key phrases together with example CCs which contained these phrases. We then reviewed the translations from these physicians to make sure that translations are consistent.

5.2. *A system design for Chinese chief complaint processing*

Fig. 1 depicts the design of our Chinese CC classification system. Our Chinese CC classification system follows six major stages. Stages 0.1–0.3 separate Chinese and English text strings in CCs, perform word segmentation for Chinese text strings, and map symptom-related phrases to English. At the end of Stage 0.3, CC records are in English. In the following three stages (Stages 1–3), the BioPortal CC classifier is invoked

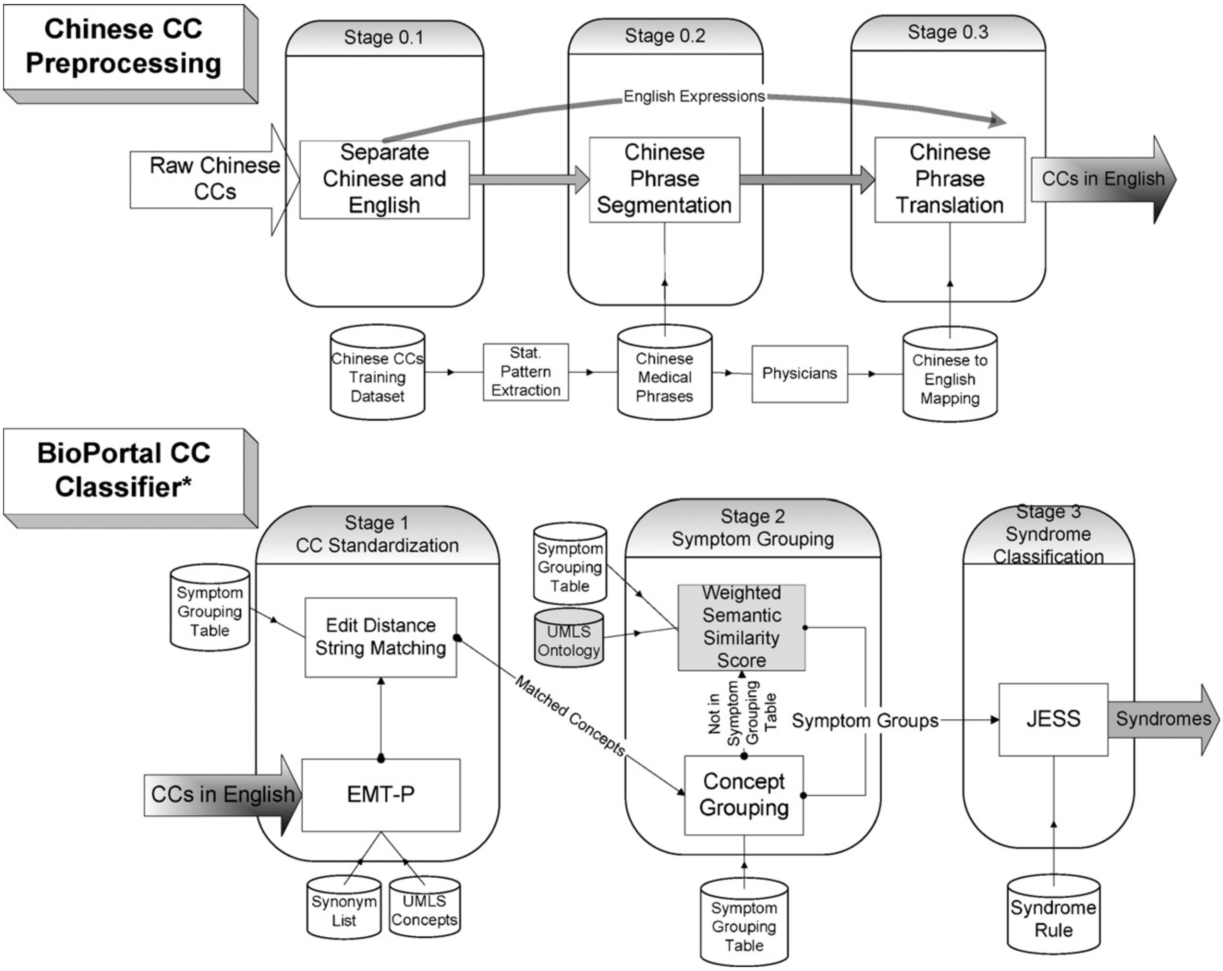

Fig. 1 – Chinese chief complaint extraction classification process. *The system design of BioPortal CC Classifier is reproduced from Lu et al. [11].

