Maximum Margin Learning with Incomplete Data: Learning Networks instead of Tables

Sandor Szedmak\textsuperscript{1}, Yizhao Ni\textsuperscript{2} and Steve R. Gunn\textsuperscript{1}

\textsuperscript{1}University of Southampton

\textsuperscript{2}University of Bristol

Pascal, WAPA 2010
Problem statement

Motivation

- Almost all real life experiments suffer from the loss of information.
- Some items are too expensive to collect in all experiments.
- The failure of the measuring devices leads to incomplete data items.

Data table example

<table>
<thead>
<tr>
<th></th>
<th>Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X^1$</td>
<td>$X^2$</td>
</tr>
<tr>
<td>$X^3$</td>
<td>$X^4$</td>
</tr>
<tr>
<td>$x^1_1$</td>
<td>$\emptyset$</td>
</tr>
<tr>
<td>$x^1_2$</td>
<td>$\emptyset$</td>
</tr>
<tr>
<td>$x^1_3$</td>
<td>$\emptyset$</td>
</tr>
<tr>
<td>$\vdots$</td>
<td>$\vdots$</td>
</tr>
<tr>
<td>$\emptyset$</td>
<td>$x^2_m$</td>
</tr>
</tbody>
</table>

$\emptyset$ missing items
Bayesian network type learning via convex optimization
Relation graph defined on the variables or subset of them

Hypothesis

\[ \begin{align*}
\phi_r(x_r) & \leftarrow \Psi \rightarrow \phi_s(x_s) \\
\Psi : & \mathcal{H}_r \times \mathcal{H}_s \rightarrow \mathbb{R},
\end{align*} \]

Multilinear similarity function

\[ \psi(\phi_r(x^r), \phi_s(x^s)) = \langle W_e, \phi_s(x^s) \otimes \phi_r(x^r) \rangle \]

\( \mathbf{W}_e^{(r,s)} \) tensor, \( \langle \rangle \) dot- tensor product

leads to an convex decomposable optimization problem.

\( x^r \in X^r \), \( x^s \in X^s \) observations

 Leads to an convex decomposable optimization problem.

Szedmak (UoS)
Learning links via MaxMargins
WAPA 2010
Experiment
Large scale

- MovieLens dataset: **10 million** items, an example of recommender systems,
- Data structure: huge but very sparse matrix,
- Task solved: matrix competition,
- 5-fold cross validation,
- Most recent result:
  - RMSE: 0.8607, MAE: 0.6594,
  - (Lawrence(2009), RMSE: 0.8740(*))

(*) unfortunately no standard test procedure exists, Lawrence result is computed on fixed test sets.