Linking and Analyzing Rabobank Group ICT's ITIL Subprocesses based on the Synergetic Capabilities of Process Mining and Data Mining

The BPI Challenge 2014

Seung Won Hong, Ji Yun Hwang, Dan Bi Kim, Hyeoung Seok Choi, Seo Jin Choi and Suk Hyun Hong

APPS(Advanced Process Performance reSearch center), Department of MIS, School of Business Administration, Myongji Univ., Seoul, Korea
{extensive.h, jiyoon930527, aw30304, choihs91}@gmail.com, csj821@naver.com, reginaxoxo@nate.com

Abstract. An increasing number of organizations have been analyzing and improving their ITIL (information technology infrastructure library) process using advanced process mining techniques. Typically, ITIL process consists of several subprocesses such as Interaction → Incident → Change management. Although we admit that it is helpful to analyze its subprocesses separately, we strongly believe that understanding and analyzing it in its entirety can provide more insights for improvement action and serve as a stepping stone for addressing the given three questions regarding the effects of Change implementations. Therefore, before answering the three questions, we analyzed the ITIL process in its entirety and found several interesting results (our creativity challenge). To preprocess and analyze more than a million of events, we used a variety of data analysis techniques and powerful tools such as Disco, Weka, Oracle DBMS. We offer evidence-based answers to the questions and demonstrate the potential benefits of process mining and other data analytics-based understanding and analysis of ITIL process. Finally, concluding remarks and recommendations for improvement are discussed.

1. Introduction

Since the BPR (business process reengineering) movement in 1990s [2], an increasing number of organizations have been making sizable investments in IT systems to achieve radical performance improvements through the automation of and IT support for business processes [3,4]. The movement has contributed to organizations' performance improvements in part. However, in recent years, it has resulted in very complex business processes that are difficult to control and manage. Therefore, it is one of the principal challenges of our day to understand, analyze, and improve the complex business processes [6].

How can we deal with this challenge? Fortunately, a lot of IT systems are accumulating invaluable (big) process data, which often records in detail which activities were executed when and by whom [7,8]. For example, SAP Process Observer records this process data, which is used for process monitoring, process mining, and process analytics*. If we put the data to good use by using process mining and other data analytics, managers can understand and analyze the complex business processes and gain insights for controlling and improving them.

According to the situation depicted in the BPI Challenge 2014, Rabobank Group ICT also has to deal with the similar challenge. The situation focuses on the ITIL process consisting of three subprocesses: Interaction → Incident → Change management subprocesses. The subprocesses are supported by ITIL Service Management tool called HP Service Manager. Four datasets for the three subprocesses were provided for this BPI Challenge. We addressed Question 4 (Creativity Challenge) at the very beginning: by linking the three subprocesses as an end-to-end ITIL process, we can compare it with a reference ITIL process and find insights that cannot be gained from analyzing the subprocesses separately. After dealing with Question 4, we attempted to address the given three questions by data analytics such as data mining. To achieve this goal, we sought to understand the data and create relevant datasets for the questions. Specifically, to offer evidence-based answers for the questions, we attempted to address the following questions in detail:

- Question 0: Analyzing an ITIL End-to-end Process
- Question 1: Identification of Impact-Patterns
- Question 2: Addressing Parameters for Every Impact-Patterns
- Question 3: Addressing Change in Average Steps to Resolution

Since we received the datasets without extracting them and could not get direct feedback on analysis results from Rabobank's stakeholders, we could not perform all the important activities proposed by Heijden [9]'s methodology. However, in order to prevent redundant work, we have tried to follow its six phases and perform key activities that are suggested during each phase. Therefore, we believe that our answers and analysis results are plausible and thus validate the synergetic capabilities of process mining and other data analytics such as data mining. We hope that our analysis results provide guidance on how to control and improve organizations' complex business processes in a big data world.

The rest of the paper is structured as follows. The next section shows which tools we used and our understanding of the data. We briefly introduce the methodology we adopted in section 3. In section 4, we explain our analysis approaches and give evidence-based answers for the given questions. The next section provides the concluding remarks and we provide recommendations for improvement based on the analysis results in the final section.

* http://scn.sap.com/docs/DOC-24983

2. Tools and Understanding of the Data

2.1 Used Tools

There are several tools used in this analysis: process mining toolkit (Disco), Data mining toolkit (Weka 3.6.11), a database management system (Oracle database 11g), Statistics toolkit (R 3.1.1) and a spreadsheet application (Microsoft Excel 2010).

2.1.1 Disco

To understand and analyze the three processes within Rabobank ICT; Interaction Management, Incident Management, and Change Management, Disco, the process mining toolkit, was used. With filtering, one of its key functions that helps users to clean up their process data and to focus their analysis, we found out not only basic statistics and process maps for each process, but also the exact measures and the processes that we wanted to identify

2.1.2 Oracle Database 11g

Using Disco as mentioned above, it was necessary for us to extract the data at more intensive levels such as the link between tables as well. Therefore, SQL with Oracle Database System was used

2.1.3 Java

To make segmented and intensive statistical data, the database including new data columns was created, with Oracle Database System.

2.1.4 R3.1.1

With Oracle Database System, R 3.1.1 was used to analyze more segmented and intensive statistics. From this tool, statistics and graphs were made.

2.1.5 Weka 3.6.11

To predict the future beforehand, it was necessary to find out patterns of the existing activities and defined them as a model. Of many data mining tools, Weka, free data mining tool, was chosen. Compared to charged softwares such as SPSS Clementine and Enterprise Miner, it is less convenient and visualized, however, it can produce the same results of the data mining analysis. Therefore, Weka 3.6.11 was used.

2.1.6 Microsoft Excel

Microsoft Excel was used to analyze and organize the given data, and also visualize them with graphs or charts.

2.2 Understanding of the Data

As mentioned before, four datasets for the three subprocesses were provided to us. To understand the datasets, we preprocessed and imported them into Oracle DBMS. After analyzing them, we found out that there are 147004, 46606, 466737, and 30275 events in the 'Detailed Interaction', 'Detailed Incident', 'Detailed Incident Activity', and 'Detailed Change' datasets respectively. Summary information about each dataset's attributes is provided in <Fig. 1-4.>. Based on this understanding of the data, we have created relevant datasets for address the questions. In particular, to understand and analyze the ITIL process in its entirety, we have made a lot of effort to link these four datasets.

