Differential Prediction Using Inductive Logic Programming

Houssam Nassif

Thesis Proposal
14 January 2011
Outline

1 Motivation
   - Differential Prediction (DP)
   - Inductive Logic Programming (ILP)
   - Applications

2 Preliminary Results
   - Predicting Hexose Binding Sites
   - DP for Invasive/In-Situ
   - BI-RADS Information Extraction

3 Proposed Work
   - Differential Predictive Rules Definition
   - DP within the ILP Framework
   - Randomizing Recall
   - BI-RADS Terms Annotation

4 Wrap-Up
Outline

1 Motivation
   • Differential Prediction (DP)
   • Inductive Logic Programming (ILP)
   • Applications

2 Preliminary Results
   • Predicting Hexose Binding Sites
   • DP for Invasive/In-Situ
   • BI-RADS Information Extraction

3 Proposed Work
   • Differential Predictive Rules Definition
   • DP within the ILP Framework
   • Randomizing Recall
   • BI-RADS Terms Annotation

4 Wrap-Up
Breast-Cancer Stages

Figure: In-Situ Cancer Stage
Breast-Cancer Stages

Figure: Invasive Cancer Stage
Cancer Stage Features

- In Situ can develop into Invasive
  - Current practice: Always treat In Situ
- Time to spread may be very long
  - Over-diagnosis (unnecessary treatment)
  - Patient may die of other causes

- What features characterize In Situ in older patients?
- What features change between older and younger?
Cancer Stage Features

- In Situ can develop into Invasive
  - Current practice: Always treat In Situ
- Time to spread may be very long
  - Over-diagnosis (unnecessary treatment)
  - Patient may die of other causes

- What features characterize In Situ in older patients?
- What features change between older and younger?
Differential Prediction (DP): Classifier exhibits significant performance differences over particular instance subgroups.
**Differential Prediction**

**Definition**

Differential Prediction (DP): Classifier exhibits significant performance differences over particular instance subgroups.
Differential Prediction

Definition

Differential Prediction (DP): Classifier exhibits significant performance differences over particular instance subgroups.
Differential Prediction (DP): Classifier exhibits significant performance differences over particular instance subgroups.
Using Regression to Detect DP

- Validate educational and psychological tests
- Detect discrepancies related to race or gender
Using Regression to Detect DP

- Validate educational and psychological tests
- Detect discrepancies related to race or gender
Using Regression to Detect DP

- Validate educational and psychological tests
- Detect discrepancies related to race or gender
DP in Machine Learning

- Byproduct of classification
- Detected by:
  - Comparing classifiers built on distinct data subgroups
  - Checking classifier performance on multiple subgroups
- Differential misclassification cost: incorporating different misclassification costs into a cost sensitive classifier

Aim

- Classifier to maximize DP over specific data subsets
- Insight into DP features
DP in Machine Learning

- Byproduct of classification
- Detected by:
  - Comparing classifiers built on distinct data subgroups
  - Checking classifier performance on multiple subgroups
- Differential misclassification cost: incorporating different misclassification costs into a cost sensitive classifier

Aim

- Classifier to maximize DP over specific data subsets
- Insight into DP features
Outline

1 Motivation
   - Differential Prediction (DP)
   - Inductive Logic Programming (ILP)
   - Applications

2 Preliminary Results
   - Predicting Hexose Binding Sites
   - DP for Invasive/In-Situ
   - BI-RADS Information Extraction

3 Proposed Work
   - Differential Predictive Rules Definition
   - DP within the ILP Framework
   - Randomizing Recall
   - BI-RADS Terms Annotation

4 Wrap-Up
Inductive Logic Programming

Definition

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

1. Generates easy to interpret if-then rules
2. Allows user interaction through background knowledge
3. Operates on relational datasets
4. Can investigate the performance of each rule, selecting for DP over given subsets
Inductive Logic Programming

**Definition**

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

1. Generates easy to interpret if-then rules
2. Allows user interaction through background knowledge
3. Operates on relational datasets
4. Can investigate the performance of each rule, selecting for DP over given subsets
Inductive Logic Programming

Definition

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

1. Generates easy to interpret if-then rules
2. Allows user interaction through background knowledge
3. Operates on relational datasets
4. Can investigate the performance of each rule, selecting for DP over given subsets
ILP Example

Example

\[ P(A), \text{red}(A), \text{big}(A), \text{round}(A) \]

\[ \text{sibling}(A, B) \]
ILP Example

P(A), red(A), big(A), round(A), sibling(A, B)
ILP Example

\[ P(A), \text{red}(A), \text{big}(A), \text{round}(A), \text{sibling}(A, B) \]
ILP Example

Example

\[ P(A), red(A), big(A), round(A), sibling(A, B) \]
ILP Example

Example

\[ P(A), \text{red}(A), \text{big}(A), \text{round}(A), \text{sibling}(A, B) \]

- \( P(X) \) if \( \text{square}(X) \)
- \( P(X) \) if \( \text{red}(X) \land \text{big}(x) \)
  - 1 false positive
- \( P(X) \) if \( \text{sibling}(X, Y) \land \text{square}(Y) \)
  - 1 false negative
- Form theory
ILP Example

Example

\( P(A), \text{red}(A), \text{big}(A), \text{round}(A) \)
\( \text{sibling}(A, B) \)
ILP Example

Example

$P(A), \text{red}(A), \text{big}(A), \text{round}(A), \text{Sibling}(A, B)$

- $P(X)$ if $\text{square}(X)$
- $P(X)$ if $\text{red}(X) \land \text{big}(x)$
  - 1 false positive
- $P(X)$ if $\text{Sibling}(X, Y) \land \text{square}(Y)$
  - 1 false negative
- Form theory
Motivation

