# MASTERARBEIT im Studiengang

# Wirtschaftsinformatik - Master of Science

gestellt von: Professor Dr. Peter Loos

Thema: Event Log Analysis for Operational Decision Support in BPM


bearbeitet von

Name: Frederik Leonhardt

Adresse: Schutzbergstraße 28
66119 Saarbrücken


Abgabetermin: .....

Spätester Beurteilungstermin: .....

**Abstract**

This work aims to improve the execution of business processes by developing an operational decision support system to help human decision makers. Historical event data is used for predictive analysis, the training of a decision tree and logistic regression, while a process monitoring engine acts as data aggregator.

The proposed approach was implemented as a proof-of-concept prototype, and evaluated by performing simulation runs and collecting the accumulated data. A small Business Process Execution and Monitoring environment connected simulation, recommendation service and process monitoring and allowed the exportation of the collected event logs. Additionally it stored prediction and classification results, which then were analysed together with the event logs with third party software.

The evaluation has shown that it is possible to achieve a good recommendation performance in regards to metrics such as root-mean squared error, recall and precision, even with a few input features. However, it has also shown that these findings only apply for process instances which have reached the last part of their life-time. When focusing on the development of the performance indicators over time, it is obvious that young cases do not have enough data associated to provide confident recommendations early on.

# Acknowledgements

I am particularly grateful to Michael zur Mühlen for his hands on supervision, and many helpful discussions. Additionally I would like to thank Peter Loos for support. Andreas Emrich contributed significantly to the project, his expertise and advice were invaluable. Alexandra Theobalt's organisational support was also much appreciated.

I also would like to thank Lucy Weston-Taylor for the proofreading and continuous support and encouragement throughout the project.

*Frederik Leonhardt, May 2014*

# Contents

# List of Figures

# List of Tables

# List of Abbreviations

| | |
|---|---|
| **ABC** | Activity-Based Costing |
| **BAM** | Business Activity Monitoring |
| **BI** | Business Intelligence |
| **BPM** | Business Process Management |
| **BPMN** | Business Process Model and Notation |
| **BPMS** | Business Process Management System |
| **BPIC** | Business Process Intelligence Challenge |
| **BPS** | Business Process Simulation |
| **DCM** | Dynamic Case Management |
| **DES** | Discrete Event Simulation |
| **DSS** | Decision Support System |
| **eEPC** | Extended Event-driven Process Chain |
| **EPC** | Event-driven Process Chain |
| **ERP** | Enterprise Resource Planning |
| **FTE** | Full-time equivalent |
| **KPI** | Key Performance Indicator |
| **MDP** | Markov Decision Process |
| **MIS** | Management Information System |
| **MXML** | Mining eXtensible Markup Language |
| **PAIS** | Process-Aware Information System |
| **PCA** | Principal Component Analysis |

| | |
|---|---|
| **UML** | Unified Modeling Language |
| **WFM** | Workflow Management |
| **WFMS** | Workflow Management System |
| **XES** | eXtensible Event Stream |

# Chapter 1

# Introduction

## 1.1 Motivation

IT systems of today's organisations are growing increasingly complex due to the requirements of modern project management. The globalisation has lead to a previously unknown economic interconnectedness of autonomous organisations. With this, a need of cooperation and information exchange between their IT systems emerges.[1] To face these challenges, businesses leverage concepts like Workflow Management (WFM).

WFM, or the automation of business processes, has already been in scientific focus since the early nineties,[2] and subsequently was widely adapted by companies in the real world. In recent years the focus has shifted to Business Process Management (BPM), often referred to as an extension of WFM.[3] BPM not only aims at process automation, but also includes means of process analysis and extends automation into the area of real-time operational business support – or short: "BPM is all about making choices".[4]

1. Frame, J. D., *The New Project Management: Tools for an Age of Rapid Change, Complexity, and Other Business Realities*, 2nd ed. (Jossey-Bass, 2002), p. 4, isbn: 978-0-7879-5892-3; North, M. J. and Macal, C. M., *Managing Business Complexity: Discovering Strategic Solutions with Agent-Based Modeling and Simulation* (New York, NY, USA: Oxford University Press, 2007), p. 3, isbn: 978-0195172119.

2. Medina-Mora, R. et al., "The action workflow approach to workflow management technology," in *Proceedings of the Conference on Computer-Supported Cooperative Work (CSCW '92)* (1992), 281–288, isbn: 0897915437.

3. van der Aalst, W. M. P., "Business Process Management: A Comprehensive Survey," *ISRN Software Engineering* 2013 (2013): p. 1, issn: 2090-7680, doi:10.1155/2013/507984.

4. Ibid., p. 10.

While WFM is a rather mechanical approach, BPM tries to account for human factors as well. Business processes involving human actors might need different handling than a pure machine-to-machine interaction. When considering humans in the management of operations, the traditional concept of resources has to be redefined. Traditionally resources include mostly mechanical numbers, i.e. funds, available machines, number of Full-time equivalents (FTEs) or production inputs and outputs. Those resources usually act as *anticipated* constraints for any planning and optimisation and sometimes can be represented in units of each other. Monetary resources usually can be converted into other resources like FTEs or available machines. [5]

A possible human-centered resource is defined by Davenport and Beck as *attention*. In their point of view the attention of its managers is a very valuable – because limited – resource for any company.[6] Based on findings in the psychology humans only have a limited pool of attention. When flooded with too much information this pool is drained quickly, leading to inefficient decisions. Process management can take those new findings into account to provide automated decision support systems which in this example could help managers shift their attention to actual and important problems in the company's workflow.

