Using expert review to calibrate semi-automated adjudication of vital sign alerts in Step Down Units

Madalina Fiterau (mfiterau@cs.cmu.edu)
Donghan Wang (donghanw@cs.cmu.edu)
Artur Dubrawski (awd@cs.cmu.edu)
Gilles Clermont (cler@pitt.edu)
Marilyn Hravnak (mhra@pitt.edu)
Michael R. Pinsky (pinsky@pitt.edu)

Society of Critical Care Medicine Annual Congress

January 2016
Disclosures

• Funded by awards:
  – National Science Foundation
    • awards 0911032 and 1320347
  – National Institutes of Health
    • award R01NR013912

• No commercial conflicts of interest
Background

- Patients are monitored using non-invasive vital sign (VS) data
- Alerts issued when a VS exceeds predefined thresholds
- Many alerts are artifacts, due to threshold-based issuance
- Artifacts cause alarm fatigue
- Machine Learning has proven useful in classifying clinical data
- Training data requires laborious expert annotation
Objective

Reduce expert annotation effort through semi-automatic adjudication of VS alerts as real or artifacts, while maintaining high accuracy.
Data Description

- Heart Rate < 40 or > 140
- Respiratory Rate < 8 or > 36
- Systolic Blood Pressure < 80 or > 200
- Diastolic Blood Pressure > 110
- $SPO_2 < 85\%$

Alerts

some are artifacts, not true alerts

Features computed from time series include common statistics of each VS: mean, stdev, min, max, range of values, duty cycle ...
Expert Review System

Artifact adjudication models

- **SPO₂** model trained on 91 committee-labeled events
- **RR** model trained on 194 committee-labeled events
Informative Projections

- Extract **simple projections** of data in which alerts appear as either convincingly correct or easily dismissible.
Expert Review System

Artifact adjudication models

- **SPO$_2$** model trained on 91 committee-labeled events
- **RR** model trained on 194 committee-labeled events
Review based on projections

New alert can be **confidently adjudicated** with the informative projection.
Chart-based review
Study Results

Expert review based on Informative Projection (80 alerts)

Experts agree with each other regarding label. (32 alerts)

Experts disagree (48 alerts)

Chart review

Experts agree (31 alerts)

Experts disagree (17 alerts)

32 alerts are confidently adjudicated.

31 alerts are confidently adjudicated.

17 alerts are ambiguous.
Adjudication error

Number of alerts

Correct classification from projections
Correct classification from chart
Incorrect classification
Alerts marked ambiguous

Semi-automated adjudication reduces error
Conclusions

- Half of alerts that can be classified are handled automatically
- 3 ways ML reduces expert labeling effort
  - Use of ML models for semi-automatic adjudication
  - Active sample selection for expert review
  - Threshold adjustment maximizes confident adjudication
- 1/5 of alerts could not be classified by system or reviewers
- Semi-automated adjudication model filters out artifactual alerts, helping to reduce alarm fatigue