Preliminary Results

Proposed Work

Wrap-Up

ILP Example

Example

\[ P(A), \text{red}(A), \text{big}(A), \text{round}(A), \text{sibling}(A, B) \]
ILP Example

- \( P(X) \) if square(\( X \))
- \( P(X) \) if \( red(X) \land big(x) \)
  - 1 false positive
- \( P(X) \) if sibling(\( X, Y \)) \land square(\( Y \))
  - 1 false negative
- Form theory

Example

\( P(A), red(A), big(A), round(A) \)

\( sibling(A, B) \)
Outline

1 Motivation
   - Differential Prediction (DP)
   - Inductive Logic Programming (ILP)
   - Applications

2 Preliminary Results
   - Predicting Hexose Binding Sites
   - DP for Invasive/In-Situ
   - BI-RADS Information Extraction

3 Proposed Work
   - Differential Predictive Rules Definition
   - DP within the ILP Framework
   - Randomizing Recall
   - BI-RADS Terms Annotation

4 Wrap-Up
Breast-Cancer Stage Modeling

- Identify patient subgroups that would benefit most from treatment
- Invasive and In Situ characteristics in older and younger women
- Data is mostly in free-text

Tasks
- DP features for Invasive and In Situ
- Information extraction from free-text
Breast-Cancer Stage Modeling

- Identify patient subgroups that would benefit most from treatment
- Invasive and In Situ characteristics in older and younger women
- Data is mostly in free-text

Tasks
- DP features for Invasive and In Situ
- Information extraction from free-text
Hexose-Binding Modeling

- Galactose, glucose, mannose
- High specificity to diverse protein families
- Interesting to uncover differential binding patterns

Tasks
- Glucose-binding model
- Data-driven empirical validation of biochemical findings
Hexose-Binding Modeling

- Galactose, glucose, mannose
- High specificity to diverse protein families
- Interesting to uncover differential binding patterns

**Tasks**
- Glucose-binding model
- Data-driven empirical validation of biochemical findings
Outline

1 Motivation
   - Differential Prediction (DP)
   - Inductive Logic Programming (ILP)
   - Applications

2 Preliminary Results
   - Predicting Hexose Binding Sites
   - DP for Invasive/In-Situ
   - BI-RADS Information Extraction

3 Proposed Work
   - Differential Predictive Rules Definition
   - DP within the ILP Framework
   - Randomizing Recall
   - BI-RADS Terms Annotation

4 Wrap-Up
Hexose Binding-Site Representation
Hexose Binding-Site Features

1: procedure EXTRACTFEATURES(binding site center)
2: for all concentric layers do
3: for all PDB atoms do
4: get distance from center
5: get charge
6: get hydrophobicity
7: get hydrogen-bonding
8: get residue
9: end for
10: end for
11: end procedure
Glucose Binding-Site Classifier (*Proteins*)

- Random Forests for feature selection
- Support Vector Machines for classification

<table>
<thead>
<tr>
<th>Features</th>
<th>L1</th>
<th>L2</th>
<th>L3</th>
<th>L4</th>
<th>L5</th>
<th>L6</th>
<th>L7</th>
<th>L8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative Charge</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Neutral Charge</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non H-Bonding</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H-Bonding</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hydrophilic</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hydroneutral</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hydrophobic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Neutral Residue</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acidic Residue</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>
Validating Hexose-Binding Knowledge (*ILP’09*)

- Use ILP system Aleph
- Extract rules from data without prior biochemical knowledge
- Compare resulting rules with known biochemical rules
- Induce most of the known hexose-binding biochemical rules
- Find a previously unreported dependency between TRP and GLU
Validating Hexose-Binding Knowledge (ILP’09)

- Use ILP system Aleph
- Extract rules from data without prior biochemical knowledge
- Compare resulting rules with known biochemical rules
- Induce most of the known hexose-binding biochemical rules
- Find a previously unreported dependency between TRP and GLU
Outline

1 Motivation
   - Differential Prediction (DP)
   - Inductive Logic Programming (ILP)
   - Applications

2 Preliminary Results
   - Predicting Hexose Binding Sites
   - DP for Invasive/In-Situ
   - BI-RADS Information Extraction

3 Proposed Work
   - Differential Predictive Rules Definition
   - DP within the ILP Framework
   - Randomizing Recall
   - BI-RADS Terms Annotation

4 Wrap-Up
Figure: In-Situ Cancer Stage
Breast-Cancer Stages

Figure: Invasive Cancer Stage
Age Matters

- Apply linear logistic regression
- Uncover a differential ability in predicting invasive and in-situ cancer in older vs. younger women
- Stratify our data:
  - Younger: < 50 years, pre-menopausal
  - Middle: [50, 65) years, peri-menopausal
  - Older: >= 65 years, post-menopausal
Age Matters

- Apply linear logistic regression
- Uncover a differential ability in predicting invasive and in-situ cancer in older vs. younger women
- Stratify our data:
  - Younger: < 50 years, pre-menopausal
  - Middle: [50, 65) years, peri-menopausal
  - Older: \( \geq 65 \) years, post-menopausal
Generate-then-Test DP Method (IHI’10)

