Multi-Label Classification with Conditional Tree-structured Bayesian Networks

Iyad Batal, Charmgil Hong, Milos Hauskrecht / Department of Computer Science, University of Pittsburgh

**Intro & Motivation**

- **Traditional Classification**
  - Each data instance is associated with a single binary class variable
- **Multi-Label Classification**
  - Each data instance is associated with multiple binary class variables
  - An image may contain multiple objects
  - A news article may cover multiple topics
  - A gene may be associated with several biological functions

An illustration of multi-label classification

---

**Goal & Challenge**

- **Goal**: Find the most probable assignment of the class variables
  - I.e., maximize the joint distribution of \( Y \) given observation \( X = x \)
- **Challenge**: The number of possible label assignments is exponential in the number of labels

---

**Method**

- **Representation**
  - Use conditional tree-structured Bayesian network (CTBN) to efficiently model and predict \( P(Y|X) \)

\[
P(y_1, ..., y_d | x) = \prod_{i=1}^{d} P(y_i | x, y_1, ..., y_{i-1})
\]

- The parent of \( y_i \) in CTBN \( T \)

- A class variable can have at most one other class variable as a parent
- The feature vector \( X \) is the common parent for all class variables
- An example CTBN

---

**Parameter Learning**

- **Goal**: Learn the CPDs of the class variables
- Represent the CPDs using probabilistic classifier functions
  - E.g., logistic regression, naïve Bayes, relevance vector machine

---

**Structure Learning**

- **Goal**: Find the dependency relationships among the class variables
- **Procedure**
  1. Draw a complete directed graph \( G \), where vertex \( v_i \) represent class variable \( Y_i \)
  2. Compute the edge weights of \( G \) using conditional log-likelihood of the data
  3. Find the optimal tree structure by solving the maximum branching problem

---

**Experiments**

- We compare CTBN with 6 state-of-the-art MLC methods on 10 publicly available datasets from different domains

**Results**

- **Exact match accuracy**: The probability of all classes being predicted correctly (higher is better)
  - Red/blue indicates whether CTBN is statistically superior/inferior to the compared method (t-test at 0.05 significance level)

---

**Table**

<table>
<thead>
<tr>
<th>Dataset</th>
<th>BR</th>
<th>CHF</th>
<th>MLKNN</th>
<th>IBLR</th>
<th>CC</th>
<th>MMOC</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enron</td>
<td>0.147</td>
<td>0.162</td>
<td>0.179</td>
<td>0.204</td>
<td>0.194</td>
<td>0.214</td>
<td>0.195</td>
</tr>
<tr>
<td>Scene</td>
<td>0.521</td>
<td>0.166</td>
<td>0.629</td>
<td>0.644</td>
<td>0.633</td>
<td>0.684</td>
<td>0.626</td>
</tr>
<tr>
<td>RCV1_sub1</td>
<td>0.162</td>
<td>0.169</td>
<td>0.078</td>
<td>0.163</td>
<td>0.173</td>
<td>-</td>
<td>0.168</td>
</tr>
<tr>
<td>RCV1_sub2</td>
<td>0.315</td>
<td>0.322</td>
<td>0.165</td>
<td>0.316</td>
<td>0.323</td>
<td>0.448</td>
<td>0.329</td>
</tr>
<tr>
<td>TMC</td>
<td>0.279</td>
<td>0.332</td>
<td>0.205</td>
<td>0.279</td>
<td>0.429</td>
<td>0.448</td>
<td>0.329</td>
</tr>
<tr>
<td>RCV1_sub3</td>
<td>0.476</td>
<td>0.466</td>
<td>0.288</td>
<td>0.417</td>
<td>0.517</td>
<td>not</td>
<td>0.531</td>
</tr>
<tr>
<td>RCV1_sub4</td>
<td>0.412</td>
<td>0.352</td>
<td>0.354</td>
<td>0.491</td>
<td>0.579</td>
<td>-</td>
<td>0.59</td>
</tr>
<tr>
<td>RCV1_sub5</td>
<td>0.401</td>
<td>0.357</td>
<td>0.276</td>
<td>0.411</td>
<td>0.497</td>
<td>-</td>
<td>0.538</td>
</tr>
</tbody>
</table>

---

*Appeared in ACM Conference of Information and Knowledge Management 2013 (CIKM-13), Burlingame, CA, USA. Copyright by Iyad Batal, Charmgil Hong, Milos Hauskrecht (iylad@cs.pitt.edu) and Milos Hauskrecht (mhauskrecht@cs.pitt.edu).*