Boolean Decision Rules via Column Generation

Sanjeeb Dash*
Oktay Günlük*
Dennis Wei*

IBM Research
Problem Statement

Learn Boolean rules for binary classification

- Disjunctive normal form (DNF, OR of ANDs)
- Conjunctive normal form (CNF, AND of ORs)

Rules with few clauses and conditions are interpretable

Optimize accuracy vs. simplicity using integer programming (IP)
DNF Boolean rule = Decision rule set

IF A THEN Y=1
IF B AND C THEN Y=1
IF D AND E THEN Y=1
ELSE Y=0

Decision tree

Decision list
IF A THEN Y=1
ELSE IF B AND C THEN Y=1
ELSE IF D AND E THEN Y=1
ELSE Y=0
Assume non-binary features have been binarized

- Categorical: “one-hot” coding (e.g. color=red, color=blue)
- Numerical: comparison with thresholds (e.g. blood pressure ≤ 130, >130)
Main Challenge

**Exponentially many** possible clauses

- e.g. # accounts, # accounts AND debt, # accounts AND debt AND months since delinquency, ...

Previous works limited search using heuristics
Column Generation

Select clauses from exponentially large set

clause complexity costs

clause data matrix

Master IP/LP
Column Generation

Solve only over small subsets

clause complexity costs

clause data matrix

Restricted Master LP
Column Generation

Solve only over small subsets

Restricted Master LP

clause complexity costs

clause data matrix

Augment with improving clauses (columns)
Column Generation

Solve only over small subsets

- Clause complexity costs
- Clause data matrix

Restricted Master LP

- Augment with improving clauses (columns)

Pricing IP

- Also generate columns using heuristic
Procedure and Optimality Guarantees

- Improving column generated?
  - Yes: Re-solve Restricted MLP
  - No: IPs solved using CPLEX

5 min time limit overall
Procedure and Optimality Guarantees

1. Improving column generated?
   - Yes: Re-solve Restricted MLP
   - No: Non-existence proven?
     - Yes: MLP solved
       - Solution integral?
         - Yes: MIP solved
         - No: Lower bound on MIP
           - Solution matches lower bound?
             - Yes: Report solution and gap
             - No: Weaker lower bound on MIP
     - No: Obtain feasible solution to MIP
       - Lower bound on MIP
         - Yes: Report solution and gap
         - No: Weaker lower bound on MIP
Accuracy-Complexity Trade-Off

Lines connect Pareto-efficient points

Column generation (CG) dominates on 8 of 16 datasets and is close on 2 others
Accuracy Maximization

CG competitive with RIPPER [Cohen 1995]

CG can find simpler rules that are no less accurate (adult, bank, magic, FICO)

<table>
<thead>
<tr>
<th>dataset</th>
<th>CG</th>
<th>BRS</th>
<th>AM</th>
<th>BCD</th>
<th>RIPPER</th>
<th>CART</th>
<th>RF</th>
</tr>
</thead>
<tbody>
<tr>
<td>adult</td>
<td>83.5</td>
<td>81.7</td>
<td>83.0</td>
<td>82.4</td>
<td>83.6</td>
<td>83.1</td>
<td>84.7</td>
</tr>
<tr>
<td>bank</td>
<td>90.0</td>
<td>87.4</td>
<td>90.0</td>
<td>89.7</td>
<td>89.9</td>
<td>89.1</td>
<td>88.7</td>
</tr>
<tr>
<td>gas</td>
<td>98.0</td>
<td>92.2</td>
<td>97.6</td>
<td>97.0</td>
<td>99.0</td>
<td>95.4</td>
<td>99.7</td>
</tr>
<tr>
<td>magic</td>
<td>85.3</td>
<td>82.5</td>
<td>80.7</td>
<td>80.3</td>
<td>84.5</td>
<td>82.8</td>
<td>86.6</td>
</tr>
<tr>
<td>mushroom</td>
<td>100.0</td>
<td>99.7</td>
<td>99.9</td>
<td>99.9</td>
<td>100.0</td>
<td>96.2</td>
<td>99.9</td>
</tr>
<tr>
<td>musk</td>
<td>95.6</td>
<td>93.3</td>
<td>96.9</td>
<td>92.1</td>
<td>95.9</td>
<td>90.1</td>
<td>86.2</td>
</tr>
<tr>
<td>FICO</td>
<td>71.7</td>
<td>71.2</td>
<td>71.2</td>
<td>70.9</td>
<td>71.8</td>
<td>70.9</td>
<td>73.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>dataset</th>
<th>accuracy</th>
<th>complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>adult</td>
<td>88.0</td>
<td>15.0</td>
</tr>
<tr>
<td>bank</td>
<td>9.9</td>
<td>6.8</td>
</tr>
<tr>
<td>gas</td>
<td>123.9</td>
<td>62.4</td>
</tr>
<tr>
<td>magic</td>
<td>93.0</td>
<td>11.5</td>
</tr>
<tr>
<td>mushroom</td>
<td>17.8</td>
<td>15.4</td>
</tr>
<tr>
<td>musk</td>
<td>123.9</td>
<td>101.3</td>
</tr>
<tr>
<td>FICO</td>
<td>13.3</td>
<td>8.7</td>
</tr>
</tbody>
</table>
Conclusion

Accurate and interpretable Boolean classification rules

Column generation to efficiently search space of rules without restrictions

Optimality guarantees on training set

Superior accuracy-simplicity trade-offs

Poster #79, Room 210, 10:45 – 12:45 today