Automated methods of adverse event detection: A critical review of the literature

Christian Rochefort, PhD
Junior 1 Investigator, FRQ-S
May 30 2013, Vancouver

Presentation outlines

- Background
  - Adverse events (AEs):
    - Incidence
    - Mortality, morbidity and costs
  - Overview of current methods of AE detection
  - Novel methods of AE detection
    - Natural language processing (NLP)
    - Machine learning (ML)
- Objectives
- Methods
- Results
- Discussion
  - Direction for future research
  - Recommendations for practice

Background

- Definition: AEs are injuries that are caused by the medical management rather than the underlying condition of the patient.
  - Incidence: 2.9% to 16.6% of all hospitalizations
  - Prevention: 30% to 58%;
  - Mortality: 44,000 to 98,000 death per year.
  - Costs: $17 to $29 billion.

- To prevent AEs, there is a need for accurate, timely and efficient methods for monitoring event rates.

Current methods of AE detection

Table 1. Advantages and disadvantages of methods used to identify adverse events (AEs) in acute care hospitals

Method	Avantages	Inconvénients
Manual chart review	• 'Gold standard'	• Costly • Resource-intensive • Time-consuming
Prevalence surveys (direct observation)	• Potentially accurate and precise • Provides data otherwise unavailable • Detect more active errors than other methods.	• Costly • Difficult to train reliable observers • Potential Hawthorne effect • Not good for detecting latent errors
Incident and accident reports (Spontaneous and voluntary reporting)	• Can detect latent errors	• Reporting bias • Resource intensive as manual review is needed • Underestimate true incidence of AEs by a factor of about 20.
Administrative data (discharge diagnostic codes)	• Not expensive • Easy to access and use • Available for large populations of patients	• Low to moderate sensitivity • Low to moderate positive predictive value • Not dated with accuracy.

Novel methods of AE detection

- **<u>Method 1</u>**: Natural language processing (NLP) / Electronic triggers
  - Definitions:

NLP	Natural language processors are systems that use linguistic (e.g. semantic analysis) and pattern matching methods (e.g., regular expressions) to extract from narrative text reports information that suggest the presence or the absence of an AE.
Electronic triggers	Abnormal lab values, use of antidotes drugs.

  - Example:
    - CDC criteria and nosocomial pneumonia;
      - CXR / CT Chest with consolidation, infiltration or cavitation;
      - WBC > 12 000/mm^3 ou WBC < 4000/mm^3
      - Temperature > 38°C

Automated method of AE detection

- **<u>Method 2</u>**: Statistical text categorization (Machine learning)

<u>Step 1</u>: Manual coding of the narrative reports

[redacted] TYPE/EXAM [redacted]
CT/CT THORAX WO CON

Indications: Pneumonia. History of allergy to contrast.

Technique: Spiral CT of the chest was performed with no intravenous contrast.

Findings: There is marked artifact from the right shoulder replacement. A nasogastric tube and endotracheal tube are noted. There is mild mediastinal adenopathy. Lack of intravenous contrast limits the sensitivity for hilar adenopathy. There is cardiomegaly. There is a moderate right pleural effusion and a small left pleural effusion. There are extensive areas of consolidation in the right lung and patchy areas of infiltrate in the left lung.

There is ill-defined 1.5 cm area of hypodensity in the left lobe of liver image 40.

Atherosclerotic coronary artery calcifications are noted.

Impression:

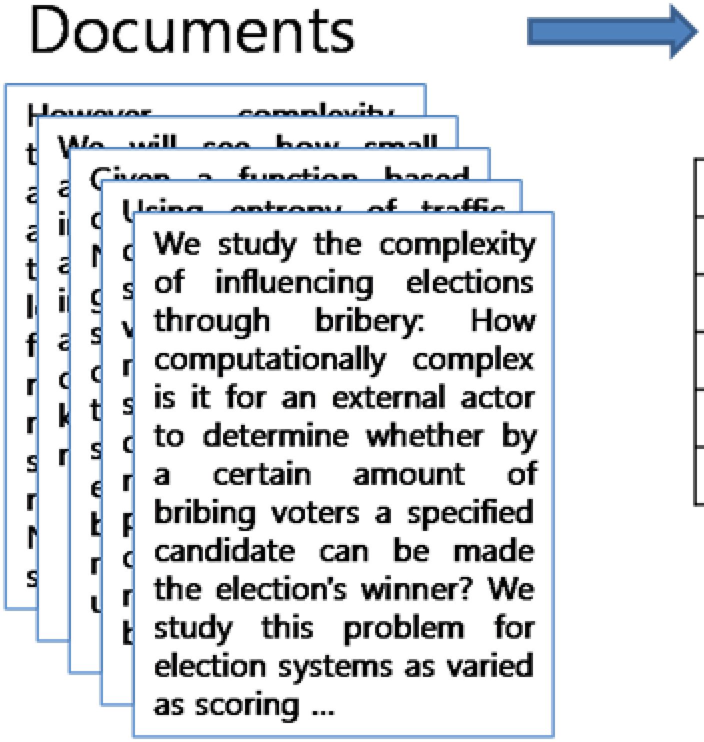
1. Bilateral pleural effusions and bilateral infiltrates right much worse than left with mediastinal adenopathy

2. Cardiomegaly and atherosclerotic coronary artery calcifications

Statistical text classification

<u>Step 2</u>: Generate a term-by-document matrix (TDM)

