

# Improvement in Performance of Ngram Classifiers with Frequent Updates

Philip Brown<sup>1</sup>, Sylvia Halasz<sup>1</sup> PhD, Colin Goodall<sup>1</sup> PhD, Dennis G. Cochrane<sup>2,3</sup> MD, John R. Allegra<sup>2,3</sup> MD, PhD

<sup>1</sup>ATT Research Labs; <sup>2</sup>Morristown Memorial Hospital Residency in Emergency Medicine, Morristown, New Jersey;

<sup>3</sup>Emergency Medical Associates Research Foundation

## Background

Syndromic surveillance of emergency department (ED) visit data is often based on computerized classifiers which assign patient chief complaints (CC) to syndromes. These classifiers may need to be updated periodically to account for changes over time in the way the CC is recorded or because of the addition of new data sources. Little information is available as to whether more frequent updates would actually improve classifier performance significantly. It can be burdensome to update classifiers which are developed and maintained manually. We had available to us an automated method for creating classifiers that allowed us to address this question more easily. The "Ngram" method, described previously, creates a CC classifier automatically based on a training set of patient visits for which both the CC and ICD9 are available. This method measures the associations of text fragments within the CC (e.g. 3 characters for a "3-gram") with a syndromic group of ICD9 codes. It then automatically creates a new CC classifier based on these associations. The CC classifier thus created can then be deployed for daily syndromic surveillance.

## Objective

Our objective was to determine if performance of the Ngram classifier for the GI syndrome was improved significantly by updating the classifier more frequently.

## METHODS

Design: Retrospective cohort. Setting: 19 EDs in NJ and NY. Participants: All visits from 1-1-2000 to 12-31-2005. Protocol: We used a lower GI grouping of ICD9 codes created by the ESSENCE-CDC project. We then used an N-Gram method adapted from AT&T Labs' technologies to build CC classifiers for lower GI. The N-Gram CC classifiers were built from training sets chosen in three different and progressively more frequent ways: (1) No updates: test set 2001-2005, training set all visits in year 2000. (2) Yearly updates: test sets each of the years 2001-2005 with training set in each case the year before. (3) Monthly updates: test sets each individual month for the years 2001-2005 and the training set in each case

was the 12 months previous. We next applied each classifier to its corresponding test set to produce daily volume estimates and compared these volumes against those found by the ICD9 classifier. We then determined the correlation coefficients for these comparisons as a measure of the performance of the CC classifier. We used Fisher's z transformation to test for statistical significance in the differences between the correlation coefficients.

## Results

The database contained 3.5 million visits.

Updates	Correlation coefficients (R)
None	0.86
Yearly	0.88
Monthly	0.93

Each pair of correlation coefficients was significantly different at the  $p < 0.01$  level.

## CONCLUSION

Performance of the classifier created by the Ngram method improved significantly when it was updated more frequently. We speculate that this may generalize to manually created classifiers, and that it may be advantageous to update manual CC classifiers on a more regular basis. Methods which update the classifier automatically such as the Ngram method may offer a way to more easily check for "drift" in the way CC is recorded in a given population.