DETECTING ARTIFACTS IN CLINICAL DATA THROUGH PROJECTION RETRIEVAL

Madalina Fiterau (mfiterau@cs.cmu.edu)
Artur Dubrawski (awd@cs.cmu.edu)
Karen Chen (lujiec@cs.cmu.edu)

Marilyn Hravnak (mhra@pitt.edu)
Gilles Clermont (cler@pitt.edu)
Michael R. Pinsky (pinskymr@upmc.edu)

IMPORTANCE OF ARTIFACT DETECTION

- Clinical monitoring systems are designed to process multiple sources of information about the current health condition of a patient and issue an alert whenever a change of status, typically an onset of some form of instability, requires attention of medical personnel.
- In practice, a substantial fraction of these alerts are triggered by malfunctions or inaccuracies of the monitoring equipment. Accidentally detached ECG electrodes, transient readings from a dislocated blood oxygenation probe yield instability alerts.
- Frequency of false detections leads to lowering sensitivity of personnel to alerts. In order to maintain and enhance effectiveness of care, it is important to reliably identify and explain the non-consequential artifacts.

DATA DESCRIPTION

A prospective longitudinal study recruited admissions across 8 weeks to a 24 bed trauma and vascular surgery stepdown unit. Noninvasive vital sign (VS) monitoring consisted of 5-lead electrocardiogram to determine:

- heart rate (HR)
- respiratory rate (RR; bioimpedance)
- systolic (SBP) and diastolic (DBP) blood pressure (oscillometric)
- peripheral arterial oxygen saturation (SPO2) by finger plethysmography

Vital signs were analyzed beyond local instability criteria:

- HR<40 or >140, RR<8 or >36, systolic BP <80 or >200, diastolic BP>110, SpO2<85%.

PROBLEM FORMULATION

We generalize the **Informative Projection Retrieval** problem (IPR) for a learning task:

- an optimization over a model class containing a set of solvers each using a small number of features;
- the solvers are used alternatively – each sample is assigned to a solver;

$$\mathcal{M} = \{\ P = \{\pi : \pi \in \Pi, dim(\pi) \leq d\},$$ Small set of projections

$$T = \{\tau_i : \tau_i \in \mathcal{T}, \tau_i : \mathcal{X} \to, \mathcal{Y}\ \forall i = 1 \dots |P|\},$$ Solvers

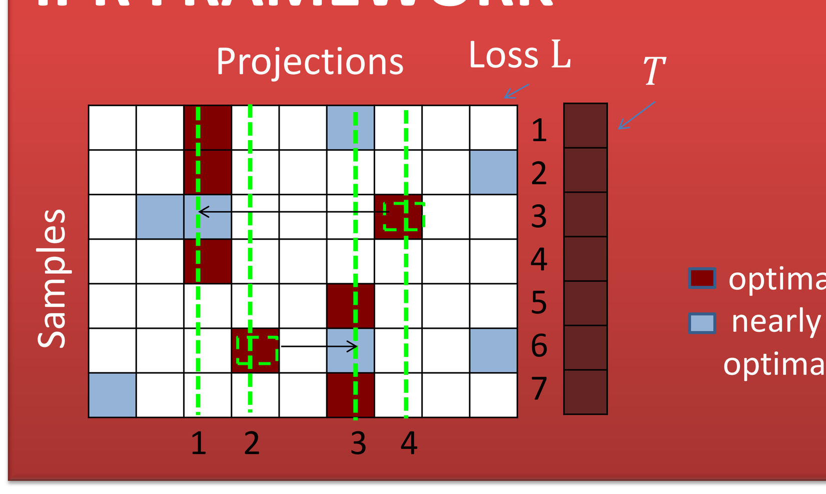
$$g \in \{f : \mathcal{X} \to \{1 \dots |P|\}\}$$ Selection function

Target model

- best model obtained by minimizing expected risk over $\mathcal{M}$ given the task-specific loss ℓ;

$$M^* = argmin_{M \in \mathcal{M}_d} \mathbb{E}_{\mathcal{X}}\, \ell\left(\tau_{g(x)}\left(\pi_{g(x)}\right), y\right)$$