Detailed Interaction		
Interaction ID	147,004	Case
CI Name (aff)	4,153	Case
CI Type (aff)	14	Case
CI SubType (aff)	67	Case
Service Comp WBS (aff)	289	Case
Status	2	Case
Impact	5	Case
Urgency	5	Case
Priority	5	Case
Category	6	Case
KM number	2,360	Case
Closure Code	24 (null)	Case
First Call Resolution	2	Case
Open Time (First Touch)	2011-09-09 9:23	Min
Close Time	2014-03-31 22:47	Max
Handle Time (secs)	22,530 / 0	Max / Min
Related Incident	46,088 (null)	Case

Fig. 1. Attributes of 'Detailed Interaction' Dataset

Detailed Incident Activity		
Incident ID	466,616	Case
Interaction ID	46,444	Case
IncidentActivity Number	466,737	Case
IncidentActivity Type	39	Case
Assignment Group	242	Case
Km number	1,825	Case
DateStamp	2014-01-01 12:31	Max

Fig. 2. Attributes of 'Detailed Incident Activity' dataset

Detailed Incident		
Incident ID	46,606	Case
CI Name (aff)	3,019	Case
CI Name (Cby)	3,652	Case
CI Type (aff)	13	Case
CI Type (Cby)	14	Case
CI Subtype (aff)	65	Case
CI Subtype (CBy)	63	Case
Service Comp WBS (aff)	274	Case
Service Comp WBS (Cby)	275	Case
Status	5	Case
Impact	5	Case
Urgency	5	Case
Priority	5	Case
Category	4	Case
KM number	1,825	Case
Alert Status	1	Case
Open Time	2012-02-05 13:32	Min
Reopen Time	2013-10-01 06:45:36 (null)	Min
Resolved Time	2013-10-01 06:45:36 (null)	Min
Close Time	2014:03:31 22:47	Max
Handle Time (Hours)	9.999.722.222	Max
Closure Code	15 (null)	Case
Related Interaction	43,060	Case
Related Change	233 (null)	Max
# Reassignments	46 / 0 (null)	Max
# Related Interactions	370 / 1 (null)	Max / Min
# Related Incidents	63 / 1 (null)	Max / Min
# Related Changes	9 / 1 (null)	Max / Min

Fig. 3. Attributes of 'Detailed Incident' dataset

Detailed Change		
Change ID	18,000	Case
CI Name (aff)	10,193	Case
CI Type (aff)	13	Case
CI Subtype (aff)	74	Case
Service Comp WBS (aff)	286	Case
Change Type	240	Case
Risk Assessment	3	Case
Emergency Change	2	Case
CAB-approval needed	2	Case
Originated from	3	Case
Planned Start	2011-06-01 7:00	Min
Planned End	2021-02-20 17:30 (null)	Max
Scheduled Downtime Start	2015-03-21 01:00 (null)	Min
Scheduled Downtime End	2012-10-16 13:09 (null)	Max
Actual Start	2012-10-16 13:09 (null)	Min
Actual End	2021-03-21 00:01 (null)	Max
Requested End Date	2028-02-20 17:00	Max
Change record Open Time	2011-09-01 09:13	Min
Change record Close Time	2014-03-31 23:53	Max
# Related Interactions	6 / 5 (null)	Max / Min
# Related Incidents	279 / 1 (null)	Max / Min

Fig. 4. Attributes of 'Detailed Change' dataset

3. Method

A number of methodologies have been developed to help perform process mining projects [e.g., 1, 5, 7, 9]. Among them, van der Heijden [9]'s PMPL(process mining project methodology) is one of comprehensive methodologies. Furthermore, the methodology was developed using System Engineering Process and validated by a real case study. Therefore, we adopted it.

It is composed of six main phases: scoping, data understanding, event log creation, process mining, evaluation, and deployment. Since we received the datasets without extracting them and could not get direct feedback on analysis results from

Rabobank's stakeholders, we could not perform all the import activities proposed in the methodology. However, in order to prevent redundant work, we have tried to follow its six phases and perform key activities that are suggested during each phase.

4. Analysis from of the Questions

4.1 Question 0: Analyzing ITIL End-to-End Process

It is about the given Creativity Challenge from Question 4 which deals with linking the given four tables and creating End-to-end process. However, because we needed to describe it first, it is set as Question 0.

The value of organization is created by an interconnected process. It other words, we can find a new insight by analyzing end-to-end process which links its subprocesses. The ITIL process of Rabobank consists of the three subprocesses (Interaction Management, Incident Management and Change Management). We were provided with the four data sets which recorded the executions for each process. To compare with the reference process, we sought to link the three subprocesses and analyze the linked End-to-end process in this chapter.

4.1.1 Linking the Three Sub Processes

To analyze the End-to-end process, first of all, it is necessary to link the three subprocesses. Incident activity dataset was used between the two datasets(Incident, Incident activity) of incident management process. It was not easy to link the three subprocesses together.

4.1.1.1 Removing Incomplete Cases

As shown below, we removed incomplete cases from each subprocess of The ITIL process of Rabobank.

- Interaction : Removing cases whose status are 'Open-Linked'
- Incident Activity: Removing cases whose incident activity type doesn't include 'Closed'.
- Change: Removing cases whose either actual start time or actual end time has null value.

4.1.1.2 Selecting Attributes Required to Link the Three Sub Processes

To link the three sub processes, we used the value combined 'CI Name (aff)' with 'Related Incident' of interaction dataset and the value combined 'CI Name (aff)' with 'Incident ID' of incident dataset. Also, we used the value combined 'CI Name (aff)' with 'Change ID' of change dataset. (refer to <Fig. 5>)

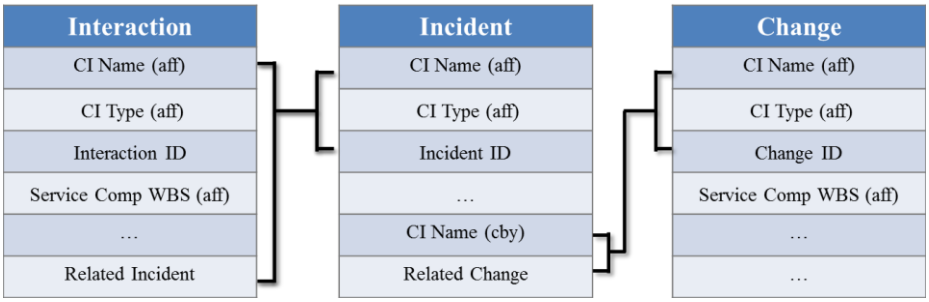


Fig. 5. Data Attributes required to link the three sub processes

4.1.1.3 An Issue of Defining Process Instance of the Linked ITIL Process

<Fig. 6> shows an example of the linked ITIL process instance. Each circle, square, and triangle indicates interaction subprocess, incident subprocess, and change subprocess respectively. As shown in <Fig. 6>, one incident can be related to several interactions. Also, one incident can be related to several changes. In these cases, a problem can occur when define the instance that links the three subprocesses. To solve this problem, as shown in <Fig. 7>, we added the events of incident subprocess and change subprocess as many as the number of the instances of Interaction subprocess. It is defined that the statistics such as frequencies or durations related to incident and change subprocesses are inaccurate when process maps are created with the pre-processed event log.

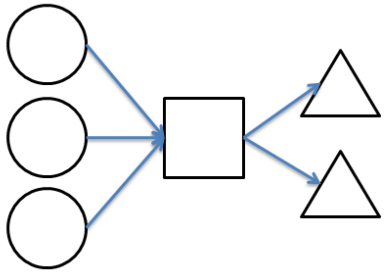


Fig. 6. An Issue of defining process instance of the linked ITIL process

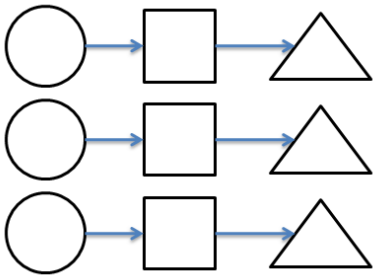


Fig. 7. Solution to the defining process instance issue.

