PROJECTION RETRIEVAL FOR CLASSIFICATION

Carnegie Mellon University

Madalina Fiterau
mfiterau@cs.cmu.edu

Artur Dubrawski
awd@cs.cmu.edu

Auton Lab

MOTIVATION

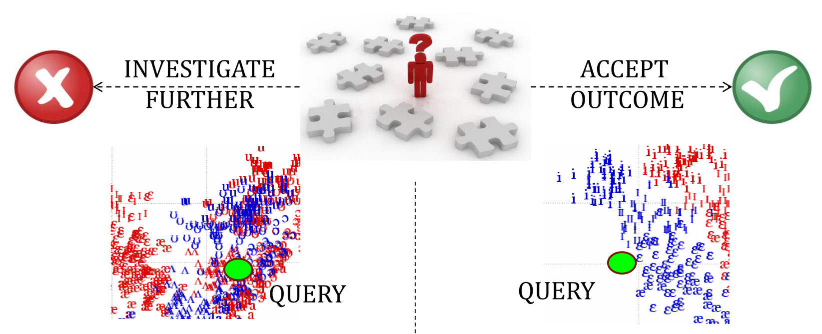
Automated Decision Support Systems

Our method targets applications where a **human operator** is involved in the decision. The process must be:

- Transparent
- Comprehensible

Thus, the problem of **finding subspaces** where data is classified with **high accuracy** but which also give operators **confidence** in the predictions.

We call this the **Projection Retrieval for Classification** (PRC) problem.

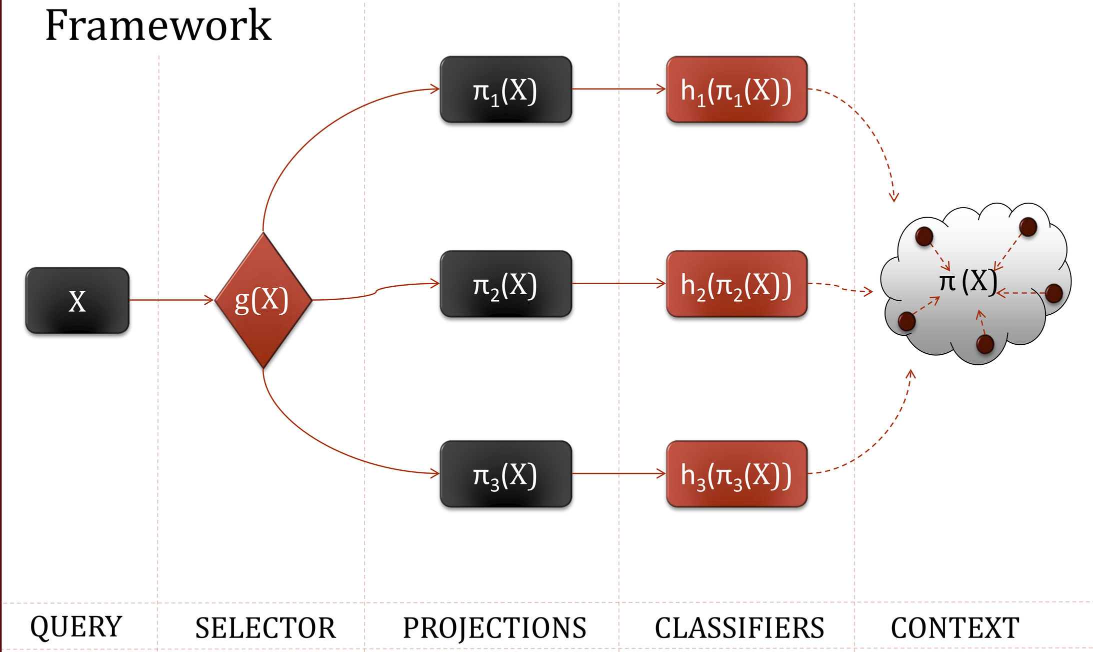


User is in control of the choice:

- Investigate Further – expensive
- Accept Outcome - assume responsibility

PROBLEM FORMULATION

Framework



User-System Interaction:

1. the user provides the system a **query** point;
2. system finds a **query-specific projection**
3. system displays result and **illustration** of how the label was obtained

Model components:

- Set of d-dimensional, axis-aligned sub-spaces of the original feature space $P \in \Pi$
- Each projection has a corresponding discriminator from the hypothesis class $\mathcal{H}$.

