Business Process Deviation Prediction
Predicting Non-Conforming Process Behavior
Michael Grohs, Peter Pfeiffer, Jana-Rebecca Rehse
General Setting

To-Be Model

Undesired deviations

Event Log
Running Example
To-Be Model: BPIC 12 only “A_” activities

Managers want to know in which traces which deviations can be expected and accordingly introduce preventive measures

Our Approach
Running Example
To-Be Model: BPIC 12 only “A_” activities

Complete Trace $t_1$

25,000€ Requested Amount
Running Example
To-Be Model: BPIC 12 only “A_” activities

Complete Trace $t_1$

Grohs, Pfeiffer, Rehse – BPDP @ ICPM 2023
26.10.2023
Running Example
To-Be Model: BPIC 12 only “A_” activities

Complete Trace $t_2$

A_ SUBMITTED → A_PARTLY SUBMITTED → A_PRE ACCEPTED → A_ ACCEPTED → A_ FINALIZED → A_ APPROVED → A_ REGISTERED → A_ ACTIVATED

Lea Lea Marc Lea Marc Marc

10,000€ Requested Amount

No more deviations

A_ SUBMITTED → A_PARTLY SUBMITTED → A_PRE ACCEPTED → A_ ACCEPTED → A_ FINALIZED

Lea

No more deviations

A_ SUBMITTED → A_PARTLY SUBMITTED → A_PRE ACCEPTED → A_ ACCEPTED → A_ FINALIZED → A_ APPROVED

Lea

No more deviations

A_ SUBMITTED → A_PARTLY SUBMITTED → A_PRE ACCEPTED → A_ ACCEPTED → A_ FINALIZED → A_ APPROVED → A_ REGISTERED

Lea

No more deviations

Grohs, Pfeiffer, Rehse – BPDP @ ICPM 2023

26.10.2023
Challenges for Deviation Prediction

$C_1$: Explicit process knowledge required

Prediction targets are clearly defined (unlike in, e.g., anomaly detections)

usage of to-be model necessary
Challenges for Deviation Prediction

$C_2$: Prediction targets have specific nature

- $C_{2.1}$: Multi-label targets
  one trace might deviate in more than one way (e.g., $d_1$ and $d_2$ in $t_1$)

- $C_{2.2}$: Imbalanced targets
  deviating traces are rather infrequent, leading to highly imbalanced data

- $C_{2.3}$: Dynamic targets
  labels change over the duration of the trace (i.e., after deviation happened)
Challenges for Deviation Prediction

\( C_3: \) Context is of importance

Complete Trace \( t_1 \)

\[
\begin{align*}
A_{\text{SUBMITTED}} & \rightarrow A_{\text{PARTLY SUBMITTED}} & A_{\text{PRE ACCEPTED}} & \rightarrow A_{\text{ACCEPTED}} & \rightarrow A_{\text{FINALIZED}} & \rightarrow A_{\text{REGISTERED}} & \rightarrow A_{\text{APPROVED}} & \rightarrow A_{\text{REGISTERED}} & \rightarrow A_{\text{ACTIVATED}}
\end{align*}
\]

Requested Amount: 25,000€

Complete Trace \( t_2 \)

\[
\begin{align*}
A_{\text{SUBMITTED}} & \rightarrow A_{\text{PARTLY SUBMITTED}} & A_{\text{PRE ACCEPTED}} & \rightarrow A_{\text{ACCEPTED}} & \rightarrow A_{\text{FINALIZED}} & \rightarrow A_{\text{APPROVED}} & \rightarrow A_{\text{REGISTERED}} & \rightarrow A_{\text{ACTIVATED}}
\end{align*}
\]

Requested Amount: 10,000€

Grohs, Pfeiffer, Rehse – BPDP @ ICPM 2023
26.10.2023
Challenges for Deviation Prediction

$C_4$: Action orientation requires specific focus

Process Manager

Want **high recall** (correctly recognizing all deviations) **over high precision** (not misclassifying conforming instances)

Default Classifiers
Optimize accuracy
Goal

Predict which deviations will happen in the future of incomplete traces

**Offline Component**
- To-Be Model $B$
- Event Log $L$
- Labelling
- Encoding
- Encoded Prefixes
- Labelled Prefixes
- Model Learning
- Trained Classifiers

**Online Component**
- Trace Prefix
- Trained Classifiers
- Predictions

$d_1$: Deviation
$d_2$: Deviation
$\ldots$
$d_m$: No Deviation
Define labels

To consider explicit process knowledge, we define all labels based on alignments of traces with to-be models

