Medical Xpress

New computer models aim to classify, help reduce injury accidents

September 2 2009, by Emil Venere

Researchers are developing computer models to comb through thousands of injury reports in large administrative medical datasets or insurance claims data to automatically classify them based on specific words or phrases.

"One goal is to identify the most important causes of injuries so that efforts could be directed toward reducing the burden of injuries in society," said Mark Lehto, an associate professor in Purdue University's School of Industrial Engineering.

The reports, usually filled out by employers, health-care professionals or claimants themselves, are currently classified by manual coders hired by users such as the National Center for Health Statistics, hospital staff or insurance industry handlers who review thousands of "injury narratives" included in reports.

"This is obviously very labor-intensive," Lehto said.

The Purdue engineer and researchers at the Liberty Mutual Research Institute for Safety in Hopkinton, Mass., assigned codes to injury reports from workers' compensation claims using two different models developed with a technique called "Bayesian methods."

"The predictions were quite good," Lehto said. "The results were comparable to the human coders. The accuracy is surprising considering all of the misspellings, run-on words, abbreviations and inconsistent or

missing punctuations seen in these workers' compensation claim narratives."

An example of an injury-claim narrative included in the paper is: "HUSB. & SON WERE REARENDED AT RED TRAFFIC LIGHT BY DRUNKEN DRIVER DRIV-ING AT LEAST 45 MPH INFULL SIZE PICK-UP TRUCK//N."

"Can you imagine reading through 10,000 of these narratives and trying to interpret what the cause of injury is and assign different codes?" said Lehto, the 2008 Liberty Mutual Research Institute for Safety visiting scholar.

Research findings were detailed in a paper published in August in the journal *Injury Prevention*. The paper was written by Lehto and Liberty Mutual research scientists Helen Marucci-Wellman and Helen Corns.

Insurance companies enter, maintain and manage tens of thousands of claims annually. The study examined approaches for efficient assignment of each claim using a computer approach with one and two-digit "event code" categories developed by the U.S. Bureau of Labor Statistics.

"So now we are trying to take these vast sets of data, which have been limited in their utility due to the large expense in hiring manual coders, and we are able to glean important information from the injury narratives and come up with new knowledge on the potential causes and prevention of injuries," Lehto said.

The new models might lead to programs that automatically code reports as they are being filed.

"These models can be easily updated to deal with new types of accidents

they haven't encountered before," Lehto said.

The models calculated the probability that reports would be classified by human coders in specific categories. One model, called "naive," reviewed individual words, and the other, called "fuzzy," looked at sequences of words and phrases in the narratives, such as "fell off a ladder."

The researchers used a database of 14,000 claim cases, with 11,000 used to develop the models and 3,000 used to test the models.

"It's important to distinguish that we predicted 3,000 cases that were different than the ones used to develop the models," Lehto said. "These were cases the models hadn't seen before, and the models accurately predicted how these cases would be classified by human coders."

Source: Purdue University ([news](#) : [web](#))