4.1.1.4 Setting Case ID

First of all, we set the case id of each subprocess with the value used when we link subprocesses. But if we analyze the process in this condition, each part(interaction, incident, change) of a process instance of linked process will be recognized as different instances(refer to <Fig. 8>). To resolve this problem, we need to select a case id among a couple of possible values. In this reason, we selected a case id by the possible ways of execution of the ITIL process.(refer to <Fig. 9>)

Interaction	Incident	Change
SD0000002-HMD000002	IM0000006-HMD000002	C00000003-HMD000002

Fig. 8. Many different Case IDs that the linked ITIL process instances can have

Number of Scenarios	Case ID
Interaction, Incident, Change process has occurred	Interaction ID - CI NAME
Interaction, Incident process has occurred	Interaction ID - CI NAME
Incident, Change process has occurred	Incident ID - CI NAME
Only Interaction process has occurred	Interaction ID - CI NAME
Only Incident process has occurred	Incident ID - CI NAME
Only Change process has occurred	Change ID - CI NAME

Fig. 9. Selecting Case ID by the possible ways of execution of the linked ITIL process

4.1.1.5 Converting Data Format for Process Mining

After looking through the given datasets, it was necessary to convert interaction and incident datasets into right format for process mining. <Fig. 10> and <Fig 11> show before and after conversions of interaction and change datasets. Incident Activity dataset was properly recorded for process mining, however, CI Name(aff) was not recorded. Hence, we referred to the CI Name(aff) per incident id of incident dataset.

CI Name (aff)	Interaction ID	Open Time (First Touch)	Close Time
SBA000243	SD0000001	09-09-2011 09:23	14-02-2014 09:05
SUB000443	SD0000002	29-09-2011 14:59	13-12-2013 16:27

Case ID	DateStamp	Activity
SD0000001-SBA000243	09-09-2011 09:23	Open
SD0000001-SBA000243	14-02-2014 09:05	Close
SD0000002-SBA000443	29-09-2011 14:59	Open
SD0000002-SBA000443	13-12-2013 16:27	Close

Fig. 10. Before(above) and after(below) conversion of Interaction dataset.

CI Name	DateStamp	Actual Start	Actual End	Change record Open Time	Change record Close Time
HMD000002	C0000003	18-12-2013 14:00	18-12-2013 16:15	01-09-2011 09:13	18-12-2013 16:16

Case ID	DateStamp	Activity
C00000003-HMD000002	01-09-2011 09:13	Change record Open
C00000003-HMD000002	18-12-2013 14:00	Actual Start
C00000003-HMD000002	18-12-2013 16:15	Actual End
C00000003-HMD000002	18-12-2013 16:16	Change record Close

Fig. 11. Before(above) and after(below) conversion of incident dataset.

4.1.2 Analyzing the Linked ITIL Process

Through 4.1, we created the process linked three subprocesses. Also, by analyzing the linked process, we compared the ITIL reference process of Rabobank with the real process and discovered a meaningful insight.

4.1.2.1 The Linked Process Model

<Fig. 12> shows the entire ITIL process. To make this process model simple, we got rid of 29 incident activities whose frequencies are less than 1% among 45 activities. Also, from Disco's slider function, we only included the paths whose frequencies are 50%.

As mentioned in 4.1.3, to link the three subprocesses, we added the events of incident and change subprocesses as many as we needed. Therefore, because the statistics such as frequencies and durations related to the activities and two subprocesses can have inaccurate value, we took no count of these statistics.

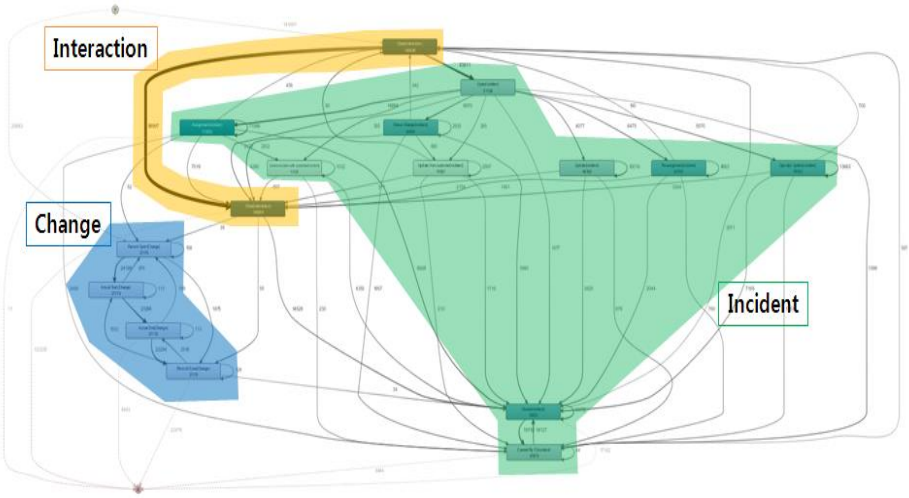


Fig. 12. The Real ITIL process of Rabobank (100% activities, 50% paths).

4.1.2.2 Finding of Unexpected Process Flow

According to the ITIL reference process of Rabobank (refer to <Fig. 13>), similar incidents which reoccur more often than usual have to go through problem analysis and be connected to change subprocess. In other words, 'Change Open Activity' of all the process instances is needed to be executed after doing 'Incident Closed Activity', because change process has to be started after all of similar incidents are closed. However, the actual process map has unexpected paths. We decided to look into the unexpected flow between incident and change subprocesses.

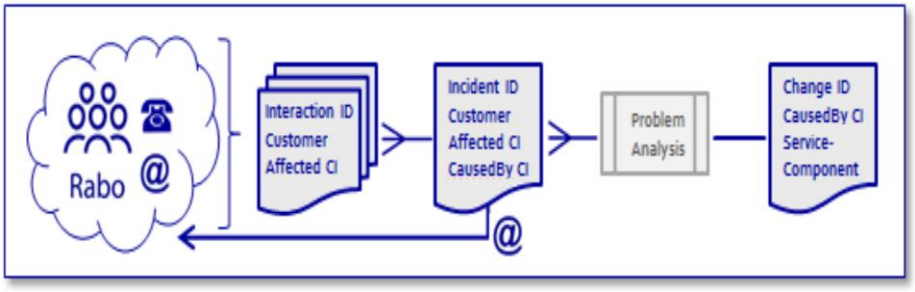


Fig. 13. The ITIL Reference process of Rabobank.

<Fig. 12> shows the path from 'Assignment (Incident)' to 'Record Open (Change)'. To get more information, with Disco's filtering function, we selected cases whose change record are open right after incident is assigned (refer to <Fig. 14>). Also, we looked through process variants to figure out characteristics of the selected cases (refer to <Fig. 15>).

