Lehto, Teemu; Hinkka, Markku; Hollmén, Jaakko

Focusing business process lead time improvements using influence analysis

Published in:
Data-driven Process Discovery and Analysis

Published: 01/01/2017

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

Please cite the original version:
Lehto, T., Hinkka, M., & Hollmén, J. (2017). Focusing business process lead time improvements using influence analysis. In P. Ceravolo, M. Van Keulen, & K. Stoffel (Eds.), *Data-driven Process Discovery and Analysis: Proceedings of the 7th International Symposium on Data-driven Process Discovery and Analysis (SIMPDA 2017), Neuchatel, Switzerland, December 6-8, 2017* (pp. 54-67). (CEUR Workshop Proceedings ; Vol. 2016). CEUR. http://ceur-ws.org/Vol-2016/

Focusing Business Process Lead Time Improvements Using Influence Analysis

Teemu Lehto[1,2], Markku Hinkka[1,2], and Jaakko Hollmén[2]

[1] QPR Software Plc, Finland
[2] Aalto University, School of Science, Department of Computer Science, Finland

Abstract. Shortening lead times in a business process is important for meetings service level agreements, decreasing inventories and working capital, keeping customers satisfied and in short: staying in business. Process mining methods make it possible to generate a large amount of transaction event data and case attributes that are useful for analysing lead times. However, finding root causes for long lead times is not so straightforward with current process mining methods. In this paper we extend our prevously presented influence analysis methodology by providing alternative treatment for continuous target variables like lead times and making it possible to give weights for each process case. We extend our contribution measure by presenting the definitions for binary/continuous as well as weighted/non-weighted needs. Using a publicly available real-life case study from Rabobank's service desk process we demonstrate the effect of using either continuous or binary approach combined with possible weighting.

Keywords: process analysis, process improvement, process mining, lead times, root cause analysis, data mining, influence analysis, contribution, working capital

1 Introduction

Every process owner and business leader in the world would be happy to hear that their own process or at least some part of it can be made faster. Reduced operational costs, better customer satisfaction and more sales are all potential benefits from reducing the process lead times. Since real-life business processes are often very complex and produce a lot of data, we need to consider many potential root causes for lead time related problems, including for example customer specific requirements, available resources, required competences for process workers, different business models, delivery options and products.

Our previously published influence analysis methodology shows how the root causes can be identified for generic process related problems [5]. The limitation of the already presented generic method is that it only supports binary classification where each case must be considered either success or failure. Analysing lead times is possible using binary classification by defining for example that every case taking more than 7 days is failure. However, in many business situations it is important to take into account the actual duration so that the longer the lead time is the bigger the problem. Another limitation of our previously published method is that it did not support case specific weights. In practice for some use cases like quality and internal auditing purposes it is acceptable to have equal weights for all cases since all cases should comply with regulations. For some other cases like sales order it might be much more important to deliver the large customer orders in time compared to delivering the small orders.

In this paper we will present a methodology to systematically analyse and provide actionable root causes for lead times issues in current business processes. We identify the root causes why some cases have very long lead times and others are very short. Our method analyses each case attribute and value separately.

The rest of this paper is organized as follows: Section 2 introduces relevant background in process mining and data analysis. Section 3 presents our extension to the influence analysis methodology introducing the contribution measures for continuous variables and case-specific weights. Section 4 shows a real-life example followed by a section for Discussions and Summary.

2 Related work

This paper extends the influence analysis methodology that we have published earlier [5]. We consider influence analysis as a practical actionable analysis which utilizes extensively the experiences and ideas from process mining [10], including specifically enriching and transforming process-based logs for the purpose of root cause analysis [9] and correlating business process characteristics [4]. Specifically the generic framework presented in [4] benefits from using the formulas and methods presented in this paper. For example considering the four additional use cases presented Table 5 in [4] the limitation of generic framework illustrations is that the presented decision tree analysis only tries to show the positive root causes for given process problem. Our methods as presented in this paper gives the results in a form of comparative benchmarking thus showing both the most influencial root causes for the "bad behavior" as well as most influencial root causes for avoiding the "bad behavior". Ability to show simultaneously the root causes for bad and good behavior makes it possible to quickly see whether the problem cases have a clear root cause or maybe the good behavior cases have a common root cause for their good behaviour. There has also been more work in detection of differences between groups [13] and finding contrast sets [1]. Our methodology is based on deviations management [6].