[11]. The terms are mapped into concepts in the UMLS ontology in Stage 1. Related concepts are then gathered and put into symptom groups in Stage 2. In Stage 3, a set of rules are used to map symptom groups to the final syndrome categories. Below we discuss each of these steps in detail.

5.2.1. *Stage 0.1: Separating Chinese and English expressions*

Stage 0.1 separates Chinese from English text strings. Since the BioPortal CC classifier can process English CCs, any existing English text strings are kept. The positions of the Chinese and English strings are also marked for future reference. For example, the chief complaint record "Dyspnea, SOB 早上開始坐骨神經痛 解尿困難 " is first divided into two parts: "Dyspnea, SOB," which will skip subsequent Chinese CC preprocessing steps; and "早上開始坐骨神經痛 解尿困難 ," which will be sent to Stage 0.2 for word segmentation.

5.2.2. *Stage 0.2: Chinese expression segmentation*

In this stage, Chinese expressions are segmented using the Chinese symptom key phrase list discussed in the previous section. The longest possible phrases in the phrase list are used for segmentation. For example, the Chinese CC "車輛導致下巴裂傷 (verbatim translation: jaw laceration caused by a car)" is a combination of the following phrases: "車輛 (c car)," "導致 (cause)," "下 巴 (jaw)," and "裂 傷 (laceration)." They are concatenated without any punctuation marks and thus require segmentation before further processing. Using the key phrase list constructed earlier, the original Chinese CC is segmented as: "[車輛]–[導致]–[下巴]–[裂傷] (verbatim translation: [a car]–[cause]–[jaw]–[laceration])." Each text string in square brackets is a phrase segmented from the original text. The verbatim translations in square brackets are the meanings of Chinese phrases segmented from the original text string. Although the combination of individual phrase translation does not constitute a complete sentence with a correct grammatical structure, they do carry valuable information about the syndrome associated with the CC.

Note that since our key phrase list is relative small, many proper nouns are not included. As a result, segmentation results may not be accurate if proper nouns are involved. For example, the Chinese CC "左手被吉娃娃咬傷 (verbatim translation: left hand bitten by a Chihuahua)" is segmented as: "[左]–[手]–[被]–[吉]–[娃]–[娃]–[咬傷] (verbatim translation: [left]–[hand]–[by]–[auspicious]–[baby]–[baby]–[bite])." In this case, the phrase "吉娃娃 (Chihuahua)" is not correctly segmented. The phrase is segmented as three individual Chinese characters because Chihuahua is not included in the key phrase list. This error, nevertheless, does not prevent us from recognizing the syndrome-related information from the inaccurately segmented result.

5.2.3. *Stage 0.3: Chinese phrase translation*

The segmented phrases generated from the previous step are used in Chinese-English symptom mapping. Phrases not recognized are omitted. For example, the segmented Chinese expression "[與]–[人]–[打架], [用]–[鍋]–[鏟]–[打到頭]–[部], [流鼻血]" is mapped to the following English expressions: "[N/A]–[N/A]–[fighting], [N/A]–[N/A]–[N/A]–[head injury]–[N/A], [epistaxis]." "N/A" indicates the term is unavailable in the mapping table. The final translated result thus is "fighting, head injury, epistaxis."

Note that the translation in this stage only depends on the 470 key phrases extracted using Extended Mutual Information. Compared to the number of commonly used Chinese characters (about 6000; see for example [42]), this key phrase list is fairly small. As shown in Section 6, this key phrase-based translation approach led to good overall syndromic classification performance. This positive finding has practical implications in syndromic surveillance. First, it indicates that triage nurses usually use a relatively small, well-defined set of phases to describe symptoms. Second, it is practical and efficient to develop a standardized vocabulary which can further facilitate the processing, aggregation, and analysis of Chinese CCs.

5.2.4. *Stages 1–3: English-based chief complaint classification*

After substituting the Chinese text strings with the translated English strings in the CCs, we proceed to use the BioPortal CC classifier. There are three major Stages in the BioPortal CC classifier: CC standardization, symptom grouping, and syndrome classification. In Stage 1, the acronyms, truncations and abbreviations are expanded using synonym lists and the SPECIALIST lexicon tool developed by the National Library of Medicine (NLM). CCs are divided into symptoms and mapped to standard Unified Medical Language System (UMLS, also developed by NLM) concepts using the Emergency Medical Text Processor (EMT-P) [45,46]. Strings not recognized by EMT-P are mapped to the closest UMLS concept using edit distance string matching.