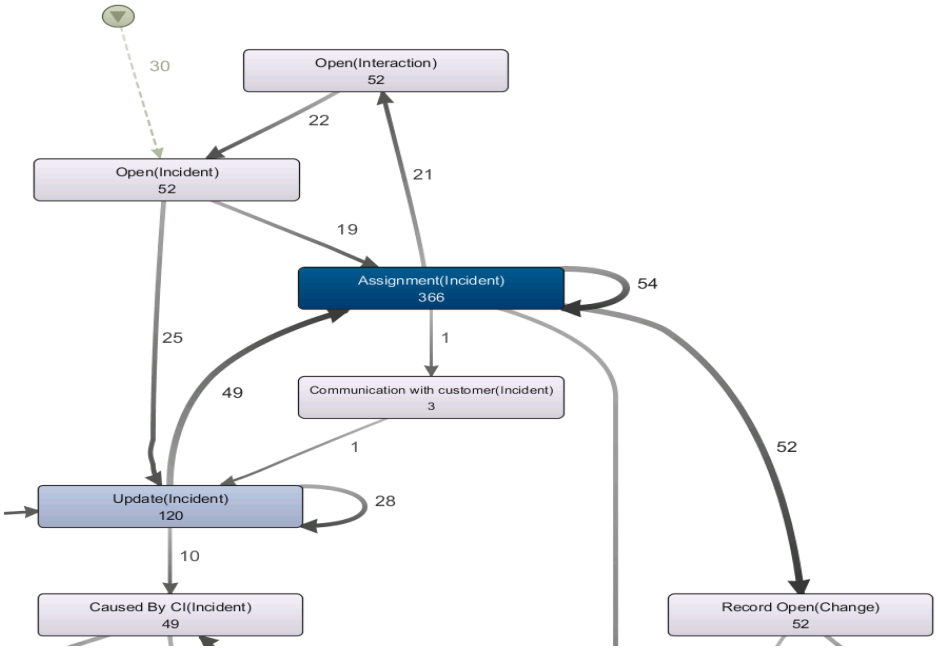


Fig. 14. Path from 'Assignment (Incident)' to 'Record Open (Change)'.

<Fig. 15> shows the variant(14 cases among 52 cases) which occurred most frequently among all the variants of cases. On a closer view of how the activity of these variants is executed, 'Assignment' and 'Operator Update' are executed repeatedly in incident process. Then, incident and interaction subprocesses are closed after change process is closed. These cases had to be resolved by change management subprocess, because interaction or incident management subprocess couldn't solve them. We can make a conclusion that the reference process model which performs similar incidents repeatedly, makes them go through problem analysis, and implements change process has a problem if this fact is found from the interviews with the person in charge as well as from data. After all, we are asking for improving the reference process model, based on the conclusion.

Fig. 15. An example of cases whose change record is open right after incident is assigned.

<Fig. 16> and <Fig. 17> show the process model of cases whose change record is open right after updating its operator and the most frequent variant. We figured out that these cases also executed 'Assignment' and 'Operator Update' repeatedly and incident is closed after change subprocess is closed.

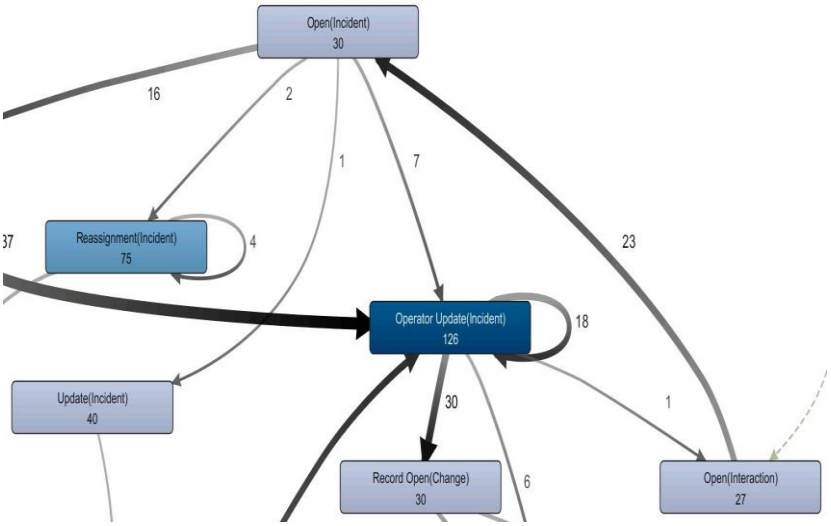


Fig. 16. Path from 'Operator Update (Incident)' to 'Record Open (Change)'

	Activity	Date	Time
1	Open(Incident)	06.10.2013	21:48:59
2	Operator Update(Incident)	06.10.2013	22:03:23
3	Assignment(Incident)	07.10.2013	10:19:28
4	Update(Incident)	07.10.2013	10:45:32
5	Update(Incident)	07.10.2013	13:21:54
6	Assignment(Incident)	07.10.2013	13:21:54
7	Reassignment(Incident)	07.10.2013	13:21:54
8	Operator Update(Incident)	07.10.2013	13:50:25
9	Reassignment(Incident)	07.10.2013	13:50:25
10	Assignment(Incident)	07.10.2013	14:45:42
11	Update(Incident)	07.10.2013	15:00:21
12	Update(Incident)	08.10.2013	09:32:15
13	Assignment(Incident)	08.10.2013	09:32:15
14	Reassignment(Incident)	08.10.2013	09:32:15
15	Assignment(Incident)	08.10.2013	10:58:59
16	Operator Update(Incident)	08.10.2013	15:22:35
17	Assignment(Incident)	08.10.2013	15:22:35
18	Reassignment(Incident)	08.10.2013	15:22:35
19	Status Change(Incident)	08.10.2013	15:41:46
20	Assignment(Incident)	08.10.2013	15:41:46
21	Assignment(Incident)	08.10.2013	15:47:32
22	Update(Incident)	08.10.2013	15:47:32
23	Assignment(Incident)	09.10.2013	09:42:50
24	Reassignment(Incident)	09.10.2013	09:42:50
25	Update(Incident)	09.10.2013	09:42:50
26	Assignment(Incident)	09.10.2013	09:55:37
27	Operator Update(Incident)	09.10.2013	10:55:55
28	Record Open(Change)	09.10.2013	10:57:00
29	Actual Start(Change)	09.10.2013	11:00:00
30	Actual End(Change)	09.10.2013	11:05:00
31	Closed(Incident)	09.10.2013	11:06:36
32	Record Close(Change)	09.10.2013	11:10:00
33	Caused By CI(Incident)	10.10.2013	08:50:25

Fig. 17. An example of cases whose change record is open right after updating its operator.

On the other hand, analyzing the end-to-end process, we found such an incomprehensible process flow. <Fig. 18> shows an example of the cases that the ITIL process workflow is executed reversely. First of all, Change management subprocess is executed, and then incident and interaction management subprocess is executed. To find and resolve the root causes why these cases occur, domain knowledge from expert who knows the system or the work is needed.

