An exact approach to learning Probabilistic Relational Model

Nourhene Ettouzi\textsuperscript{1}, Philippe Leray\textsuperscript{2}, Montassar Ben Messaoud\textsuperscript{1}

\textsuperscript{1} LARODEC, ISG Tunis, Tunisia
\textsuperscript{2} LINA, Nantes University, France

International Conference on Probabilistic Graphical Models
Lugano (Switzerland), Sept. 6-9, 2016
Motivations

- **Probabilistic Relational Models** (PRMs) extend Bayesian networks to work with relational databases rather than propositional data.

<table>
<thead>
<tr>
<th>Movie.Genre</th>
<th>Votes.Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drama, M</td>
<td>0.5 0.5</td>
</tr>
<tr>
<td>Drama, F</td>
<td>0.3 0.7</td>
</tr>
<tr>
<td>Horror, M</td>
<td>0.2 0.8</td>
</tr>
<tr>
<td>Horror, F</td>
<td>0.9 0.1</td>
</tr>
<tr>
<td>Comedy, M</td>
<td>0.5 0.5</td>
</tr>
<tr>
<td>Comedy, F</td>
<td>0.6 0.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>User.Gender</th>
<th>M</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4</td>
<td>0.6</td>
<td></td>
</tr>
</tbody>
</table>
Motivations

- Our goal: **PRM structure learning** from a relational database

- Only few works, inspired from classical BNs learning approaches, were proposed to learn PRM structure

**Our proposal**

- **an exact approach** to learn (guaranteed optimal) PRMs
Motivations

- Our goal: PRM structure learning from a relational database

- Only few works, inspired from classical BNs learning approaches, were proposed to learn PRM structure

<table>
<thead>
<tr>
<th>Approaches</th>
<th>BNs</th>
<th>PRMs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score-based</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Approximate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exact</td>
<td>✓</td>
<td>x</td>
</tr>
<tr>
<td>Constraint-based</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Hybrid</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Our proposal

- an exact approach to learn (guaranteed optimal) PRMs
Motivations

- Our goal: PRM structure learning from a relational database
- Only few works, inspired from classical BNs learning approaches, were proposed to learn PRM structure

<table>
<thead>
<tr>
<th>Approaches</th>
<th>BNs</th>
<th>PRMs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score-based</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Approximate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exact</td>
<td>✓</td>
<td>☒</td>
</tr>
<tr>
<td>Constraint-based</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Hybrid</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Our proposal

- an exact approach to learn (guaranteed optimal) PRMs
Outline ...

1 Learning optimal BN
   - Score-based approaches
2 Learning PRM
   - Definitions
   - Probabilistic relational models
   - Learning
3 Learning optimal PRM
4 Conclusion
Main issue: find the **highest-scoring** network

1. decomposable scoring function (BDe, MDL/BIC, AIC, ...)
2. optimal search strategy
Main issue: find the highest-scoring network

1. decomposable scoring function (BDe, MDL/BIC, AIC, ...)
2. optimal search strategy

**Decomposability**

Let us denote $V$ a set of variables. A scoring function is decomposable if the score of the structure, $\text{Score}(BN(V))$, can be expressed as the sum of local scores at each node.

$$\text{Score}(BN(V)) = \sum_{X \in V} \text{Score}(X \mid Pa_X)$$

Each local score $\text{Score}(X \mid Pa_X)$ is a function of one node and its parents.
Exact score-based approaches

Main issue: find the **highest-scoring** network

1. decomposable scoring function (BDe, MDL/BIC, AIC, ...)
2. optimal search strategy

- Spanning Tree [Chow et Liu, 1968]
- Mathematical Programming [Cussens, 2012]
- Dynamic Programming [Singh et al., 2005]
- **A* search** [Yuan et al., 2011]
  - variant of Best First Heuristic search (BFHS) [Pearl, 1984]
  - BN structure learning as a shortest path finding problem
  - evaluation functions $g$ and $h$ based on local scoring function
Exact score-based approaches

Main issue: find the **highest-scoring** network

1. decomposable scoring function (BDe, MDL/BIC, AIC, ...)
2. optimal search strategy

- Spanning Tree [Chow et Liu, 1968]
- Mathematical Programming [Cussens, 2012]
- Dynamic Programming [Singh et al., 2005]
- **A* search** [Yuan et al., 2011]
  - variant of Best First Heuristic search (BFHS) [Pearl, 1984]
  - BN structure learning as a shortest path finding problem
  - evaluation functions $g$ and $h$ based on local scoring function
A* search for BNs [Yuan et al., 2011]

Order graph: the search space

Optimal BN
A* search for BNs [Yuan et al., 2011]

\[ g(U) : \text{sum of edge costs on the best path from the start node to } U. \]
\[ g(U \rightarrow \{U \cup X\}) = \text{BestScore}(X, U) \]
A* search for BNs [Yuan et al., 2011]

h(U) enables variables not yet present in the ordering to choose their optimal parents from V.

\[ h(U) = \sum_{X \in V \setminus U} \text{BestScore}(X \mid V \setminus \{X\}) \]

\( g(U) \) : sum of edge costs on the best path from the start node to U.