In Stage 2, standardized symptoms are grouped together using a symptom grouping table. Symptoms that cannot be found in the existing symptom grouping table but are closely related to known symptom groups according to the UMLS ontology are grouped using the weighted semantic similarity score (WSSS) method. The UMLS contains about 2.5 million English terms and their semantic relations. By exploring the relations in the UMLS, known symptoms can be expanded to cover unseen symptoms.

Finally, in Stage 3, a rule engine (implemented in JESS, http://herzberg.ca.sandia.gov/jess/) uses a rule-set based on the Early Aberration Reporting System (EARS, developed by Centers for Disease Control and Prevention) symptom mapping rules to map symptom groups to syndromic categories. In the context of chief complaint classification, the rule-based method requires less training data and is flexible in incorporating new syndromic categories. For details of the BioPortal CC classifier, readers are referred to Lu et al. [11].

6. An evaluation study

This section reports an evaluation study. To the best of our knowledge, there is no publicly available CC classification system for Asian languages. Therefore, there is no existing system that can be directly used as a benchmark in our evaluation study. Instead of conducting a system-level evaluation study, we compare the core component of our Chinese CC preprocessing approach against other Chinese-English mapping methods (i.e., bilingual dictionary translation and machine translation methods) [24,28] in terms of the final syndromic classification performance. In addition to assessing the efficacy of our approach, this comparative study can provide insights about the unique characteristics of the multilingual CC classification problem and provide directions for future improvements.

In this section, we first summarize the syndrome definitions and the gold standard dataset used in this study. The translation methods used as benchmarks are described next. Finally, the empirical findings are presented with examples that illustrate the difference between these translation methods.

6.1. *Syndromic definitions and the gold standard*

We used eight syndrome categories chosen by five local collaborating physicians: constitutional, gastrointestinal, rash, respiratory, upper respiratory, lower respiratory, fever, and other. "Other" was a miscellaneous category for CCs that did not fit into any of the rest syndromes. One chief complaint could be assigned to more than one syndrome. For example, if the upper respiratory or lower respiratory was assigned, the respiratory syndrome automatically applied as well. These categories were similar to those reported in previous studies [4,11,47].

To the best of our knowledge, there is no publicly available dataset with labeled Chinese CCs. Therefore gold standard for system evaluation had to be constructed for this study. The gold standard dataset was a random sample of 1884 CC records from the MK Hospital in Taiwan. Three experts including two physicians and one nurse in Taiwan were given the syndrome definitions and the set of 1884 testing CCs. They were asked to assign CCs to syndromes independently. After collecting the assignments from the experts, a majority rule was used to determine the final syndrome assignments of each CC. On average, one CC was assigned to 1.44 syndromes. According to the final gold standard, gastrointestinal syndrome had the highest prevalence of 31.28%. About 20% of CCs contained fever syndrome. The prevalence of constitutional and respiratory syndromes is about 15%.

Kappa statistic was calculated to determine the assignment agreement among the three experts. The overall agreement was good (kappa = 0.83). All syndromic categories had kappa higher than 0.85 except for the constitutional syndrome, which had kappa of 0.56. Only syndromes with excellent agreement (kappa higher than 0.75) were used in the evaluation study ([48], p. 218).

6.2. *Performance benchmarks: Bilingual dictionary and Google translation*

Several alternative approaches could provide Chinese-English translations. Translations using a bilingual dictionary provided a simple and reasonable performance baseline. For terms with more than one translation in the bilingual dictionary, the first translation was used. A popular and publicly available Chinese-English dictionary was used to provide translations in this setting (http://www.mandarintools.com/cedict.html). There are about 220,000 entries in the collection. This setting is referred to as Bilingual Dictionary translation.

Machine translation is often more sophisticated. We adopted the machine translation method as another benchmark for our evaluation. We used Google Language Tools to provide the translations (http://www.google.com/language_tools?hl=EN). According to a recent machine translation evaluation study conducted by the National Institute of Standards and Technology (NIST) in 2006, the machine translation system developed by Google was one of the best systems among 46 participants for Chinese-English translation [49]. As such, the web-accessible Google machine translation system provided an excellent professional benchmark. After collecting translations from Google Language Tools, the same BioPortal CC classifier was used to provide syndrome classification results. This setting is referred to as Google Translation in the subsequent section.

For our approach, we used an extended mutual information measure to construct a key phrase list for Chinese-English mapping. Our approach is referred to as Mutual Information-based Mapping (MIM).

