Lehto, Teemu; Hinkka, Markku; Hollmén, Jaakko

Analyzing business process changes using influence analysis

Published in:
CEUR Workshop Proceedings

Published: 01/01/2018

Document Version
Publisher's PDF, also known as Version of record

Please cite the original version:
Analyzing Business Process Changes Using Influence Analysis

Teemu Lehto\textsuperscript{1,2}, Markku Hinkka\textsuperscript{1,2}, and Jaakko Hollmén\textsuperscript{2}

\textsuperscript{1} QPR Software Plc, Finland
\textsuperscript{2} Aalto University, School of Science, Department of Computer Science, Finland

Abstract. Real world business operations are continuously changing. Periodical business performance review sessions typically focus on monitoring changes in key performance indicator (KPI) measures. However, the detection and review of activity level changes in actual business processes is often based on subjective manual observations. This means that many changes are not detected in timely manner making the organization slower to adapt to changes. In this paper we present a systematic method for detecting business process changes for business review purposes based on transaction level data. Our method uses process mining principles and is based on our previously published influence analysis methodology. Unlike most process mining change detection algorithms which operate on case level our method analyzes changes in the individual event level. We show how case level data can be used to construct features to the event level. Our method detects changes in timely manner since there is no need to wait for the cases to be completed. We present two alternative ways, binary approach and continuous event-age approach, for dividing events into recent and old for business review purpose. We also demonstrate the method with data from a real-life case.

Keywords: process analysis, process improvement, change detection, concept drift, process mining, performance management, key performance indicator, root cause analysis, data mining, influence analysis, contribution

1 Introduction

The ability to detect changes is crucial for developing and improving agile business operations. Unwanted changes need to be mitigated quickly and desired changes need to be reinforced and shared as best practices. In this paper we present a systematic approach for analyzing business process data using process mining principles\cite{18} and specifically our previously published influence analysis methodology\cite{11}, \cite{12}. Our methodology is capable of detecting business process changes and describing for each change the important business relevant attributes of what changed, how big is the change, what may be the cause for the changes and what might be the effect and outcome of the change.

Our method is particularly useful for periodic business review situations which take place in most of the business organizations globally. During the business review, managers typically review the performance of business operations
using Key Performance Indicators (KPIs). One problem is that managers typi-
ically do not have an accurate fact-based understanding and analysis of what
has changed during the review period. Instead they typically rely on subjective
comments, views and suggestions biased by acute business challenges and crises.
Using our method in this situation the managers easily see what has changed
during the review period by comparing the new process mining data against data
from previous business review periods. Our method will discover changes that
take place very fast as well as more gradual changes that occur in the course of
several years giving managers accurate data about changes and trends.

The rest of this paper is organized as follows: Section 2 introduces relevant
background in process mining, concept drift and business process management.
Section 3 presents our methodology for analyzing business process changes. Sec-
tion 4 shows a real-life example of using the methodology on the loan application
process followed by a section for Discussions and Summary.

2 Related work

This paper is based on our previously published influence analysis methodol-
ogy which shows how the root causes can be identified for generic process rel-
ated problems [11] and [12]. In this paper we discuss about the concept drift
that occurs over the time and show how influence analysis can be used to dis-
cover changes. Data preparation for influence analysis is based on process mining
methodologies [18] where data from ERP systems is transformed into event log
format containing events with event attributes connected to cases with case at-
tributes.

Handling concept drift in process mining has been discussed in detail in [2],
[3] and [5]. These papers present that operational processes change and suggest
three main problems to be studied: detection of change points, characterization
of change and insight to the process evolution. This work is excellent for under-
standing how complete process executions have changed. However, during the
business review situation the management is reviewing a fixed period of time and
trying to identify as early signals as possible hinting how the processes might
be changing at this very moment. In effect the change point is set to be the
beginning of the review period and question is "show us the things that have
changed after the start of review period as compared to the things that took
place before the review period". As [2], [3] and [5] analyze complete cases they
can be categorized as off-line analysis of changes. If cases take 6 months to com-
plete the analysis results based on complete cases are at least 6 months old. In
this paper we will present a method for on-line analysis of changes.

Concept drift in relation to machine learning has been studied a lot for exam-
ple in [8], [15] and [19]. The objective of those studies is to increase the accuracy
of predictions by utilizing machine learning algorithms that discover the changes
in the process. Instead of making accurate predictions our method is tailored to
discover and explain changes as part of the systematic periodical business review.
A novel Trace Clustering algorithm [10] presents an approach to analyze attribute data from events and cases in addition to the traditional business process data. The approach is based on Markov cluster (MCL) algorithm [7] for finding similar cases. Although the results look promising the challenge of this approach is that it uses complete cases and is thus more useful for off-line analysis than for periodical on-line analysis.