Even though business process performance has been studied a lot most of the studies only cover the usage of binary conditions or decision tree approach [8]. Wetzstein et al. have presented a framework for monitoring and analyzing influential factors of business process performance [14]. However their method requires the usage of binary contribution measure and in this paper we will present the option of using a continuous contribution formula. Grger et al. demonstrate very relevant data mining approaches for manufacturing process optimization [3] using binary and decision tree approach.

Basic idea of influence analysis [5] is to find root causes for deviations in the business process. Influence analysis has been used successfully for improving incoming invoice handling process [12]. Examples of root causes for long lead times found with influence analysis in these four case study companies include 'Contract number blank', 'Currency GBP', 'Business Unit X', 'Invoice Type EV' and 'Invoice status cancelled' [12]. Without data analysis tools it would be very difficult to find this kind of root causes. Our previously published influence analysis methodology consists of following steps:

1. Identify the relevant business process and define the case
2. Collect event and case attribute information
3. Create new categorization dimensions
4. Form a binary classification of cases such that each case is either problematic or successful
5. Select a corresponding interestingness measure based on the desired level of business process improvement effect
6. Find the best categorization rules and attributes
7. Present the results to business people

In this paper we extend the previous step 4. so that classification can be either binary as previously or we can use a continuous variable for representing the goodness or badness of a case. Regarding step 5. we only use the *as-is average* as the Change Type in this paper as that measure has proven to be most useful. However, it is also possible to use *ideal* and *other average* Change Types. Regarding step 6. we add new calculation formulas to cover also weighted versions of both binary and continuous contribution.

3 Analysis Types for Influence Analysis

In this section we present four different formulas that are to be used as contribution measures for influence analysis methodology. These measures are listed in Table 1. Depending on the performance indicator the contribution formula can be either binary or continuous. Depending on relative importance of cases the contribution formula can be weighted or not weighted. In typical business process analysis situations an actual business problem can often be formulated with any of these four formulas. Since the formulas give potentially different results it is important to understand that seemingly small differences in formulating the problem may lead to large differences in the analysis results. Thus it is often beneficial to use multiple contribution formulas for double-checking that suggested business process improvement areas are correct.

Our previous paper [5] presented the binary performance indicator with equal weights, corresponding to the contribution formula *Binary Contribution (BiCo)*. The contribution of this paper is to present three other formulas: *Continuous Contribution (CoCo)*, *Weighted Binary Contribution (wBiCo)* and *Weighted Continuous Contribution (wCoCo)*.

Table 1. Analysis Types

		Performance Indicator	
		Binary	Continuous
Weights for individual cases	Equal weights	Binary Contribution (BiCo)	Continuous Contribution (CoCo)
	Different weights	Weighted Binary Contribution (wBiCo)	Weighted Continuous Contribution (wCoCo)

Our method and calculations start from understanding the initial size of the business process problem. As presented in [5] one should focus development resources to improving issues where the size of the problem is large and the size of required investment is small. Problem size and an example lead time process for each Contribution Formula is shown in Table 2. When considering business process lead times we typically want to make the process generally faster (continuous variable) or then we want to ensure that the lead time of each instance is shorter than a given target (binary variable). Continuous is used when faster performance is always better and there is no lower bound. Binary approach is used for example when each process instance is categorized as successful if it meets a Service Level Agreement (SLA) and unsuccessful if it exceeds SLA. Following the power-law distributions in empirical data [2] principle we can use binary approach by selecting about 20% of worst performing cases to find explanations for bad performance.