6.3. *Performance comparison*

In our study, system performance was measured using widely used metrics, including sensitivity (recall), specificity, positive predictive value (PPV or precision), F measure, and F2 measure [4,5,8,13,14]. The performance of all methods under consideration was measured using the same gold standard. McNemar's test [50,51] could be applied for accuracy and sensitivity comparison. However, McNemar's test could not be used to compare PPV, F measure, and F2 measure. Standard paired and independent comparisons were not applicable in this situation as their assumptions did not hold. We thus applied a bootstrapping method to calculate the confidence intervals of the performance differences for all measures so that the experimental results could be interpreted in terms of formal hypothesis testing [11].

6.4. *Experimental results*

6.4.1. *Performance results*

Performance comparison results between MIM and Google Translation can be found in Table 4. The second column of Table 4 lists the positive cases in each syndromic category. The third through the 7th columns list the performance in terms of PPV, sensitivity, specificity, F measure, and F2 measure. In most syndromic categories, the MIM method generates PPV, sensitivity, specificity, F measure and F2 measure higher than 0.9. Rash syndrome has the worst performance with F measure of 0.82. The fever syndrome has the best performance with F measure of 0.97.

Compared to Google Translation, the MIM method has significantly higher PPV, sensitivity, and specificity in most syndromic categories. Given the significant differences in PPV and sensitivity, it is not surprising to find that the MIM method has significantly higher F measure and F2 measure than those of the Google Translation, as these two measures are the functions of PPV and sensitivity. It is interesting to note that MIM has significantly higher F measure and F2 measure in all syn-

Table 4 – Performance comparison for MIM and Google Translation

Syndrome	TP + FN	PPV	Sensitivity	Specificity	F	F2
Mutual Information-based Mapping (MIM)						
GI	592	0.97***	0.97***	0.98***	0.97***	0.97***
RASH	45	0.87**	0.77	0.99**	0.82***	0.80***
RESP	331	0.89***	0.96***	0.97**	0.93***	0.94***
URESP	132	0.86***	0.91**	0.98***	0.88***	0.89***
LRESP	272	0.93	0.98***	0.98	0.95***	0.96***
FEVER	413	0.99**	0.96	0.99**	0.97	0.97
Google Translation						
GI	592	0.91	0.90	0.96	0.91	0.91
RASH	45	0.76	0.73	0.99	0.75	0.74
RESP	331	0.84	0.83	0.96	0.83	0.83
URESP	132	0.70	0.83	0.97	0.76	0.78
LRESP	272	0.96**	0.80	0.99***	0.87	0.84
FEVER	413	0.98	0.96	0.99	0.97	0.97

Statistical test is based on 3000 bootstrappings. **p*-value <0.1; ***p*-value <0.05; ****p*-value <0.01.

Table 5 – Performance comparison for MIM and Bilingual Dictionary

Syndrome	TP + FN	PPV	Sensitivity	Specificity	F	F2
Mutual Information-based Mapping (MIM)						
GI	592	0.97***	0.97***	0.98***	0.97***	0.97***
RASH	45	0.87***	0.77	0.99***	0.82***	0.80***
RESP	331	0.89	0.96***	0.97	0.93***	0.94***
URESP	132	0.86***	0.91***	0.98*	0.88***	0.89***
LRESP	272	0.93	0.98***	0.98	0.95**	0.96***
FEVER	413	0.99	0.96***	0.99	0.97	0.97
Bilingual Dictionary						
GI	592	0.36	0.36	0.70	0.36	0.36
RASH	45	0.54	0.77	0.98	0.64	0.68
RESP	331	0.88	0.79	0.97	0.83	0.82
URESP	132	0.43	0.16	0.98	0.24	0.20
LRESP	272	0.95**	0.90	0.99**	0.93	0.92
FEVER	413	NA	0.00	1.00**	NA	NA

Statistical test is based on 3000 bootstrappings.

* *p*-value <0.1.

** *p*-value <0.05.

*** *p*-value <0.01.

dromic categories except the fever syndrome. MIM and Google Translation have almost the same performance for the fever syndrome. A review of translation results in this syndromic category shows that one keyword ("fever") can cover more than 90% of all true positive cases. As a result, providing good translation for this category is relatively easier than that of other categories. Overall the experimental results indicate that the MIM method provides better syndrome classification performance comparing to processing Chinese CCs using the Google machine translation system.

