DPClass: Effective but Concise Discriminative Patterns-Based Classification

Jingbo Shang, Wenzhu Tong, Jian Peng, Jiawei Han
Outline

- Motivation: Why Discriminative Patterns based?
- DPClass: Methodology
- Experimental Results
- Discussion and Future Work
Why Discriminative Patterns based?

- Single Feature v.s. Combinations of Features
  - A single feature sometimes means nothing.
  - Combinations of Features are more meaningful.
    - Example: Xor Problem, which is **not linear separable** using single features.
- Mining semantically meaningful patterns
  - Construct **high-order** interactions in features
  - **Compress** the predictive model
Classification: Why Not Use Tree-based Models?

- Single Tree Models
  - e.g. Decision Tree/Boosted Tree
  - Sensitive to training instances \( \rightarrow \) Overfitting

- Multiple Trees Models
  - e.g. Random Forest
  - Tree-independent: the growth & traditional pruning strategies
  - Model size could be very large \( \rightarrow \) Slow online prediction
  - Uninterpretable
Classification: Why Not Use PatClass/DDPMine?

- Frequent Patterns v.s. Discriminative Patterns
  - Frequent doesn’t imply discriminative
  - The number of frequent patterns might be very large
  - → a large but useless pool of frequent patterns
Outline

- Motivation: Why Discriminative Patterns based?
- DPClass: Methodology
- Experimental Results
- Discussion and Future Work
DPClass: Compatible Discriminative Patterns for Linear Models

![Diagram showing DPClass process]

- **Training Dataset**
  - Discriminative Patterns Generation
  - Multiple Tree-based Model

- **Testing Dataset**
  - Efficient Testing

- **Compressed Model**
  - Linear Model Training
  - Top-k Discriminative Patterns

**Possible discriminative patterns**:
- A non-leaf node & a discriminate pattern
- A selected discriminative pattern
- A non-selected discriminative pattern

**Equation**

\[ 0.8 + 0.5 \times b - 1 \times g + 2.1 \times f - 0.7 \times j \]
What Kind of Patterns Are of Discriminative?

- Strong signals on the specific classification task
  - E.g. A pattern with very high information gain
What Kind of Patterns Are of Top-$k$ Patterns?

- Some effects of different patterns may have a large portion of overlaps, e.g. $(v_0 \cap v_1 \cap v_2)$ and $(v_0 \cap v_1 \cap v_2 \cap v_3)$

- A set of patterns is compatible $\triangleq$ They have strong signals on the specific classification task and every single pattern has its own “significant” contributions.

**Definition 4.** *Top-$k$ Patterns* is a size-$k$ subset of discriminative patterns, which have the best performance, which is the accuracy in classification tasks, based on the training data.
DPClass I: The Specific Task

- We discuss binary classification here
- \( N \) training instances \((x_1, y_1), (x_2, y_2) \ldots (x_N, y_N)\)
- \( \forall 1 \leq i \leq N, y_i \in \{+1, -1\} \)
- \( x_i \) is the feature vector of \( i \)-th instance
  - Both numeric (continuous) and categorical (discrete) variables are acceptable
DPClass II: Discriminative Patterns Generation

- Random Forest
  - Maximize the randomness
    - Random features
    - Random partitions
    - Random instances (bootstrap)
DPClass II: Discriminative Patterns Generation

- **Parameters**
  - # of trees = $T$
  - loss function = information gain
  - depth $\leq d$
  - support $\geq \sigma$ (based on bootstrapped instances)

- We admit all prefix of these tree-paths as patterns
  - # of leaves $\leq \min \left\{ 2^d, \frac{N}{\sigma} \right\} \cdot T$
  - # of candidate patterns $\leq \min \left\{ 2^d, \frac{N}{\sigma} \right\} \cdot T \cdot d$

- Assume $T = 100$, # of candidate pattern $\sim 10^4$
DPClass III: Compatible Discriminative Patterns Selection

- Select a $k$-set of most compatible discriminative patterns

- Implementation
  - Forward Selection (Greedy)
  - LASSO (GLMNET)
Outline

- Motivation: Why Discriminative Patterns based?
- DPClass: Methodology
- Experimental Results
- Discussion and Future Work
Experiments: Synthetic Experiment

- For each patient, we have several uniformly sampled features as the following.
  - Age (A). Positive Integers no more than than 60.
  - Gender (G). Male or Female.
  - Lab Test 1 (LT1). Categorical values from (A, B, O, AB).
  - Lab Test 2 (LT2). Continuous values in [0..1].
- The positive label of the hypothesis disease will be given when at least one of the following rules holds.
  - (age > 18) and (gender = Male) and (LT1 = AB) and (LT2 ≥ 0.6)
  - (age > 18) and (gender = Female) and (LT1 = O) and (LT2 ≥ 0.5)
  - (age ≤ 18) and (LT2 ≥ 0.9)
Experiments: Synthetic Experiment

- $10^5$ random patients in train (0.1% noise), $5 \times 10^4$ random patients in test
- 99.99% Accuracy
- Top-3 Patterns:
  - (age > 18) and (gender = Female) and (LT1 = O) and (LT2 ≥ 0.496)
  - (age ≤ 18) and (LT2 ≥ 0.900)
  - (age > 18) and (gender = Male) and (LT1 = AB) and (LT2 ≥ 0.601)
Experiments: Compare to DDPMine

- Top-20 Compatible Discriminative Patterns

Table 1: Test Accuracy on UCI Datasets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>DPClass-F</th>
<th>DPClass-L</th>
<th>DDPMine</th>
</tr>
</thead>
<tbody>
<tr>
<td>adult</td>
<td>85.66%</td>
<td>84.33%</td>
<td>84.82%</td>
</tr>
<tr>
<td>crx</td>
<td>85.38%</td>
<td>83.49%</td>
<td>84.93%</td>
</tr>
<tr>
<td>hypo</td>
<td>99.58%</td>
<td>99.28%</td>
<td>99.24%</td>
</tr>
<tr>
<td>sick</td>
<td>98.35%</td>
<td>98.87%</td>
<td>98.36%</td>
</tr>
</tbody>
</table>
Experiments: Train/Test Accuracy v.s. top-K

Figure 2: The impact of top-k patterns in DPClass-Forward.
Experiments: TODO

- Larger and newer datasets
Outline

- Motivation: Why Discriminative Patterns based?
- DPClass: Methodology
- Experimental Results
- Discussion and Future Work
Conclusions

- DPClass can compress the model and thus the online prediction is extremely fast
- DPClass have comparable performance as before
  - Even better in experiments
- DPClass can learn the interpretable patterns
  - Shown in the synthetic experiment
Future Work

- Extend DPClass to DPLearn
- Task Oriented Discriminate Patterns Learning
  - Classification
  - Multi-class classification
  - Regression
  - Survival Analysis