Dataset → $\{(x_1, y_1) \dots (x_n, y_n)\} \in \mathcal{X}^n \times \mathcal{Y}^n\}$ Small set of projections

$$\mathcal{M}_d = \{P = \{\pi : \pi \in \Pi, dim(\pi) \le d\},\; H = \{h_i : h_i \in \mathcal{H}, h_i : \pi_i \to \mathcal{Y}, \forall i = 1 \dots |P|\},\; g \in \{f : \mathcal{X} \to \{1 \dots |P|\}\}\}$$

Target model

Discriminators

Selection function

- A selection function g, which yields, for a query point x, the projection/discriminator pair $(\pi_{g(x)}, h_{g(x)})$ for the point. $h_{g(x)}(\pi_{g(x)}(x))$ represents the predicted label for x.

The aim is to minimize the expected classification error over $\mathcal{M}$.

Formulation of the Projection Retrieval problem:

$$M^* = argmin_{M \in \mathcal{M}_d} \mathbb{E}_{\mathcal{X}}[y \ne h_{g(x)}(\pi_{g(x)}(x))]$$

ENTROPY-BASED OBJECTIVE

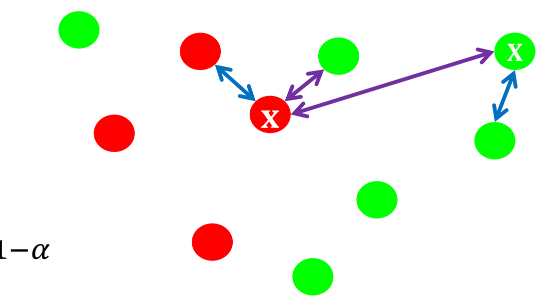
It is expected that each projection would benefit different areas of the feature space: $\mathcal{A}(\pi) = \{x \in \mathcal{X} : \pi_{g(x)} = \pi\}$

$$M^* = argmin_{M \in \mathcal{M}_d} \sum_{\pi \in \Pi} p(\mathcal{A}(\pi))\, \underbrace{H(Y|\pi(X); X \in \mathcal{A}(\pi))}_{\text{independent of } \mathcal{H}}$$

Adapted objective by substituting **conditional entropy** for **prediction error**.

REGRESSION ON ENTROPY CONTRIBUTIONS FOR INFORMATIVE PROJECTIONS

1. Local Entropy Estimators



The neighbor-based estimator for conditional entropy:

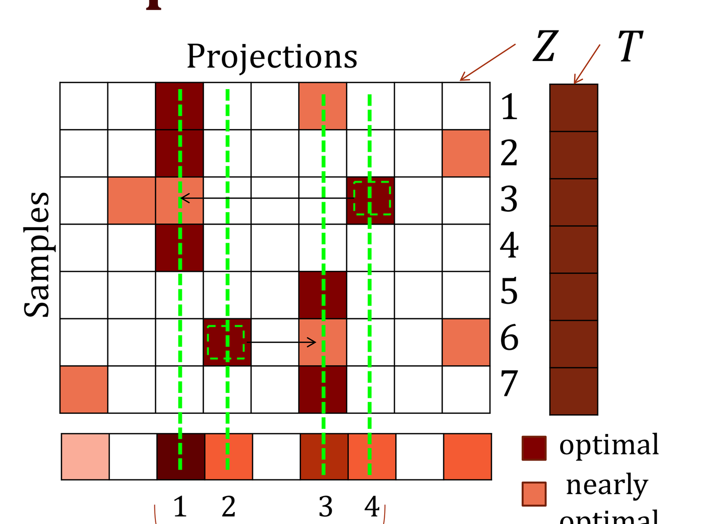
Based on the divergence estimator by Poczos and Schneider, *"On the estimation of alpha-divergences"* (AI Statistics 2011)

$$\hat{H}(Y|X \in \mathcal{A}) \propto \frac{1}{n}\sum_{i=1}^{n} I[x_i \in \mathcal{A}] \left(\frac{n-1}{n}\left(\frac{dist_{k+1}(x_i, X_{y_i})}{dist_k(x_i, X_{\neg y_i})}\right)^{\dim(X)}\right)^{1-\alpha}$$

For a projection π, we'll use the estimator $\hat{H}(Y|\pi(X); X \in \mathcal{A}(\pi))$.
The optimal model can be computed through the minimization:

$$\widehat{M} = argmin_{M \in \mathcal{M}_d} \sum_{\pi_j \in \Pi} \sum_{i=1}^{n} \underbrace{I[g(x_i) \to \pi_j]}_{B_{ij}\text{- selection matrix}} \underbrace{\left(\frac{dist_{k+1}(\pi_j(x_i), \pi_j(X_{y_i}))}{dist_k(\pi_j(x_i), \pi_j(X_{\neg y_i}))}\right)^{\dim(\pi_j)(1-\alpha)}}_{Z_{ij}\text{-local entropy contributions}}$$

2. Optimization Procedure



RECIP limits the number of projections in the model

RECIP aims to find a set of **few projections** for which the entropy contributions are **close to the optimum**.