Older Stratum Reports

ILP Classifier

Invasive v/s In Situ Rules

Younger Stratum Reports

Differential Prediction

Older-Specific Invasive/In Situ Rules
Middle-Cohort Precision Comparison

<table>
<thead>
<tr>
<th>Rule</th>
<th>Invasive Older Prediction</th>
<th>In-Situ Older Prediction</th>
<th>Invasive Younger Prediction</th>
<th>In-Situ Younger Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Older Cohort (p-value)</td>
<td></td>
<td>Younger Cohort (p-value)</td>
<td></td>
</tr>
<tr>
<td>Rule 1</td>
<td>0.04*</td>
<td></td>
<td>0.00*</td>
<td></td>
</tr>
<tr>
<td>Rule 2</td>
<td>0.01*</td>
<td></td>
<td>0.00*</td>
<td></td>
</tr>
<tr>
<td>Rule 3</td>
<td>0.05</td>
<td></td>
<td>0.00*</td>
<td></td>
</tr>
<tr>
<td>Rule 4</td>
<td>0.26</td>
<td></td>
<td>0.27</td>
<td></td>
</tr>
<tr>
<td>Rule 5</td>
<td>0.48</td>
<td></td>
<td>0.32</td>
<td></td>
</tr>
<tr>
<td>Invasive Older Prediction</td>
<td>0.50</td>
<td></td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>Rule 1</td>
<td>0.00*</td>
<td></td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>In-Situ Older Prediction</td>
<td>0.06</td>
<td></td>
<td>0.06</td>
<td></td>
</tr>
</tbody>
</table>

* Statistically significant at the 95% confidence level.
# Mammography Features

<table>
<thead>
<tr>
<th>Structured</th>
<th>Extracted using NLP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Family breast cancer history</td>
<td>Mass margin</td>
</tr>
<tr>
<td>Personal breast cancer history</td>
<td>Mass shape</td>
</tr>
<tr>
<td>Prior surgery</td>
<td>Calcification distribution</td>
</tr>
<tr>
<td>Palpable lump</td>
<td>Calcification morphology</td>
</tr>
<tr>
<td>Screening v/s diagnostic</td>
<td>Architectural distortion</td>
</tr>
<tr>
<td>Indication for exam</td>
<td>Associated findings</td>
</tr>
<tr>
<td>Breast Density</td>
<td>Mammary lymph node</td>
</tr>
<tr>
<td>BI-RADS code left</td>
<td>Asymmetric breast tissue</td>
</tr>
<tr>
<td>BI-RADS code right</td>
<td>Focal asymmetric density</td>
</tr>
<tr>
<td>BI-RADS code combined</td>
<td>Tubular density</td>
</tr>
<tr>
<td>Principal finding</td>
<td>Mass size</td>
</tr>
</tbody>
</table>
Motivation
- Differential Prediction (DP)
- Inductive Logic Programming (ILP)
- Applications

Preliminary Results
- Predicting Hexose Binding Sites
- DP for Invasive/In-Situ
- BI-RADS Information Extraction

Proposed Work
- Differential Predictive Rules Definition
- DP within the ILP Framework
- Randomizing Recall
- BI-RADS Terms Annotation

Wrap-Up
Breast Imaging Reporting & Data System (BI-RADS)

**Motivation**
- Shape:
  - Round
  - Oval
  - Lobular
  - Irregular
- Margins:
  - Circumscribed
  - Microlobulated
  - Obscured
  - Indistinct
  - Spiculated
- Density:
  - High
  - Equal
  - Low
  - Fat Containing
- Special Cases:
  - Tubular Density
  - Intramammary Lymph Node
  - Asymmetric Breast Tissue
  - Focal Asymmetric Density

**Preliminary Results**
- Associated Findings:
  - Skin Retraction
  - Nipple Retraction
  - Skin Thickening
  - Trabecular Thickening
  - Skin Lesion
  - Axillary Adenopathy

**Proposed Work**
- Distribution:
  - Grouped
  - Linear
  - Segmental
  - Regional
  - Diffuse

**Wrap-Up**
- Higher Probability of Malignancy:
  - Pleomorphic
  - Fine, Linear
- Intermediate:
  - Amorphous
- Typically Benign:
  - Skin
  - Vascular
  - Coarse
  - Rod-Like
  - Round
  - Lucent-Centered
  - Eggshell
  - Milk of Calcium
  - Suture
  - Dystrophic
  - Punctate
Lexicon specifies synonyms
- E.g.: Equal density, Isodense

Lexicon allows for ambiguous wording

<table>
<thead>
<tr>
<th>Text</th>
<th>Concept</th>
</tr>
</thead>
<tbody>
<tr>
<td>indistinct margin</td>
<td>indistinct margin</td>
</tr>
<tr>
<td>indistinct calcification</td>
<td>amorphous calcification</td>
</tr>
<tr>
<td>indistinct image</td>
<td>not a BI-RADS concept</td>
</tr>
</tbody>
</table>
Algorithm Flowchart (*ICDM-W’09*)

- Context Free Grammar
- Straight-forward negation
- Negation-deactivation triggers
Rule Generation Example

- Aim: Skin Thickening concept
- Lexicon specifies “skin thickening”
- Try “skin” and “thickening” in same sentence
  - thickening of the overlying skin
  - marker placed on the skin overlying a palpable focal area of thickening in the upper outer right breast
- Experts suggest “skin” and “thickening” in close proximity

- Start with a large scope
  - Assess number of true and false positives
- Move to smaller scopes
  - Assess number of false negatives
- Experts decide on the best distance
Rule Generation Example

- Aim: Skin Thickening concept
- Lexicon specifies “skin thickening”
- Try “skin” and “thickening” in same sentence
  - thickening of the overlying skin
  - marker placed on the skin overlying a palpable focal area of thickening in the upper outer right breast
- Experts suggest “skin” and “thickening” in close proximity