Table 2. Problem size and example lead time for Analysis Types

Analysis Type	Problem size	Example lead time
BiCo	Amount of problematic cases	In service desk process a lead time longer than 7 days could be considered a problem case.
wBiCo	Sum of value of problematic cases	Free-of-charge pizza if delivery takes more than 45 minutes. Problem size is equal to the monetary value of late pizza deliveries.
CoCo	Sum of positive overtime compared to average lead time	Lead time from the customer calling a helpdesk to the moment the call is answered. The shorter the lead time the better it is.
wCoCo	Sum of overtime for each case compared to the weighted average lead time multiplied by the weight separately for each case	Lead time from the sending an invoice to the moment the payment arrives. When this lead time is multiplied by the value of the invoice we get the working capital, ie. using value of invoice as the weight for each case.

In this paper we consider an actionable business process improvement in area X, which is a subset of the whole population, as an improvement that will change the performance of future cases in area X to be improved so that the performance of area X reaches the the current as-is average performance of whole population. For example a US company may have delivery challenges in Dallas region and the improvement project would then be to improve the performance in Dallas to the same level as other regions. For each measure we show the formula that we call *Contribution%* which gives result as a percentage figure between *-100%* and *+100%*. Positive *Contribution%* indicates how large part of the current lead time problem can be improved by making the selected business area to perform on the same level as the initial average for the whole business. Negative *Contribution%* tells how much bigger the lead time problem will become if performance in selected business area is weakened to the current average level.

Weighting Weighted contributions can be used in contribution analysis. Simply we need Weight attribute for each case and we need to replace the 'amount of cases' values with 'sum of Weights of cases'. Now as an example we could have a total amount of 13 million EUR orders in the analysis and 2 million EUR orders are being delivered after the requested delivery date. So we will then run the contribution analysis to find the case attributes and values that have the biggest contribution in terms of EUR to the 2 million that is being delivered late. If there is one single order of 1.99 million EUR that was delivered late, then obviously the characteristics of that single order will overrule all other possible findings, even though if 100 other orders were delivered late. But that is definitely just the wanted finding because in real life if situation is like that then the one order (almost) fully explains the orders being late and there may be no point in trying to find more root causes for late deliveries.

3.1 Common Definitions

Here we present the common definitions used in all contribution formulas.

Definition 1. *Let $C = \{c_1, \ldots, c_N\}$ be a set of cases in the process analysis. Each case represents a single business process execution instance.*

Definition 2. *Let $C_p = \{c_{p_1}, \ldots, c_{p_N}\}$ be a set of problematic cases. $C_p \subseteq C$.*

Definition 3. *Let $C_a = \{c_{a_1}, \ldots, c_{a_N}\}$ be a set of cases belonging to business process improvement segment A. $C_a \subseteq C$.*

Definition 4. *Let d_{c_j} be the duration of the case c_j.*

Definition 5. *Let w_{c_j} be the weight of the case c_j. We consider linear weights so that double weight always means double importance. If $w_{c_j} = 0$ then case c_j will have no effect in the analysis when calculating weighted results.*

Definition 6. *Let pr be the size of the problem in the original situation before any business process improvement: BiCo: amount of problem cases, wBiCo: sum of weights of problem cases, CoCo: sum of overtime compared to average duration, wCoCo: sum of overtime per case multiplied with weight of the case compared to the weighted average duration.*

3.2 BiCo - Binary Contribution

For binary contribution the problem size is the amount of problematic cases. Every case needs to be classified as problematic or successful as shown in [5], ie. in order to analyse the process lead time one needs to specify a limit such that exceeding the limit classifies the case as problematic and otherwise it should be successful.. Definitions for *BiCo* have already been presented in [5]. However we have adopted a new syntax for the definitions in order to make it easier for the reader of this paper to compare binary/continuous and weighted/non-weighted to each other.

