Mixtures of Markov trees:

- Composed of a set \( \mathcal{T} = \{ T_1, \ldots, T_m \} \) of \( m \) elementary Markov tree densities and a set \( \{ \mu_i \}_{i=1}^m \) of weights.
- Convex combination of tree predictions:
  \[
  P_f(x) = \sum_{k=1}^m \mu_k P_{T_k}(x).
  \]

Key points:

- Trees \( \rightarrow \) efficient algorithms.
- Mixture \( \rightarrow \) improved modeling.

We attempt to combine both.

Motivation:

- Variance reduction methods are good on low samples sets.
- Maximum-likelihood methods partition the data set.

→ Is it possible to reduce the variance of the EM mixture by combining both methods?

Concept:

Building a mixture of ensemble of Markov trees:

\[
  P_f(x) = \sum_{k=1}^m \mu_k P_{T_k}(x)
\]

\[
  P_R(x) = \sum_{j=1}^N \lambda_j P_{T_j}(x) \quad \forall k \in [1, m].
\]

Experiments on synthetic distributions of 200 variables show combining the methods can improve accuracy.

The proposed approach improves over an EM mixture when recovering a mixture of Markov trees:

- The original mixture has 3 Markov trees, 200 binary variables, and uniform weights.
- The number of trees in the target mixture is known.
- The accuracy of the EM mixture is always improved by replacing each tree by an ensemble.
- The first ensemble seems better for low sample sizes, the second better when \( N \) increases.

There are 2 approaches to improve over a single Chow-Liu tree:

Bias reduction, e.g. EM algorithm [1]

- Learning the mixture is viewed as a global optimization problem aiming at maximizing the data likelihood.
- There is a bias-variance trade-off associated with the number of terms.
- It leads to a partition of the learning set: each tree models a subset of observations.

Variance reduction, e.g. perturb and combine [2]

- This approach can be viewed as an approximation of Bayesian learning in the space of Markov tree structures.
- A sequence of trees is generated by a randomized Chow-Liu algorithm:
  - pure random structure, edge subsampling, bootstrapping...

More terms might be necessary in the ensembles when estimating more complex probability distributions:

<table>
<thead>
<tr>
<th>Data set</th>
<th>( N = 200 )</th>
<th>( N = 500 )</th>
<th>( N = 1000 )</th>
<th>( N = 2500 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHILD</td>
<td>200 24 2521</td>
<td>1 24 24 24</td>
<td>4 24 24 24</td>
<td>4 24 24 24</td>
</tr>
<tr>
<td>PIPS</td>
<td>411 3 9757 21</td>
<td>1 3 16 9</td>
<td>1 3 16 9</td>
<td>1 3 16 9</td>
</tr>
<tr>
<td>ALARM10</td>
<td>371 2 4496 3</td>
<td>5 17 6 19</td>
<td>5 17 6 19</td>
<td>5 17 6 19</td>
</tr>
<tr>
<td>GINE</td>
<td>801 2 8348 25</td>
<td>- 9 16 8 2</td>
<td>- 9 16 8 2</td>
<td>- 9 16 8 2</td>
</tr>
<tr>
<td>LUNG CANCER</td>
<td>809 2 8452 25</td>
<td>- 8 17 8 2</td>
<td>- 8 17 8 2</td>
<td>- 8 17 8 2</td>
</tr>
<tr>
<td>LUNG</td>
<td>724 2 1423 3</td>
<td>7 15 13 12</td>
<td>7 15 13 12</td>
<td>7 15 13 12</td>
</tr>
<tr>
<td>INSURANCE10</td>
<td>270 2 1405 3</td>
<td>1 22 1 15</td>
<td>1 22 1 15</td>
<td>1 22 1 15</td>
</tr>
<tr>
<td>MUNIN</td>
<td>189 1 16 156 2</td>
<td>15 9 6 5 14</td>
<td>- 5 5</td>
<td>- 5 5</td>
</tr>
<tr>
<td>HAILFINDER</td>
<td>569 3 11 9748 25</td>
<td>25 25 25</td>
<td>- 25</td>
<td>- 25</td>
</tr>
<tr>
<td>ALL</td>
<td>105 30 40 62 123</td>
<td>8 58 50</td>
<td>8 58 50</td>
<td>8 58 50</td>
</tr>
</tbody>
</table>

These experiments were performed on 5 randomly generated target distributions \( \times 5 \) learning sets (for each sample size).

References


Acknowledgement

This work was funded by the Belgian Fund for Research in Industry and Agriculture (FRIA), the Biomagnet IUAP research project of the Belgian Science Policy Office and the Pulsar2 network of excellence of the EC. It is not under those organisms scientific responsibility.