\( g(U \rightarrow \{U \cup X\}) = \text{BestScore}(X, U) \)
A* search for BNs [Yuan et al., 2011]

Parent graph of $X_1$
Outline ...

1. Learning optimal BN
   - Score-based approaches

2. Learning PRM
   - Definitions
   - Probabilistic relational models
   - Learning

3. Learning optimal PRM

4. Conclusion
Relational schema $\mathcal{R}$

Definitions

- classes + attributes, $X.A$ denotes an attribute $A$ of a class $X$
- reference slots = foreign keys (e.g. Vote.Movie, Vote.User)
- inverse reference slots (e.g. User.User$^{-1}$)
- slot chain = a sequence of (inverse) reference slots
  e.g. Vote.User.User$^{-1}$.Movie, all the movies voted by a particular user.
Definitions

- classes + attributes, $X.A$ denotes an attribute $A$ of a class $X$
- reference slots = foreign keys (e.g. Vote.Movie, Vote.User)
- inverse reference slots (e.g. User.User$^{-1}$)
- slot chain = a sequence of (inverse) reference slots
  - ex: Vote.User.User$^{-1}$.Movie: all the movies voted by a particular user
**Relational schema $\mathcal{R}$**

**Definitions**

- classes + attributes, $X.A$ denotes an attribute $A$ of a class $X$
- reference slots = foreign keys (e.g. $Vote.Movie$, $Vote.User$)
- inverse reference slots (e.g. $User.User^{-1}$)
- slot chain = a sequence of (inverse) reference slots
  - ex: $Vote.User.User^{-1}.Movie$: all the movies voted by a particular user.
Definitions

- classes + attributes, $X.A$ denotes an attribute $A$ of a class $X$
- reference slots = foreign keys (e.g. Vote.Movie, Vote.User)
- inverse reference slots (e.g. User.User$^{-1}$)
- slot chain = a sequence of (inverse) reference slots
  - ex: Vote.User.User$^{-1}$.Movie: all the movies voted by a particular user
Probabilistic Relational Models

[Koller & Pfeffer, 1998]

Definition

A PRM $\Pi$ associated to $\mathcal{R}$:

- A qualitative dependency structure $\mathcal{S}$ (with possible long slot chains and aggregation functions)
- A set of parameters $\theta_\mathcal{S}$

Aggregators

- $\text{Mode}(\text{Vote.User.User}^{-1}.\text{Movie.genre}) \rightarrow \text{Vote.rating}$
- Movie rating from one user can be dependent with the most frequent genre of movies voted by this user
PRM structure learning

Relational variables

- finding new variables potentially dependent with each target variable, by exploring the relational schema and the possible aggregators
  
- ex: Vote.Rating, Vote.user.user\(^{-1}\).Rating, Vote.movie.movie\(^{-1}\).Rating, ...

⇒ adding another dimension in the search space
 ⇒ limitation to a given maximal slot chain length

Constraint-based methods

- relational PC [Maier et al., 2010] relational CD [Maier et al., 2013], rCD light [Lee and Honavar, 2016]

Nourhene Ettouzi, Philippe Leray, Montassar Ben Messaoud

An exact approach to learning PRM
Learning optimal BN
Learning PRM
Learning optimal PRM
Conclusion

PRM structure learning

Relational variables

⇒ adding another dimension in the search space
⇒ limitation to a given maximal slot chain length

Constraint-based methods

- relational PC [Maier et al., 2010]
- relational CD [Maier et al., 2013]
- rCD light [Lee and Honavar, 2016]

Score-based methods

- Greedy search [Getoor et al., 2007]

Hybrid methods

- relational MMHC [Ben Ishak et al., 2015]
PRM structure learning

Relational variables

- adding another dimension in the search space
- limitation to a given maximal slot chain length

Constraint-based methods

- relational PC [Maier et al., 2010]
- relational CD [Maier et al., 2013], rCD light [Lee and Honavar, 2016]

Score-based methods

- Greedy search [Getoor et al., 2007]

Hybrid methods

- relational MMHC [Ben Ishak et al., 2015]
PRM structure learning

Relational variables

- adding another dimension in the search space
- limitation to a given maximal slot chain length

Constraint-based methods

- relational PC [Maier et al., 2010]
- relational CD [Maier et al., 2013]
- rCD light [Lee and Honavar, 2016]

Score-based methods

- Greedy search [Getoor et al., 2007]

Hybrid methods

- relational MMHC [Ben Ishak et al., 2015]
Outline ...