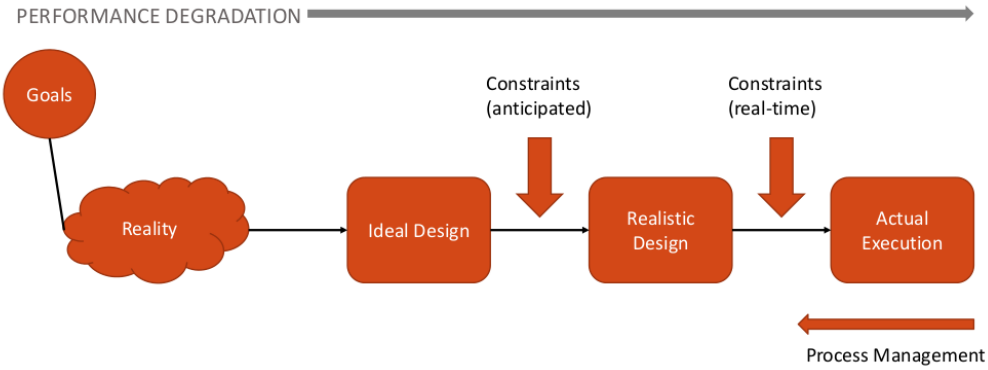



Figure 1.1: The challenge of process management.[7]

Fig. 1.1 illustrated the performance degradation encountered in the implementation of a process model. At the highest level, strategic goals lead an organisation's long-term orientation. Based on these goals, an ideal process design can be derived. At this point any anticipated constraints need to

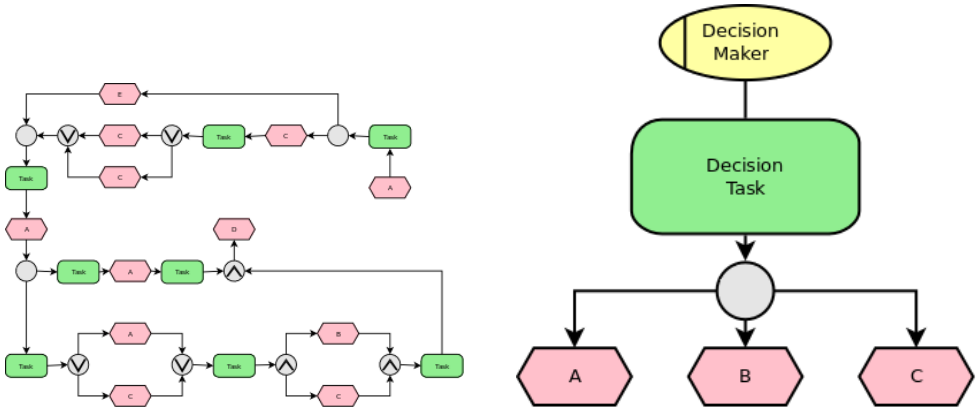
5. Naïvely neglecting other real-time constraints, e.g. a limited labour market.

6. Davenport, T. H. and Beck, J. C., *The attention economy: understanding the new currency of business* (Boston, MA: Harvard Business School Press, 2001), ISBN: 1578518717.

7. Wrt. discussion with Michael zur Mühlen; October 9,2013)

be taken into consideration for a realistic design. Furthermore, during the real-time execution unforeseen constraints again can lower the overall performance. Business Process Management as a whole aims at counteracting this performance degradation at all levels. Supporting the design of a process model can help to assess anticipated constraints and to create a better process model. Operational support can specifically help at the lowest level, the actual process execution, to mitigate performance penalties originating from real-time constraints.

### 1.1.1 Perspectives on Decision Making Tasks in BPM




(a) Designer Perspective (b) User Perspective

Figure 1.2: Traditional perspectives on Workflow Management (own illustration).

A process model plays a central role in BPM, and this thesis distinguishes three perspectives on decision making tasks in workflow management: The designer perspective, the user perspective and the management perspective.

The *Designer Perspective* is a holistic view on the underlying model of a workflow, illustrated in Fig. Fig. 1.2a It is widely used for approaches of data and process mining. Data mining describes the knowledge discovery on large data sets, with the goal of extracting information into an understandable structure like patterns or decision rules. Process mining extends those methods to the domain of process management. It aims at gaining a deep insight into raw event data collected by a business IT system and often leads to the automatic creation of process models.[8] The designer perspective is useful to analyse and visualise the process structure, to carry out

8. VAN DER AALST, W. M. P., *Process Mining: Discovery, Conformance and Enhancement of Business Processes* (Heidelberg: Springer Berlin / Heidelberg, 2011), pp. 1-10, ISBN: 9783642193446, doi:10.1007/978-3-642-19345-3.

performance evaluations and to test process compliance. It can be described as *the big picture* on the workflow of an organisation.

The *User Perspective* as depicted in Fig. 1.2b concentrates on decisions within a workflow. Every time a decision point within the workflow is reached, the involved actors need to decide on a possible course of action. Actors include both humans and automated IT systems. It is used to develop local decision support by means of decision mining (data mining) or recommender systems. However this perspective is restricted to a single process instance in the workflow and requires prior identification of decision tasks with their possible outcomes. To provide informed decision support at such points, the underlying process model can be utilised.