Table 5 summarizes performance comparison between MIM and Bilingual Dictionary. In general, MIM performs much better than Bilingual Dictionary in terms of PPV, sensitivity specificity, F and F2 measures. Most of the performance difference is significant at a 99% confidence level. Note that Bilingual Dictionary has zero sensitivity in fever syndrome. The reason behind the low performance is because fever was translated to "have a high temperature" by the definition of the bilingual dictionary. The BioPortal CC classifier failed to recognize the phrase as related to fever syndrome. A review of individual

Table 6 – Example 1: Raw Chinese CC, translations and classification results

Translation method	Translation outcome	Syndrome outcome	Gold standard
Raw Chinese CC: 全身酸痛 喉嚨痛今早始 (verbatim translation: whole body soreness and sore throat. began this morning).			
MIM	Soreness, sore throat	UPPER RESP, RESP	CONST, RESP, UPPER RESP
Bilingual Dictionary	Ache, today early begin	UNKNOWN	
Google Translation	General soreness sore throat this morning before.	UPPER RESP, RESP	

Table 7 – Example 2: Raw Chinese CC, translations and classification results

Translation method	Translation outcome	Syndrome outcome	Gold standard
Raw Chinese CC: 吐 晚上開始 (verbatim translation: vomiting. began this evening).			
MIM	Vomiting	GI	GI
Bilingual Dictionary	To spit, in the evening begin	UNKNOWN	
Google Translations	Spit at the beginning	UNKNOWN	

translated CC records indicated that there was a gap between the terms covered by the bilingual dictionary and the terms that were commonly seen in our Chinese CC dataset.

6.4.2. *Examples*

A few examples may help us understand the performance difference among these translation methods. Table 6 provides an example of the input and output of the syndromic classification system. The raw Chinese CC “全身酸痛 喉嚨痛今早始 (verbatim translation: whole body soreness and sore throat. began this morning)” has two important keywords: soreness and sore throat. The MIM method caught both keywords. Google translated the CC as “general soreness sore throat this morning before,” which was accurate. The translation result from Bilingual Dictionary, nevertheless, failed to provide any meaningful information for syndromic surveillance. As mentioned above, the major reason behind the poor translation results of Bilingual Dictionary was the lack of medically related terminologies in the dictionary collection.

Another example can be found in Table 7. The raw Chinese CC “吐 晚上開始 (verbatim translation: vomiting. began this evening)” contains symptoms related to gastrointestinal syndrome. The MIM method did a better job by giving the translation “vomiting.” Google translated it as “spit at the beginning,” which is incorrect. Surprisingly, the translation of Bilingual Dictionary was very similar to that of Google. The poor performance of Google may be due to the concise nature of CCs. There is no context for the machine translation system to disambiguate “吐 ” as vomiting instead of spit.

Finally, in Table 8, the Chinese CC “昨天開始發燒 喘 (verbatim translation: fever and dyspnea. began yesterday)” is related to fever and respiratory syndrome. The MIM method gave a correct translation while the Bilingual Dictionary translated “喘 (gasping)” and “發燒 (fever)” as “to gasp” and “have a high temperature.” “to gasp” is recognized by the BioPortal CC classifier as related to respiratory syndrome. But “have a high temperature” could not be linked to fever syndrome in subsequent processing. The Bilingual Dictionary indeed had “have a fever” as its second translation. However, there was no simple way to decide when other translations instead of the first one should be used ex ante. Google Language Tool provided the correct translation for fever but gave “surge” as the translation for “喘 (gasping).” The translation for “喘 (gasping)” was wrong and we could not find any relation between the translated term “surge” and the original Chinese expression. A possible explanation is that the training dataset for Google translation system did not include documents in medical context and thus it has problem providing high quality medical translation.

The above examples help confirm the discussion about the shortcomings of bilingual dictionary and machine translation approaches for multilingual syndromic classification in our literature review. Bilingual dictionaries often lack terminologies that are commonly seen in Chinese CCs. Machine translation performs better but may provide translations that are meaningless in medical context. The proposed MIM method constructs terminologies bottom-up using a statistical pattern extracting method thus can provide the best translation results for Chinese CCs.

7. Conclusions and future directions

We studied the importance of Chinese CCs and the feasibility of extending an existing English-based CC classification system for Chinese syndromic surveillance. From our empirical study based on about one million CC records, the prevalence of Chinese CC is about 25% and more than half of Chinese phrases appeared in CC records are symptom-related.

