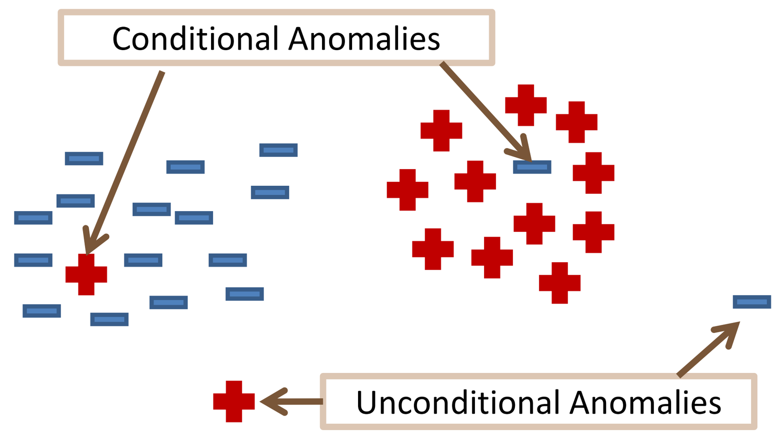
Conditional Anomaly Detection Using Soft Harmonic Functions: An Application to Clinical Alerting

MICHAL VALKO, HAMED VALIZADEGAN, BRANISLAV KVETON, GREGORY F. COOPER, MILOS HAUSKRECHT

Motivation



- traditionally: anomalies in the data
- we want to detect **anomalies in responses**
- conditioning on the remaining features/covariates
- very useful for medical applications
  - action anomalies: lab orders and medications
  - budget control, overspending

Background

Goal: Conditional Anomaly Detection

- detect anomalous decisions
- robust to traditional outliers

Problem statement (★): Given a set of n past observed examples $(\mathbf{x}_i, y_i)_{i=1}^{n}$ (with possible label noise), check if any instance i in recent m examples $(\mathbf{x}_i, y_i)_{i=n+1}^{n+m}$ is unusual.

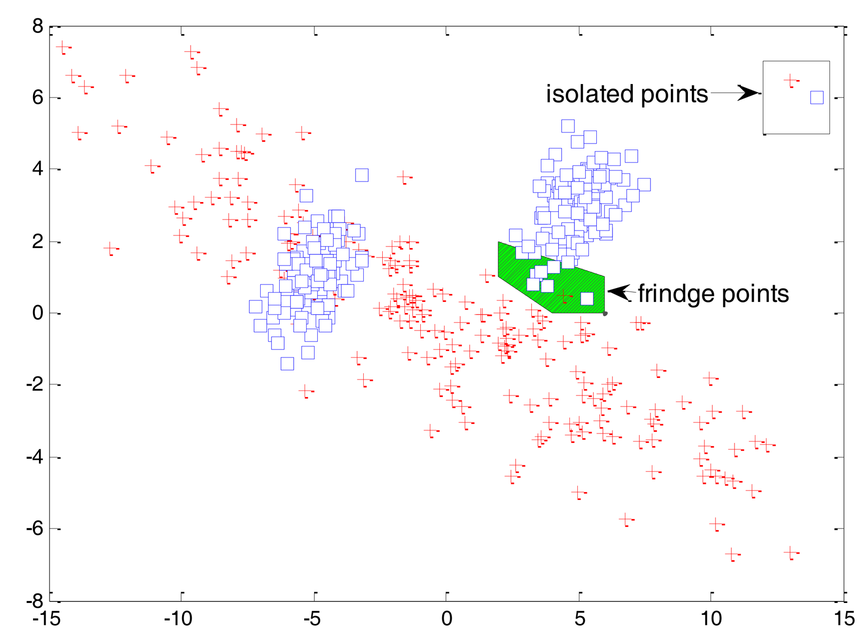
Alternative methods:

- class outlier approach
  - take traditional anomaly detection method
  - detect anomalies within the same class
  - cons: ignores the other classes
- discriminative approach
  - difference between predictions and labels
  - cons: sensitive to fringe and isolated points

*Our method takes all classes into account and uses **regularization** to avoid unwanted behavior.*

Challenges

- underlying density is often unknown
- high-dimensional and non-linear data
- fringe points (on the boundary support)
- isolated points (unconditional outliers)



Algorithm

- graph-based representation

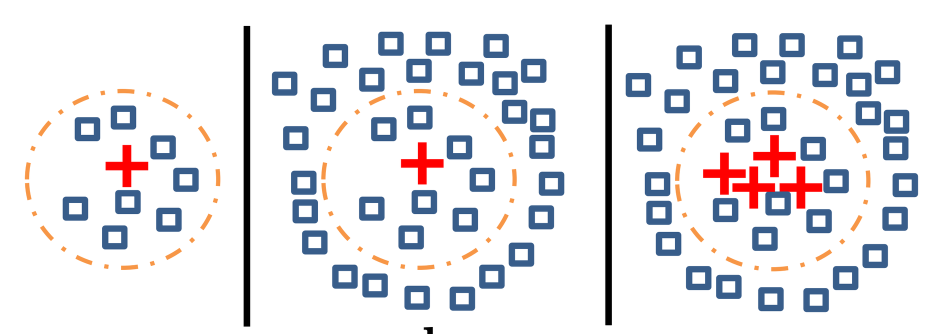
$$w_{ij} = \exp\left[-\left(\|\mathbf{x}_i - \mathbf{x}_j\|^2_{2,\psi}\right)/\sigma^2\right]$$