- Start with a large scope
  - Assess number of true and false positives
- Move to smaller scopes
  - Assess number of false negatives
- Experts decide on the best distance
Outline

1 Motivation
   - Differential Prediction (DP)
   - Inductive Logic Programming (ILP)
   - Applications

2 Preliminary Results
   - Predicting Hexose Binding Sites
   - DP for Invasive/In-Situ
   - BI-RADS Information Extraction

3 Proposed Work
   - Differential Predictive Rules Definition
   - DP within the ILP Framework
   - Randomizing Recall
   - BI-RADS Terms Annotation

4 Wrap-Up
DP Rules Generation Paradigm

Aim

Formally define the differential predictive rules generation paradigm

Definition

DP Rule/Concept: Given a stratified dataset, a rule/concept whose performance is significantly better over one stratum as compared to the others
## DP Rules Generation Paradigm

### Aim
Formally define the differential predictive rules generation paradigm

### Definition
**DP Rule/Concept:** Given a stratified dataset, a rule/concept whose performance is significantly better over one stratum as compared to the others
Definition (Stratified Dataset)

Let $c$ be a concept defined over the set of instances $X$, and let $D = \{\langle x, c(x)\rangle\}$ be a set of training examples labeled according to $c$. Let $D_i$ be $Q$ disjoint subsets of $D$, with $Q \geq 2$, and let $D_i^l$ be the training examples of $D_i$ that have class label $l$, such that:

$$(\forall (i, j) \in [1, Q], i \neq j) \quad D_i \subset D, \quad D_i \cap D_j = \emptyset, \quad \forall l \ D_i^l \neq \emptyset. \quad (1)$$

A $K$-stratified dataset $\mathcal{D}$ over the set of instances $X$ is the union of $K$ such subsets $D_i$, with $2 \leq K \leq Q$, such that:

$$\mathcal{D} = \{D_i \mid 1 \leq i \leq K\}. \quad (2)$$
**Definition (Differential Predictive Concept)**

Let $c$ be a concept over the set of instances $X$, and let $\mathcal{D}$ be a $K$-stratified dataset. Let $S(c, D_i)$ be the classification performance score for $c$ over the subset $D_i$. A **stratum-$j$ specific differential predictive concept** is a concept $c_j$ such that:

$$S(c_j, D_j) \gg S(c_j, D_i), \ (\forall i \neq j).$$

(3)

- The score difference can be evaluated using statistical significance tests or by setting a threshold.
<table>
<thead>
<tr>
<th>Outline</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Motivation</strong></td>
</tr>
<tr>
<td>- Differential Prediction (DP)</td>
</tr>
<tr>
<td>- Inductive Logic Programming (ILP)</td>
</tr>
<tr>
<td>- Applications</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Preliminary Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Predicting Hexose Binding Sites</td>
</tr>
<tr>
<td>- DP for Invasive/In-Situ</td>
</tr>
<tr>
<td>- BI-RADS Information Extraction</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Proposed Work</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Differential Predictive Rules Definition</td>
</tr>
<tr>
<td>- DP within the ILP Framework</td>
</tr>
<tr>
<td>- Randomizing Recall</td>
</tr>
<tr>
<td>- BI-RADS Terms Annotation</td>
</tr>
</tbody>
</table>

| Wrap-Up |
DP within the ILP Framework

**Aim**
Implement DP rules generation within ILP

- Generate-then-test approach
- Test-incorporation approach, more rigorous
- Alter the ILP search
- Alter evaluation function to score a clause according to its DP performance over stratified training set
- Return rules selected for their DP score
Generate-then-Test DP Method (*IHI’10*)

- **Older Stratum Reports**
- **ILP Classifier**
- **Invasive v/s In Situ Rules**
- **Younger Stratum Reports**
- **Differential Prediction**
- **Older-Specific Invasive/In Situ Rules**

Motivation

Preliminary Results

Proposed Work

Wrap-Up
Test-Incorporation DP Method

Motivation

Preliminary Results

Proposed Work

Wrap-Up

Older Stratum Reports

DP Sensitive ILP Classifier

Younger Stratum Reports

Stratum-Specific Invasive/In Situ Rules
DP-Sensitive Scoring Function

Definition (DP-Sensitive Scoring Function)

Let $R$ be a clause over the set of instances $X$, and let $\mathcal{D}$ be a 2-stratified dataset over $X$. Let $S(R, D_i)$ be the classification performance score for $R$ over the subset $D_i$. We define the differential-prediction-sensitive scoring function $Q$ as

$$Q(R, D_1, D_2) = S(R, D_1) - S(R, D_2).$$ (4)

Advantages
- Any classification scoring function $S$ can be used
- Generates a set of rules as a consistent theory
**Definition (DP-Sensitive Scoring Function)**

Let $R$ be a clause over the set of instances $X$, and let $\mathcal{D}$ be a 2-stratified dataset over $X$. Let $S(R, D_i)$ be the classification performance score for $R$ over the subset $D_i$. We define the differential-prediction-sensitive scoring function $Q$ as

$$Q(R, D_1, D_2) = S(R, D_1) - S(R, D_2).$$

(4)

**Advantages**

- Any classification scoring function $S$ can be used
- Generates a set of rules as a consistent theory
Coverage Scoring Function

- Rule coverage score: $\text{Cover}(P) - \text{Cover}(N)$
- DP: $(\text{Cover}(P_1) - \text{Cover}(N_1)) - (\text{Cover}(P_2) - \text{Cover}(N_2))$
Coverage Scoring Function