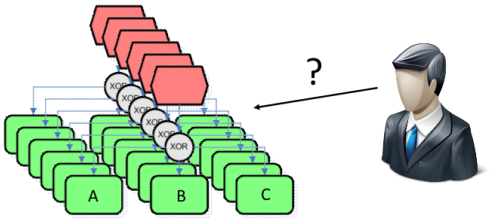



Figure 1.3: Management perspective (own illustration).

The *Management Perspective* illustrated in Fig. 1.3 describes a global perspective on all running processes. In real life systems, hundreds of processes are active simultaneously and depend on each other. Given a scarce resource availability within the organisation, both in terms of time (which is ultimately money) and human attention, managers need to decide on which area they should focus on. To make this decision, they need information about current bottlenecks and process instances which might develop to be problematic in the future. A global decision support system needs to incorporate this perspective. The management perspective is characterised by a focus on multiple process instances and their inter-dependencies, which often are not obvious to the observer. It aims at solving the question *What should I focus on?*.

This thesis focuses on the identification of problematic instances in the management perspective with the goal to develop a support systems for managers. The system should help them distribute their attention span on the most pressing issues which need manual resolving. Additionally it should give hints on how to mitigate the identified problems.

### 1.1.2 Pattern matching trap

When taking humans into consideration, apart from a new perspective on resources another aspect arises. While humans have excellent capabilities in pattern matching, they often fail to fully assess a new situation. Instead they tend to match upcoming decision situations with patterns experienced in the past. This behaviour is called *pattern matching trap*[9] and can lead to biased decisions (e.g. the retailer who only using historic sales forecast, when in reality there are additional other predictor values).

An automated support system offers an additional objective, unbiased view onto decision making tasks. By combining both the humans capability of matching patterns and the systems ability to learn from history data, a forecasting process can be optimised which subsequently leads to better decisions.

Such real-time approaches are rendered possible by recent adaptations of Process-Aware Information Systems (PAISs) in real company environments, which allow the collection of vast amounts of process data. This development is also known as *Big Data*.[10]

## 1.2 Problem Statement

### 1.2.1 Research Questions

In reality the full potential of real-time optimisation has not been tapped yet. Past approaches have often ignored the human factor in BPM, and concentrated solely on technical process optimisation and automation. Ideally, however, human decision makers should be the focus of BPM. These considerations lead to the following questions:

- How can human-centered operational decision support in BPM environment be provided?
- Which previous approaches already exist in this area?

Subsequently, in order to advance decision support technology in BPM, requirements for such decision support need to be established. Most human decision makers interact with BPM through concrete process instances, and

9. Hoch, S. J. and Schkade, D. A., "Psychological Approach Support to Decision Systems," *Management Science* 42, no. 1 (1996): pp. 52 ff.

10. Manyika, J. et al., *Big data: The next frontier for innovation, competition, and productivity*, technical report June (McKinsey Global Institute, 2011), pp. 15-37.

therefore this actual process execution should be improved. To do so, firstly problematic instances have to be identified. This requires well-defined, automatically detectable characteristics and generic performance scores. They can be derived by analysing real process logs. The second challenge is to identify those instances in real time. At the end of this process, the human decision makers need to be informed of any such problematic instances. This consideration leads to the following questions:

- What are problematic process instances and how to identify them?
- How to ensure that these instances get the necessary attention by decision makers?
- How to use existing event data to support future decision making?
- What recommendations can be given when encountering a problematic process instance?

### 1.2.2 Goals

In the course of this thesis a generic approach to providing operational decision support in BPM should be developed. To achieve this, firstly an overview of existing approaches in the area of operational decision support needs to be created. This should result in a suitable classification, a description of the shortcomings and the extraction of requirements for the conceptual background.

The conceptual approach needs to tackle the identification of problematic process instances, possibly by means of identifying reliable real-time indicators to recognise such processes. Furthermore, a decision support strategy aimed to help human decision makers should be introduced.

After that, the approach should be implemented as a proof-of-concept prototype and evaluated on a suitable data set. The evaluation should be aimed to verify whether the approach developed in the course of this thesis is feasible and actually improves the performance of business processes.

## 1.3 Methodology and Structure

### 1.3.1 Methodological approach

To tackle those problems this work follows the design science paradigm. It aims at gaining scientific insight through creation and evaluation of IT solutions in form of models, methods or prototypes.[11] Hevner defines seven guideline for design science research, each of which this work follows.[12] The following paragraphs give an overview over those guidelines, and later on the chapters of this thesis are related to each one of them.

The first guideline states that "design-science research must produce a viable artifact in form of a construct, a model, a method or instantiation"[13] and the second guideline stresses about the need of concentrating upon an important and relevant business problem. Guideline 3 states that the quality of the design artifact must be "demonstrated via well-executed evaluation methods",[14] while guideline 4 demands that "effective design-science research must provide clear contributions in the areas of the design artifact".[15]

In guideline 5 he encourages a research rigour by using "rigorous methods in both the construction and evaluation of the design artifact".[16] Guideline 6 states that design science is a iterative search process. The approach presented in this thesis makes use of certain assumptions and simplifications, which may not be realistic enough to make a significant impact on practice, and can therefore only represent a starting point. Lastly, the presentation must be understandable for both technology-oriented and management-oriented audiences, a principle required by the seventh guideline.