Total problem size for BiCo is the amount of problematic cases $pr_{BiCo} = \left|C_p\right| = \sum\limits_{c_j \in C_p} 1$ as shown in equation 1 in Table 8 in Appendix A. Average function for BiCo is the average problem density $rho = \frac{\left|C_p\right|}{|C|} = \frac{\sum_{c_j \in C_p} 1}{\sum_{c_j \in C} 1}$ as shown in equation 2. Similarly the average problem density for BiCo of subset C_a is $\rho_a = \frac{\left|C_p \cap C_a\right|}{|C_a|} = \frac{\sum_{c_j \in (C_p \cap C_a)} 1}{\sum_{c_j \in C_a} 1}$ as shown in equation 3. Finally the$Contribution\%$ for BiCo of subset C_a is $con_{BiCo} = \frac{(\rho_a - \rho) \sum_{c_j \in C_a} 1}{pr_{BiCo}} = \frac{\left|C_p \cap C_a\right|}{\left|C_p\right|} - \frac{|C_a|}{|C|} = \frac{\sum_{c_j \in (C_p \cap C_a)} 1}{\sum_{c_j \in C_p} 1} - \frac{\sum_{c_j \in C_a} 1}{\sum_{c_j \in C} 1}$ as shown in equation 4

3.3 wBiCo - Weighted Binary Contribution

Weighted Binary Contribution extends the previous sigma-based formulas by replacing the static equal weight with case specific weights w_{c_j}. Problem size as defined in equation 5 in Table 8 in Appendix A is the sum of weights of all problem cases. Average problem density as defined in equation 6 in Table 8 is the sum of weights of all problem cases divided by the sum of weights of all cases, and in a similar way the average problem density in equation 6 in Table 8 is the sum of weights of all problem cases in subset C_a divided by the sum of weights of all cases in subset C_a.

3.4 CoCo - Continuous Contribution

Continuous Contribution allows analysing the lead time variables as continuous without the need for a fixed separation of cases into long and short cases, ie.

without the need of having a binary value for each case. For continuous analysis we consider a case as problematic if the value of continuous target variable is bigger than the average in the population as shown in equation 9 in Table 8. The bigger the positive difference is the worse the behaviour. On the other hand, if the continuous target value is less than average then the case is a better-than-average. Using this approach the sum of positive deviations is always the same as the absolute value of the sum of negative deviations, meaning that the problem size for the whole population C is always zero. Problem size of any subset C_a may be nonzero meaning that the cases in subset C_a either have higher or smaller values for the target variable than the whole population. For analysing lead times the continuous target variable is any lead time variable of the business process cases, for example the total end-to-end lead time or any partial lead time from one activity to another. Average function for continuous analysis types is the average lead time, which is defined for the whole population with equation 10 and subset using equation 11 in Table 8.

Contribution measure for each possible subset C_a for CoCo is calculated as follows: subtract the average lead time of whole population C from the average lead time of the subset C_a, multiply this by the amount of cases in subset C_a. This gives an absolute value of how much more or less time is spent on the cases in subset C_a as a total compared to average of C. Final step is to divide this figure by the problem size, ie by the total sum of positive (or negative) cases in the population, giving the definition for equation 12 in Table 8.

3.5 wCoCo - Weighted Continuous Contribution

In this subsection we extend the previous defined continuous contribution formulas to supports case specific weights. It is good to note that weighted continuous contribution corresponds exactly to the working capital need in a business process. As an example lets consider the process of building houses where each case is one house. Working capital needed is proportional to the total cost of each house and the lead time from starting the constructions to selling the house. Weighted Continuous Contribution gives this measure when the cost of house is used as case specific weight and building time is used as the lead time. The business improvement activity for reducing working capital for this construction company then corresponds to conducting influence analysis using weighted continuous contribution analysis type to find our the those subsets that should be the focus for process improvements.

Average weighted lead time using case specific weights is calculated with equation 14 in Table 8. Difference to the non-weighted formula is that the lead time of each case is multiplied by the case specific weight and finally the result is divided by the total sum of weights. This weighted lead time is then used to calculate the total problem size according to equation 13 in Table 8 so that the absolute difference of lead time for each case is multiplied by the case specific weight and then summed up. In business terms this corresponds to calculating the extra working capital (positive) or unneeded working capital (negative) for each case and then summing them together. According to our approach if the lead time for every case is equally long then the problem size is zero and there is no extra working capital in the process.