An approach more targeted for on-line business process drift detection is presented in [14]. It uses the concept of partial order runs to run statistical tests in order to find the exact point in time for the change. A somewhat similar method for concept-drift detection in event log streams has been studied in [1] which present a method for detecting actual concept-drift time and individual anomalies using histograms and clustering. However, these methods do not take into account the attribute data and are not aimed to provide insight to the business review question of what has changed during the current business review period in comparison to the operations before.

Since it is difficult to detect the changes by using only traditional statistical measures, an interesting set of visual analytics tools enabling interactive process analysis and process mining is presented in [4]. Plotting all events to a stacked area graph with absolute calendar time on horizontal axis this paper presents a visualization for detecting concept drift and changes in business process and case attribute data. Even though the presented visual analytics are useful they do not clearly provide a concrete answer to what has changed during the past review period. Presented visualization techniques also have challenges when the amount of case attributes is so large that all case attributes cannot be included in the visualizations at the same time.

Yet another approach to make business people aware of changes that require active intervention is presented in [17]. The method uses a sophisticated cost model for optimize the generation of alarms for business people. Challenge with this active intervention method is that it requires a lot of settings and detailed level knowledge about the importance of various process issues. These settings must be beforehand so that the algorithm can then suggest active intervention when needed. Our understanding from actual business operations is that this kind of settings and detailed information is not available or it is very difficult and expensive to maintain over the time. On the other, practically all organizations do have so kind of business reviews, so it is beneficial to present the discoveries as part of the business review meetings.

Summary of related work:

- Process mining and concept drift has been studied a lot.
- Most of the presented studies can be categorized as off-line analysis. They are related to detecting and analyzing changes based on completed cases.
- Limited tools exist for on-line analysis that could be used to compare fixed business period (like previous month, week, day, quarter or year) with past performance.
Some methods are tailored for detecting process flow changes and some methods detect case attribute data changes, our approach can detect both changes and comparing them with a uniform scale for reporting purposes.

Machine learning is mostly used for making predictions and not used so much for supporting business review analysis.

Our previously published influence analysis methodology shows how the root causes can be identified for generic process related problems [11]. In this paper we discuss about the concept drift that occurs over the time and show how influence analysis can be used to discover changes. As shown in [12] the influence analysis can be used as binary analysis or continuous analysis. In this paper we show the both approaches and discuss the benefits and challenges of each approach.

3 Analyzing Business Process Changes

Idea of this paper is to analyze the variance and deviations in business processes on individual transaction level in order to discover and explain changes that have taken place. Using the influence analysis methodology our objective is to find areas that have more variance compared to the average areas. Our method is based on the idea that if there is no changes in the operations then the data in ERP system for the review period is similar to the data for the previous periods. On the other hand, if there is changes, then the data will be different than in the past. In this chapter we refine and augment the previously presented influence analysis steps.

3.1 Identify the relevant business process and define the case

Our approach detects changes from one business process at a time. A large organization with multiple processes needs to run the analysis separately for each business process to detect the changes in all business operations. Typically, the business reviews are based on consolidated data, for example a dashboard report can contain several Key Performance Indicators (KPI). The ERP system in large organization can easily contain 1 billion new database level transactions (ie. database rows) per month. If the review is based on 10 KPIs with 100 consolidated drill-down measures each, then then we could say that we use 1 000 out of 1 billion, i.e. 0.0001% of the available data for making findings in business review. However, if we set-up 10 process mining models that contain an average of 1 million transaction level events per business review period of 1 month, then we use 10 000 000 out of 1 billion transaction, i.e. 1% of total data for supporting the business review. In this example we would use 10 000 times more data for supporting business review compared to the previous situation with only the KPI data. Based on these ideas we propose organizations to analyze as many processes as possible and include as many events as possible in order to get a wide view into the changes in business operations. We also suggest to the data to be prepared so that it covers as long as possible end-to-end processes in order to facilitate identifying root causes for the discovered process changes.
3.2 Collect event and case attribute information

In this paper we propose the idea of collecting and analyzing the data on event level. In practice this means that data is not consolidated from individual cases to the event level but rather the event level data is used as it is and case level data is copied to each event.