Besides those guidelines Wilde and Hess identify methodologies commonly used in Information Systems Research and Business Informatics.[17] Based on their catalogue this work uses a simulation approach to evaluate a prototypical recommender system.

11. Wilde, T. and Hess, T., *Methodenspektrum der Wirtschaftsinformatik: Überblick und Portfoliobildung*, technical report 2 (München: Institut für Wirtschaftsinformatik und Neue Medien, 2006), p. 3.

12. See Hevner, A. R. et al., "Design science in information systems research," *MIS Quarterly* 28, no. 1 (2004): p. 11-25, ISSN: 0276-7783, `http://www.hec.unil.ch/ypigneur/HCI/articles/hevner04.pdf`.

13. Ibid., p. 12.

14. Ibid.

15. Ibid., p. 23.

16. Ibid., p. 24.

17. Wilde, T. and Hess, T., "Forschungsmethoden der Wirtschaftsinformatik - Eine empirische Untersuchung," *Wirtschaftsinformatik* 49, no. 4 (2007): p. 282.

### 1.3.2 Structure of this thesis

In the following paragraphs is explained how the chapters relate to the design science paradigm. This thesis aims at improving operational support by developing a real-time approach to decision optimisation. The motivation in chapter 1 gives an introduction into the relevance of this topic as a business problem, and summarises the problem statement.

Chapter 2 explains the knowledge base – the theoretical foundations of the problem areas. The approach including the assumptions and simplifications made in this work are introduced in chapter 3. Understanding design science as iterative approach, those assumptions may not be realistic enough to make a significant impact on practice, but the approach should represent a starting point. Additionally, in this chapter performance metrics are listed to measure the quality of the artifact.

The implementation described in chapter 4 serves as viable design artifact. At the same time, as a novel prototype, this design artifact forms the core contribution of the research. Together with raw data discussed the same chapter, it provides a clear and verifiable research contribution. The evaluation carried out in chapter 5 utilises an experimental approach by conducting an controlled experiment on real process data.

Together chapters 2 to 5 aim at satisfying the fifth guideline, achieving a research rigour by justifying why the developed artifact works or does not work. A conclusion about the findings and possible future work scenarios are outlined in chapter 6.

In the end, this written work as a whole acts as communication of the research conducted therein, both for technology-oriented and management-oriented audiences.

# Chapter 2

# Basic Terminology & Related Work

The first part of this chapter gives an overview of Business Process Management associated concepts like the BPM life-cycle, and briefly discusses potential perspectives on BPM. This overview focuses mostly on the run-time parts of BPM, where operational decision support can be provided. At the end process mining techniques are briefly described.

The second part introduces basic definitions and requirements of decision support. After that, related approaches to operational decision support in BPM environments are discussed in detail. The result is a classification of these approaches based on their level of automation and type of support.

At the end a summary of the findings follows along with a description of limitations and shortcomings of the approaches described before.

## 2.1 Business Process Management

### 2.1.1 Basic Terminology

This section introduces the basic terminology used in this thesis. To begin with the concept of a business process and its relates to information systems is described. After that a brief overview of the history of BPM follows, covering the major evolutionary steps.

#### 2.1.1.1 Definition of Business Process Management

The concept of BPM is based around business processes. A business process can be defined as a set of activities "that are performed in coordination in an organizational and technical environment".[1]

> Workflow management technology aims at the automated support and coordination of business processes to reduce cost and flow times, and increase quality of service and productivity.[2]

More specifically, WFM refers to the automation of business processes[3] in a rather mechanistic manner.[4] A software system implementing the work flow, a workflow management system, directs the control flow based on control data while executing processes.[5] That is to say, it is responsible for the (automated) coordination and distribution of concurrent activities and detection of exceptional situations. BPM can be seen as an extension of WFM.[6]

> Business Process Management (BPM) is the discipline that combines knowledge from information technology and knowledge from management sciences and applies this to operational business processes.[7]

Some authors argue that WFM and BPM in practice are used interchangeably,[8] but for the academic purpose of this work the difference between them stands. BPM is considered to have a broader scope, since it does not exclu-

1. Weske, M., *Business process management: concepts, languages, architectures.*, 2nd ed. (Berlin, Heidelberg: Springer Berlin Heidelberg, 2012), p. 5, ISBN: 978-3-642-28616-2, doi:10.1007/978-3-642-28616-2.

2. van der Aalst, W. M. P. and Jablonski, S., "Dealing with workflow change: identification of issues and solutions," *Computer systems science and engineering* 15, no. 5 (2000): p. 267.

3. Jablonski, S. and Bussler, C., *Workflow Management: Modeling Concepts, Architecture and Implementation* (London, UK: International Thomson Computer Press, 1996), p. 7 ff. ISBN: 9781850322221.

4. van der Aalst, "Business Process Management: A Comprehensive Survey," p. 1.

5. Becker, J., Kugeler, M., and Rosemann, M., eds., *Prozessmanagement: Ein Leitfaden zur prozessorientierten Organisationsgestaltung*, 7th ed. (Berlin, Heidelberg: Springer Berlin Heidelberg, 2012), p. 202, ISBN: 978-3-642-33843-4, doi:10.1007/978-3-642-33844-1.

6. van der Aalst, "Business Process Management: A Comprehensive Survey," p. 1.