Finally the contribution calculations for weighted continuous analysis are done similarly than in non-weighted analysis, ie subtract the weighted average duration of subset C_a from the total weighted average and multiply this by the sum of weights in subset C_a. When this is divided by the total problem size we get the amount of working capital that would be freed if the lead times for cases C_a could be reduced to the average weighted lead time in the whole population C as shown in equation 16 in Table 8.

3.6 Strengths and Weaknesses of Analysis Types

It is not trivial to decide which analysis type should be used in a particular business process analysis situation. Table 3 shows the strengths and weaknesses for binary and continuous analysis types and Table 4 respectively for weights. Often it is desirable to select one analysis type as the primary type for a particular analysis and then use the other analyses for reviewing, double-checking and confirming results from a perspective.

Table 3. Strengths and Weaknesses for Binary/Continuous Analysis Types

Type	Strengths	Weaknesses
Binary	– Can be applied to every lead time problem by separating cases into good cases and bad cases based on a selected cut-of threshold. – Can be controlled by setting the cut-of threshold for duration. – Manages outliers very well because every case is just considered good or bad and the amount of extra lead time is not considered at all.	– Requires decision for the cut-of threshold. If customer of the process would like to get the result in 10 days and average duration currently is 6 days, should we consider all cases taking more than 10 days as bad, or should the cut-of threshold be 9 days in order to improve cases that are close to be missed, or should be threshold be 20 days to allow identification of areas that have severe problems. Influence analysis gives potentially different results with every cut-of threshold.
Continuous	– Does not need any separate cut-of threshold. Continuous variable like lead time is used directly by the algorithm and overtime is calculated from the average duration.	– Is sensitive for outliers. If one case takes million times longer than the other cases then the whole analysis is likely to suggest improvement in all the subsets C_a containing that particular case.

Table 4. Strengths and Weaknesses for using weights

Type	Strengths	Weaknesses
Equal weights	– No need to define and calculate weights. – Every case simply has equal weight. – Not sensitive to outliers regarding weights.	– In real life it is often a bigger problem to loose a large customer, fail Service Level Agreement of important customer request or have quality issues with expensive products. Using equal weights does not take importance into account.
Different weights	– Using the sales value, profit, importance or similar as weights makes results more aligned with business value and importance.	– It is not easy or straightforward to define the weight for each case. In real data some cases may have small or even zero value but they could be part of a bigger project and as such very important for a major customer. – Using weights makes the analysis sensitive to outliers.

4 Case Study: Rabobank Group ICT

In this section we show a real life example of using the presented analysis types with a publicly available data from Rabobank Group ICT from BPI Challenge 2014 [11]. The data contained 46 616 cases and a total of 466 737 events. As a lead time we consider the total duration of each case including all cases found in the dataset. Typical process mining analysis discovers that the average duration for cases is 5.07 days and median duration is 18 hours. For the purpose of comparing the analysis we set the threshold of problematic cases in the binary analyses to be 7 days which results in a total of 7 400 (15.9%) problematic cases.

As the weighting for cases we use a formula $w_{c_j} = (6 - Impact_{c_j})(6 - Urgency_{c_j})(6 - Priority_{c_j})$ where Impact, Urgency and Priority all have values in (1,2,3,4,5) where 1 means highest importance and 6 is the lowest importance. With this formula the highest possible weight is $w_{high} = (6-1)(6-1)(6-1) = 125$ and lowest possible weight is $w_{low} = (6-5)(6-5)(6-5) = 1$. Using these weights the weighted average lead time drops to 3.97 days. This means that on average the lead time is shorter for more important cases than for less important cases. Same finding can also be made from the binary results since the average problem density is 15.9% and weighted average problem density is 11.1%.