Since our goal is to create new insight for business people, we encourage to use all possible event and case attribute data that is available. Generation of suitable log files with extended attributes is well studied area [13]. There also exists methods for enriching and aggregating event logs to case logs [16]. We summarize a key method for constructing event and case logs for event level with following steps:

- Starting point is the event which is typically a result of one business transaction, for example the relational database table whose rows correspond to transactions $E$.
- Use the properties of each event $e_i$ in $E$ as event attributes.
- Identify for each event $e_i$ a corresponding case $c_i$ and copy all case attributes as event attributes.
- Form a event path for each event $e_i$ by concatenating the event type names of events linked to the same case sorted from oldest to newest. Event path can be expressed in many ways, for example as single event attribute containing the full path or as several attributes containing single predecessor values.
- Identify for each event $e_i$ in $E$, a set of objects $O_i$ such that every object $o_{ij}$ in $O_i$ is linked to $e_i$. Use the properties of objects $o_{ij}$ as additional event attributes for events $e_i$.
- Further augment every event $e_i$ by adding external events that have occurred at the same time. Examples of external events include machinebreak, weekend, strike, queuetoolong and badweather. Adding external events makes it possible to use this same approach for detecting changes in external circumstances as well.

3.3 Create new categorization dimensions

In this paper we will present new categorization dimensions specifically useful for the event level analysis.

The purpose of this step is to create new categorization dimensions for the cases. All these dimensions will then be used for detecting the changes, so the more dimensions we have the larger the coverage of our analysis will be. Table 1 shows examples of dimensions that can be created for every event log based on the log itself.

Categorization Dimensions form the bases for our Influence Analysis when discovering the business process changes. Without any Categorization Dimensions we could only make a discovery that in the review period there is more, less or equal amount of transactions compared to the comparison period. Having the Event types dimension enables us to detect changes in the amounts of particular
Table 1. Categorization Dimension for analyzing business process changes

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Amount</th>
<th>Dimension Identified</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Event types</td>
<td>One</td>
<td>Event Type</td>
<td>Event type name</td>
</tr>
<tr>
<td>Case attributes</td>
<td>One for each case attribute</td>
<td>CA1 + Case attribute name</td>
<td>Case attribute value</td>
</tr>
<tr>
<td>Case attributes by event type</td>
<td>One for each combination of event type and case attribute</td>
<td>CA2 + Event type name + Case attribute name</td>
<td>Case attribute value</td>
</tr>
<tr>
<td>Event attributes</td>
<td>One for each event attribute</td>
<td>EA1 + Event attribute name</td>
<td>Event attribute value</td>
</tr>
<tr>
<td>Event attributes by</td>
<td>One for each combination of event type and event attribute</td>
<td>EA2 + Event type name + Event attribute name</td>
<td>Event attribute value</td>
</tr>
<tr>
<td>Predecessor change</td>
<td>One</td>
<td>Predecessor1</td>
<td>Event type name + Predecessor event type name</td>
</tr>
<tr>
<td>Predecessor change by event type</td>
<td>One for each event type</td>
<td>Predecessor2 + Event type name</td>
<td>Predecessor event type name</td>
</tr>
<tr>
<td>Process path</td>
<td>One</td>
<td>Path1</td>
<td>Full predecessor event type path without the event itself</td>
</tr>
<tr>
<td>Process path by event type</td>
<td>One for each event type</td>
<td>Path2 + Event type name</td>
<td>Full predecessor event type path without the event itself</td>
</tr>
</tbody>
</table>

event types, for example we could find out that there was more *Ontime Delivery* kind of events and less *Customer Complaint* kind of events during the review period as compared to the comparison period. The *Case attributes* dimension in Table 1 is even more interesting since it allows us to detect changes in the background data of active cases, for example in November there was more cases from *Region* with value *Finland* compared to previous 6 months. Case attribute changes may be analyzed as specific to certain event types using the event type name in the dimension identifier or as global case attributes without the event type name, or both. In the similar manner all the dimensions in Table 1 can be added to the analysis. Total amount of dimensions, ie. feature vectors for case analysis can easily grow large if all the dimensions are taken into use. For example with 30 event types, 50 case attributes and 10 event attributes the total amount of dimensions from Table 1 would be $1 + 50 + 1500 + 10 + 300 + 1 + 30 + 1 + 30 = 1923$. In order to handle this curse of dimensionality we suggest three solutions in real life business review cases. 1. Use Influence Analysis as described in this paper since it in effect only shows those dimensions where the changes are largest. 2. Select only those dimensions that seem to be important for review purposes. Benefit of this is that business people are not overloaded with data that they cannot understand. Problem is that some dimensions may at some point of time contain very useful information about process changes and in case that dimension is taken away then naturally it is not reported to business people, so the change may be left unnoticed. 3. Use advanced feature selection algorithms as presented in [9].
3.4 Define data for review and comparison periods