- Rule coverage score: $Cover(P) - Cover(N)$
- DP: $(Cover(P1) - Cover(N1)) - (Cover(P2) - Cover(N2))$
Instance Relabeling DP Method

- Relabel Pos = P1 + N2
- Relabel Neg = P2 + N1
- Run standard ILP
- Cover(Pos) − Cover(Neg)
- Cover(P1 + N2) − Cover(P2 + N1)
- (Cover(P1) + Cover(N2)) − (Cover(P2) + Cover(N1))
- (Cover(P1) − Cover(N1)) − (Cover(P2) − Cover(N2))
Instance Relabeling DP Method

- Relabel $\text{Pos} = P1 + N2$
- Relabel $\text{Neg} = P2 + N1$
- Run standard ILP
  - $\text{Cover}(\text{Pos}) - \text{Cover}(\text{Neg})$
  - $\text{Cover}(P1+N2) - \text{Cover}(P2+N1)$
  - \((\text{Cover}(P1) + \text{Cover}(N2)) - (\text{Cover}(P2) + \text{Cover}(N1))\)
  - \((\text{Cover}(P1) - \text{Cover}(N1)) - (\text{Cover}(P2) - \text{Cover}(N2))\)
Instance Relabeling DP Method

- Relabel \( Pos = P1 + N2 \)
- Relabel \( Neg = P2 + N1 \)
- Run standard ILP
- \( Cover(Pos) - Cover(Neg) \)
- \( Cover(P1+N2) - Cover(P2+N1) \)
- \((Cover(P1) + Cover(N2)) - (Cover(P2) + Cover(N1))\)
- \((Cover(P1) - Cover(N1)) - (Cover(P2) - Cover(N2))\)
Baseline DP Method

- Include stratifying attribute as a predicate $p$
- Run ILP over whole dataset
- Select rules containing the predicate $p$
- Rules specific to the stratum the predicate $p$ refers to

Example

$P(X)$ if $\text{red}(X) \land \text{big}(X)$
Baseline DP Method

- Include stratifying attribute as a predicate $p$
- Run ILP over whole dataset
- Select rules containing the predicate $p$
- Rules specific to the stratum the predicate $p$ refers to

Example

$P(X) \text{ if } \text{red}(X) \land \text{big}(X)$
Baseline DP Method

- Include stratifying attribute as a predicate $p$
- Run ILP over whole dataset
- Select rules containing the predicate $p$
- Rules specific to the stratum the predicate $p$ refers to

Example

$P(X)$ if $\text{red}(X) \land \text{big}(X)$
Implementing $K$-Stratified DP

- Reduce a $K$-strata problem to $K$ 2-strata problems
- Keep stratum $i$, collapse others together
- Extract stratum $i$ DP rules

- Multi-strata DP-sensitive scoring function
- $f$-divergence functions?
Implementing $K$-Stratified DP

- Reduce a $K$-strata problem to $K$ 2-strata problems
- Keep stratum $i$, collapse others together
- Extract stratum $i$ DP rules

- Multi-strata DP-sensitive scoring function
- $f$-divergence functions?
Outline

1 Motivation
   - Differential Prediction (DP)
   - Inductive Logic Programming (ILP)
   - Applications

2 Preliminary Results
   - Predicting Hexose Binding Sites
   - DP for Invasive/In-Situ
   - BI-RADS Information Extraction

3 Proposed Work
   - Differential Predictive Rules Definition
   - DP within the ILP Framework
   - Randomizing Recall
   - BI-RADS Terms Annotation

4 Wrap-Up
Aleph (Top-Down)

Require: Examples \( E \), mode declarations \( M \), background knowledge \( B \), scoring function \( S \)

1: \( \text{Learned\_rules} \leftarrow \{\} \)
2: \( \text{Pos} \leftarrow \text{all positive examples in } E \)
3: \( \text{while } \text{Pos} \text{ do} \)
4: \( \text{Select example } e \in \text{Pos} \)
5: \( \text{Construct bottom clause } \bot_e \text{ from } e, M \text{ and } B \) \( \triangleright \) Saturation step
6: \( \text{Candidate\_literals} \leftarrow \text{Literals}(\bot_e) \)
7: \( \text{New\_rule} \leftarrow \text{pos}(X) \) \( \triangleright \) Most general rule
8: \( \text{repeat} \) \( \triangleright \) Top-down reduction step
9: \( \text{Best\_literal} \leftarrow \text{argmax}_{L \in \text{Candidate\_literals}} S(\text{New\_rule with precondition } L) \)
10: \( \text{Add } \text{Best\_literal} \text{ to preconditions of } \text{New\_rule} \)
11: \( \text{until No more } S(\text{New\_rule}) \text{ score improvement} \)
12: \( \text{Learned\_rules} \leftarrow \text{Learned\_rules} + \text{New\_rule} \)
13: \( \text{Pos} \leftarrow \text{Pos} - \{\text{members of } \text{Pos} \text{ covered by } \text{New\_rule}\} \)
14: \( \text{end while} \)
15: \( \text{return } \text{Learned\_rules} \)
ProGolem (Bottom-Up)