Documents

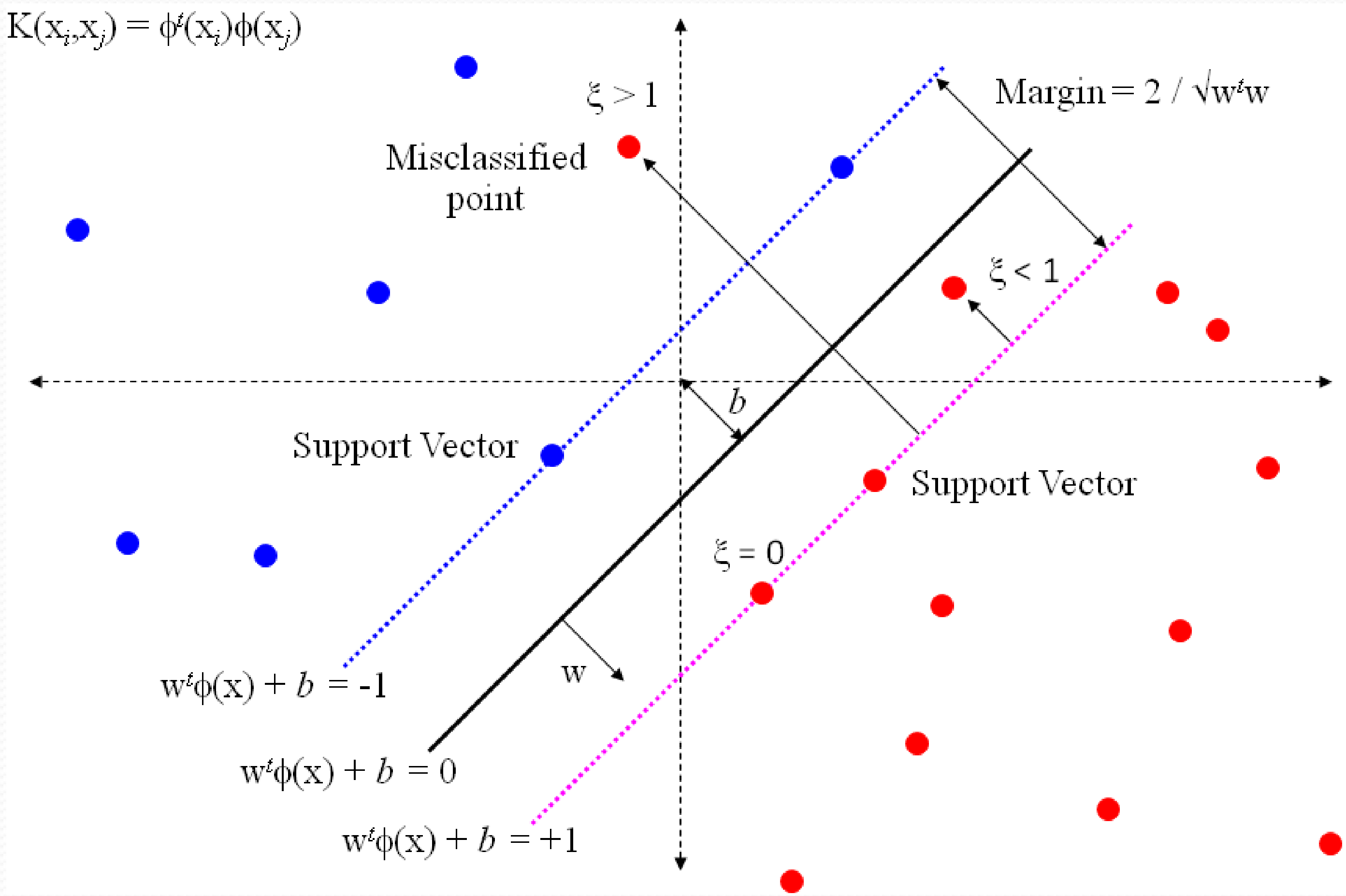


Vector-space representation

	D1	D2	D3	D4	D5
complexity	2		3	2	3
algorithm	3			4	4
entropy	1			2	
traffic		2	3		
network		1	4		

Term-document matrix

Statistical text classification
Étape 3: Machine learning (SVM)



Objective

- To perform a systematic review of the studies examining the accuracy of natural language processing (NLP) systems or machine learning (ML) models for the purpose of identifying AEs in acute care hospitals.

Methods

- Design:
  - Literature review
- Data sources:
  - PubMed: from January 1980 to May 2013
  - ‘Related article’ feature of PubMed
  - Bibliographic review of the retrieved articles
- Inclusion criteria:
  a) conducted in an inpatient setting,
  b) described an NLP or ML method of AE detection, and
  c) assessed the accuracy of this method in comparison with a reference standard assessment of the medical chart (or equivalent).

Results

- From 76 articles retrieved, 20 met the inclusion criteria.
- The methods evaluated:
  - 15 (75%) used natural language processing (NLP)
  - 5 (25%) used machine learning (ML) models
- The events detected:
  - 56% - general adverse events (e.g., fall, DVT, PE, patient ID)
  - 38% - nosocomial infections (e.g., pneumonias)
  - 6% - other specific events (e.g., renal failure, myocardial infarction).

Results

- Studies using natural language processing (NLP)
  - Accuracy:
    - Sensitivity: .23 - .70
    - Specificity: .48 - .97
    - Positive predictive value: .41 - .84
    - Negative predictive value: .74 - .96
  - Accuracy varies with:
    - Complexity of the task
    - Number and type of data sources used

Results

- Studies using machine learning (ML) models
  - Accuracy
    - Sensitivity: .85 - .98
    - Specificity: .78 - .93
    - Positive predictive value: .72 - .98
    - Negative predictive value: .96 - .99
  - Accuracy:
    - Appears to be more homogenous than with NLP systems
    - Does not seem to be influenced by the complexity of the task
    - Does vary with the type of data sources
      - Discharge summaries result in lowest accuracy
      - Progress notes and radiology reports result in highest accuracy.

Discussion

- Machine learning (ML) models may be more accurate than natural language processing (NLP) systems.
  - It is possible that ML models are simply more accurate than NLP systems for the purpose of identifying adverse events;
  - It is also plausible that ML classifiers trained for a specific tasks (e.g., identifying DVTs) are also more accurate than general-purpose NLP classifiers (e.g., MedLEE, SymText).

- Additional studies are required to test these hypotheses.

Discussion

- Studies have compared the accuracy of NLP with that of the AHRQ Patient Safety Indicators (diagnostic codes)
  - Adverse events selected:
    - Acute renal failure
    - Pulmonary embolism / deep vein thrombosis
    - Sepsis
    - Pneumonia
    - Myocardial infarction
  - Main findings:
    - NLP has significantly higher sensitivity
    - NLP has slightly but significantly lower specificity

Murff et al., JAMA, 2011; 306(8):848-855

Conclusions

- Studies suggest that the accuracy of:
  - NLP > discharge diagnostic codes
  - ML may be > than NLP
- There is a need for:
  - comparative studies of these methods
  - studies examining the external validity of these methods
- Both NLP and ML:
  - Require integrated electronic medical record (EMR)
  - Are technically challenging to implement