Define: $T_i = min_j Z_{ij}$

In the example, the optimal values are marked with dark red. Using 4 projections, the minimal value for the entropy estimator is obtained, however projections 2 and 4 are only useful for one point each.

Clearly, some of the points will need to settle for sub-optimal projections; in the example, points 3 and 6.

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

ITERATE UNTIL CONVERGENCE

1. Get estimate of selection matrix B — $min_B \parallel T - Z \otimes B\mathcal{J}_{|\Pi|,1} \parallel_2^2 + \lambda \sum_{k=1}^{|\Pi|} |B_k|_1$
2. Compute multiplier δ inversely proportional with projection popularity — $\delta_k = |B_k|_1, \quad \delta = 1 - \delta/|\delta|_1$
3. Obtain new selection matrix B penalizing Bδ — $min_B \parallel T - Z \otimes B\mathcal{J}_{|\Pi|,1} \parallel_2^2 + \lambda |B\delta|_1$

where $Z_{ij} \otimes B_{ij} = Z_{ij} B_{ij}$

3. Selection Function

After obtaining a small, stable set of projections P, there is still the question of selecting the appropriate one for a specific query q. An immediate solution is to select the projection in P for which the entropy contribution at x is the smallest considering all possible labels of q.

$$(\hat{k}, \hat{y}) = argmin_{(k,y)} \left(\frac{dist_k(\pi_k(q), \pi_k(X_y))}{dist_k(\pi_k(q), \pi_k(X_{\neg y}))}\right)^{\dim(\pi_k)(1-\alpha)}, \text{ where } k \in \{1 \dots |P|\} \text{ and } \alpha \approx 1$$

CONCLUSIONS

- RECIP is a principled, regression-based algorithm that solves PRC
- RECIP optimizes the selection using point-specific entropy estimators
- RECIP recovers intuitive projections that aid operator decisions

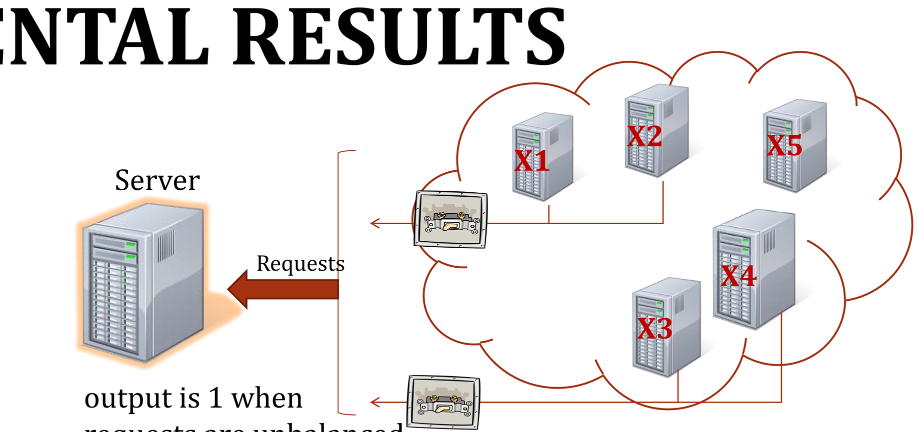
RESEARCH DIRECTIONS

- Efficient computation of entropy contribution using metric trees
- Adapting and distributing RECIP systems for specific applications
- Projection Retrieval for clustering, regression, multitask classification