The original Influence Analysis that is presented in [11] uses a binary classification that specifies each case as either problematic or successful. In this paper we alter this classification in three ways. First, instead of analyzing cases we run the analysis on the transaction/event level. Second, instead of specifying cases as problematic or successful we specify events as belonging to the review period or as belonging to the reference period. Third, in addition to using only the Binary approach as presented in [11] we also use an additional Continuous approach as presented in [12].

**Binary approach**  Figure 1 shows how the analysis data is divided into four different periods in order to identify process changes

![Fig. 1. Business review periods using fixed periods](image)

- **Review period** (c). All events occurring during this period are taken into consideration when discovering changes. If these events, their quantities and event attributes are similar to the comparison period events, then there is no big changes. In real life something is always changing so our target is to detect the most important changes. Example review period could be one month like November.

- **Comparison period** (b). All events occurring during this period are also taken into consideration when discovering changes. Typical setup would be to use the 6 months prior to the review period as comparison period so as an example the review period could be May to October. If a business is very seasonal then one option is to use year-to-year comparison period so that the Comparison period could be same month last year.

- **History period** (a). The events that occurred during the History period are not used as separate events for the review and/or comparison sets. However, these events will be used for constructing the business process path (trace) for each event in both review and comparison periods. For example: Review period and comparison periods both contain events `OntimeDeliveryFailed`. 

38
In order to understand the root causes for these failures we want to include a full process path for each \textit{OntimeDeliveryFailed} event so that we would see the difference of how cases are ending up to the \textit{OntimeDeliveryFailed} process step. For this reason we need to use the predecessor events also from the History period when constructing this path for review and comparison period events.

- \textbf{Most Recent Data period (d)}. All events occurring after the review period are excluded completely from the analysis. As an example, the typical business review for November is done in early December when the data from November is complete. We do not want to use the recent data from December as it becomes available because that data will be analyzed in next month business review. Naturally it is possible to set-up the review period as last 30 days so that all recent data from last 30 days is regarded as the review data and the Most Recent Data period would then be empty.

Benefit of using a binary approach is that it is typically easy to use for business people who have prior knowledge about the operations for both review period and comparison period. Since the history period has no weight in the analysis, they can fully ignore exceptional transactions, projects and cases that have been completed during history period. Binary approach also guarantees that all discovered changes indeed have taken place exactly during the well-defined review period.

\textbf{Continuous approach} Another approach for defining review period and comparison period is to use a continuous measure to determine which period any particular event belongs. Benefits of using Continuous approach include: 1. If the analysis is done by an analyst as part of a one-time process analysis then there is no continuous review process and it would be easier to just let the system divide event into review and comparison periods. 2. Continuous approach give the analyst more freedom for setting weights for individual events, for example events occurring during past 6 months may have a certain weights and events occurring 6-12 months ago could have even higher weight. One straightforward way to split events into Review and Comparison periods could be: Use half of the available data for History period. This ensures that most of the events have proper history and we should not discover changes that result from predecessor events not being included in the dataset. The other half could then again be used as 50% Comparison period and 50% Review period. This approach gives a nice 50% ratio so that for each dimension and analysis finding there should be an equal amount of that in both Comparison and Review data. If we want to specifically detect changes that have occurred over the time then we could consider giving the oldest and newest cases more weight than for the events that take place when Comparison period ends and Review period starts, since we are not really reviewing a specific calendar month in business review style. One way to achieve this is to calculate an \textit{Age} attribute for each event. \textit{Age} would be equal to the elapsed time between the actual time of the event and current time. We then use the \textit{Age} attribute as the lead time measure for continuous
contribution formulas as defined in Table 8, in [12]. In practice the continuous approach using Age gives the largest weight for the events that take place in the beginning of the comparison period and in the end of the review period. Events that take place in the middle have very small weight so the analysis tells how changes have taken place during the whole period. This is particularly good approach for analyzing small gradual changes that occur over a longer period of time.

