

NRCC - CNRC

*Institute for
Information
Technology*

Identifying ILI Cases from Chief Complaints

Comparing Keyword and Support Vector
Machine Methods



National Research
Council Canada

Conseil national
de recherches Canada

Canada

NRCC - CNRC

*Institute for
Information
Technology*

Authors

- **National Research Council of Canada, Institute for Information Technology**
 - Darrell Ferguson, MCS
 - Norman Vinson, PhD
 - Jason Morin, MSc
 - Joel Martin, PhD
- **University of Ottawa Heart Institute**
 - Susan McClinton, BScN
 - Richard Davies, MD PhD

Objective

- **Ottawa Public Health required a system to monitor Emergency Department Chief Complaints for “Influenza Like Illnesses” (ILI).**
- **We took on the task of identifying ILI from Chief Complaints.**
- **Two approaches were tested:**
 - Keywords: Scanning for significant keywords and phrases.
 - SVM: A machine learning text classifier based on Support Vector Machines.

Background: Using Chief Complaints

- **Easily available**
 - Already in electronic form.
- **Received early**
 - Early indicator.
- **Classifying as ILI**
 - ILI doesn’t mean a positive lab test.
 - Age is important.
- **Gold Standard classified by experts:**
 - May 2008 – June 2009.
 - 149910 case records.
 - Contained free text.

Background: What is an “ILI” Case?

- “Inclusive” (include any case that could be seen as ILI).
- Age ranges (less than 5, greater than 4) included as
 - Some symptoms depend on age: GI in young children.

Keyword Search

- **Experts provided:**
 - Symptoms
 - Synonyms
 - Rules for their application.
- **Used partially automated techniques to discover:**
 - Additional synonyms,
 - Misspellings (occurring more than once),
 - Abbreviations.
- **Achieved:**
 - Precision: 96.8%
 - Recall: 97.0%
 - Specificity: 99.6%

Support Vector Machines

- **A supervised learning method of classification.**
- **Features used:**
 - Single words from chief complaints.
 - Age range.
- **Achieved:**
 - Precision: 97.4%
 - Recall: 97.5%
 - Specificity: 99.7%

Keyword versus SVM

Method	Precision	Recall	Specificity
Keywords	96.8%	97.0%	99.6%
SVM	97.4%	97.5%	99.7%

- **Other comparisons:**
 - Keyword can be predictable and results are easier to decipher/"debug".
 - SVM can learn misspellings, synonyms, truncations, and when to apply while keyword has to be done by human and rules to fix can be complicated to set up for the user.
 - SVM requires feature selection, parameter searching.
 - SVM can be applied to other syndromes easily.

Future Work

- **French data.**
- **More data feeds:**
 - Test Results
 - Final Diagnosis
 - Patient Record
- **More Syndromes:**
 - Respiratory
 - Gastrointestinal
 - Constitutional
 - ...

Science
— at work for —
Canada