**Require:** Examples E, mode declarations M, background knowledge B, Scoring function S

1: Learned_rules ← {}  
2: Pos ← all positive examples in E  
3: while Pos do  
4: Select example \(e \in Pos\)  
5: Construct bottom clause \(\bot_e\) from \(e, M\) and \(B\) \(\triangleright\) Saturation step  
6: New_rule ← \(\bot_e\) \(\triangleright\) Most specific rule  
7: repeat \(\triangleright\) Bottom-up reduction step  
8: Select a different example \(e' \in Pos\)  
9: Blocking_literals ← ARMG(New_rule, e')  
10: Remove Blocking_literals from preconditions of New_rule  
11: until No more \(S(New_rule)\) score improvement  
12: Learned_rules ← Learned_rules + New_rule  
13: Pos ← Pos − \{members of Pos covered by New_rule\}  
14: end while  
15: return Learned_rules
Bottom-Up Search Advantages

- **Omitted Variable Problem**
- Not considering a DP variable
- Bottom-up starts with all attributes

- **Myopia Effect**
- Top-down search assumes literals conditionally independent given target class
- If features highly correlated, searches very similar hypotheses
Bottom-Up Search Advantages

- Omitted Variable Problem
- Not considering a DP variable
- Bottom-up starts with all attributes

- Myopia Effect
- Top-down search assumes literals conditionally independent given target class
- If features highly correlated, searches very similar hypotheses
Non-Determinacy and Recall

Example

\textit{legalName}(Joe, X); \textit{parent}(Joe, Y); \textit{Sibling}(Joe, Z)

Definition

\textbf{Predicate Non-Determinacy:} The number of possible solutions of a given predicate

\textbf{Determinate Predicate:} At most one solution

Definition

Recall: Imposed bound on predicate non-determinacy
Non-Determinacy and Recall

Example

\[ \text{legalName}(\text{Joe}, X); \text{parent}(\text{Joe}, Y); \text{sibling}(\text{Joe}, Z) \]

Definition

**Predicate Non-Determinacy:** The number of possible solutions of a given predicate

**Determinate Predicate:** At most one solution

Definition

**Recall:** Imposed bound on predicate non-determinacy
Randomized ProGolem

- Highly non-determinate data
- Exponential learning time for bottom-up learner
- ProGolem: limit bottom clause to first \textit{recall} instantiations

**Aim**
- Randomize ProGolem recall
- Use it for DP

**Example (Bottom Clause (A))**
\[
\text{red}(A), \text{big}(A), \text{round}(A), \\
\text{sibling}(A, B), \\
\text{red}(B), \text{big}(B), \text{round}(B)
\]
Randomized ProGolem

- Highly non-determinate data
- Exponential learning time for bottom-up learner
  - ProGolem: limit bottom clause to first *recall* instantiations

**Example (Bottom Clause (A))**

\[ \text{red}(A), \text{big}(A), \text{round}(A), \text{sibling}(A, B), \text{red}(B), \text{big}(B), \text{round}(B) \]
Randomized ProGolem

Example (Bottom Clause (A))

\[ \text{red}(A), \text{big}(A), \text{round}(A), \]
\[ \text{ sibling}(A, B), \]
\[ \text{red}(B), \text{big}(B), \text{round}(B) \]

- Highly non-determinate data
- Exponential learning time for bottom-up learner
- ProGolem: limit bottom clause to first \textit{recall} instantiations

Aim

- Randomize ProGolem recall
- Use it for DP
Randomized ProGolem

- Highly non-determinate data
- Exponential learning time for bottom-up learner
- ProGolem: limit bottom clause to first *recall* instantiations

**Aim**
- Randomize ProGolem recall
- Use it for DP

**Example (Bottom Clause (A))**
- $\text{red}(A)$, $\text{big}(A)$, $\text{round}(A)$,
- $\text{_sibling}(A, B)$,
- $\text{red}(B)$, $\text{big}(B)$, $\text{round}(B)$
Outline

1 Motivation
   - Differential Prediction (DP)
   - Inductive Logic Programming (ILP)
   - Applications

2 Preliminary Results
   - Predicting Hexose Binding Sites
   - DP for Invasive/In-Situ
   - BI-RADS Information Extraction

3 Proposed Work
   - Differential Predictive Rules Definition
   - DP within the ILP Framework
   - Randomizing Recall
   - BI-RADS Terms Annotation

4 Wrap-Up
BI-RADS Terms Annotation

Aim

Improve BI-RADS extraction from free-text

- Current method maps words to concepts
- Extend to term annotation
  - Create first BI-RADS annotation tool
  - Attempt new term/concept discovery
- Transfer method to other languages (Portuguese)
## BI-RADS Terms Annotation

### Aim

- Improve BI-RADS extraction from free-text

- Current method maps words to concepts
- Extend to **term annotation**
  - Create first BI-RADS annotation tool
  - Attempt new term/concept discovery
- Transfer method to other languages (Portuguese)
BI-RADS Terms Annotation

Aim

Improve BI-RADS extraction from free-text

- Current method maps words to concepts
- Extend to term annotation
  - Create first BI-RADS annotation tool
  - Attempt new term/concept discovery
- Transfer method to other languages (Portuguese)
BI-RADS Annotator Template
Outline

1. Motivation
   - Differential Prediction (DP)
   - Inductive Logic Programming (ILP)
   - Applications

2. Preliminary Results
   - Predicting Hexose Binding Sites
   - DP for Invasive/In-Situ
   - BI-RADS Information Extraction

3. Proposed Work
   - Differential Predictive Rules Definition
   - DP within the ILP Framework
   - Randomizing Recall
   - BI-RADS Terms Annotation

4. Wrap-Up
## Timeline

<table>
<thead>
<tr>
<th>Year</th>
<th>Task</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall 2010</td>
<td>Formally define DP rules</td>
</tr>
<tr>
<td></td>
<td>Translate rules into Portuguese</td>
</tr>
<tr>
<td>Spring 2011</td>
<td>Randomize and test ProGolem recall</td>
</tr>
<tr>
<td></td>
<td>Implement BI-RADS annotator</td>
</tr>
<tr>
<td>Fall 2011</td>
<td>Implement and test ILP-based DP methods</td>
</tr>
<tr>
<td></td>
<td>Extract breast cancer DP rules</td>
</tr>
<tr>
<td>Spring 2012</td>
<td>Wrap-up work</td>
</tr>
<tr>
<td></td>
<td>Write and defend thesis</td>
</tr>
</tbody>
</table>
Prediction of Protein-Glucose Binding Sites Using SVMs.