Top-3 positive and negative contributions for each Analysis Type are shown in Table 5. For BiCo the highest contribution% is 4.2% for case attribute value

Table 5. Comparison of top-three root causes based on different values for case attribute *ServiceComp WBS(CBy)* for for all analysis types

ServiceComp WBS(CBy)	BiCo	wBiCo	CoCo	wCoCo
WBS000091	+4,2% (+1)	+3,4% (+2)	+1,7% (+8)	+1,2% (+7)
WBS000072	+2,8% (+2)	+0,8% (+10)	+3,4% (+4)	+0,6% (+12)
WBS000088	+2,4% (+3)	+3,7% (+1)	+10,4% (+1)	+12,7% (+1)
WBS000162	+2,2% (+4)	+2,9% (+3)	+8,6% (+2)	+10,0% (+2)
WBS000055	+0,7% (+9)	+1,2% (+7)	+3,5% (+3)	+4,2% (+4)
WBS000043	+0,2% (+21)	+1,2% (+8)	+1,3% (+10)	+4,8% (+3)
.....				
WBS000228	-1,0% (-4)	+0,1% (+50)	-1,7% (-3)	-0,2% (-39)
WBS000146	-0,5% (-9)	-2,2% (-3)	-0,7% (-11)	-2,8% (-4)
WBS000095	-1,7% (-3)	-0,6% (-8)	-2,5% (-2)	-0,9% (-8)
#N/B	-1,7% (-2)	-8,3% (-1)	+2,6% (+5)	-5,9% (-2)
WBS000073	-9,3% (-1)	-4,1% (-2)	-17,4% (-1)	-10,9% (-1)

WBS000091 and lowest contribution% is -9.3% for case attribute value *WBS000091*. When considering the most beneficial focus are for process improvement reducing the lead time most we see that BiCo results in *WBS000091* and all other Contribution Formulas result in *WBS000088*. According to the figures the best performing area regarding lead time is *WBS000073* in all Contibution Formulas except that in Weighted Binary Contribution the best practice area is *#N/B*.

Some interesting results include the behaviour of cases whose attribute *ServiceComp WBS(CBy)* has the value *WBS000091* which contributes to 4.2% (Top 1) of the total problem in BiCo, 3.4% (Top 2) in wBiCo, but only 1.7% (Top 8) in CoCo and only 1.2% (Top 7) for wCoCo. Reason for higher contribution in BiCo and lower in CoCo is that average lead time for area *WBS000091* is only 6.14 which is only little longer than the average for the whole process 5.07. This means that there are many *WBS000091* -cases that have lead time a little bit longer than 7 days. On the other hand the behaviour of area *WBS000088* is the opposite, since it only contributes 2.4% (Top 3 value) of the total problem in BiCo, 3.7% (Top 1) in wBiCo, much more 10.4% (Top 1) in CoCo and even more 12.7% (Top 7) in wCoCo. Reason for this behaviour is that the average lead time for cases in area *WBS000088* is 39.2 days which is much longer than the average lead time 5.07

Very interesting results include the behaviour of area *#N/B* which is listed as best practice area with negative contribution -1.7% in BiCo (Top -2), -8.3% in wBiCo (Top -1) and -5.9% in wCoCo (Top -2). However it is listed as problem area with positive contribution 2.6% in CoCo (Top 5 Problem Area!). There are at least two reasons for this result: first the high weight cases in *#N/B* perform much better than the low weight cases, ie. BiCo contribution gets 6.6 percentage points better with weighting than without and CoCo contribution gets 8.5 percentage points better. Second reason is that area *#N/B* performs consistently worse in Continuous analysis compared to binary analysis, which is caused by the higher than average lead time of 6.2% in CoCo, which again is caused by certain amount of very long lead time cases.

Activity occurrences. In this subsection we show the Rabobank root cause analysis for long lasting cases using activity occurrence data. As a preprosessing step we add a new case attribute for each different activity name and use the amount of activity occurrences as the value for that case attribute in each case. For example if activity *Status Change* occurs twice for a certain case, then the value of case attribute Status Change will 2 for that particular case.