7. Ibid.

8. zur Muehlen, M. and Hansmann, H., "Workflowmanagement," in *Prozessmanagement: Ein Leitfaden zur prozessorientierten Organisationsgestaltung*, 7th ed., ed. Becker, J., Kugeler, M., and Rosemann, M. (Berlin, Heidelberg: Springer Berlin Heidelberg, 2012), p. 367, ISBN: 978-3-642-33843-4, doi:10.1007/978-3-642-33844-1_11.

sively aim at process automation, but also extends this automation into the area of real-time operational support. To achieve this, BPM utilizes Operations Research approaches like process analysis. A crucial part of BPM is the underlying process model. It aims to capture different ways in which a process instance (a *case*), can possibly be handled. Such a model can be expressed in various languages,[9] the most commonly used are Business Process Model and Notation (BPMN), Unified Modeling Language (UML) and the Extended Event-driven Process Chain (eEPC).[10]

A type of software that manages, controls and supports operational processes is called Business Process Management System. Along with Enterprise Resource Planning (ERP) systems and high-level middleware, WFM and BPM systems are both PAISs. These systems share the characteristic of an explicit process notation, and the system is aware of its processes. This enabled the systems to collect structured process data (event logs), which can later be used for analysis and model enhancements.

#### *2.1.1.2 A Brief History of BPM*

Adam Smith (1723-1790) and Frederick Taylor (1856-1915), with their concepts of *division of labour* and *scientific management*, are often named as the early pioneers of BPM, as it is easy to see that these ideas are used in modern BPM systems.[11] The enabling of mass production through assembly lines is deemed another early milestone, which is attributable to Henry Ford (1863-1947).[12]

In the following century the rapid development of IT systems, and advances in the research of modelling languages helped to develop sophisticated BPM

9. For a more comprehensive overview, see Dumas, M., van der Aalst, W. M., and ter Hofstede, A. H., *Process-Aware Information Systems* (Hoboken, NJ, 2005), isbn: 9780471663065; zur Muehlen, M., Recker, J., and Indulska, M., "Sometimes Less is More: Are Process Modeling Languages Overly Complex?," in *The 3rd International Workshop on Vocabularies, Ontologies and Rules for The Enterprise* (IEEE Publishers, 2007)

10. See Meyer, S. et al., "Towards Modeling Real-World Aware Business Processes," in *Proceedings of the Second International Workshop on Web of Things*, June (New York, NY, USA: ACM, 2011), p. 2, isbn: 9781450306249, doi:10.1145/1993966.1993978.

11. See Davenport, T. H. and Short, J. E., "The New Industrial Engineering: Information Technology and Business Process Redesign," *MIT Sloan Management Review* (Cambridge, MA, USA) 31, no. 4 (1990): p. 1, http://sloanreview.mit.edu/article/the-new-industrial-engineering-information-technology-and-business-process-redesign/; van der Aalst, "Business Process Management: A Comprehensive Survey," p. 3.

12. See van der Aalst, "Business Process Management: A Comprehensive Survey," p. 3.

systems. Milestones include the invention of the Turing machine in 1936, which can be viewed as data-enabled process model; the introduction of computers in the late 1940s, which started to influence business processes; and the introduction of Petri Nets in 1962, which is one of the de-facto standards in mathematical and business process modelling today. The development of Database Management Systems started in the 1970s, enabling the development of Workflow Management Systems in the 1990s, predecessors to today's integrated Enterprise Resource Planning (ERP) and BPM systems.[13]

The complexity of business processes in today's organisations is ever increasing, with the rise of cross-organisational communication and global collaboration. To cope with this complexity, process models and BPM systems are widely used in today's organisations. This leaves the management of organisations and the users of BPM systems with a plethora of choices regarding all aspects of their processes, the creation and management of them – "BPM is all about making choices".[14]

### 2.1.2 Concepts of BPM

This section introduces the *Business Process Life-Cycle*, a five-phased approach to BPM, and the five *Perspectives on BPM*. At the end, the the key concerns and use cases of BPM are discussed.

#### *2.1.2.1 Business Process Life-Cycle*

Fig. 2.1 shows one of many versions of the Business Process Life-Cycle in the literature. All life-cycles are based on the same concepts and only differ in the perspective and/or granularity. The key concepts originate from the *Deming Wheel*, an iterative PDCA cycle (plan – do – check – act).[15] Such an iterative cycle can be applied for improvement processes in any area, including production processes or business processes. The BPM life-cycle depicted here describes five distinct phases of managing a business process.[16]

13. van der Aalst, "Business Process Management: A Comprehensive Survey," pp. 3–4.
14. Ibid., p. 10.
15. Moen, R. and Norman, C., *Evolution of the PDCA Cycle*, technical report (2006), 1–11.
16. See van der Aalst, *Process Mining: Discovery, Conformance and Enhancement of Business Processes*.
17. Wrt. Ibid., p. 8

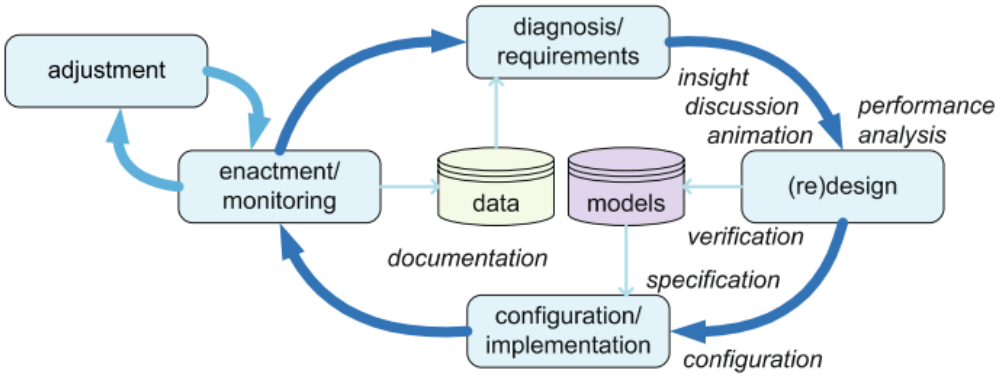



Figure 2.1: Five phases of the Business Process Life Cycle.[17]