Uncovering Age-Specific Invasive and DCIS Breast Cancer Rules Using ILP.

An ILP Approach to Validate Hexose Binding Biochemical Knowledge.

Information Extraction for Clinical Data Mining: A Mammography Case Study
Summary

- First glucose-binding model
- Validate hexose-binding knowledge
- BI-RADS extractor
- First DP rules generation
- Formally define DP rules generation paradigm
- Implement DP rules within ILP
- Randomize ProGolem recall
- Improve BI-RADS extraction from free-text
## Hexose Features

<table>
<thead>
<tr>
<th>Atomic Feature</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Charge</td>
<td>Negative, Neutral, Positive</td>
</tr>
<tr>
<td>Hydrogen-bonding</td>
<td>Non-hydrogen bonding, Hydrogen-bonding</td>
</tr>
<tr>
<td>Hydrophobicity</td>
<td>Hydrophilic, Hydroneutral, Hydrophobic</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Residue Grouping</th>
<th>Amino Acids</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aromatic</td>
<td>HIS, PHE, TRP, TYR</td>
</tr>
<tr>
<td>Aliphatic</td>
<td>ALA, ILE, LEU, MET, VAL</td>
</tr>
<tr>
<td>Neutral</td>
<td>ASN, CYS, GLN, GLY, PRO, SER, THR</td>
</tr>
<tr>
<td>Acidic</td>
<td>ASP, GLU</td>
</tr>
<tr>
<td>Basic</td>
<td>ARG, LYS</td>
</tr>
</tbody>
</table>
## Atomic Chemical Properties I

<table>
<thead>
<tr>
<th>PDB atom symbol</th>
<th>Residues</th>
<th>Partial Charge</th>
<th>Hydrophobicity</th>
<th>Hydrogen Bonding</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Amino acid oxygen atoms</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>All amino acids</td>
<td>0</td>
<td>HPHIL</td>
<td>HB</td>
</tr>
<tr>
<td>OXT</td>
<td>All amino acids</td>
<td>-ve</td>
<td>HPHIL</td>
<td>HB</td>
</tr>
<tr>
<td>OE1, OE2, OD1, OD2</td>
<td>GLU, ASP</td>
<td>-ve</td>
<td>HPHIL</td>
<td>HB</td>
</tr>
<tr>
<td>OE1, OD1</td>
<td>GLN, ASN</td>
<td>0</td>
<td>HPHIL</td>
<td>HB</td>
</tr>
<tr>
<td>OG, OG1, OH</td>
<td>SER, THR, TYR</td>
<td>0</td>
<td>HPHIL</td>
<td>HB</td>
</tr>
<tr>
<td><strong>Amino acid carbon atoms</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>All amino acids</td>
<td>0</td>
<td>HNEUT</td>
<td>NHB</td>
</tr>
<tr>
<td>CA</td>
<td>All amino acids</td>
<td>0</td>
<td>HNEUT</td>
<td>NHB</td>
</tr>
<tr>
<td>CB, CG, CD, CE</td>
<td>ALA, SER, THR, CYS, ASP, ASN, GLU, GLN, ARG, LYS, PRO</td>
<td>0</td>
<td>HNEUT</td>
<td>NHB</td>
</tr>
<tr>
<td>CB, CG, CD, CE</td>
<td>LEU, VAL, ILE, MET</td>
<td>0</td>
<td>HPHOB</td>
<td>NHB</td>
</tr>
<tr>
<td>CG1, CG2, CD1, CD2, CD1</td>
<td>LEU, VAL, ILE</td>
<td>0</td>
<td>HPHOB</td>
<td>NHB</td>
</tr>
<tr>
<td>CG, CD1, CD2, CE1, CE2, CZ, CG,CD1, CD2, CE2, CE3, CZ2, CZ3, CH2</td>
<td>PHE, TYR, TRP</td>
<td>0</td>
<td>HPHOB</td>
<td>NHB</td>
</tr>
<tr>
<td>CG, CD2, CE1</td>
<td>HIS</td>
<td>0</td>
<td>HPHOB</td>
<td>NHB</td>
</tr>
</tbody>
</table>
### Appendix B: Mammography