3.5 Detecting changes using Influence Analysis

The Influence Analysis has been presented in [11] and [12] for analyzing root causes for process mining cases. In this paper we use the same formulas with the exception that instead of using process mining cases we do the analysis on process mining event level. Another change is that the concept of comparing problematic cases with normal cases is replaced with the definition of comparing review period data with comparison period data in order to detect changes that have occurred in the time dimension. To reflect these changes the Influence Analysis definitions and equations are presented as follows:

**Definition 1.** Let \( E = \{e_1, \ldots, e_N\} \) be a set of events in the process analysis. Each event represents a single transaction that happened at a particular time and is related to single business process instance.

**Definition 2.** Let \( E_a = \{e_{a1}, \ldots, e_{aN}\} \) be a set of events sharing a same characteristics as defined in segment A. \( E_a \subseteq E \). These characteristics are derived from different values for the Categorization Dimensions.

**Definition 3.** Let \( E_p = \{e_{p1}, \ldots, e_{pN}\} \) be a set of Review Period events. \( E_p \subseteq E \) to be used in Binary Approach.

**Definition 4.** Let \( d_{e_j} \) be the age of the event \( e_j \) to be used in Continuous Approach.

**Definition 5.** Let \( pr \) be the problem size in the original situation before any business process improvement. According to the terminology in [12] for Binary Contribution (BiCo) the problem size is equal to the total number of events in review period, and for Continuous Contribution (CoCo) it is the sum of distance between Age and average Age for each event separately. The practical meaning of problem size is to define the number of events that belong to the review period.

**Binary Change Window** In Binary Change Window analysis each event is either included in the set of Review Period events or the set of Comparison period events. Note that events belonging to History period and Most Recent Data period have already been excluded from the analysis.

Converting the formulas from Table 8 in [12] from cases to events we get: Total problem size for BiCo is the number of problematic events \( pr_{BinaryChange} = \)
\[ |E_p| = \sum_{e_j \in E_p} 1 \] as shown in equation 1. Average function for BiCo is the average problem density \( \rho = \frac{|E_p|}{|E|} = \frac{\sum_{e_j \in E_p} 1}{\sum_{e_j \in E} 1} \) as shown in equation 2. Similarly the average problem density for BiCo of subset \( E_a \) is \( \rho_a = \frac{|E_p \cap E_a|}{|E_a|} = \frac{\sum_{e_j \in (E_p \cap E_a)} 1}{\sum_{e_j \in E_a} 1} \) as shown in equation 3. Finally the Contribution% for BiCo of subset \( E_a \) is

\[
\text{ConBiCo} = \frac{(\rho_a - \rho)}{\rho} \sum_{e_j \in E_a} 1 = \frac{|E_p \cap E_a|}{|E_p|} = \frac{\sum_{e_j \in (E_p \cap E_a)} 1}{\sum_{e_j \in E_p} 1} - \frac{\sum_{e_j \in E_a} 1}{\sum_{e_j \in E} 1}
\]
as shown in equation 4 in Table 8. in [12].

**Continuous Change Window** For Continuous Change Window the formulas from Table 8 in [12] are written as: Total problem size for CoCo is the sum of distance between Age and average Age for each event separately

\[
p_{\text{ContinuousChange}} = \frac{1}{2} \sum_{e_j \in E} |d_{e_j} - \bar{d}|
\]
as shown in equation 9. Average function for CoCo is the average age \( \bar{d} = \frac{\sum_{e_j \in E} d_{e_j}}{\sum_{e_j \in E} 1} \) as shown in equation 10. Similarly the average problem density for CoCo of subset \( E_a \) is \( \bar{d}_a = \frac{\sum_{e_j \in E_a} d_{e_j}}{\sum_{e_j \in E_a} 1} \) as shown in equation 11. Finally the Contribution% for CoCo of subset \( E_a \) is

\[
\text{ConCoCo} = \frac{(\bar{d}_a - \bar{d})}{p_{\text{CoCo}}} \sum_{e_j \in E_a} 1
\]
as shown in equation 12 in Table 8. in [12].

**4 Case Study: BPI Challenge 2017 Dataset**

In this section we show a real life example of using the presented methodology on the loan applications process data from a Dutch Financial Institute. The data is publicly available as BPI Challenge 2017 Dataset [6] and contains 31 509 cases and a total of 1 202 267 events. The original dataset has been prepared in a way that it contained full cases. Since the purpose of our analysis is to show business process changes within a continuous monitoring situation, we have taken the following steps in preparing a setup for business review analysis.

