Multilingual Chief Complaint Classifier

Hsin-Min Lu, Hsinchun Chen, Daniel Zeng,
Chwan-Chuen King, Lea Trujillo, and Ken Komatsu

Management Information Systems Department, the Eller College of Management, University of Arizona, Tucson, Arizona 85721, USA; the Institute of Automation, the Chinese Academy of Sciences; Arizona Department of Health Services, Phoenix, Arizona, 85007; Graduate Institute of Epidemiology, National Taiwan University, Taipei, Taiwan

OBJECTIVE

This paper describes the effort to design and implement a chief complaint (CC) classification system that is capable of processing CCs in both English and Chinese.

BACKGROUND

Free text chief complaints (CCs), which may be recorded in different languages, are an important data source for syndromic surveillance systems. For automated syndromic surveillance, CCs must be classified into predefined syndromic categories to facilitate subsequent data aggregation and analysis. However, CCs in different languages pose technical challenges for the development of multilingual CC classifiers. We addressed the technical challenges by first developing a ontology-enhanced CC classifier which exploits semantic relations in the Unified Medical Language System (UMLS) to expand the knowledge of a rule-based CC classifier. Based on the ontology-enhanced English CC classifier, a translation module was incorporated to extract symptom-related information in Chinese CCs and translate it into English. This design thus enables the processing of CCs in both English and Chinese.

METHODS

There are three major stages in the ontology-enhanced CC classifier [1]: CC standardization, symptom grouping, and syndrome classification. In Stage 1, the acronyms and abbreviations are expanded. CCs are divided into symptoms and mapped to standard UMLS concepts using EMT-P [2]. 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 WSSS method can expand existing grouping knowledge and improve overall performance. Finally, in Stage 3, a rule-based engine is used to map from symptom groups to syndromic categories.

To facilitate the processing of CCs in Chinese, we coupled a Chinese-English translation module and the ontology-enhanced CC classifier. The design of the translation module can be found in Figure 1. There are three major steps involved in the Chinese-English CC translation: separating Chinese and English text strings, Chinese phrase segmentation, and Chinese phrase translation. The first step separates Chinese and English text strings. Since the ontology-enhanced CC classifier can process English CCs, existing English terms are preserved. Relative positions of text strings are marked and recorded. In the second step, Chinese expressions in the Chinese CCs are segmented using the Chinese terms available from the phrase table, which was constructed using a statistical pattern extraction method based on the concept of mutual information. The longest possible phrases in the phrase list are used for segmentation. Finally, the segmented phrases output from the previous step are used as the basic elements for Chinese-English symptom mapping.

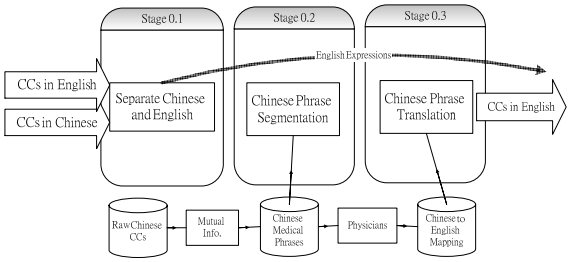

Figure 1 – System design of Chinese-English CC translation module.

RESULTS

The evaluation study, based on English CCs collected from a hospital in Phoenix and Chinese CCs collected from a hospital in Taipei, Taiwan, shows promising results. The system is able to process both CCs in English and Chinese with satisfactory sensitivity, positive predictive value, and specificity. This system design may be expanded to process CCs in other languages.

REFERENCES

[1] Lu H-M, Zeng D, Chen H. Ontology-enhanced Automatic Chief Complaint Classification for Syndromic Surveillance. Journal of Biomedical Informatics, forthcoming.

[2] Travers, D.A., & Haas, S.W. (2004). Evaluation of emergency medical text processor, a system for cleaning chief complaint textual data. Academic Emergency Medicine, 11(11), 1170-1176.

Further Information:
Hsin-Min Lu, hmlu@email.arizona.edu; http://ai.arizona.edu