Set of deviation types $D$

Assign labels

To account for multiple targets that change dynamically throughout the trace, we label each prefix individual

Label for each prefix for each $d \in D$

Addresses $C_1$

Addresses $C_{2.1}$ & $C_{2.3}$
Business Process Deviation Prediction
Labelling - Example

Complete Trace $t_1$

Corresponding Labels

<table>
<thead>
<tr>
<th>Trace Prefix</th>
<th>Labels</th>
</tr>
</thead>
<tbody>
<tr>
<td>(SUB)</td>
<td>$d_1 : (\to, \text{APP})$, $d_2 : (\text{APP}, \to)$</td>
</tr>
<tr>
<td>(SUB, PAR)</td>
<td>1</td>
</tr>
<tr>
<td>(SUB, PAR, PRE)</td>
<td>1</td>
</tr>
<tr>
<td>(SUB, PAR, PRE, ACC)</td>
<td>1</td>
</tr>
<tr>
<td>(SUB, PAR, PRE, ACC, FIN)</td>
<td>1</td>
</tr>
<tr>
<td>(SUB, PAR, PRE, ACC, FIN, REG)</td>
<td>0</td>
</tr>
<tr>
<td>(SUB, PAR, PRE, ACC, FIN, REG, APP)</td>
<td>0</td>
</tr>
<tr>
<td>(SUB, PAR, PRE, ACC, FIN, REG, APP, ACT)</td>
<td>0</td>
</tr>
</tbody>
</table>
Apply and compare two context-aware encodings

- Complex Index-Based Encoding (CIBE)
- Learned Encoding (MPPN)

Addresses $C_3$
We apply under-sampling to the training split

Applied Under-Sampling: One-Sided Selection
- Combines Tomek Links and Condensed Nearest Neighbor (CNN) Rule
- Tomek Links removes ambiguous samples
- CNN removes redundant samples
We train one classifier per $d \in D$.

Addresses $C_{2.1}$
We use a weighted cross-entropy loss function ($WCEL$)

$$\text{loss}_{WCEL} = \begin{cases} 
16 & \text{if False Negative} \\
1 & \text{if False Positive}
\end{cases}$$

- Penalizes FN more
- Optimizes Recall

Addresses $C_4$
### Evaluation

#### Datasets

<table>
<thead>
<tr>
<th>Log L</th>
<th>Traces</th>
<th>Events</th>
<th>Trace Attr.</th>
<th>Trace Length</th>
<th>Dev. Types</th>
<th>Deviating Traces</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>min.</td>
<td>avg.</td>
<td>max.</td>
<td>min.</td>
<td>avg.</td>
<td>max.</td>
</tr>
<tr>
<td>BPIC 12</td>
<td>A</td>
<td>O</td>
<td>1</td>
<td>3</td>
<td>4.7</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>13,087</td>
<td>5,015</td>
<td>60,849</td>
<td>31,244</td>
<td>3</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>8</td>
<td>399</td>
<td>927</td>
<td>1,191</td>
<td>1,761</td>
</tr>
<tr>
<td>BPIC 20</td>
<td>Dom. Dec.</td>
<td>Int. Dec.</td>
<td>10,500</td>
<td>56,437</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>6,449</td>
<td>72,151</td>
<td>6,886</td>
<td>36,796</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>RIF</td>
<td>Prep.</td>
<td>2,099</td>
<td>18,246</td>
<td>16</td>
<td>1</td>
</tr>
<tr>
<td>MobIS</td>
<td>3,354</td>
<td>55,809</td>
<td>11</td>
<td>16.6</td>
<td>49</td>
<td>43</td>
</tr>
</tbody>
</table>

- **Differing numbers of prefixes**
- **High imbalance of deviations**
- **Differing amount of context**
- **Many deviation types \( d \in D \)**
## Evaluation

### Results

<table>
<thead>
<tr>
<th>Log</th>
<th>Baselines</th>
<th>BPDP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Genga et. al.</td>
<td>BPDP&lt;sub&gt;CIBE&lt;/sub&gt;</td>
</tr>
<tr>
<td></td>
<td>CatBoost</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Suffix Prediction</td>
<td></td>
</tr>
</tbody>
</table>

- **Approach with similar goal based on statistics**
- **Approaches that could be used instead of BPDP**
- **BPDP with both CIBE and MPPN encodings**