Table 6 shows the top-5 positive and negative root causes for binary analysis types and 7 for the continuous analysis types. From these tables we make following observations:

- Lack of reassignments is the most important negative root cause for a case to exceed 7 day SLA (BiCo analyses) or generally take a long time (CoCo analysis). In other words, having zero reassignments makes a case very fast.
- Contribution values for activity occurrence amounts are much higher than they are for the case attribute *ServiceComp WBS(CBy)*, which means that these activity amounts correlate more with the total duration than the case attribute *ServiceComp WBS(CBy)*.
- *Update from customer(1)* is most important positive root cause for long case duration as can be seen in continuous contributions in table 7. However for binary contributions in table 6 the *Status Change(2)* is more important positive root cause which means that having two occurences of *Status Change* is a bigger risk for failing SLA than getting an update from customer.

Table 6. Comparison of top-five root causes based on amount of occurrences of activities for binary analysis types

BiCo	wBiCo
Closed(2) 9,7%	Status Change(2) 10,3%
Status Change(2) 8,9%	Communication with customer(1) 10,1%
Communication with customer(1) 8,8%	Closed(2) 9,9%
Reopen(1) 7,8%	Update from customer(1) 9,5%
Update from customer(1) 7,1%	Assignment(3) 9,3%
...	...
Status Change(0) -20,8%	Status Change(0) -22,0%
Assignment(1) -26,4%	Assignment(1) -26,8%
Update(0) -29,6%	Update(0) -32,0%
Operator Update(0) -35,8%	Operator Update(0) -41,5%
Reassignment(0) -37,9%	Reassignment(0) -43,2%

Table 7. Comparison of top-five root causes based on amount of occurrences of activities for continuous analysis types

CoCo	wCoCo
Update from customer(1) 14,9%	Update from customer(1) 15,6%
Closed(2) 13,1%	Status Change(2) 12,8%
Status Change(2) 12,2%	Update from customer(2) 12,5%
Reopen(1) 11,8%	Update(2) 12,1%
Description Update(1) 11,1%	Description Update(1) 12,0%
...	...
Update from customer(0) -38,6%	Assignment(1) -43,0%
Assignment(1) -45,8%	Update from customer(0) -46,3%
Update(0) -51,3%	Update(0) -50,1%
Operator Update(0) -53,2%	Operator Update(0) -55,8%
Reassignment(0) -62,0%	Reassignment(0) -65,0%

5 Summary and Conclusions

In this paper we have presented a method for focusing business process improvement to reduce lead times. We have defined four different analysis types and shown how they can be used with actual data. Summary of our key findings is:

1. Influence analysis methodology is able to find root causes for long lead times.
2. Root causes for long lead times may be substantially different when using a predefined lead time limit for problematic/successful cases (binary) compared to when using continuous lead time values.
3. Case specific weighting can be easily used when analysing both binary and continuous contribution.
4. When weighting is used together with continuous contribution the analysis can be directly used as working capital analysis solution.

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.

