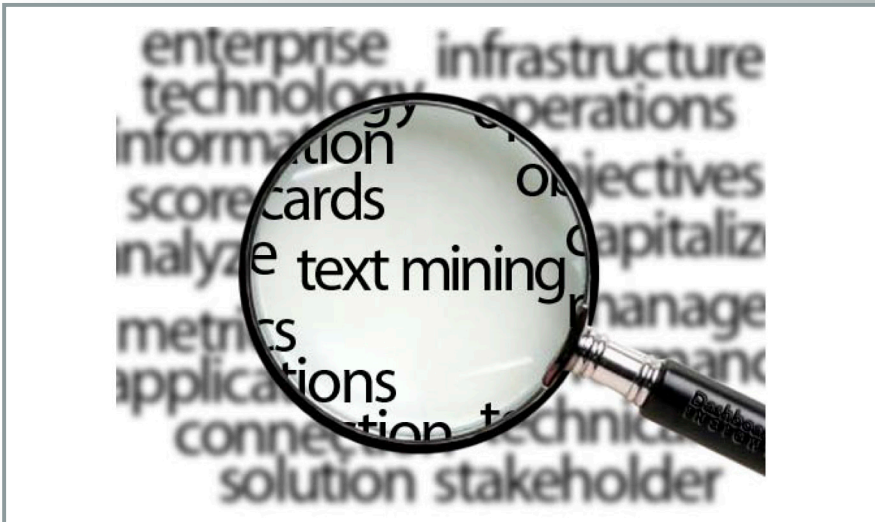


Detecting Temporal Precursor Words and Phrases Using a Learning Algorithm and Wavelet Analysis

UT-B IDs 200902235, 20102377



Technology Summary

Most research on mammography focuses on image data, not textual reports. However, the reports associated with patient visits offer a valuable set of observations. To take advantage of these sequential writings, a robust ORNL learning algorithm assembles, searches, and analyzes cue phrases in radiology reports to determine if they define normal or abnormal traits in mammograms over time.

Specifically, this system learns phrase patterns (skip bigrams) from textual documents and separates the documents into two distinct classes. The algorithm then performs longitudinal scans of mammogram records from patient visits, using the phrase patterns and a new wavelet analysis technique to detect precursors to breast abnormalities. Using this method, researchers found common phrases in both the normal and abnormal reports and were able to successfully detect common phrase patterns that uniquely identify two classes of documents.

In a follow-up system, the researchers combined the textual analysis algorithm with a discrete wavelet transform—a function in mathematics—to do a temporal analysis of precursor words in medical records. A critical feature of this method is that it will not only identify frequencies in a sequence, but also the point in time in which they occur.

Advantages

- Earlier detection of cancers and abnormalities over time
- Identification of patterns that define abnormal and normal physician reports
- Use of a single learning algorithm that can be used for an intelligent software agent

Potential Applications

- Early detection of breast cancer and other breast abnormalities
- Cyber security and text mining applications

Patents

Robert M. Patton, Thomas E. Potok, and Barbara G. Beckerman, *Detecting Temporal Precursor Words in Text Documents Using Wavelet Analysis*, U.S. Patent Application 61/310,351, filed March 4, 2010.

Robert M. Patton and Thomas E. Potok, *Method for Learning Phrase Patterns from Textual Documents*, U.S. Patent Application 61/331,941, filed May 6, 2010.

Lead Inventor

Robert M. Patton
Computational Sciences and Engineering
Division
Oak Ridge National Laboratory

Licensing Contact

David L. Sims
Technology Commercialization Manager,
Building, Computational, and
Transportation Sciences
UT-Battelle, LLC
Oak Ridge National Laboratory
Office Phone: 865. 241.3808
E-mail: simsdl@ornl.gov