We used a statistical pattern extraction method based on the mutual information to extract important phrases from Chinese CCs and constructed mappings to English. The UMLS-based BioPortal CC classifier, which was designed to process CCs in English, was used to process translated CCs. We compared the syndrome classification performance of the proposed translation method with those using the machine translation system provided by the Google Language Tool and a bilingual dictionary. Compared to Google Translation, our approach delivered significantly higher PPV, sensitivity, specificity, F measure, and F2 measure for most syndromic categories. We found similar results in the comparison between our approach and the translations provided by the bilingual dictionary.

Table 8 – Example 3: Raw Chinese CC, translations and classification results

Translation method	Translation outcome	Syndrome outcome	Gold standard
Raw Chinese CC: 昨天開始發燒 喘 (verbatim translation: fever and dyspnea. began yesterday)			
MIM	Fever, dyspnea	RESP, LRESP, FEVER, CONST	RESP,
Bilingual Dictionary	Yesterday begin have a high temperature, to gasp	LRESP, RESP	LRESP,
Google Translation	Surge began yesterday fever	FEVER, CONST	FEVER

Summary points

What was already known in this field?

- Emergency department free-text chief complaints are an important data source for syndromic surveillance.
- Free-text chief complaints need to be classified into syndrome categories to facilitate subsequent aggregation and analysis.
- Chief complaint classification for English has been widely studied and can deliver reasonable performance.

What this study has added to our knowledge?

- Chinese chief complaint records contain useful syndrome-related information.
- Symptom-related information in Chinese can be effectively extracted and translated into English using a small set of key phrases.
- Chinese chief complaints can be successfully classified by adding a translation module to an existing English chief complaint classifier.

The observed superior performance of our proposed Chinese-English mapping approach indicates that the 470 key phrases extracted from about one million Chinese CCs could cover common triage usage. We believe that with a more comprehensive study of Chinese CC records, a set of standardized vocabulary could be constructed and our approach can be adopted in real-world applications. We do caution that languages are constantly evolving. Periodic reviews of extracted key phrases would be necessary to ensure inclusion of new phases.

The syndrome definitions used in this study only cover those mostly commonly used by public health practitioners in Taiwan. We are currently working on identifying other useful syndromes and developing proper training and testing data. We also plan to extend our MIM-based approach and develop an approach that can be flexible enough for international public health situational awareness. In addition to technical research, we are currently working with selected hospitals in Taipei to operationalize and validate our multilingual BioPortal system for syndromic surveillance. We expect that running the Chinese CC classification system in real-world settings (use original phrases) will validate of our ideas and offer new technical insights to motive further research.

Acknowledgements

This work was supported in part by the U.S. National Science Foundation through Grant #IIS-0428241 (“A National Center of Excellence for Infectious Disease Informatics”). It also draws on earlier work supported by the Arizona Department of Health Services. Chwan-Chuen King acknowledges support from the Taiwan Department of Health (DOH95-DC-1021). Daniel Zeng wishes to acknowledge support from the National Natural Science Foundation of China (60573078 and 60621001), the Chinese Academy of Sciences (2F05N01 and 2F07C01) and the Ministry of Science and Technology (2006AA010106).

REFERENCES

[1] H. Liang, Y. Xue, Investigating public health emergency response information system initiatives in China, Int. J. Med. Inf. 73 (2004) 675–685.

[2] J.G. Bellika, T. Hasvold, G. Hartvigsen, Propagation of program control: a tool for distributed disease surveillance, Int. J. Med. Inf. 76 (2006) 313–329.

[3] W.W. Chapman, J.N. Dowling, M.M. Wagner, Generating a reliable reference standard set for syndromic case classification, J. Am. Med. Inf. Assoc. 12 (6) (2005) 618–629.

[4] W.W. Chapman, L.M. Christensen, M.M. Wagner, P.J. Haug, O. Ivanov, J.N. Dowling, et al., Classifying free-text triage chief complaints into syndromic categories with natural language processing, Artif. Intell. Med. 33 (1) (2005) 31–40.

[5] O. Ivanov, M.M. Wagner, W.W. Chapman, R.T. Olszewski, Accuracy of three classifiers of acute gastrointestinal syndrome for syndromic surveillance, in: AMIA Symposium, 2002, pp. 345–349.

[6] J.U. Espino, M.M. Wagner, The accuracy of ICD-9 coded chief complaints for detection of acute respiratory illness, in: Proceedings of the AMIA Annual Symposium, 2001, pp. 164–168.

[7] K. Marko, S. Schulz, U. Hahn, Automatic lexeme acquisition for a multilingual medical subword thesaurus, Int. J. Med. Inf. 76 (2–3) (2007) 184–189.

