Ontology-Based Automatic Chief Complaints Classification
for Syndromic Surveillance

Ken Komatsu, Lea Trujillo
Arizona Department of Health Services

Hsin-Min Lu, Daniel Zeng, Hsinchun Chen
University of Arizona

OBJECTIVE

This paper presents a novel approach of using a medical ontology to classify free-text chief complaints into syndrome categories.

BACKGROUND

Free-text emergency department (ED) triage chief complaints (CCs) are a popular data source used by many syndromic surveillance systems because of their timeliness, availability, and relevance. The lack of standardization of CC vocabulary poses a major technical challenge to any automatic CC classification approach. This challenge can be partially addressed by several methods, for example, medical thesaurus, spelling check, manually-created synonym list, and supervised machine learning techniques that directly operate on free text. Current approaches, however, ignore the fact that medical terms appearing in CCs are often semantically related. Our research exploits such semantic relations through a medical ontology in the context of automatic CC classification for syndromic surveillance.

METHODS

In our approach, the syndromic classification process is divided into three stages: CC standardization, symptom grouping, and syndrome classification. The system first standardizes a CC into one or more symptoms/concepts in Unified Medical Language System (UMLS) using Emergency Medical Text Processor (EMT-P) [1]. These symptoms are then grouped into existing symptom groups through a Symptom Grouping Table (SGT), a component of the EARS system [2]. For symptoms falling outside the SGT, we have developed a novel UMLS-based approach to map symptoms to the appropriate symptom groups by identifying semantic relationships among terms. Our approach uses the concept tree from the UMLS as the input and relies on semantic similarity, which is defined over the tree path distance between UMLS concepts, to quantify the "fitness" between a symptom and a candidate symptom group. In the last stage, the system maps the symptom group to a predefined syndrome category using a rule inference engine, JESS (www.jessrules.com).

RESULTS

We have conducted an initial computational evaluation of our approach using a CC dataset collected from one hospital in the State of Arizona. Early results show that, compared to using SGT
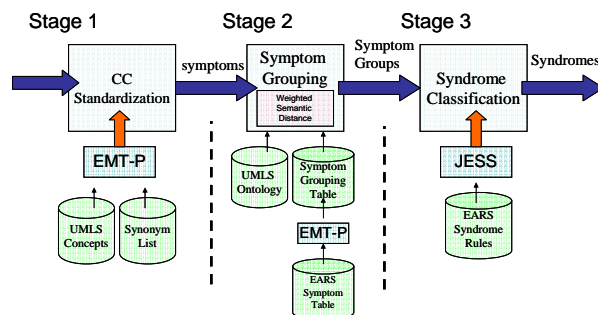

Fig. 1. System Design for Ontology-based Chief Complaints Classification

alone, our approach is able to improve the probability that a unique symptom can be correctly mapped to a symptom group from 24.7% to 54.0%.

In addition, we are creating a gold-standard CC dataset based on the data collected in Arizona hospitals using several domain experts to manually map raw CCs into syndromes. We will then compare the performance of our method with that of other major CC classification systems to assess the accuracy of our method of syndrome classification.

CONCLUSIONS

SGTs can provide a systematic method to compose disease or event-specific syndrome categories. Our ontology-based symptom group mapping method is very helpful in suggesting symptom groups for symptoms not contained in the predetermined SGT.

A computational study designed to evaluate our ontology-based approach using real-world data is expected to be completed by the end of July and we plan to include a discussion of the detailed findings at the SSC 2006 if this abstract is accepted.

REFERENCES

[1] Travers DA and Haas SW. (2004) Evaluation of emergency medical text processor, a system for cleaning chief complaint text data. Academic Emergency Medicine, 11(11): 1170-1176.

[2] Hutwagner L, Thompson W, Seeman GM, and Treadwell T. (2003) The Bioterrorism Preparedness and Response Early Aberration Reporting System (EARS). Journal of Urban Health: Bulletin of the New York Academy of Medicine. 80(2): Suppl 1; i89-i96.