A Natural Language Processing Algorithm for Improving Efficiency of Breast Cancer Surveillance Abstraction

David Carrell, PhD[1]; Scott Russell Halgrim, MA[1]; Diem-Thy Tran, BS[1]
[1]Group Health Research Institute, Seattle, WA

GroupHealth®

INTRODUCTION

An ongoing breast cancer research project at our institution employs trained abstractors to track malignant and benign breast pathology in a large patient population. To reduce the number of reports requiring human review, this surveillance project uses a keyword search in SAS designed with high sensitivity to select potentially relevant reports for human review. Abstractors assign each selected report to one of three categories: 1) breast malignancy, 2) breast benign, or 3) other/unrelated.

We present as an alternative a natural language processing (NLP) algorithm for identifying breast *malignancies* in the same reports and comparatively assess its performance.

METHODS

The NLP algorithm was developed in cTAKES.[1] It uses a lexicon and a template-filling approach to determine whether mentions of breast malignancy are present or absent in pathology reports (Figure 1).

For this experiment a sample of reports from all 11,115 reports generated 6/1-7/31/2009 was used. In that period the keyword search identified 323 reports for review. All 323 reports were reviewed to determine actual mentions of malignancy.

The NLP algorithm processed the same 323 reports selected by keyword search plus a random sample of 1,000 of the remaining reports. All reports flagged by NLP were reviewed to determine actual mentions of malignancy.

Reports identified by the keyword search and NLP approaches were compared with respect to number of reports flagged and number of actual malignancies.
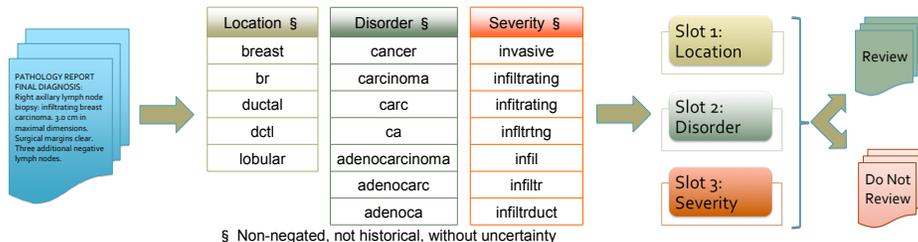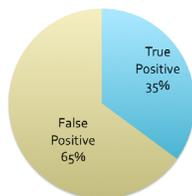
RESULTS

§ Non-negated, not historical, without uncertainty

Figure 1. Rule-based NLP template filling process to identify breast malignancy mentions

Table 1. Efficiency of two approaches for identifying breast malignancies

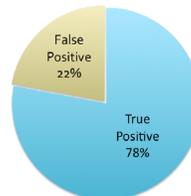| Approach	Flagged for Review	Actual Malignancies	Sensitivity	Efficiency
Keyword search	323	114	100%	35%
NLP algorithm	145	113	99.1%	78%

Figure 2. Comparative efficiency of the keyword and NLP approaches

RESULTS

Of the 11,115 reports produced during the period 323 were flagged for review by keyword search, of which 114 (35%) mentioned malignancies (Table 1).

Of the 1,323 reports processed by NLP (323 + 1,000), 145 were flagged for review, of which 113 (78%) mentioned malignancies. All of the 143 flagged by NLP were among the 323 also flagged by keyword search. None of the 1,000 randomly selected reports not flagged by keyword search were flagged by NLP.

The NLP algorithm produced one false negative.

Using the traditional keyword search as a gold standard, the NLP approach achieved 99.1% sensitivity, and reduced the number of reports requiring review by more than half (Figure 2).

CONCLUSION

NLP algorithms offer significant efficiencies for surveillance projects that require identifying malignancy mentions in pathology reports.

Future work should explore performance of the NLP approach at the *patient level* as a possible solution to false negatives at the *report level* (e.g., if a patient has two reports mentioning malignancy and one is flagged by NLP and the other is not, the false negative would not affect surveillance as this patient would be properly identified).

References
1. Savova GK, Masanz J, Ogren PV, et al. Mayo clinical text analysis and knowledge extraction system (cTAKES): architecture, component evaluation and applications. J Am Medical Inform Assoc Sept 2010;17(5):507-513.