PREDICTIVE MODELING FOR BUSINESS PROCESSES

DESIGNING AND EVALUATING AN INTERPRETABLE PREDICTIVE MODELING TECHNIQUE FOR BUSINESS PROCESSES
MOTIVATION

Past

- Ex-post analysis

Present

- Process monitoring

Future

- Predictive analytics
MOTIVATION

PREDICTIVE MODELING FOR BUSINESS PROCESSES? WHY?

▪ Early warning systems
  ▪ Predict future behavior
  ▪ Warn managers if future is bad
  ▪ Intervention possible

▪ Anomaly detection systems
  ▪ Predict future behavior
  ▪ Warn managers if a surprising future has happened
  ▪ Analysis / Intervention possible
LITERATURE
TRANSITION SYSTEM MINING
LITERATURE
FREQUENCIES ANNOTATED
LITERATURE
COMPLETION TIMES ANNOTATED
ABSTRACT PRINCIPLE

Event log

Event Sequence

Dimensionality Reduction

State

Classification / Regression / ...

Prediction

P (Future | State, Other Information)

A \rightarrow 0.33\%

B \rightarrow 0.67\%
MAIN PROBLEM

DIMENSIONALITY REDUCTION <-> PROCESS DISCOVERY

- Dimensionality reduction
  - Map event log to a useful feature set

- Question: What is a good process discovery algorithm for predictive modeling applications?
TWO APPROACHES

GRAMMATICAL INFERENCE THEORY

- Process = set of valid event sequences
- Process = probability distribution over event sequences
TWO APPROACHES
GRAMMATICAL INFERENCE THEORY

- Strong language bias necessary
- Weaker language bias possible

Choice by language bias

Choice by comparison based on data
PROBABILISTIC MODELS

PROBABILISTIC MODELS

- Hidden Markov Model (HMM)
- Probabilistic Finite Automaton (PFA)
PFA ESTIMATION

3 FAMILIES OF METHODS

- Bayesian inference
  - Do not estimate a single model (e.g., Gibbs sampling)
  - But: effective!

- Parameter estimation
  - Estimate parameters (often: ML estimation)
  - Quite effective too

- State merging
  - Iteratively merge states, starting with prefix tree
  - Least effective

PFA MODIFICATIONS

Predictive Modeling for Business Processes
Dominic Breuker
2014-09-08
PFA MODIFICATIONS

\[ P(Z_0) \sim \text{Categorical}(\pi_0, ..., \pi_K) \]

\[ P(X_t | Z_t = k) \sim \text{Categorical}(b_{k0}, ..., b_{kE}) \]

\[ P(Z_t | Z_{t-1} = k, X_{t-1} = e) \sim \text{Categorical}(a_{ke0}, ..., a_{keK}) \]

\[ P(\pi_1, ..., \pi_K) \sim \text{Dirichlet}(\rho_1, ..., \rho_K) \]

\[ P(b_{k1}, ..., b_{kE}) \sim \text{Dirichlet}(s_{k1}, ..., s_{kE}) \]

\[ P(a_{ke1}, ..., a_{keK}) \sim \text{Dirichlet}(r_{ke1}, ..., r_{keK}) \]
# EVALUATION (PREDICTION)

<table>
<thead>
<tr>
<th>Event log</th>
<th>Predictor</th>
<th>Accuracy</th>
<th>ØSensitivity</th>
<th>ØSpecificity</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td>EM</td>
<td>0.719</td>
<td>0.578</td>
<td>0.955</td>
<td>11.183</td>
</tr>
<tr>
<td>W</td>
<td>5-gram</td>
<td>0.728</td>
<td>0.588</td>
<td>0.957</td>
<td>Infinity</td>
</tr>
<tr>
<td>2012</td>
<td>EM</td>
<td>0.801</td>
<td>0.723</td>
<td>0.980</td>
<td>3.093</td>
</tr>
<tr>
<td>A</td>
<td>4-gram</td>
<td>0.801</td>
<td>0.723</td>
<td>0.980</td>
<td>2.839</td>
</tr>
<tr>
<td>2012</td>
<td>EM</td>
<td>0.811</td>
<td>0.647</td>
<td>0.973</td>
<td>4.513</td>
</tr>
<tr>
<td>O</td>
<td>3-gram</td>
<td>0.811</td>
<td>0.647</td>
<td>0.973</td>
<td>4.180</td>
</tr>
<tr>
<td>2013</td>
<td>EM</td>
<td>0.714</td>
<td>0.383</td>
<td>0.974</td>
<td>12.041</td>
</tr>
<tr>
<td>Incidents</td>
<td>4-gram</td>
<td>0.635</td>
<td>0.377</td>
<td>0.967</td>
<td>Infinity</td>
</tr>
<tr>
<td>2013</td>
<td>EM</td>
<td>0.690</td>
<td>0.521</td>
<td>0.945</td>
<td>7.231</td>
</tr>
<tr>
<td>Problems</td>
<td>3-gram</td>
<td>0.699</td>
<td>0.564</td>
<td>0.948</td>
<td>Infinity</td>
</tr>
</tbody>
</table>
DEMOnSTRATION

- Visualization is possible
- Threshold
  - Cut out improbable transitions
- Also possible: Petri net synthesis
  - -> Petrify
# EVALUATION (PROCESS DISCOVERY)

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Fitness</th>
<th>Advanced behavioral appropriateness</th>
</tr>
</thead>
<tbody>
<tr>
<td>EM + Petrify</td>
<td>0.998</td>
<td>0.908</td>
</tr>
<tr>
<td>AGNES-Miner</td>
<td>0.995</td>
<td>0.813</td>
</tr>
<tr>
<td>α+</td>
<td>0.969</td>
<td>0.873</td>
</tr>
<tr>
<td>α++</td>
<td>0.984</td>
<td>0.879</td>
</tr>
<tr>
<td>DT Genetic Miner</td>
<td>0.996</td>
<td>0.778</td>
</tr>
<tr>
<td>Genetic Miner</td>
<td>0.998</td>
<td>0.737</td>
</tr>
<tr>
<td>Heuristics Miner</td>
<td>0.973</td>
<td>0.809</td>
</tr>
<tr>
<td>ILP Miner</td>
<td>1.000</td>
<td>0.786</td>
</tr>
</tbody>
</table>

Conclusion

- Goal: Develop a good “event sequence -> state” reduction for predictive modeling in BPM
  - Probabilistic approach
  - Weak language bias

- Probabilistic finite automaton (PFA)
  - Modified (start/end state + regularization)
  - Estimation with EM
  - Can be used as process discovery algorithm

- PFA can be better than n-gram approaches...
- ... but does not have to be!
Questions?

Slides available at:
http://goo.gl/Bi99Ck