1. Learning optimal BN
   - Score-based approaches

2. Learning PRM
   - Definitions
   - Probabilistic relational models
   - Learning

3. Learning optimal PRM

4. Conclusion
Strategy similar to [Yuan et al., 2011], but adapted to the relational context

**Two key points**

- **search space**: how to deal with "relational variables"? 
  ➞ relational order graph

- **parent determination**: how to deal with slot chains, aggregators, and possible "multiple" dependencies between two attributes? 
  ➞ relational parent graph
  ➞ evaluation functions
Strategy similar to [Yuan et al., 2011], but adapted to the relational context

Two key points
- search space: how to deal with "relational variables"? ⇒ relational order graph
- parent determination: how to deal with slot chains, aggregators, and possible "multiple" dependencies between two attributes? ⇒ relational parent graph ⇒ evaluation functions
Definition

- lattice over $2^{\mathcal{X} \cdot \mathcal{A}}$, powerset of all possible attributes
- no big change wrt. BNs
Relational parent graph

**Definition (one for each attribute X.A)**

- lattice over the candidate parents for a given maximal slot chain length (+ local score value)
- the same attribute can appear several times in this graph
- one attribute can appear in its own parent graph, e.g. gender
Evaluation functions

Relational cost so far  

- $g(U → U ∪ \{X.A\}) = \text{interest of having a set of attributes } U \text{ (and } X.A\text{) as candidate parents of } X.A$
- $= \text{BestScore}(X.A | \{CPa_i/A(CPa_i) ∈ A(U) ∪ \{A\}\})$

Relational heuristic function  

- $h(U) = \sum_X \sum_{A ∈ A(X) \setminus A(U)} \text{BestScore}(X.A | CPa(X.A))$
Evaluation functions

**Relational cost so far**

- \( g(U \rightarrow U \cup \{X.A\}) = \) interest of having a set of attributes \( U \) (and \( X.A \)) as candidate parents of \( X.A \)
- \( = \text{BestScore}(X.A \mid \{CPa_i/\mathcal{A}(CPa_i) \in A(U) \cup \{A\}\}) \)

**Relational heuristic function**

\[
h(U) = \sum_{X} \sum_{A \in A(X) \setminus A(U)} \text{BestScore}(X.A \mid CPa(X.A))
\]
Relational BFHS: example

(a) The shortest path resulting from $A^*$ search
(b) An optimal PRM related to the relational schema
Outline ...

1. Learning optimal BN
   - Score-based approaches
2. Learning PRM
   - Definitions
   - Probabilistic relational models
   - Learning
3. Learning optimal PRM
4. Conclusion
Visible face

An exact approach to learn optimal PRM, inspired from previous works dedicated to Bayesian networks [Yuan et al., 2011; Malone, 2012; Yuan et al., 2013] whose performance was already proven.

To do list

- Implement this approach on our software platform PILGRIM
- Provide an anytime PRM structure learning algorithm, following the ideas presented in [Aine et al., 2007; Malone et al., 2013] for BNs

Thank you for your attention
Visible face

An exact approach to learn optimal PRM, inspired from previous works dedicated to Bayesian networks [Yuan et al., 2011; Malone, 2012; Yuan et al., 2013] whose performance was already proven

To do list

- **Implement** this approach on our software platform PILGRIM
- Provide an **anytime** PRM structure learning algorithm, following the ideas presented in [Aine et al., 2007; Malone et al., 2013] for BNs

Thank you for your attention
Visible face

An exact approach to learn optimal PRM, inspired from previous works dedicated to Bayesian networks [Yuan et al., 2011; Malone, 2012; Yuan et al., 2013] whose performance was already proven.

To do list

- Implement this approach on our software platform PILGRIM
- Provide an anytime PRM structure learning algorithm, following the ideas presented in [Aine et al., 2007; Malone et al., 2013] for BNs

Thank you for your attention