EXPERIMENTAL RESULTS

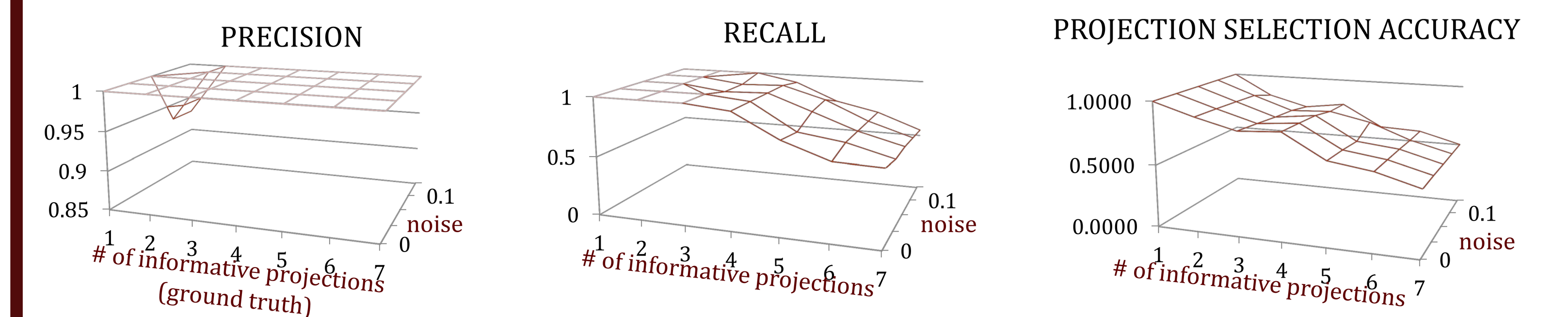
Artificial Data

The scenario involves a server which receives requests from pairs of machines. In this case, at a given time, requests cannot come from X1 and X2 simultaneously. Thus, (X1, X2) is an informative projection. The server is occasionally overloaded (Y=1). We would like to predict whether a given workload configuration will overload the server. We generated sets of such data with 10 features and 1 to 7 informative projections and varying noise levels.



output is 1 when requests are unbalanced

$$f(X_1, X_2, X_3, X_4, X_5) = \begin{cases}(X_1 > 0.5) \oplus (X_2 > 0.5)\\ (X_3 > 0.5) \oplus (X_4 > 0.5)\end{cases}$$



The retrieved projections are informative with high probability.

Most informative projections are recovered.

RECIP is generally capable of selecting the correct projection to deal with a query.

The problem of retrieving informative projections is difficult because of the feature overlap. The recovery success rate decreases as the number of relevant projections - and implicitly their feature overlap – increases.
The dataset noise represents the proportion of points that do not follow the model. Noise does not have a significant impact on the performance – even when 20% of the data points are noisy, the right projections are still recovered.

UCI Data

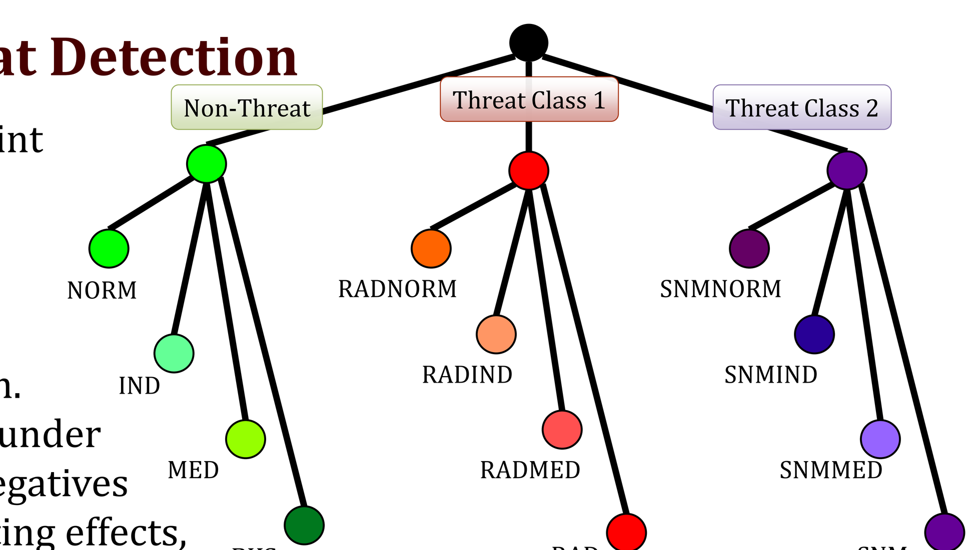
Dataset	KNN (all dimensions)	Number of features	RECIP	# of RECIP projections	d. of RECIP projections	% change in accuracy	% reduction in max dim.
MiniBOONE	0.7896	10	0.7396	1	1	6.33	90
Breast Cancer Wis	0.8415	11	0.8275	4	1	1.66	90.90
Spam	0.7680	10	0.7680	5	{1,2,3,3,3}	0	70
Vowel	0.9839	10	0.9839	1	10	0	0
Breast Tissue	1.0000	10	1.0000	1	2	0	80
Canes	0.7788	50	0.7807	5	1	**-0.24**	**98**
Cell	0.7072	6	0.7640	4	{1,2,2,2}	**-8.03**	**66.67**

