Relational Differential Prediction

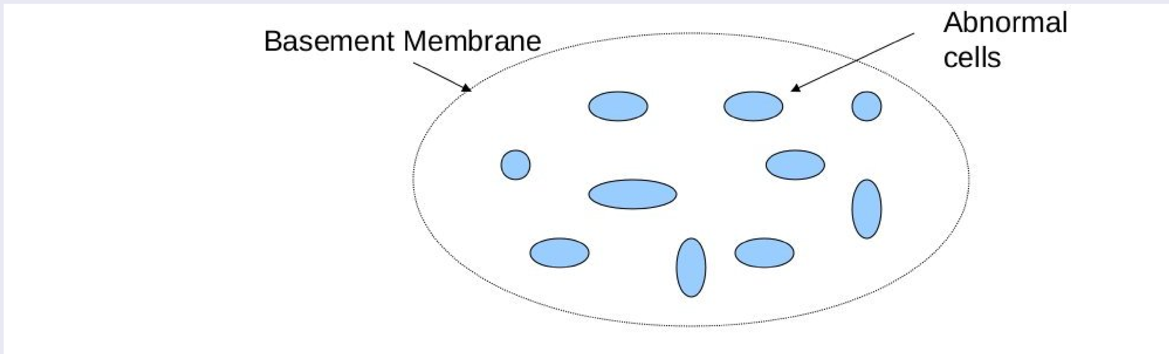
Houssam Nassif[1], Vítor Santos Costa[2], Elizabeth S. Burnside[1], and David Page[1]
1 University of Wisconsin, Madison, USA 2 University of Porto, Portugal

Motivation: Breast Cancer Stages

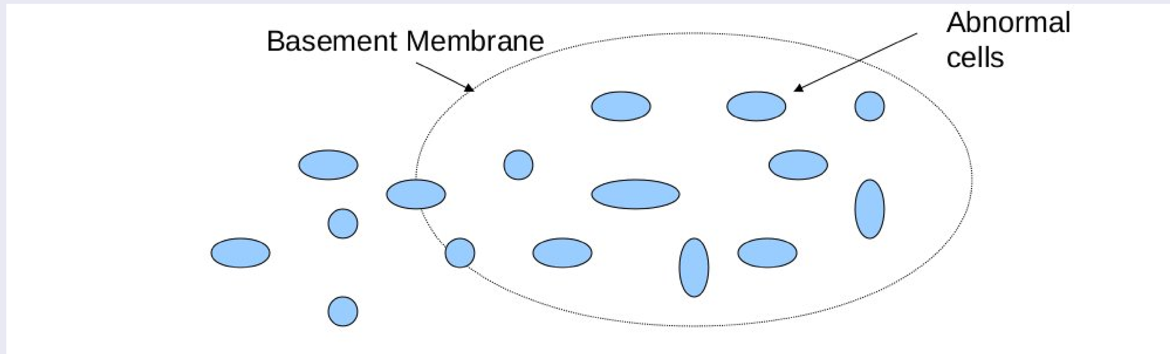
- Two basic stages of breast cancer:

  In Situ : Cancer cells still confined within the ducts and lobules

  Invasive : Cancer cells broken through, invade surrounding tissue



In Situ cancer stage



Invasive cancer stage

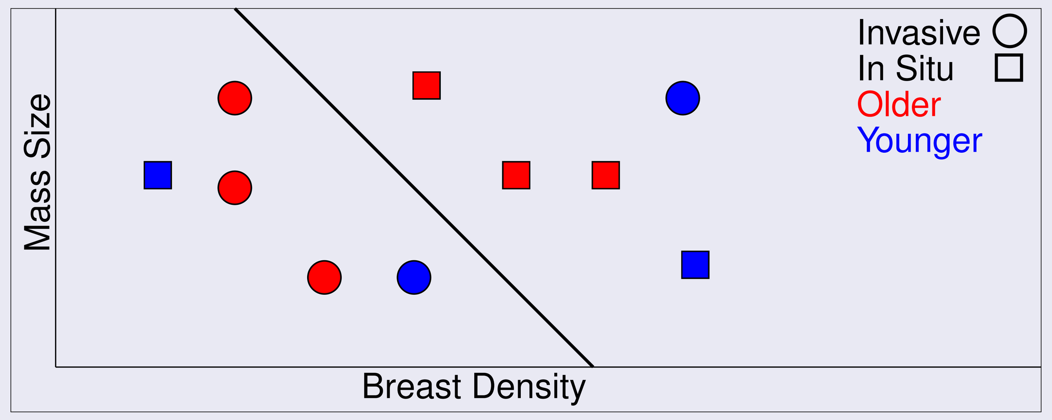
In situ can develop into invasive. Current practice is to always treat in situ. But time required for in situ to spread may be very long, the patient may die of other causes. Besides, treatment may generate undesirable side-effects. This is especially true in older patients.