The cycle usually starts with a **design phase**, where a business process model is originally designed. If this phase is entered in a later iteration, an existing model is improved instead. Those improvements are guided by the insights gathered in the preceding diagnosis phase.

The resulting design acts as specification for the following **configuration phase**. There, the model is implemented either as a WFM system or into an already existing system. If the design is already modelled in a language supported by the WFM, this phase can be completed very quickly. Basic testing of the process should be carried out in this phase as well. The first two phases can be referred to as *build time*.

After the system is running with the implemented design the **enactment phase** starts. Here, running processes are monitored to identify any re-configurations which might be necessary at run-time. Those changes are made in the **adjustment phase**, and are not implemented by re-designing the original model, but rather by utilising existing controls of the running system. Those phases are usually called *run time*.

The collected data is analysed in depth in the **diagnosis phase**. This allows examination of model weaknesses, resource bottlenecks and any other problems. Any insights gathered in this diagnosis influence the following redesign phase, where the model is adapted based on the conclusions drawn from the observations in the monitoring phase.

As mentioned before, other representations of the BPM life-cycle exist in the literature. While the presented model contains an adjustment phase, another representation focusing on the design part might contain a business process re-engineering phase instead.[18] Alternatively, the BPM life-cycle can

18. See Becker, Kugeler, and Rosemann, *Prozessmanagement: Ein Leitfaden zur*

be understood as containing the three main phases design, configuration and enactment. The diagnosis phase is now decoupled, and can be seen as it's own dimension. There model-based analysis is conducted parallel to the design phase, while data-based analysis happens parallel to the enactment phase.[19] In contrast to seeing the analysis as a separate phase downstream, this approach has the advantage of integrating the analysis into both build and run time of the process. Such a view allows the resulting insights to be used directly for process or model optimisation, without the need to re-iterate the whole cycle.

#### *2.1.2.2 Five Perspectives on Business Process Management*

To gain a comprehensive view on BPM, modern literature differentiates between five perspectives.[20] The **control-flow perspective** (also referred to as process perspective) models the ordering of activities and their routing and control flow. It is often the backbone of a process model. The **resource perspective** (also known as organisational perspective) contains roles and their relations, organisational units and other resources. The **data perspective** (also known as case perspective) contains modelling decisions and is responsible for data creation. The **time perspective** is concerned with timing and frequency of events, as well as modelling durations and deadlines. Lastly, the **function perspective** describes activities and related application.

#### *2.1.2.3 Key Concerns and Use Cases of BPM*

In his recent survey about BPM, van der Aalst identifies six key concerns of BPM and 20 use cases.[21] The following list provides an overview and a short explanation for each key concern, going into more details where necessary for this thesis (i.e enactment infrastructures and data-based analysis).

**Process Modeling Languages**

The choice of a language to represent an organisation's business pro-

*prozessorientierten Organisationsgestaltung*, p. 314.

19. See VAN DER AALST, "Business Process Management: A Comprehensive Survey," p. 5.

20. VAN DER AALST and JABLONSKI, "Dealing with workflow change: identification of issues and solutions"; VAN DER AALST, *Process Mining: Discovery, Conformance and Enhancement of Business Processes*; VAN DER AALST, "Business Process Management: A Comprehensive Survey."

21. VAN DER AALST, "Business Process Management: A Comprehensive Survey," pp. 23-28.

cesses is essential due to the central role of process modeling and analysis in BPM.[22]

**Process Enactment Infrastructures**

The core of any BPM system is a process enactment service. It provides the run-time environment and is responsible for workflow control and execution. *Process definition tools* are used at build time to specify the workflow process definitions, while *administration and monitoring tools* are used to monitor and control the workflows at run-time. This includes allocating people (or resources) and handling of exceptions.[23]

**Process Model Analysis**

A process model can be analysed in two ways: *verification* and *performance*. Verification measures the correctness of a process, while performance analysis is focused on performance indicators like flow times, utilisation, etc. Neither approach uses event data, but perform analysis only using the model. From a management point of view, performance analysis is more relevant.[24]

**Process Mining**

The goal of *process mining* is to exploit accumulating event data of an information system in a meaningful way. It can be used to discover new process models, or to check conformance of an existing model.[25]

**Process Flexibility**

The ability to deal with both foreseen and unforeseen changes of the business environment is called *process flexibility*.[26]

**Process Reuse**

Since organisations adapted BPM in their day-to-day operations, large process model repositories have accumulated. This leads to problems maintaining and re-using those models. BPM systems need to ease the search for models, and their maintenance (e.g. updating models or merging models).[27]

For each key concern a couple of use cases have been identified. The following paragraph describes the relevant use cases for this thesis: *Adapt while running (adaWR)*, *Monitor (Mon)* and *Enact model (EnM)* in detail.

