Abstract

Learning probabilistic graphical models from high-dimensional datasets is a computationally challenging task. In many interesting applications, the domain dimensionality is such as to prevent state-of-the-art statistical learning techniques from delivering accurate models in reasonable time. We introduce a hybrid random field (HRF) model for pseudo-likelihood estimation in high-dimensional domains. A theoretical analysis proves that the class of pseudo-likelihood distributions representable by HRFs strictly includes the class of joint probability distributions representable by Bayesian networks (BNs). In order to learn the structure of HRFs from data, we develop the Markov Blanket Merging (MBM) algorithm. Theoretical and experimental evidence shows that MBM scales up very well to high-dimensional datasets. As compared to other statistical learning techniques, MBM delivers accurate results in a number of link prediction tasks, while achieving also significant improvements in terms of computational efficiency. Our software implementation of the models investigated in this work is publicly available at http://www.dii.unisi.it/~freno/.

Hybrid Random Fields

Let \( X \) be a set of random variables \( X_1, \ldots, X_n \). A HRF for \( X_1, \ldots, X_n \) is a set of BNs \( \mathcal{BN}_1, \ldots, \mathcal{BN}_n \) (with graphs \( \mathcal{G}_1, \ldots, \mathcal{G}_n \)) such that:

i. Each \( \mathcal{BN}_i \) contains \( X_i \) plus a subset \( \mathcal{M}(X_i) \) of \( X \setminus \{X_i\} \);

ii. For each \( X_i \), \( P(X_i|X \setminus \{X_i\}) = P(X_i|\mathcal{M}(X_i)) \), where \( \mathcal{M}(X_i) \) is the set containing the parents, the children, and the parents of the children of \( X_i \) in \( \mathcal{G}_i \). \( \mathcal{M}(X_i) \) is a Markov blanket (MB) of \( X_i \) within \( \mathcal{BN}_i \). The elements of \( \mathcal{R}(X) \) are called ‘relatives of \( X \).

![Graphical components of a HRF](image)

Pseudo-Likelihood Distributions

- The joint distribution of a set \( X = X_1, \ldots, X_n \) of random variables is estimated using the pseudo-likelihood measure:

\[
P^*(X = x) = \prod_{i=1}^{n} P(X_i = x_i|\mathcal{M}(X_i) = mb(X_i))
\]

**Theorem 1:** For each Bayesian network \( \mathcal{BN} \), there exists a HRF representing the same joint distribution represented by \( \mathcal{BN} \).

**Theorem 2:** For any non-chordal graph \( \mathcal{G} \), there exist two conditional independence statements such that no BN can entail both statements at the same time, while a HRF can always be constructed so as to entail both statements.

Structure Learning: The Markov Blanket Merging Algorithm

- **Model Initialization:** Assign each variable \( X_i \) as relatives the \( k \) nodes with highest \( \gamma^2 \) dependence on \( X_i \). Learn a BN \( \mathcal{BN} \) for each \( X_i \) together with \( \mathcal{R}(X) \).

- **Search Operator:** For each \( X_i \), if \( X_i \) appears in any \( \mathcal{BN} \) other than \( \mathcal{BN}_i \), take the union \( \mathcal{U} \) of all MBs of \( X_i \). If \( \mathcal{U} \) does not exceed size \( k^* \), use its elements as relatives for \( X_i \) and learn a new BN \( \mathcal{BN}_i \).

- **Search Heuristic:** For each new BN \( \mathcal{BN}_i \), if \( \mathcal{BN}_i \) increases the model pseudo-likelihood, accept \( \mathcal{BN}_i \), otherwise stick to \( \mathcal{BN}_i \).

- **Stopping Criterion:** Keep searching until the model pseudo-likelihood stops increasing.

![An iteration of MBM](image)

Experimental Evaluation

- We assess both the accuracy of HRMs in three link prediction applications and the scalability of the MBM learning algorithm:

  - **Link Prediction:** The first two tasks require to predict references in scientific papers (CiteSeer and Cora datasets), while the third task requires to predict preferences for movies (MovieLens dataset).

  - **Computational Burden of (Structure) Learning:** The scalability of MBM is evaluated using a number of synthetic datasets of growing dimensionality, where each dataset contains 1000 independent and identically distributed patterns.

  - HRMs are compared to BNs, dependency networks (DNs), Markov random fields (MRFs), and naive Bayes (NB). The structure of BNs is learned either with a general hill-climbing strategy searching a space of possible graphs, or with the K2 algorithm. The initialization step of MBM is used as structure learning algorithm for DNs and MRFs.

**Link Prediction Applications**

<table>
<thead>
<tr>
<th></th>
<th>CiteSeer</th>
<th>Cora</th>
</tr>
</thead>
<tbody>
<tr>
<td>BN</td>
<td>0.2823 ± 0.0470</td>
<td>0.2916 ± 0.0229</td>
</tr>
<tr>
<td>DN</td>
<td>0.0130 ± 0.0010</td>
<td>0.0417 ± 0.0034</td>
</tr>
<tr>
<td>HRF</td>
<td>0.2865 ± 0.0251</td>
<td>0.2647 ± 0.0168</td>
</tr>
<tr>
<td>MRFF</td>
<td>0.1771 ± 0.0306</td>
<td>0.0704 ± 0.0096</td>
</tr>
<tr>
<td>NB</td>
<td>0.0537 ± 0.0108</td>
<td>0.1519 ± 0.0131</td>
</tr>
</tbody>
</table>

Table: Average MRR ± standard deviation on the CiteSeer and Cora datasets (5-fold cross-validation) for BN, DN, HRF, MRFF, and NB.

**Computational Burden of Structure Learning**

![Learning time for BN, DN, HRF, and MRFF as the domain size increases](image)