[8] R.T. Olszewski, Bayesian classification of triage diagnoses for the early detection of epidemics, in: FLAIRS Conference, Menlo Park, California, 2003, pp. 412–416.

[9] J.U. Espino, J. Dowling, J. Levander, P. Sutovsky, M.M. Wagner, G.F. Copper, SyCo: a probabilistic machine learning method for classifying chief complaints into symptom and syndrome categories, in: Syndromic Surveillance Conference, Baltimore, Maryland, 2006.

[10] S. Sadeghi, A. Barzi, N. Sadeghi, B. King, A Bayesian model for triage decision support, Int. J. Med. Inf. 75 (5) (2006) 403–411.

[11] H.-M. Lu, D. Zeng, H. Chen, Ontology-enhanced Automatic Chief Complaint Classification for Syndromic Surveillance, J. Biomed. Inf. 41 (2) (2008) 340–356.

[12] H.-M. Lu, D. Zeng, H. Chen, Ontology-based automatic chief complaints classification for syndromic surveillance, in: IEEE International Conference on Systems, Man, and Cybernetics, Taipei, Taiwan, 2006.

[13] C.J. van Rijsbergen, Information Retrieval, Butterworths, London, 1979.

[14] S.V.S. Pakhomov, J.D. Buntrock, C.G. Chute, Automating the assignment of diagnosis codes to patient encounters using example-based and machine learning techniques, J. Am. Med. Inf. Assoc. 13 (2006) 516–525.

[15] L. Hutwagner, W. Thompson, G.M. Seeman, T. Treadwell, The bioterrorism preparedness and response early aberration reporting system (EARS), J. Urban Health 80 (2 Suppl. 1) (2003) i89–i96.

[16] T.-S. Wu, Establishing emergency department-based infectious disease syndromic surveillance system in Taiwan–aberration detection methods, epidemiological characteristics, system evaluation and recommendations. Master Thesis, National Taiwan University, 2005.

[17] T.-S. Wu, F.-Y. Shih, M.-Y. Yen, J.-S. Wu, S.-W. Lu, K. Chang, et al., Establishing a nationwide emergency department-based

syndromic surveillance system for better public health responses in Taiwan, BMC Public Health 8 (1) (2008) 18.

[18] F.S. Fisher, F.S. Whaley, W.M. Krushat, D.J. Malenka, C. Fleming, J.A. Baron, et al., The accuracy of Medicare's hospital claims data: progress has been made, but problems remain, Am. J. Public Health 82 (2) (1999) 243–248.

[19] F.C. Day, D.L. Schriger, M. La, Automated linking of free-text complaints to Reason-for-Visit categories and International Classification of Diseases diagnoses in emergency department patient record databases, Ann. Emerg. Med. 43 (3) (2004) 401–409.

[20] K.W. Li, C.C. Yang, Conceptual analysis of parallel corpus collected from the web, J. Am. Soc. Inf. Sci. Technol. 57 (5) (2006) 632–644.

[21] J. Qin, Y. Zhou, M. Chau, H. Chen, Multilingual web retrieval: an experiment in English-Chinese business intelligence, J. Am. Soc. Inf. Sci. Technol. 57 (5) (2006) 671–683.

[22] H.-H. Chen, Cross-language information retrieval: theories and technologies, J. Libr. Inf. Sci. 28 (1) (2002) 19–32.

[23] D. Arnold, L. Balkan, S. Meijer, R. Humphreys, L. Sadler, Machine Translation: An Introductory Guide, Blackwells-NCC, London, 1994.

[24] T. Sakai, MT-based Japanese-English cross-language IR experiments using the TREC test collections, in: Fifth International Workshop on Information Retrieval with Asian Language, Hong Kong, China, 2000.

[25] D.W. Oard, Adaptive Vector Space Text Filtering for Monolingual and Cross-Language Applications, University of Maryland, College Park, 1996.

[26] R.D. Brown, Example-based machine translation in the Pangloss system, in: Proceedings of the 16th International Conference on Computational Linguistics (COLING-96), Copenhagen, Denmark, 1996, pp. 169–174.

[27] T. Talvensaari, M. Juhola, J. Laurikkala, K. Jarvelin, Corpus-based cross-language information retrieval in retrieval of highly relevant documents, J. Am. Soc. Inf. Sci. Technol. 58 (3) (2007) 322–334.

[28] P. Daumke, K. Marku, M. Poprat, S. Schulz, R. Klar, Biomedical information retrieval across languages, Inf. Health Social Care 32 (2) (2007) 131–147.