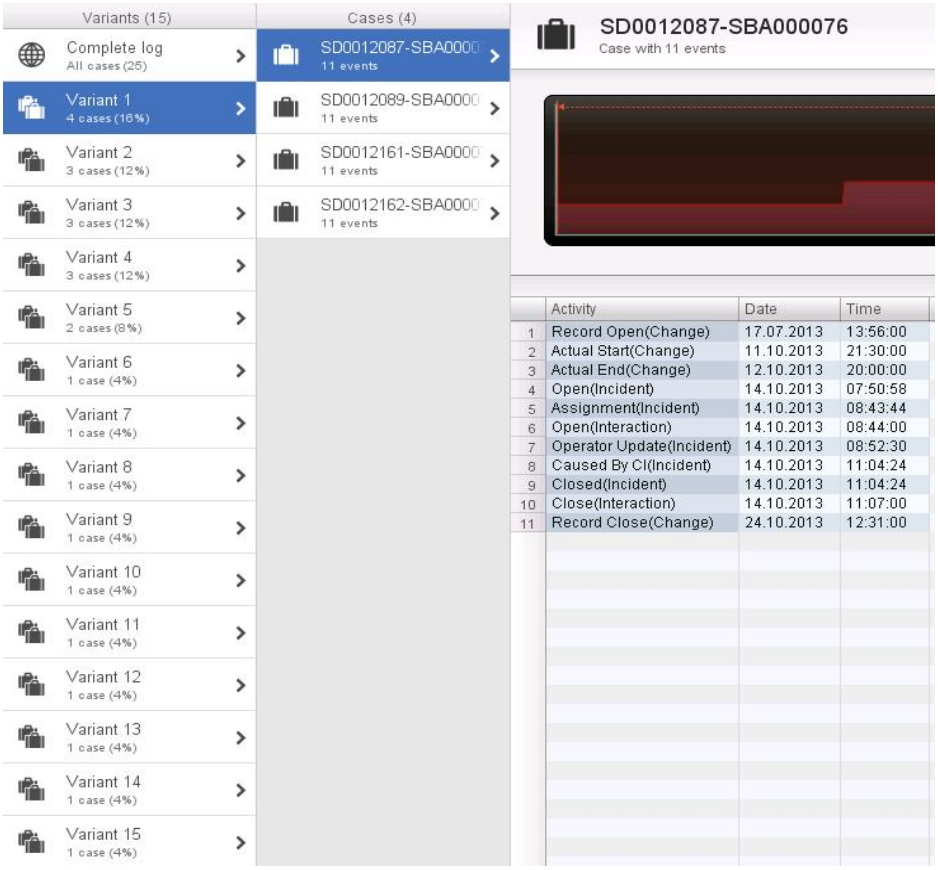


Fig. 18. An example of the cases that the ITIL process is executed reversely.

We managed to create the end-to-end process by linking the given four tables. However, the process analysis on the duration of effects and the degree of effects of each Change implementation didn't work out, because there was no link between Change implementations and Interactions.
Therefore, it was impossible to use them as analysis method to address the first, second, and third question.
Because of the limit above, we analyzed the first, second, and third question, by using the analysis method each question requires.

4.2 Analysis Assumptions

4.2.1 Underlying Assumptions for Analyzing the Questions

4.2.1.1 Events, which you can figure out interconnections among Interaction, Incident, and Change Management Process, are chosen for the Analysis Range of our Report.

Change Management is done when particular service disruptions reoccur more often than usual during the process of Interaction Management, or when Interaction Management handles their tasks over to Incident Management because they cannot deal with the problems they are in charge at the level of the Service Desk (SD). After doing Changes, to catch the impact of particular Changes through the changes in the workload of Interaction or Incident, we needed to select events related to Changes, and because of that the premise of our report is to reflect events which can be used to confirm the interrelations among Interaction, Incident, and Change Management Processes. Of course, there's an exception that a change is done by itself; Change Management – Originated from Problem values. However, it is not excluded because it has an influence on the CI Name, or WBS equally as well.
As stated above, we needed to extract samples from each event data, and we set standards for extracting samples; the problem received (aff) for Interaction, the problem which is the real cause after solving it (cby) for Incident, and the problem affected after a Change (aff) for Change. Then, we connected these samples together and analyzed them.

4.2.1.2 The cases, which have particular values Irrelevant to the Problems of the System from the Parameters of Interaction, are Excluded.

Of the parameters of Interaction, we excluded the cases which have particular values irrelevant to the system's problems. For example, 'No error – works as designed', 'Operator error', 'User error', 'User manual not used', 'Questions', 'Auto Closed', and 'Solved by User Instruction' from Closurecode and 'request for information' from Catetory are not caused by the system, but by simple mistakes, or human resources. This means that the interactions with these values are nothing to do with handling the problems of the system over to Change. Also, the workload is not fluctuated by the impact of Change. That's why we excluded these cases to exactly find out the impact of Change, or the fluctuations of Interaction and Incident.

4.2.1.3 The cases, which have few values, or pass the bounds of Common sense of the Overview of the Process given, are also Excluded.

The cases which have less than ten Interactions and have the number of Change more than the number of Interaction are excluded. Because cases with few data barely fluctuate, so it is difficult to verify that the values of the fluctuations and the patterns

are of great significance. Also, cases, having the number of Change more than the number of Interaction, are excluded because they are considered that they are not caused by Interaction immediately, but by Change with specific reason. The cases such as above accounted for about 9%, however, we figured out that the 9% of the data would not affect the reliability of the analysis and could still use the rest of it, even if we removed them. Therefore, we decided to get rid of it.

4.3 Question 1: Identification of Impact-Patterns

4.3.1 Understanding the Question

Rabobank Group ICT posits that the implementation of change is related to the increased/decreased workload of Service Desk (SD) and IT Operations (ITO).This implementation affects one or more configuration items. By analyzing the log, the organization aims to identify any patterns of this relationship for various service components to which a configuration item is related to. Finding these patterns is likely to contribute to predicting the workload at the SD and/or ITO after future change implementations.

The ITIL process consists of the three big subprocesses; Interaction Management → Incident Management → Change Management. The Interaction Management subprocess is ended if the problems brought up in this process are closed. If not, the problems are transferred to the Incident Management subprocess. Therefore, as the number of the problems brought up in Interaction Management subprocess decreases, the number of the problems that need to be closed in Incident Management subprocess decreases, as well. Based on this logical inference, we sought to analyze the increase/decrease workload of SD and ITO on the number of events of Interaction Management subprocess.

4.3.2 Levels of Analysis

To answer the first question, it is important to set levels of analysis in Change Management subprocess. As shown below, we chose ‘CI Name (aff)’ as level of analysis among those four attributes (refer to <Fig. 19>). Change Management subprocess could be most specifically understood when CI Name is set as Case ID. This detailed understanding is helpful to measure and analyze the increase/decrease workload by change implementations to be specific.

Of Unique Values

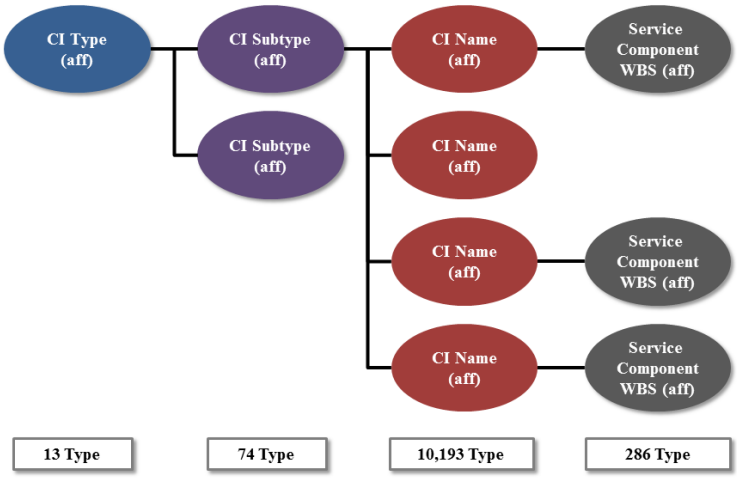


Fig. 19. Alternative Data Fields for Determining Analysis Level

4.3.3 Key Data Columns for Addressing Question 1

Key data attributes for addressing Question 1 are as follows.