- label propagation on graph

$$\boldsymbol{\ell}^\star = \min_{\boldsymbol{\ell}\in\mathbb{R}^n} (\boldsymbol{\ell} - \mathbf{y})^\mathsf{T} C(\boldsymbol{\ell}-\mathbf{y}) + \boldsymbol{\ell}^\mathsf{T} K \boldsymbol{\ell}$$

- regularization to prevent unwanted anomalies
- checking for inconsistencies

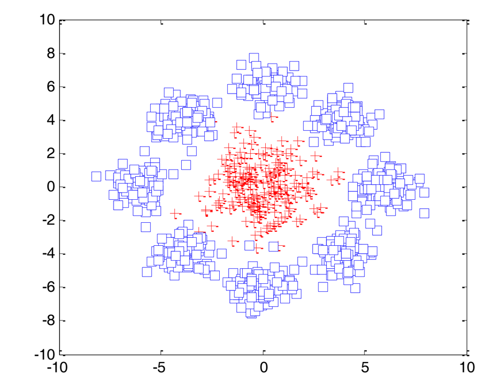
$$\boldsymbol{\ell}^\star = \left((c_l I)^{-1}(\mathcal{L}(W)+\gamma_g)+I\right)^{-1}\mathbf{y}$$



- addressing computational complexity
  - create a backbone graph
  - make the calculation on a smaller graph
  - compact computation

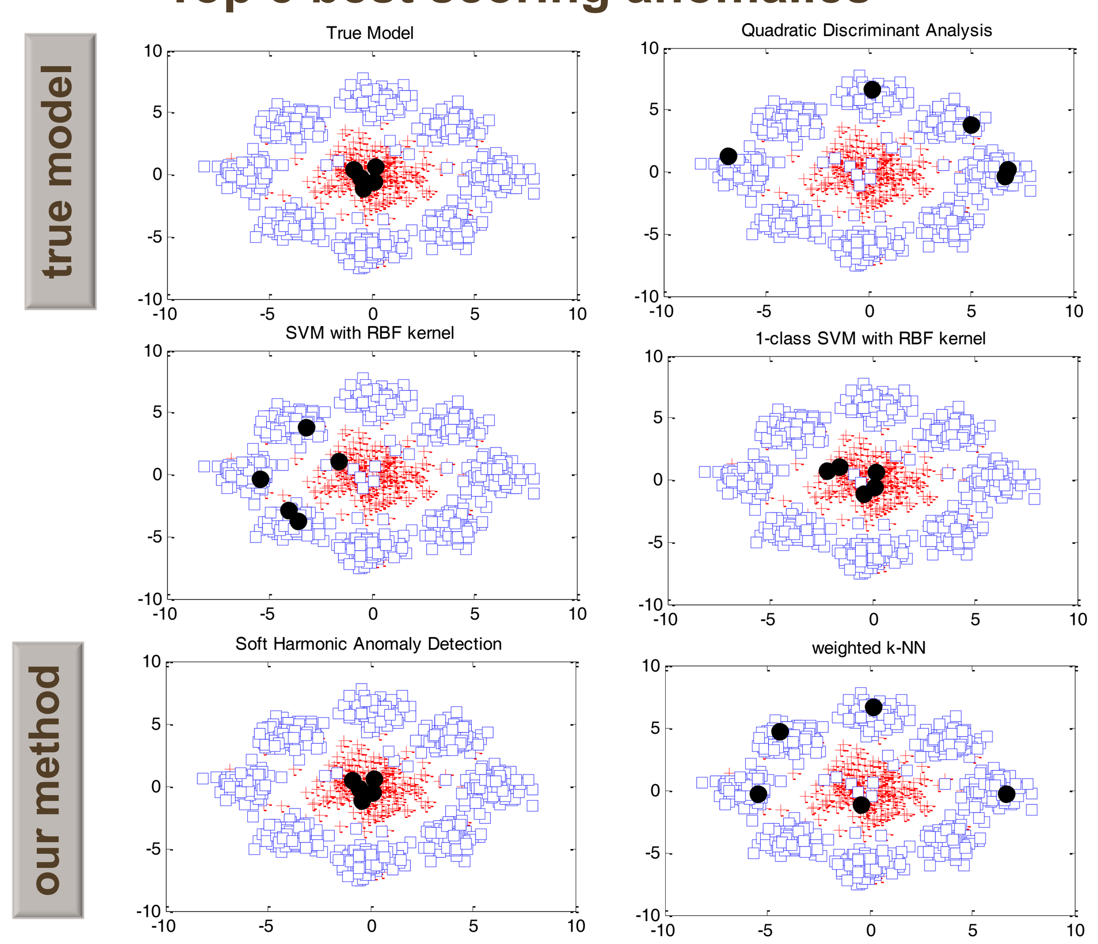
Comparison on Synthetic Data

- evaluation of (conditional) anomaly methods is very challenging
- synthetic data with known distribution
- flip 3% of the labels
- compare how the anomaly score agrees with true score