IPR FRAMEWORK



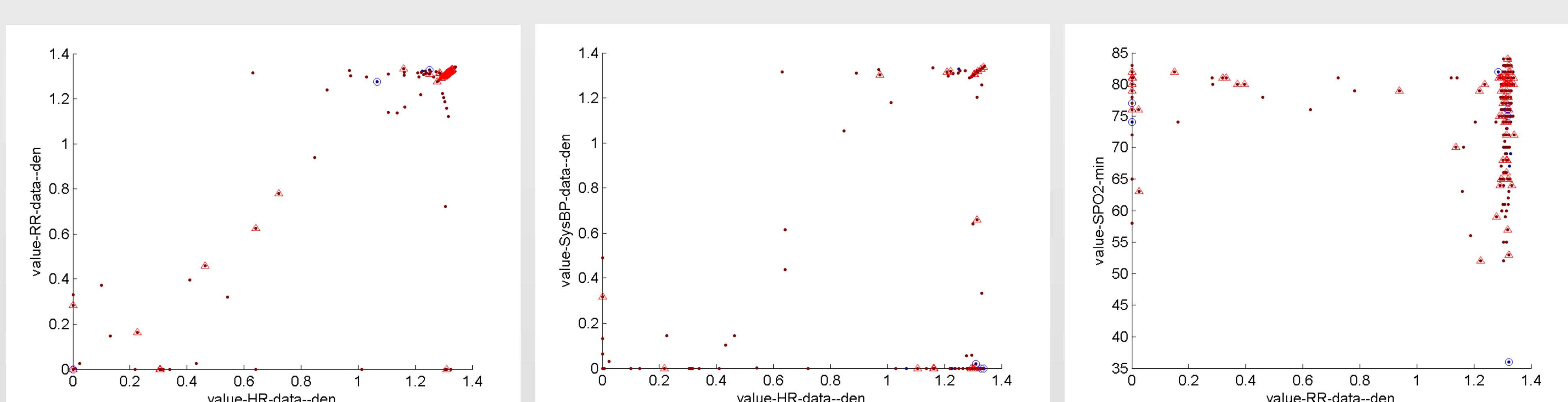
The aim is to find a set of **few projections** for which the entropy contributions are **close to the optimum**.

Define: $T_i = min_j L_{ij}$

The algorithm biases the projection selection toward 'popular' projections through a multiplier δ.

$\delta = [1 \dots 1]$
repeat
  $B = \arg\min_B ||L^* - L \odot B||_2^2 +$
   $\lambda_1 \sum_{j=1}^{d^*} ||B_{.,j}||_{\ell_1} + \lambda_2 |B\delta|_{\ell_1}$
   subject to
   $||B_{k,.}||_{\ell_1} = 1 \qquad k = 1 \dots n$
  $\delta_k = ||B_{.,j}||_{\ell_1} \qquad j = 1 \dots d^*$ (update muliplier)
  $\delta = (||\delta||_{\ell_1} - \delta)/||\delta||_{\ell_1}$
until δ converges
$\Pi = \{\pi_i; \quad |B_{.,i}|_{\ell_1} > 0 \quad \forall i = 1 \dots d^*\}$
return Π