Open time of Interaction Management: Time received to resolve the service disruption from customers.
Change record close time of Change Management: Time the change is closed on the system. It is defined that the change affects the system after this time.
CI Name: It is a data attribute which is very important that links the three subprocesses. That is, this attribute shows configuration item (CI Name (aff)) where a disruption of an ICT Service is noticed in Interaction Management subprocess and configuration item (CI Name (CBy)) which caused the disruption of an ICT service in Incident Management subprocess. Lastly, it also shows configuration item (CI Name (aff)) which will be affected by this change in Change Management subprocess.

4.3.4 Analysis

To measure the increase/decrease workload after Change implementations, we measured the number of the configuration item (CI Name (aff)) which is affected by Interaction Management subprocess by the selected Case ID('CI Name (aff)' of Change Management subprocess) on a daily basis. Based on 'Change record close time of Change Management', the increase/decrease workload is described in <Fig. 20-22>. Through data, on the other hand, it was found that there are no calls/mails which are newly registered on weekends. We, therefore, excluded weekends to avoid incorrect analysis results and analyzed data. Also, for your better understanding of analysis results, a trend line is added to the result graph (refer to <Fig. 20-22 Black Lines>).

4.3.5 Analysis Results

As shown in <Fig. 20-22>, the number of unique values of 'CI Name (aff)' is 10193. Due to space restrictions, we cannot display all the analysis results and graphs per case id. Hence, three exemplary patterns are displayed in this report (refer to <Fig.20-22>).

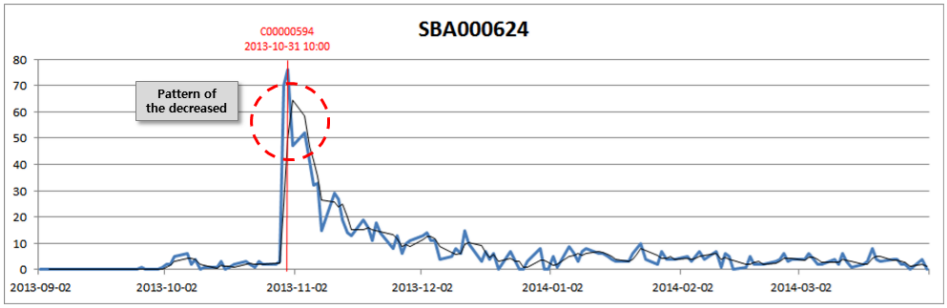


Fig. 20. Shows the increase/decrease after change implementations of the CI Name (aff) SBA000624.

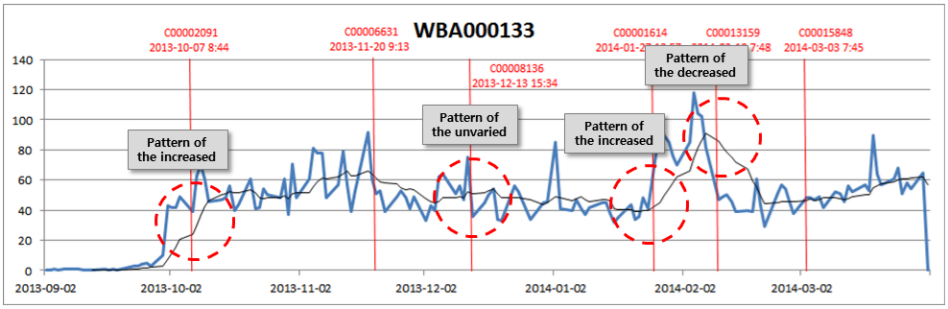


Fig. 21. Shows the increase/decrease after change implementations of the CI Name (aff) WBA000133

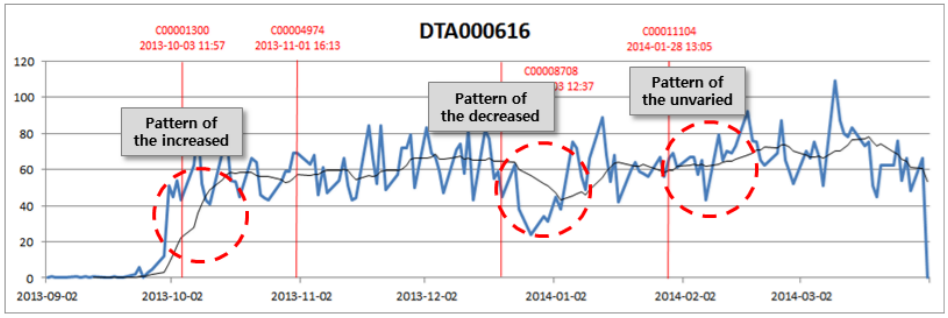


Fig. 22. Shows the increase/decrease after change implementations of the CI Name (aff) DTA000616 .

1. In case that Interaction is likely to be on the increase after Change implementations.
   - Change ID of the CI Name (aff) WBA000133 : C0002091, C00001614
   - Change ID of the CI Name (aff) DTA000616 : C00001300

2. In case that Interaction is likely to be on the decrease after Change implementations.
   - Change ID of the CI Name (aff) SBA000624 : C00000594
   - Change ID of the CI Name (aff) WBA000133 : C00006631, C00013159
   - Change ID of the CI Name (aff) DTA000616 : C00008708

3. Interaction with no change after Change implementations.
   - Change ID of the CI Name (aff) WBA000133 : C00008136, C00015848
   - Change ID of the CI Name (aff) DTA000616 : C00004974, C00011104

4.4 Question 2: Parameters for Every Impact-Pattern

4.4.1 Understanding the Question

From 4.2, we found the patterns of the increase/decrease workload of Service Desk (SD) after Change implementations. It is an important task to find parameters describing what the future Change implementations affect the workload of SD, using results of the past Change implementations. To deal with this task, we used a Decision Tree method of data mining. The fields of Change Table directly related to Change Management subprocess and the variants derived from processing these fields are used as parameters of the Decision Tree, and we sought to find such significant parameters.

4.4.2 Levels of Analysis

To answer the first question, we selected the most specific analysis unit, based on the data field ‘CI Name (aff)’. However, we made a judgment that this analysis unit is inappropriate when finding significant parameters from Decision Tree analysis. It is not easy to find significant parameters because of the small increase/decrease workload of SD at the level of configuration item after Change implementations. To answer the second question, therefore, we sought to raise the level of this analysis to the level of service component.

4.4.3 Analysis

4.4.3.1 Measuring the Increase/Decrease Workload of SD after Change Implementations

As mentioned above, we sought to measure the increase/decrease workload of SD after Change implementations per service component in the number of Closed Interaction from the service component. However, it is extremely challenging because it is impossible to measure the increase/decrease of SD only which is the number of Closed Interaction, due to the interaction between effects of Change implementations per service component.

To overcome this limitation, we only included the Change implementation when there is no other Change implementations within 8 days after one Change implementation for certain service component. To determine the Change implementation is a good case or a bad case, we compared the number of Interaction for a day which occurred the day before this Change implementation with the average daily number of Interaction for 7 days which occurred after the Change implementation. The average daily number of Interaction which occurred for 7 days was counted except for weekends and holidays.

As stated above, with this measurement method, there's a limit that all the Change implementations are not included in analysis. However, it was found that this method can contribute to preventing getting such distorted analysis results from the interaction between Change implementations.