- What features characterize In Situ in older patients when compared to younger patients?
- What features change between older and younger?

Differential Prediction

Differential Prediction (DP): The case where a classifier exhibits significant performance differences over particular instance subgroups.

In this example, the classifier yields an 80% accuracy classifying invasive vs. in situ. Looking at age subgroups, we classify older patients with 100% accuracy, while younger ones are at 50%.



- This classifier exhibits differential prediction over the age subgroups.

Aim

- Build a classifier that maximizes differential prediction over given data subsets
- Extend differential prediction to relational sets
- Gain insight into differential prediction features of the underlying problem

Definitions

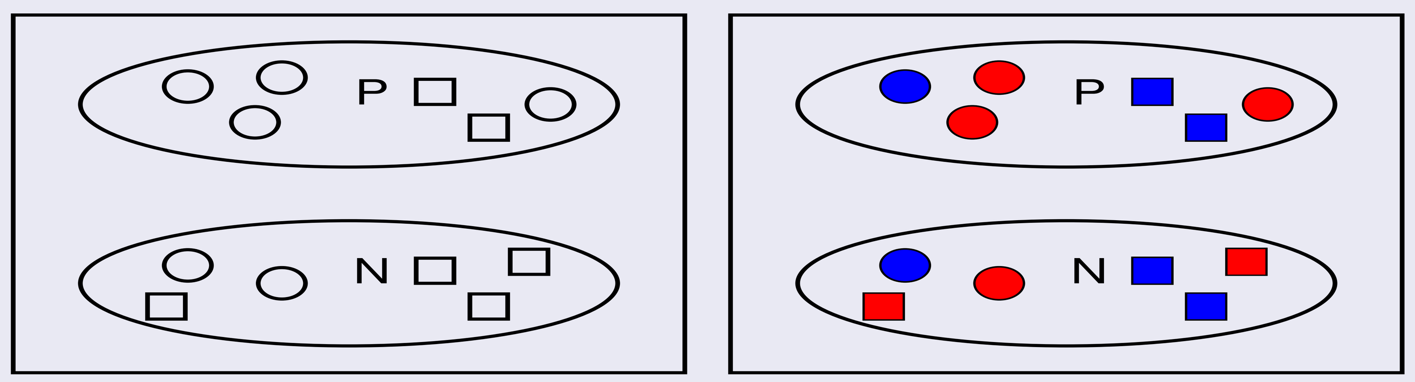
Inductive Logic Programming (ILP) : A machine learning approach that learns a set of first-order logic rules that explain the data.

Stratified Dataset : A dataset stratified into disjoint subgroups, each called a stratum. Each stratum should contain at least one example of each target class.

Differential Predictive Rule : Given a stratified data, a rule whose performance is significantly better over one stratum as compared to the others.

Method 1: Baseline (BASE)

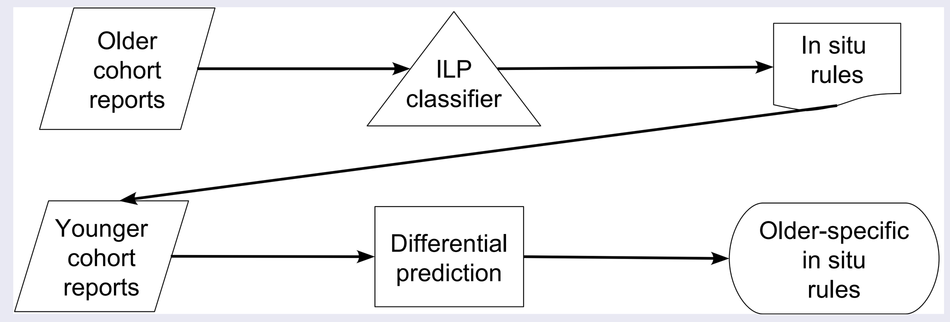
- Include the stratifying attribute as an additional predicate
- Run standard ILP over whole dataset, get resulting theory rules
- Differential predictive rules are those theory rules containing the stratifying predicate
- Example: $InSitu(X)$ *if* $older(X) \wedge calcification(X)$



