

## IDENTIFICATION OF CAUSE AND EFFECT IN CAUSAL SENTENCES OF GERIATRIC CARE DOMAIN USING CONDITIONAL RANDOM

**Saeed Mehrabi** and Anand Krishnan<sup>1</sup> (Mathew Palakal), Health Informatics Program, School of Informatics, Indiana University–Purdue University, Indianapolis. Indiana 46202

Event extraction is a key step in many text mining applications. Identified events can be used in various applications such as question-answering systems, information extraction, summarization or building the knowledge base of a clinical decision support system. In this study we used PubMed abstracts of Geriatric care domain that were manually categorized into 42 different subdomains and further divided into causal and non-causal sentences by three domain experts. There are a total of 19,677 sentences in the collected abstracts from PubMed, out of which 2,856 sentences were selected and manually annotated with cause and effect events. We used conditional random fields (CRFs) that are statistical algorithms used to sequentially tag each word in a sentence as a cause or effect event based on some input variables or features. Features used in this study are words, words categories (lowercase, uppercase, mixed of letter and digits, etc.), affixes, part of speech and phrase chunks such as noun or verb phrase. For every word, a window of features before and after each word was also considered. We tested window of size, one to five meaning one to five features before and after each word was included as the input variables. The CRF algorithm was trained and tested on data set with 2,520 sentences in training set, 252 sentences in validation and 84 sentences in test set. Window of four features before and after each word had the best performance with 75.1% accuracy and F-measure of 85% with 84.6% precision and 87% recall.

<sup>1</sup>Department of Computer Science, Indiana University-Purdue University, Indianapolis, IN 46202

We would like to thank the research and development team at Allegient, LLC. This work was supported in part by a grant titled, "Automatic Acquisition of Knowledge for Senior Care Planning" from the National Institute of Health.