4.4.3.2 Analysis Procedure

We sought to use a Decision Tree method of data mining to distinguish between Good Case and Bad Case in the same way of describing each Change implementation and to find such significant parameters that can generalize this distinction. At this time, The Change implementations whose increase/decrease in average is 0 are excluded because it is difficult to distinguish them between Good Case and Bad Case.

4.4.3.3 Alternative Parameters for Decision Tree Analysis

The three variants below derived from processing 21 data fields of the given Change Table are selected as alternative parameters for Decision Tree analysis.

LEAD_TIME: Subtracted value from 'Change Record Close Time' to 'Change Record Open Time'. It indicates the time spent on a certain Change implementation.

SATISFY_REQ: By comparing 'Request End Date' with 'Change Record Close Time', it indicates whether revisions are complete by the requested date.

INT_COUNT_1: The increase/decrease of Closed Interaction for a day after 'Change Record Close Time'.

.

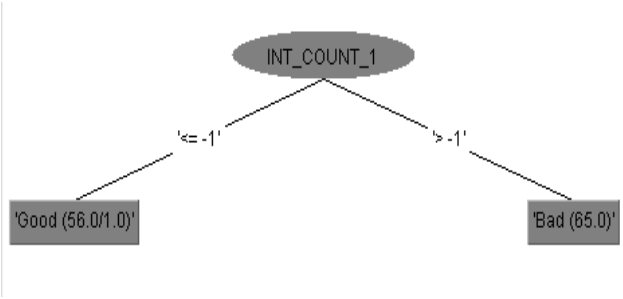
4.4.4 Exporting Data

Of 30275 cases of the Change implementations, 451 cases meet the measuring method stated above. Among 451 cases, 340 cases have Service Component included in the given data table. Lastly, of 340 cases, we excluded the cases whose average increase/decrease for 7 days after Change implementations. Therefore, 121 cases were used for analysis.

4.4.5 Analysis Results

4.4.5.1 Shows the Analysis Results from Data Mining Tool-Weka.

The result showed that only INT_COUNT_1 which indicates the increase/decrease of Closed Interactions for a day after Change implementations can be used to predict results. It is found that the cases with the decreasing number of Closed Interactions for a day after Change implementations decrease after 7 days with 98.2% reliability. It is also found that the cases with the increasing number of Closed Interactions for a day or the cases whose number of Closed Interactions has not changed after Change implementations increase after 7 days with 99.1% reliability (refer to <Fig. 23>). Taken together, it has 99.1% reliability.



Parameter	Condition	Result	Reliablility	
Int_Count_1	Under -1	Good	98.20%	55/56
	More than over 1	Bad	100%	65/65

Fig. 23. shows the analysis results from Data Mining Tool-Weka.

4.5 Question 3: Change in Average Steps to Resolution

4.5.1 Understanding the Question

The ultimate goal of this question is to find the effects of Change implementations of Rabobank. Project managers perform Change Management subprocess to lead to an improvement plan to prevent these problems to happen again. They want to offer a better service to customers after Change implementations; they sometimes want to offer the same service as before. We sought to find the effects of Change Management subprocess based on the result of the increase/decrease of Service Component as stated in 4.3.

4.5.2 Levels of Analysis

The level of analysis to find the effects of Change implementations can be a criterion of the attribute 'Change ID'. However, as we mentioned basic assumptions in 4.1, when figuring out the effects of Change implementations with the Service Components involved in the three subprocesses, all the Service Components of Change ID cannot be included (refer to <Fig. 24>). Therefore, we selected Change ID which can only consist of the Service Components involved in all the three subprocess.

Detailed Change table

Change ID	Service component WBS (aff)
C00000282	WBS000008
	WBS000028
	WBS000123

Influence After Change table

Change ID	Service component WBS (aff)	Comp Status	Complete Change	Change Effect
C00000282	WBS000008	Good	N	It is impossible to judge change status
	X	?		
	X	?		

Fig. 24. An example of Change ID excluded from analysis.

4.5.3 Explaining the Data Columns for Analysis

To address the third question, we created Table 'Influence After Change' which shows the effects of each Change implementation. <Fig. 25> shows the field definitions of this table.

Service component WBS (aff)	Service Comp WBS (aff) which is affected after Change implementations.
Change ID	The unique ID of a Influence After Change (i.e., Change ID of 'Detailed Change' Table)
Comp_Status	Indication of the increase/decrease of Service Components after Change implementations. Based on the measuring method stated in 4.3, it is determined that it is 'Good' if the service Components after Change implementations are increasing. If not, it is 'Bad'.
Complete_Change	Column which checks whether the Service Comp WBS (aff) affected by the Change ID of table Detailed Change is in there ('Y'), or not ('N').
Change_Effect	Column which judges the effects of Change implementations. It is determined that it is 'Good' if more than half of the Service components changed in a positive way. If not, it is 'Bad'. It is 'Null' if it is impossible to judge the effects of Change implementations because there's no Service Comp WBS (aff) when comparing with table Detailed Change.

Fig. 25. Field Definitions of 'Influence after Change' Table

4.5.4 Analysis

To make a judgment that whether each Change implementation is well performed, as mentioned above, the measuring method to the effects of each Change implementation is required. We used the measuring method described in 4.3 to measure the increase/decrease of Service Components. We also selected the Change ID that can consist of the Service Components involved in the three subprocesses. Based on these Service Components are Change ID, we sought to figure out the effects of Change implementations. We determined that if more than half of the Service Components of each Change ID after Change implementations have positive results, it contribute to offering a better service.

4.5.5 Analysis Results

Based on the extracted Service Components for analyzing 4.3, 258 Change IDs were found. The result showed that 205 Change IDs, accounting for 75%, have more than half of the Service Components with decreasing Interaction among Service Components involved in each Change ID. In other words, the Change implementations related to 205 change IDs led to a better service in general. However, 53 Change IDs, accounting for 21%, have more than half of the Service Components with increasing Closed Interactions among Service Components involved in each Change ID (refer to <Fig. 26>). In conclusion, the Change implementations related to 53 Change IDs led to a bad service.

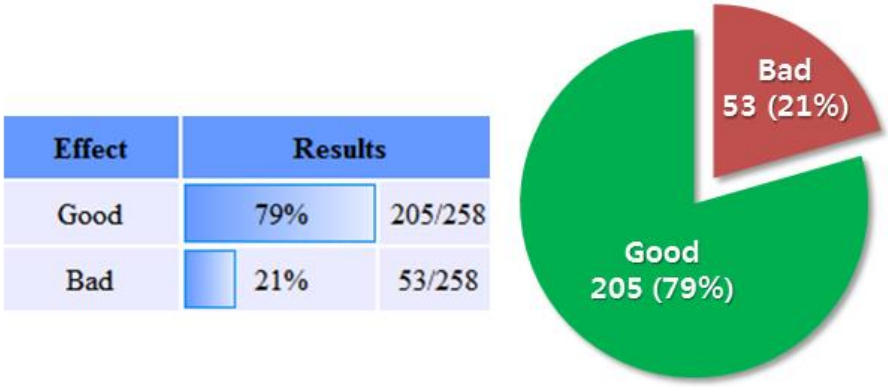


Fig. 26. Analysis Results of Question 3

5. Conclusions

Real life event logs of Rabobank ICT's ITIL process consisting of interaction/ incident/change management subprocesses were provided for the BPI Challenge 2014. These subprocesses are supported by ITIL Service Management tool called HP Service Manager. Furthermore, four questions related to the three subprocesses were raised. To address the questions, we attempted to understand the data and create relevant datasets for the questions. As we noted before, we addressed Question 4 (Creativity Challenge) at the very first. The reason is that addressing our creativity challenge (i.e., understanding the three subprocesses as an end-to-end ITIL process) can be a stepping stone for understanding and addressing the other three questions.