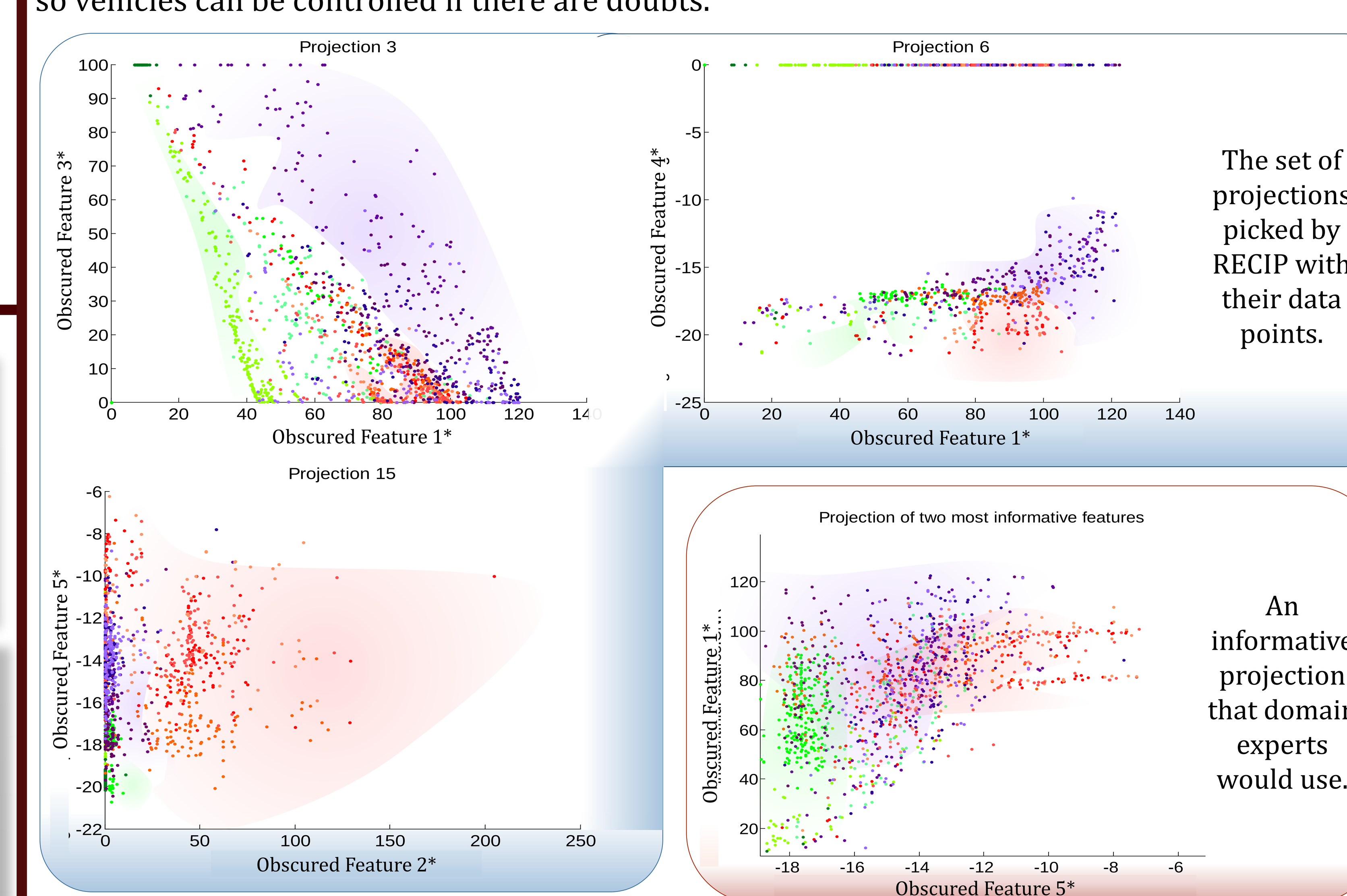
The table shows that, for real data, RECIP finds low-dimensional projections that achieve about the same performance as KNN on all dimensions. This is not always possible, as shown by the Vowel example. In other cases though, projections are 1,2 or 3 dimensional.

Case Study – Nuclear Threat Detection

- System installed at a border checkpoint
- Vehicles crossing border are scanned
  - measurements of radioactivity
  - contextual information
- Is the scanned vehicle a threat?

Border control agent validates prediction.
Positive classification rate of the system under strict bounds → increased risk of false negatives
False negatives have potentially devastating effects, so vehicles can be controlled if there are doubts.





The set of projections picked by RECIP with their data points.

An informative projection that domain experts would use.

*features are non-public, as per contractual agreement