Data without stratifying attribute Data with stratifying attribute

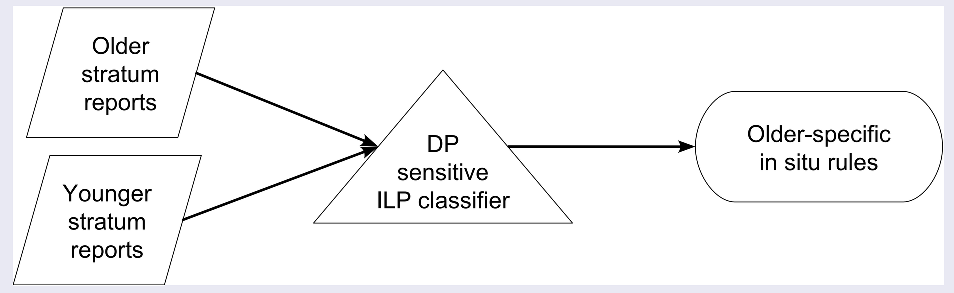
Method 2: Model Filtering (MF)

- Build ILP model on older stratum, resulting in good predicting in situ rules
- Test these rules on the younger cohort, selecting rules with bad performance
- Selected rules are older in situ differential predictive rules



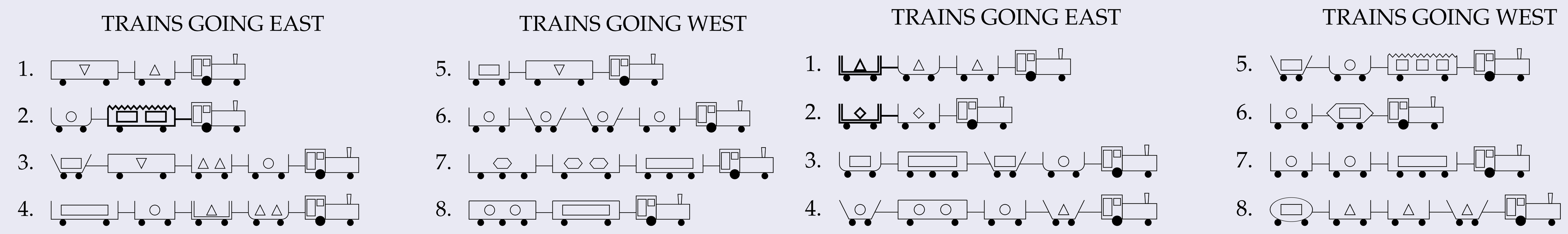
Method 3: Differential Prediction Search (DPS)

- Alter ILP to consider both strata during search space exploration and rule construction
- Define a differential-prediction-sensitive score that measures a rule performance difference over both strata
- Return rules selected for their differential prediction score



Synthetic Data: Michalski Trains

- Two categories of trains, *A* and *B*. Trains go east or west.
  - *A* and *B* trains go east if they have a short carriage in front of a long one (common rule)
  - *A* trains go east if they have a carriage with a jagged roof (stratum-A specific rule)
  - *B* trains go east if they have a carriage with a double-hull (stratum-B specific rule)



East *A* trains: if short in front of long, or jagged-roof (in bold) East *B* trains: if short in front of long, or double-hulled (in bold)

- Given a train dataset, the aim is to recover the stratum-A specific rules
  - Multiple experiments: vary number of trains, add noise, have one or multiple target rules to recover
  - Rank resulting rules by score, match to target ground truth rules, and compare using AUC-PR on recovered rules

Mean Area Under the Precision-Recall Curve (AUC-PR) for 30 experiments in each block

	One target rule						Multiple target rules					
	clean			noisy			clean			noisy		
Size	BASE	MF	DPS	BASE	MF	DPS	BASE	MF	DPS	BASE	MF	DPS
100	0.73	**0.83**	0.62	0.57	**0.62**	0.54	0.61	**0.70**	0.42	0.38	**0.52**	0.31
1000	0.87	**0.90**	0.88	0.63	0.80	**0.87**	0.75	**0.86**	0.77	0.52	0.55	**0.65**

- DPS method more appropriate for real-world (large + noisy) data

Breast Cancer Diagnosis

- Aim: discover older in situ differential predictive rules
  - BASE method didn't return any rule
  - MF method returned rules pertaining to theme number 1
  - DPS method returned rules pertaining to themes 1 – 5
- DPS provides a more complete picture than MF

- A tumor is older-specific in situ if:
  1. It has a calcification
  2. It is a class 2 breast density
  3. It has a prior in situ biopsy
  4. It's BI-RADS score increased
  5. It is a screening visit

- Since ground truth is unknown, compare methods using uplift curves

  Lift : Number of positives in top ranking fraction p

  Uplift curve : Range $p \in [0, 1]$, plot $\{p, Lift_t - Lift_o\}$
- Use theory to form TAN classifier to assign example probability
- DPS uplift curve consistently outperforms MF