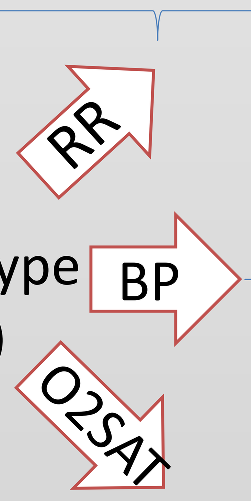
ARTIFACT CLASSIFICATION MODELS

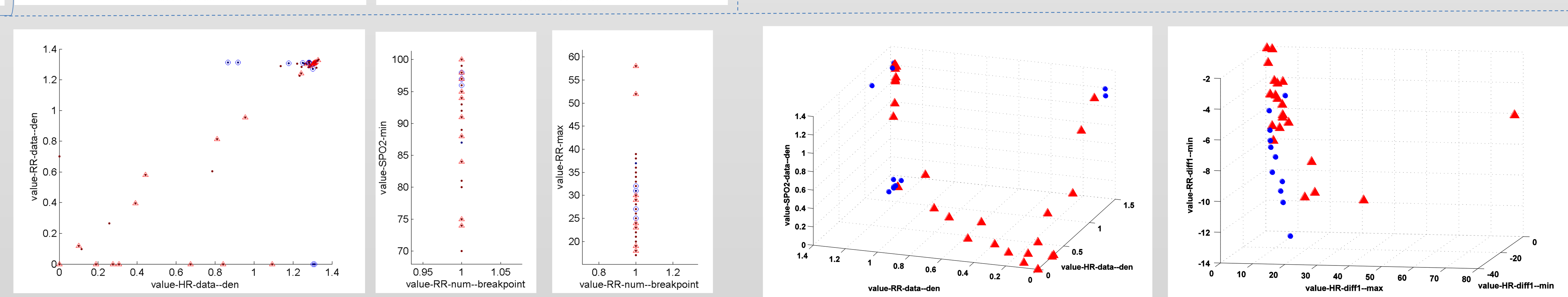


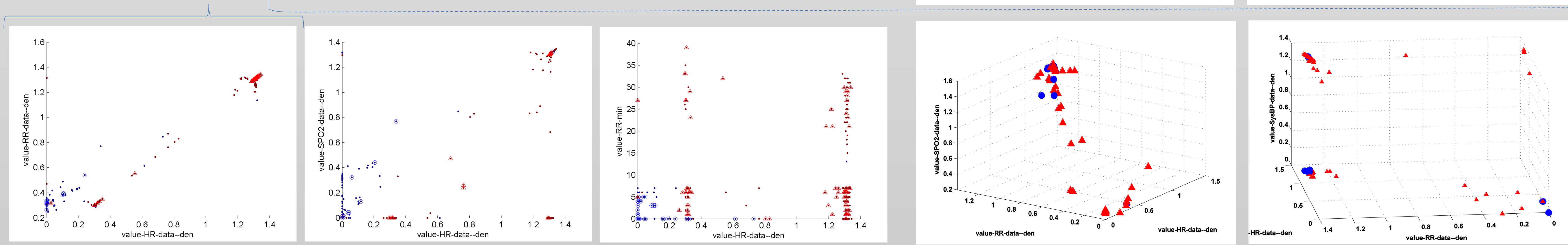
- Each alert is associated with a category indicating the first abnormal vital
- 812 alerts of 3 types: respiratory rate, oxygen saturation, blood pressure
- Features computed from each vital signal independently:
  - during the duration of each alert and a short window (of 4 minutes) preceding alert onset
  - include common statistics of each vital signal such as mean, standard deviation, minimum, maximum and range of values

Alarm Type	RR	BP		SPO2	
	2D	2D	3D	2D	3D
Accuracy	0.98	0.833	0.885	0.911	0.9151
Precision	0.979	0.858	0.896	0.929	0.9176
Recall	0.991	0.93	0.958	0.945	0.9957

Models for each alert type (2D and 3D)



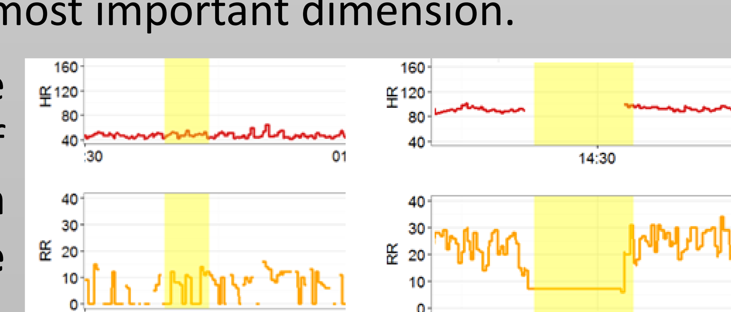




CASE STUDY: OUTLIER DETECTION

A good indication – as stated by experts – of the **invalidity of a RR alert** is the **lack of HR data**. A decision rule used by clinicians tests whether there HR data is available. In classifying RR-based alerts, the algorithm **correctly picked HR data density** as the most important dimension.

The graph marked with * contains two samples that would be classified as non artifacts. Both have continuous streams of data, but the **RR signals are irregular** – an **uncommon artifact**. Further investigation showed that **variance of the signal values** provides a reliable way to detect these outliers.



CASE STUDY: FINDING ERRORS IN DATA

Some samples were **classified by the system as artifacts** while the **domain experts considered them true alerts**. On closer inspection, they seemed to exhibit artifact-like features - with little or no recorded values in the HR signal.

When we drilled down to look at the data, we found that the **samples were actually labeled incorrectly** in the training set. Therefore, RIPR can also be useful in detecting inconsistencies due to human error.

SUMMARY

- The retrieved low-dimensional projections make it possible for domain experts to quickly validate the assigned labels
- The models aided experts in deriving labeling rules
- The method was used to point out uncommon cases and mislabeled data

Thus, the proposed framework promises to be useful to clinicians by partially annotating medical data in a human understandable and intuitive manner.

ONGOING RESEARCH

- Active labeling: using active learning to pick sets of samples to be annotated by domain experts has the potential to
  - Reduce the amount of manual labeling
  - Improve performance by quickly finding sub-models dealing with common cases and then shifting focus to difficult ones
- Multilabel learning: alerts are actually due to several vitals; considering the correlations between outputs could result in better models