[29] A. Pirkola, T. Hedlund, H. Keskustalo, K. Jarvelin, Dictionary-based cross-language information retrieval: problems, methods, and research findings, Inf. Retrieval 4 (3–4) (2001) 209–230.

[30] M. Aljlayl, O. Frieder, D. Grossman, On bidirectional English-Arabic search, J. Am. Soc. Inf. Sci. Technol. 53 (13) (2002) 1139–1151.

[31] W.J. Teahan, Y. Wen, R. McNab, I.H. Witten, A compression-based algorithm for Chinese word segmentation, Computat. Linguistics 26 (3) (2001) 375–393.

[32] H.T. Ng, J.K. Low, Chinese part-of-speech tagging: one-at-a-time or all-at-once? Word-based or character-based? in: Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP 2004), Barcelona, Spain, 2004, pp. 277–284.

[33] C. Manning, H. Schütze, Foundations of Statistical Natural Language Processing, MIT Press, Cambridge, MA, 1999.

[34] R. Sproat, C.L. Shin, A statistical method for finding word boundaries in Chinese text Comput. Process. Chin. Oriental Lang. 4 (4) (1990) 336–351.

[35] C.C. Yang, J.W.K. Luk, S.K. Yung, J. Yen, Combination and boundary detection approaches on Chinese indexing, J. Am. Soc. Inf. Sci. 51 (4) (2000) 340–351.

[36] J. Wang, Automatic thesaurus development: term extraction from title metadata, J. Am. Soc. Inf. Sci. Technol. 57 (7) (2006) 907–920.

[37] L.-F. Chien, PAT-tree-based adaptive keyphrase extraction for intelligent Chinese information retrieval, Inf. Process. Manag. 35 (4) (1999) 501–521.

[38] K.S. Cheng, G.H. Young, K.F. Wong, A study on word-based and integral-bit Chinese text compression algorithm, J. Am. Soc. Inf. Sci. 50 (3) (1999) 218–228.

[39] Z. Wu, G. Tseng, Chinese text segmentation for text retrieval: achievements and problems, J. Am. Soc. Inf. Sci. 44 (9) (1993) 532–542.

[40] C.L. Goh, M. Asahara, Y. Matsumoto, Chinese word segmentation by classification of characters, Comput. Ling. Chin. Lang. Process. 10 (3) (2005) 381–396.

[41] J.G. Cleary, I.H. Witten, Data compression using adaptive coding and partial string matching, IEEE Trans. Commun. 32 (4) (1984) 396–402.

[42] F.L. Wang, C.C. Yang, Mining web data for Chinese segmentation, J. Am. Soc. Inf. Sci. Technol. 58 (12) (2007) 1820–1837.

[43] T.-H. Ong, H. Chen, Updateable PAT-Tree approach to Chinese key phrase extraction using mutual information: a linguistic foundation for knowledge management, in: Proceedings of the Second Asian Digital Library Conference, Taipei, Taiwan, 1999.

[44] G.H. Gonnet, R. Baeza-Yates, T. Snider, New indices for text: PAT Trees and PAT arrays, in: Information Retrieval: Data Structures and Algorithms, Prentice-Hall, 1992, pp. 66–82.

[45] D.A. Travers, S.W. Haas, Evaluation of emergency medical text processor, a system for cleaning chief complaint textual data, Acad. Emerg. Med. 11 (11) (2004) 1170–1176.

[46] D.A. Travers, S.W. Haas, Using nurses' natural language entries to build a concept-oriented terminology for patients' chief complaints in the emergency department, J. Biomed. Inf. 36 (2003) 260–270.

[47] P.H. Gesteland, R.M. Gardner, F.-C. Tsui, J.U. Espino, R.T. Rolfs, B.C. James, et al., Automated syndromic surveillance for the 2002 winter Olympics, J. Am. Med. Inf. Assoc. 10 (November/December (6)) (2003) 547–554.

[48] J.L. Fleiss, Statistical Methods for Rates and Proportions, second ed., John Wiley & Sons, NY, NY, 1981.

[49] NIST, NIST 2006 Machine Translation Evaluation Official Results, 2006 (cited 2007 July 2), available from: http://www.itl.nist.gov/iad/894.01/tests/mt/doc/mt06eval_official_results.html.

[50] A. Agresti, Categorical Data Analysis, John Wiley & Sons, Hoboken, New Jersey, 2002.

[51] Q. McNemar, Note on the sampling error of the difference between correlated proportions or percentages, Psychometrika 12 (1947) 153–157.