Grohs, Pfeiffer, Rehse – BPDP @ ICPM 2023

26.10.2023
## Evaluation Results

<table>
<thead>
<tr>
<th>Log</th>
<th>Baselines</th>
<th>BPDP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Genga et. al.</td>
<td>CatBoost</td>
</tr>
<tr>
<td></td>
<td>Dev</td>
<td>No Dev</td>
</tr>
<tr>
<td>BPIC 12A</td>
<td>Precision</td>
<td>Recall</td>
</tr>
<tr>
<td>BPIC 12O</td>
<td>Precision</td>
<td>Recall</td>
</tr>
<tr>
<td>BPIC 20 Dom. Dec.</td>
<td>Precision</td>
<td>Recall</td>
</tr>
<tr>
<td>BPIC 20 Int. Dec.</td>
<td>Precision</td>
<td>Recall</td>
</tr>
<tr>
<td>BPIC 20 RIP</td>
<td>Precision</td>
<td>Recall</td>
</tr>
<tr>
<td>BPIC 20 Prep.</td>
<td>Precision</td>
<td>Recall</td>
</tr>
<tr>
<td>MobIS</td>
<td>Precision</td>
<td>Recall</td>
</tr>
</tbody>
</table>
## Evaluation Results

The table shows the performance of various models and techniques evaluated against baselines. The models under consideration are BPIC 12A, BPIC 12O, BPIC 20 Dom. Dec., BPIC 20 Int. Dec., BPIC 20 RFP, BPIC 20 Prep., and MobIS. The metrics include Precision, Recall, AUCROC, and performance metrics such as Dev and No Dev. The best performance is indicated by the highest values in each column.

<table>
<thead>
<tr>
<th>Log</th>
<th>Baselines</th>
<th>BPDP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Genga et. al.</td>
<td>CatBoost</td>
</tr>
<tr>
<td></td>
<td>Precision</td>
<td>Recall</td>
</tr>
<tr>
<td>BPIC 12A</td>
<td>0.2136</td>
<td>0.0921</td>
</tr>
<tr>
<td></td>
<td>0.0678</td>
<td>0.9684</td>
</tr>
<tr>
<td>BPIC 12O</td>
<td>0.1462</td>
<td>0.8663</td>
</tr>
<tr>
<td></td>
<td>0.1340</td>
<td>0.8403</td>
</tr>
<tr>
<td>BPIC 20 Dom. Dec.</td>
<td>0.1961</td>
<td>0.7318</td>
</tr>
<tr>
<td></td>
<td>0.2040</td>
<td>0.7205</td>
</tr>
<tr>
<td>BPIC 20 Int. Dec.</td>
<td>0.1738</td>
<td>0.8823</td>
</tr>
<tr>
<td></td>
<td>0.1648</td>
<td>0.8684</td>
</tr>
<tr>
<td>BPIC 20 RFP</td>
<td>0.1480</td>
<td>0.7352</td>
</tr>
<tr>
<td></td>
<td>0.1244</td>
<td>0.7273</td>
</tr>
<tr>
<td>BPIC 20 Prep.</td>
<td>0.1475</td>
<td>0.8705</td>
</tr>
<tr>
<td></td>
<td>0.1067</td>
<td>0.8629</td>
</tr>
<tr>
<td>MobIS</td>
<td>0.1211</td>
<td>0.8355</td>
</tr>
<tr>
<td></td>
<td>0.1245</td>
<td>0.8415</td>
</tr>
</tbody>
</table>

The models are compared against baselines and each other, with the best performance indicated by bold values.
Evaluation

Shapley Values for one specific prediction

Applying XAI to identify which features lead to a deviation prediction

unlikely to predict a deviation

likely to predict a deviation

Activity of 4th Event: A_ACCEPTED
Trace: AMOUNT_REQUIRED
Resource of 3rd Event: 112
Activity of 3rd Event: A_DECLINED
Activity of 2nd Event: A_PARTLY SUBMITTED
Activity of 3rd Event: A_PREACCEPTED
Trace Weekday Start: Tuesday
Month of 3rd Event: October
Month of 1st Event: November
Trace Weekday Start: Saturday
Further evaluation results

- **Shapley values** allow process managers to detect **features that increase likelihood of deviations**
- Additional use case evaluation showed that **BPDP is early in its predictions**

Insights into deviation prediction

- Only **both under-sampling and WCEL** leads to sufficient results
- Learned encoding **MPPN** performs **nearly as good as CIBE** although trained on next-event prediction
- **Additional context** in BPIC 20 logs could lead to **better performance for shorter prefixes**
Any Questions?
Backup
We train one classifier per $d \in D$

*EF = number of encoded features for each trace prefix