22. van der Aalst, "Business Process Management: A Comprehensive Survey," p. 12.
23. Ibid., p. 15.
24. Ibid., p. 19.
25. Ibid., p. 22.
26. Ibid., p. 25.
27. Ibid., p. 26.

*Monitor* and *Adapt while running* and *Enact model* are located in the process execution and belong to the BPM key concern **Process Enactment Infrastructures**. Monitor "refers to all measurements done at runtime without creating or using a model".[28]

At runtime, choices may be resolved by human decision making. The use case adaWR refers to the principle that "BPM is all about making choices".[29] It includes process configuration (adjustment phase), which cares about selecting a desired behaviour from a family of process variants. This use case refers to situations where the model is adapted at runtime.

Lastly, *Analyze Performance Based on Model* can be assigned to the key concern **Process Model Analysis**. It is located in the BPM phases configuration and analysis. The use case refers to analyses of expected performance in terms of response times, waiting times, flow times, utilisation, costs, etc.

### 2.1.3 Process Mining

Since this work is using process mining techniques to create a process model, this section allows a closer look into the process mining use case. Process Mining is a way to exploit existing process data in a meaningful way, for example to aid process designers with the construction of process models or to enhance existing models. Process Mining is based on event logs, traces of the former execution of a business process in an information system. Event logs can be available in different maturity levels, from poorest quality (missing events, logged by hand) up to excellent quality (trustworthy, complete, well-defined). In a typical information system, event logs can be extracted from transactional data (e.g. a database system) or from logging traces. However, this approach will usually only lead to data of medium quality. On the other hand, a PAIS purposefully collects event logs, automatically and in a systematic manner, resulting in much better data.[30] Such structured event logs may store additional information like an timestamp, information about the resource executing an activity, or other data elements recorded with the event.

28. van der Aalst, "Business Process Management: A Comprehensive Survey," p. 10.

29. Ibid., p. 13.

30. IEEE Task Force on Process Mining, "Process Mining Manifesto," in *BPM 2011 International Workshops*, ed. Daniel, F., Barkaoui, K., and Dustdar, S. (2012), p. 7, isbn: 978-3-642-28115-0.

31. ibid., p. 3

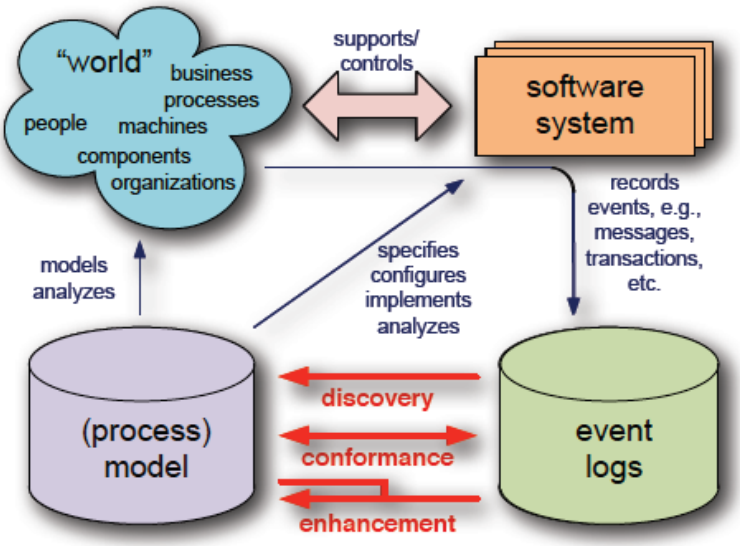



Figure 2.2: Positioning of the three main types of process mining: (a) *discovery*, (b) *conformance checking*, and (c) *enhancement*.[31]

Fig. 2.2 shows the three main types of process mining. The first and most prominent process mining technique is **process discovery**. Based on a set of event logs a fitting process model can be produced, e.g. by using the *α-algorithm*.[32] This type of process mining can be related to the design phase of the BPM life-cycle.

The second type is **conformance checking**, which is essentially a target-performance comparison of process model (target) and reality as captured in event logs (performance). Conformance checking usually relates to the diagnosis phase. Lastly, process mining can be used for model **enhancement**. The idea is to improve an already existing model using event logs from previous executions. Enhancement is be located in the diagnosis and re-design phases of the BPM life-cycle.

Process mining covers multiple perspectives on the BPM life-cyle: the control-flow perspective (by establishing an ordering of activities), the resource perspective (by classifying people into roles and organisational units), the data perspective (paths in the process), and the time perspective (discovery of bottlenecks, predict remaining processing time).

While process mining is mainly used in an offline scenario, it can also be used

32. Cf. van der Aalst, W. M. P., Weijters, T., and Maruster, L., "Workflow mining: discovering process models from event logs," *IEEE Transactions on Knowledge and Data Engineering* 16, no. 9 (September 2004): pp. 1128-1142, ISSN: 1041-4347, doi:10.1109/TKDE.2004.47.

for operational decision support by considering running cases, i.e. instances which have not completed yet.[33]

## 2.2 Operational Decision Support

### 2.2.1 Basic Terminology

Before discussing decision support, the underlying decision making process is examined. To begin with, human information processing is described with the help of a simplified model, and potentials for improvement are identified based on the attention and resource theories. Then, decision support systems are discussed and a classification in levels of automation is introduced.