A Appendix - Summary of Contribution Formulas

Table 8. Problem size, average function and contribution% definitions

Analysis Type	Total Problem size	Average function	Average function for subset C_a	*Contribution%*
BiCo	$pr_{BiCo} = \sum_{c_j \in C_p} 1 \quad (1)$	Average Problem density: $\rho = \frac{\sum_{c_j \in C_p} 1}{\sum_{c_j \in C} 1} \quad (2)$	$\rho_a = \frac{\sum_{c_j \in (C_p \cap C_a)} 1}{\sum_{c_j \in C_a} 1} \quad (3)$	$\frac{(\rho_a - \rho) \sum_{c_j \in C_a} 1}{pr_{BiCo}} \quad (4)$
wBiCo	$pr_{wBiCo} = \sum_{c_j \in C_p} w_{c_j} \quad (5)$	Weighted Average Problem density: $\rho_w = \frac{\sum_{c_j \in C_p} w_{c_j}}{\sum_{c_j \in C} w_{c_j}} \quad (6)$	$\rho_{w_a} = \frac{\sum_{c_j \in (C_p \cap C_a)} w_{c_j}}{\sum_{c_j \in C_a} w_{c_j}} \quad (7)$	$\frac{(\rho_{w_a} - \rho_w) \sum_{c_j \in C_a} w_{c_j}}{pr_{wBiCo}} \quad (8)$
CoCo	$pr_{CoCo} = \frac{1}{2} \sum_{c_j \in C} \lvert d_{c_j} - \bar{d} \rvert \quad (9)$	Average lead time: $\bar{d} = \frac{\sum_{c_j \in C} d_{c_j}}{\sum_{c_j \in C} 1} \quad (10)$	$\bar{d}_a = \frac{\sum_{c_j \in C_a} d_{c_j}}{\sum_{c_j \in C_a} 1} \quad (11)$	$\frac{(\bar{d}_a - \bar{d}) \sum_{c_j \in C_a} 1}{pr_{CoCo}} \quad (12)$
wCoCo	$pr_{wCoCo} = \frac{1}{2} \sum_{c_j \in C} w_{c_j} \lvert d_{c_j} - \bar{d}_w \rvert \quad (13)$	Weighted Average lead time: $\bar{d}_w = \frac{\sum_{c_j \in C} w_{c_j}\ d_{c_j}}{\sum_{c_j \in C} w_{c_j}} \quad (14)$	$\bar{d}_{w_a} = \frac{\sum_{c_j \in C_a} w_{c_j} d_{c_j}}{\sum_{c_j \in C_a} w_{c_j}} \quad (15)$	$\frac{(\bar{d}_{w_a} - \bar{d}_w) \sum_{c_j \in C_a} w_{c_j}}{pr_{wCoCo}} \quad (16)$

References

1. Bay, S. and Pazzani, M. "Detecting group differences: Mining contrast sets". Data Mining and Knowledge Discovery 5 (3): 213246, 2001.
2. Clauset, Aaron, Cosma Rohilla Shalizi, and Mark EJ Newman. "Power-law distributions in empirical data." SIAM review 51.4 (2009): 661-703.
3. Grger, Christoph, Florian Niedermann, and Bernhard Mitschang. "Data mining-driven manufacturing process optimization." Proceedings of the world congress on engineering. Vol. 3. 2012.
4. De Leoni, Massimiliano, Wil MP van der Aalst, and Marcus Dees. "A general process mining framework for correlating, predicting and clustering dynamic behavior based on event logs." Information Systems (2015), http://dx.doi.org/10.1016/j.is.2015.07.003
5. Lehto, T., Hinkka, M. and Hollmén, J. "Focusing Business Improvements Using Process Mining Based Influence Analysis." International Conference on Business Process Management. Springer International Publishing, 2016.
6. Piatetsky-Shapiro, Gregory, and Christopher J. Matheus. "The interestingness of deviations." Proceedings of the AAAI-94 workshop on Knowledge Discovery in Databases. Vol. 1. 1994.
7. QPR Software Plc. "Improved invoicing for Caverion through process insight." IEEE CIS Task Force on Process Mining, Case Studies, 2014.
8. Rozinat, Anne, and Wil MP van der Aalst. Decision mining in ProM. Springer Berlin Hei-delberg, 2006.
9. Suriadi, Suriadi, Ouyang, Chun, van der Aalst, Wil M.P., & ter Hofstede, Arthur (2013) Root cause analysis with enriched process logs. Lecture Notes in Business Information Processing [Business Process Management Workshops: BPM 2012 International Work-shops Revised Papers], 132, pp. 174-186.
10. Van Der Aalst, Wil, et al. "Process mining manifesto." Business process management workshops. Springer Berlin Heidelberg, 2012.
11. Van Dongen, B.F. BPI Challenge 2014. Rabobank Nederland. Dataset. http://dx.doi.org/10.4121/uuid:c3e5d162-0cfd-4bb0-bd82-af5268819c35, 2014
12. Vanjoki, V. "Automated Purchase to Pay Process Value Modeling and Comparative Process Speeds." Lappeenranta University of Technology, 2013.
13. Webb, GI and S. Butler and D. Newlands (2003). "On Detecting Differences Between Groups". KDD'03 Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.
14. Wetzstein, Branimir, et al. "Monitoring and analyzing influential factors of business process performance." Enterprise Distributed Object Computing Conference, 2009. EDOC'09. IEEE International. IEEE, 2009.