**Atomic Chemical Properties II**

<table>
<thead>
<tr>
<th>PDB atom symbol</th>
<th>Residues</th>
<th>Partial Charge</th>
<th>Hydrophobicity</th>
<th>Hydrogen Bonding</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Amino acid nitrogen atoms</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>All amino acids except PRO</td>
<td>0</td>
<td>HPHIL</td>
<td>HB</td>
</tr>
<tr>
<td>N</td>
<td>PRO</td>
<td>0</td>
<td>HPHIL</td>
<td>NHB</td>
</tr>
<tr>
<td>NE2, ND2</td>
<td>GLN, ASN</td>
<td>0</td>
<td>HPHIL</td>
<td>HB</td>
</tr>
<tr>
<td>NZ</td>
<td>LYS</td>
<td>+ve</td>
<td>HPHIL</td>
<td>HB</td>
</tr>
<tr>
<td>NE</td>
<td>ARG</td>
<td>+ve</td>
<td>HPHIL</td>
<td>NHB</td>
</tr>
<tr>
<td>NH1, NH2</td>
<td>ARG</td>
<td>+ve</td>
<td>HPHIL</td>
<td>HB</td>
</tr>
<tr>
<td>ND1, NE2</td>
<td>HIS</td>
<td>0</td>
<td>HPHIL</td>
<td>HB</td>
</tr>
<tr>
<td>NE1</td>
<td>TRP</td>
<td>0</td>
<td>HNEUT</td>
<td>NHB</td>
</tr>
<tr>
<td><strong>Amino acid sulfur atoms</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SG</td>
<td>CYS</td>
<td>0</td>
<td>HPHIL</td>
<td>HB</td>
</tr>
<tr>
<td>SD</td>
<td>MET</td>
<td>0</td>
<td>HNEUT</td>
<td>NHB</td>
</tr>
<tr>
<td><strong>Water and ions atoms</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>HOH</td>
<td>0</td>
<td>HPHIL</td>
<td>HB</td>
</tr>
<tr>
<td>O1, O2, O3, O4</td>
<td>SO4, 2HP</td>
<td>-ve</td>
<td>HPHIL</td>
<td>HB</td>
</tr>
<tr>
<td>CA, MG, ZN</td>
<td>CA, MG, ZN</td>
<td>+ve</td>
<td>HPHIL</td>
<td>HB</td>
</tr>
</tbody>
</table>
### SVM and RF Results

<table>
<thead>
<tr>
<th>Property</th>
<th>RF</th>
<th>Feature Number</th>
<th>Error (%)</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>Support Vectors (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Charge</td>
<td>false</td>
<td>24</td>
<td>24.32</td>
<td>79.31</td>
<td>73.33</td>
<td>77.03</td>
</tr>
<tr>
<td></td>
<td>true</td>
<td>5</td>
<td>14.86</td>
<td>86.21</td>
<td>84.44</td>
<td>44.59</td>
</tr>
<tr>
<td>Hydrogen Bonding</td>
<td>false</td>
<td>16</td>
<td>17.57</td>
<td>82.76</td>
<td>82.22</td>
<td>41.89</td>
</tr>
<tr>
<td></td>
<td>true</td>
<td>3</td>
<td>14.86</td>
<td>82.76</td>
<td>86.67</td>
<td>47.30</td>
</tr>
<tr>
<td>Hydrophobicity</td>
<td>false</td>
<td>24</td>
<td>16.22</td>
<td>72.41</td>
<td>91.11</td>
<td>65.57</td>
</tr>
<tr>
<td></td>
<td>true</td>
<td>15</td>
<td>12.16</td>
<td>82.76</td>
<td>91.11</td>
<td>40.54</td>
</tr>
<tr>
<td>Residue Grouping</td>
<td>false</td>
<td>48</td>
<td>21.62</td>
<td>48.28</td>
<td>97.78</td>
<td>100.0</td>
</tr>
<tr>
<td></td>
<td>true</td>
<td>19</td>
<td>09.46</td>
<td>93.10</td>
<td>88.89</td>
<td>41.89</td>
</tr>
<tr>
<td>Features Combined</td>
<td>false</td>
<td>112</td>
<td>18.92</td>
<td>75.86</td>
<td>84.44</td>
<td>79.73</td>
</tr>
<tr>
<td></td>
<td>true</td>
<td>24</td>
<td>08.11</td>
<td>89.66</td>
<td>93.33</td>
<td>40.54</td>
</tr>
</tbody>
</table>
### Age Cohorts

<table>
<thead>
<tr>
<th>Subset</th>
<th>Invasive</th>
<th>In-Situ</th>
<th>Subset Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Younger1</td>
<td>132</td>
<td>55</td>
<td>187</td>
</tr>
<tr>
<td>Younger2</td>
<td>132</td>
<td>55</td>
<td>187</td>
</tr>
<tr>
<td>Younger Total</td>
<td>264</td>
<td>110</td>
<td>374</td>
</tr>
<tr>
<td>Middle1</td>
<td>199</td>
<td>85</td>
<td>284</td>
</tr>
<tr>
<td>Middle2</td>
<td>199</td>
<td>85</td>
<td>284</td>
</tr>
<tr>
<td>Middle Total</td>
<td>398</td>
<td>170</td>
<td>568</td>
</tr>
<tr>
<td>Older1</td>
<td>200</td>
<td>66</td>
<td>266</td>
</tr>
<tr>
<td>Older2</td>
<td>201</td>
<td>66</td>
<td>267</td>
</tr>
<tr>
<td>Older Total</td>
<td>401</td>
<td>132</td>
<td>533</td>
</tr>
<tr>
<td>Grand Total</td>
<td>1063</td>
<td>412</td>
<td>1475</td>
</tr>
</tbody>
</table>
Comparing Automated and Manual Extraction

- Automated method superior to manual method ($p = 0.024$)
- Probabilistic interpretation of $F$-score with Laplace prior

<table>
<thead>
<tr>
<th>Method</th>
<th>Predicted</th>
<th>Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Feature Present</td>
<td>Feature Present</td>
</tr>
<tr>
<td>Automated</td>
<td>211</td>
<td>5</td>
</tr>
<tr>
<td>Manual</td>
<td>198</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>4074</td>
</tr>
<tr>
<td></td>
<td>23</td>
<td>4074</td>
</tr>
</tbody>
</table>