DRule: A Density-based Approach to Discovering Quantitative Association Rules

Thomas Van Brussel    Bart Goethals
Overview

Introduction

Intuition

Algorithm

Conclusion & Future work
Association Rules

- Association rules used to discover relations between attributes

\{\text{Beef, Onion}\} \Rightarrow \{\text{Bun}\}

<table>
<thead>
<tr>
<th>TID</th>
<th>Beef</th>
<th>Onion</th>
<th>Bun</th>
<th>Eggs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>2</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>3</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>4</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

- Support $= \frac{2}{4}$ and Confidence $= \frac{2}{3}$
- Well-researched
Quantitative Association Rules

- Numerical data

\[
\text{Latitude}[49.72, 50.05] \land \text{Longitude}[77.76, 79.1] \\
\Rightarrow \text{Focal depth}[0.0, 9.0].
\]
Quantitative Association Rules

- Numerical data

Latitude\[49.72, 50.05\] \land\ \text{Longitude}[77.76, 79.1] \implies \text{Focal depth}[0.0, 9.0].
Other Approaches

- Discretization

- Overlapping regions cause issues
- Can cause loss of information
Other Approaches

Mostly optimization strategies

- Image segmentation algorithms
  - Limited in the number of dimensions
- Genetic algorithms
  - Follow a rule template

Not capable of finding all rules in the data. Restrictions might be very severe.
Intuition

- Rules of the form

\[ A_i[l_i, u_i] \land \ldots \land A_k[l_k, u_k] \Rightarrow T[l_t, u_t] \]

- Support, confidence not enough
  - Use all points to maximize
- Apply different measure: **density**

\[ |N(x) = \{ y \mid d(x, y) \leq \epsilon \}| \geq \min_{\text{pts}} \]

- Only require density for dimensions separately
  - Density in all dimensions can miss good rules
Every rectangle is a dense region
Every rectangle is a dense region

1. Find all regions in dimension $A_i$
Algorithm – Simple Example

Every rectangle is a dense region

1. Find all regions in dimension $A_i$
2. Find all regions in dimension $A_j$

Found three regions for rule generation

(What if we want them all?)
Repeat starting from $A_j$
Algorithm – Simple Example

Repeat starting from $A_j$

1. Find all dense regions in dimension $A_j$
Algorithm – Simple Example

Repeat starting from $A_j$

1. Find all dense regions in dimension $A_j$
2. Scan resulting regions for dense regions in dimension $A_i$

Finds all smaller regions
Algorithm – Overlap

- Overlap in horizontal dimension
- Regions easily separated by algorithm
- Difficult for discretization
Algorithm – Overlap

- Overlap in horizontal dimension
- Regions easily separated by algorithm
- Difficult for discretization

Example rules:
- \( Y[2.5, 3.5] \Rightarrow X[2, 5] \) (red points)
- \( X[0, 3] \Rightarrow Y[1, 2] \) (blue points)
Novel method:

- exhaustively mine quantitative association rules
- while not having to preprocess the data (discretization) or
- specify what rules we want in advance and
- allowing for overlap between rules.
Future work

▶ Not enough pruning (yet)
  ▶ Many patterns
  ▶ Techniques from classical association rule mining might prove effective
▶ Efficiency is still a concern
  ▶ Search through dimensions can be exponential