- November 2016 is selected as the business review month. The data contains 104 946 events whose time stamp belongs to November, so the Review period will consist of these events.
- All events occurring later than November belong to Most Recent Data period and are excluded from analysis, consisting of 120 568 events. These events would naturally be included in later business review periods.
- Comparison period has been chosen to include the 6 months before review period, i.e. from May 2016 to October 2016. Comparison period contains 647 406 events.
History period contains 329,347 events occurring before May 2016. These events are used for constructing the process path and predecessor dimensions for History and Review events, but they are not included in the analysis as actual events belonging to either Comparison or Review sets.

Total amount of events in the analysis is 752,352 consisting of 104,946 events for Review period (13.95% of all events) and 647,406 events for Comparison period (86.05% of all events).

![Fig. 2. Changes for BPI Challenge 2017 Applications. Changes for November 2016 compared to previous 6 months](image)

**Results using Binary Change Window** Figure 2 shows the top-10 most important changes in the business process and related data for the review period. We see that there is a lot of User changes in event attribute `org:resource` so it seems like employees are changing a lot. `User_133` has conducted 7,282 events during the review period and only 4,267 in the comparison period so 33% of his events have taken place during the review period, which makes him to be the biggest increase in volume taken into account the size of his total activity (7,995 events) and the difference 33% from average 13.95% of activities which should take place in review period.

Figure 3 shows the changes in only the event type dimension. The event types `W_Call incomplete files - suspend` and `W_Call after offers - ate_abort` occur more often during the Review period whereas the event types `W_Validate application - resume` and `W_Call after offers - suspend` occur less often during the Review period than in Comparison period.

Considering the business process related changes where the order of activities is changing we limit the analysis to only the predecessor changes where a certain event takes place immediately after another event as shown in Figure 4. During the review period the control flow transition from event `W_Call after offers - suspend`
Fig. 3. Changes in Event Types for BPI Challenge 2017 Applications. Changes for November 2016 compared to previous 6 months

---

ate_abort to W_Call after offers - schedule occurs more often and the transition from event type W_Verify application - request to W_Verify application - resume less often than during the comparison period.

Results using Continuous Change Window Figure 5 shows the continuous approach versions of the same overall analysis as the previous binary approach Figure 2. The continuous analysis is configured to discover differences in events from mid-August to November with events from May to mid-August. The results of Continuous analysis are the result of giving each event a weight based on the Age of the event. The bigger the distance from average Age the bigger the weight of that particular event. In our example data the average Age of events is 103.97 days and the distance from average Age is then between +103.97 days and -103.97 days. An event taking place in either end (oldest and youngest) have about 100 times the weight compared to an event taking place 1 day after of before the average Age. Similarly an event that takes place exactly in the average Age has zero weight as it does not belong either to the old period or new period. Continuous analysis results are well in line with the binary approach results and differences are based on the different setup of Review and Comparison periods and a different weighting aproach as described. For example User_133 as the new value for org:resource is still the biggest change and both org:resource User_67 and User_65 are included in top-10 changes for both Binary and Continuous approaches as is visible in Figures 2 and 5.

5 Summary and Conclusions

In this paper we have presented a method for detecting business process changes. The method is based on our previously published Influence Analysis and it uses
the conformance measure to scale different types of changes in order to present various kinds of changes sorted by their significance. One novel idea in this paper is to use Influence Analysis on the event level instead of business process case level. Operating on the event level makes it possible to use all available data from the review period for detecting changes instead of having to wait until a business process case is completed. Summary of our key experiences when using the analysis with real-life cases include:

- Changes in business operations can be analyzed by comparing Review period events to the Comparison period events using influence analysis.
- Business people quickly learn to read the influence analysis results on monthly bases. Detecting the top-10 or top-50 changes gives a very good starting point for a more detailed periodical analysis of business process changes.
- Detected changes may also be a result of incorrect data integration between process mining system and the actual ERP system(s). The method presented in this paper serves as an easy-to-use quality assurance tool for evaluating the correctness of periodical data loads and integrations. For example, after each monthly, weekly or daily data import the system can notify business analyst about the top-10 changes so that a potential technical integration problem is detected and corrected before other business users spend a lot of time in analyzing incorrect data.

Acknowledgements. We thank QPR Software Plc for the practical experiences from a wide variety of customer cases and for funding our research. The algorithms presented in this paper have been implemented in a commercial process mining tool QPR ProcessAnalyzer.

References


Fig. 4. Changes in Predecessors for BPI Challenge 2017 Applications. Changes for November 2016 compared to previous 6 months
Fig. 5. Changes for BPI Challenge 2017 Applications. Changes for November 2016 compared to previous 6 months using continuous comparison approach.