QDA	61.0% (1.2)
SVM	46.1% (3.1)
1-class SVM	64.7% (0.7)
[redacted]	
wk–NN	61.4% (2.1)

Top 5 best scoring anomalies



Comparison on UCI ML Datasets

- ordinal response used to calculate the true score

	Wine Quality	Housing	Auto MPG
QDA	75.1% (1.3)	56.7% (1.5)	65.9% (2.9)
SVM	75.0% (9.3)	58.5% (4.4)	37.1% (8.6)
1-class SVM	44.2% (1.9)	27.2% (0.5)	50.1% (3.5)
wk–NN	67.6% (1.4)	44.4% (2.0)	61.4% (2.3)
SoftHAD	74.5% (1.5)	71.3% (3.2)	72.6% (1.7)

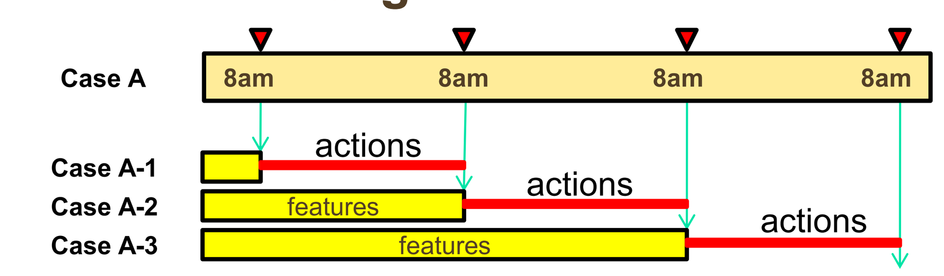
Contributions

- non-parametric and graph-based method for conditional anomaly detection
- takes advantage of the data structure
- important application for medical data
- robust to fringe and isolated points

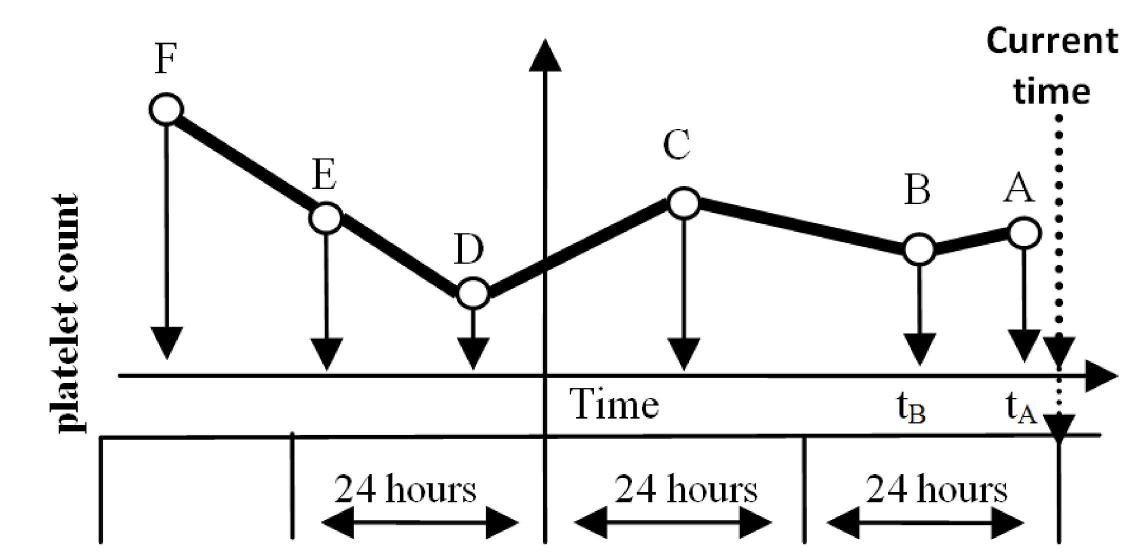
Results on Clinical Data (EHRs)

- medical health records (UMPC)
- 4486 patients (50K instances, 9K features)
- 749 laboratory tests or medication orders
- 222 instances evaluated
- panel of 15 expert clinicians (3 per instance)
- evaluation metric: area under ROC

Case Segmentation of EHR



Feature Construction from EHRs



Last slope = (B-A) / (tB-tA)   Drop from baseline = F-A

Outperforming SVM method over the range of settings of regularization parameters