5.1 Analysis Results of ITIL End-to-end Process

We found that the actual process is different from the ITIL Reference Process Model of Rabobank. According to this Reference process model, similar incidents which reoccur more often than usual have to go through problem analysis and be connected to Change Management subprocess. However, the actual process map has unexpected paths. To find and resolve the root causes why these cases occur, domain knowledge from experts who know the system and work well is required. There is need to improve the ITIL Reference Model based on the actual workflow.

5.2 Analysis Results of Implemented Changes

- Question 1: Analysis Results of the Increased/Decreased/Unvaried Patterns of Interaction

The frequency and the portion of three types are as follows. (i) Decreased: 633 (12.5%), (ii) Increased: 672 (13.2%), (iii) Unvaried: 3779 (74.3%) The result of this analysis showed the increased/decreased/unvaried patterns of Interaction after Change implementations (refer to <Fig. 20-22>). The frequencies and ratios of each pattern are described as follows:

- Question 2: The Increase/Decrease Closed Interactions Based on Service Components.

Based on the analysis of Service Component, the result of whether there is the increase/decrease of Closed Interactions showed that 55 cases, accounting for 45%, indicate the decrease of Closed Interactions after Change implementations and 66 cases, accounting for 55%, indicate the increase of Closed Interactions after Change implementations.(refer to <Fig. 27>)

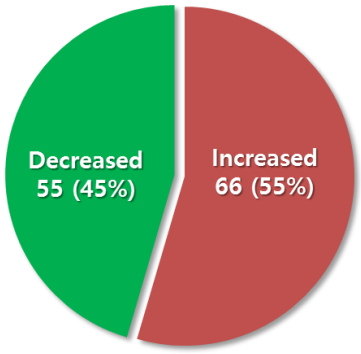


Fig. 27. The result of the increase/decrease of Closed Interactions based on Service Component.

- Question 3: Result of the Increase/Decrease of Closed Interaction, based on Change ID

Based on Change IDs, the result of the increase/decrease of Closed interactions showed that 205 Change implementations, accounting for 79%, have a positive impact on decreasing Closed interactions, however, 53 Change implementations, accounting for 21%, have a negative impact on decreasing Closed interactions (refer to <Fig. 26>).

From this analysis, it is concluded that the decrease of Closed Interactions is not always led by the result of Change Management subprocess. Of course, there are many different variants that can lead to the increase/decrease of Closed interactions, however, it is found that the ultimate goal of this process has not been achieved yet because Change implementations are the activity for the increase/decrease of Interaction and Incident. Therefore, Rabobank Group ICT needs to analyze the root causes and find the way to improve Change Management subprocess.

6. Recommendations for Improvement

6.1 Improving Log Data

6.1.1 Improving Log Data for Developing Delicate Forecasting Model

As stated in 4.3, we used 21 fields of table Detail Change and the three additional variants derived from these fields as alternative parameters for Decision Tree analysis. Among them, however, 'the increase/decrease of Closed interactions for a day after Change implementations' was the only data that can help us to predict the workload of SD. It is possible that other variants which we couldn't find can be significant parameters, however, it is determined that additional variants which can be used for data mining, such as a Decision Tree analysis, need to be recorded on Log Data.
There's another important limit of this log data that we cannot figure out how Change implementations affect the increase/decrease workload of SD, through the interaction between Change implementations. To address this problem, among the given 30275 Change implementations, we used only the part of it.

We can use a lot more data for analysis if the minimum details about defining the impacts of each Change implementation are recorded on log data. For example, it is provided that 'Impact', 'Urgency', and 'Priority' related to each registered Interaction need to be recorded in table 'Detailed Interaction'. These records are used to define Request End Date of the Interaction. By applying this approach to Change Management service process, we can have criteria for judgment of the duration of effects and the degree of effects of each Change implementation.

6.1.2 Improving Log Data of Disconnection between Sub Processes

We managed to create the end-to-end process by linking the given four tables. However, the process analysis on the duration of effects and the degree of effects of each Change implementation didn't work out, because there was no link between Change implementations and Interactions.

To address this problem, there is a need to record the link between each Change implementation and Interaction. We suggest that you search for the Change IDs related to CI Names and record it such as 'Related Incident' of table 'Detailed Interaction' registering Interactions. Through this record, the interaction between Change implementations and Interactions can be clearly seen. Also, this record can be used to overcome the process of disconnection and to analyze log data in depth. Furthermore, it is helpful to find significant parameters of a Decision Tree if it can be found how many Change IDs are related to interaction implementations.

6.2 Use of Forecasting Analysis Results

As stated in 4.4, the result showed that Closed Interactions for a day after 'Change Record Close' can predict Closed Interactions for 7 days with 99.8% reliability. Based on this analysis result, we suggest that you need to deal with further increase workload when Closed Interactions are not likely to decrease for a day after Change implementations. For example, by analyzing Change implementations or expanding human resources of SD, you can cope with increasing Interactions. Also, analyzing the increasing Interactions and handing out manuals for addressing this problem can contribute to efficiency of the workload of SD.

7. Acknowledgements

The authors are grateful to Dr. Anne Rozinat and Dr. Christian Gunther (Fluxicon) for providing them with an evaluation version of Disco and an accompanying copy of the BPIC 2014 datasets. They also thank Dr. Young Sik Kang (Myongji Univ.) and Chang Jae Kang (PMIG) for their thoughtful guidance.

8. References

[1] M. Bozkaya, J. Gabriels, J.M. van der Werf, Process diagnostics: A method based on process mining, *Proceedings of the International Conference on Information, Process, and Knowledge Management*, 2009.

[2] M. Hammer, Reengineering work: Don't automate, obliterate, Harvard Business Review 68 (4), 1990, pp. 104-112.

[3] P. Harmon, Business Process Change: A Guide for Business Managers and BPM and Six Sigma Professionals, Morgan Kaufmann, 2007.

[4] A. Mahal, How Work Gets Done: Business Process Management, Basics and Beyond, Technics Publications, LLC, 2010.

[5] Á. Rebuge, D.R. Ferreira, Business process analysis in healthcare environments: A methodology based on process mining, Information Systems 37 (2), 2012, pp. 99-116.

[6] M. Schäfermeyer, C. Rosenkranz, R. Holten, The impact of business process complexity on business process standardization, Business & Information Systems Engineering 4 (5), 2012, pp. 261-270.

[7] W.M.P. van der Aalst, Process Mining: Discovery, Conformance and Enhancement of Business Processes, Springer, 2011.

[8] W.M.P. van der Aalst, H.A. Reijers, A.J.M.M. Weijters, B.F. van Dongen, A.K. alves de Medeiros, M. Song, H.M.V. Verbeek, Business process mining: An industrial application, Information Systems 32 (5), 2007, pp. 713-732.

[9] T.H.C. van der Heijden, Process Mining Project Methodology: Developing a General Approach to Apply Process Mining in Practice, Eindhoven, 2012.