#### *2.2.1.1 What is Decision Making?*

A sub-field of cognitive psychology is concerned with the systematic analysis of the human decision making process. To understand human decision making, a holistic view on human information processing is necessary. Fig. 2.3 (overleaf) depicts the four distinct steps of human information processing according to Parasuraman et al.[34] This model only acts as a gross simplification, since most tasks involve inter-dependent stages and are often executed concurrently. However, this simplification allows the direct translation of these stages into system functions, which can be automated: information acquisition, information analysis, decision and action selection, and action implementation.[35]

The first step is sensory processing of a particular item of information – gaining awareness of the information. This stage includes initial pre-processing of data prior to full perception. After that the information resides in the working memory, allowing for conscious perception and manipulation of the

33. van der Aalst, W. M. P., Pesic, M., and Song, M., "Beyond Process Mining: From the Past to Present and Future," in *22nd International Conference on Advanced Information Systems Engineering* (Hammamet, Tunisia: Springer Berlin Heidelberg, 2010), p. 2, doi:10.1007/978-3-642-13094-6_5.

34. Parasuraman, R., Sheridan, T. B., and Wickens, C. D., "A model for types and levels of human interaction with automation.," *IEEE transactions on systems, man, and cybernetics. Part A, Systems and humans* 30, no. 3 (May 2000): pp. 287 f. issn: 1083-4427.

35. Ibid., p. 288.

36. ibid., p. 287.

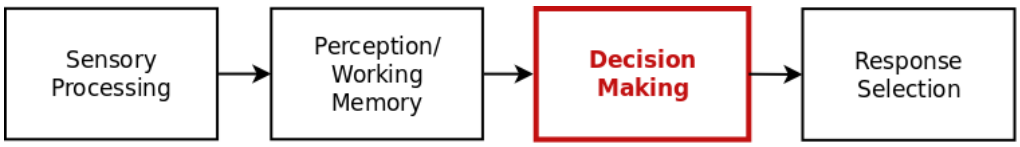



Figure 2.3: Four stage Model of human information processing: (a) *sensory processing*, (b) *perception*, (c) *decision making*, and (d) *response selection* (adapted version).[36]

information. The third stage acts as *point of decision*, where a conscious decision is made based on the previous cognitive processing. Consequently, an appropriate response or action consistent to this decision is selected.[37]

However, not all information is deemed important enough by the mind, and might be discarded as unimportant before reaching the conscious decision making phase (e.g. when hearing a familiar sound, a person usually does not direct their attention towards it). To make an efficient and informed decision, the information first has to pass this perception filter. This cognitive process is called *attention*, and it refers to selectively concentrating on one aspect of the environment while ignoring others.[38]

Both Davenport and Beck, and Sternberg and Sternberg argue that any information input requires some of our limited-capacity attentional resources.[39] In cognitive psychology, this is called *resource theory*. Although humans are able to automate certain tasks so that they require less resources, this mainly applies to repetitive tasks. Because "organizational attention involves rich, parallel processes", managers confronted with decision tasks usually are not able to automate those while maintaining good decision quality.[40]

#### *2.2.1.2 Decision Support Systems*

Management Information Systems (MISs) and Decision Support Systems (DSSs) are (semi-)automated systems supporting managers in their semi-structured or unstructured decision making activities.[41] This relates to the management perspective introduced in section 1.1.1, which describes a global

37. Parasuraman, Sheridan, and Wickens, "A model for types and levels of human interaction with automation.," p. 287.

38. Anderson, J. R., *Cognitive Psychology and Its Implications*, 6th ed. (Worth Publishers, 2004), p. 519, ISBN: 978-0716701101.

39. Davenport and Beck, *The attention economy: understanding the new currency of business*, pp. 10-50; Sternberg, R. J. and Sternberg, K., *Cognitive Psychology*, 6th ed. (Cengage Learning, 2011), ISBN: 978-1133313915.

40. Cf. Davenport and Beck, *The attention economy: understanding the new currency of business*.

41. See Eom, S. B. et al., "A Survey of Decision Support System Applications (1988-1994)," *The Journal of Operational Research Society* 49, no. 2 (1998): pp. 1 ff.

perspective on all running processes. DSSs do not aim to replace the human user, but merely supporting them in their decisions.

Considering the human attention restrictions described in the last section, a DSS should pre-filter information to avoid an information overflow, and direct the user to relevant items of information, i.e. only those which require human attention. Additionally, DSSs can support the user in the actual decision task, by providing a selection of recommended means of action.[42] This helps mitigating the *pattern matching trap*, in which a users decisions are led by wrong conclusions drawn from experiences in the past.

#### *2.2.1.3 Levels of Automation*

Automated systems can be classified into one of the ten levels of automation depicted in Fig. 2.4 (overleaf). Automation "refers to the full or partial replacement of a function previously carried out by the human operator."[43] This implies that there are various levels of automation on which automated systems can operate.

Depending on three risk factors, an assessment for a desired level of automation can be made. The primary evaluative criteria are *human performance consequences*, a system and its human operators should offer a higher performance after implementing automation. This includes factors like mental workload, situation awareness, complacency and skill degradation.[44]

A secondary evaluative criteria is *automation reliability*, since the operators mental workload will only benefit when the automation is reliable. Another secondary criteria is the *cost of decisions*, mostly relevant in the case of an incorrect or inappropriate action. This risk associated with an erroneous decision $D$ can be approximated as cost of an error multiplied by the probability of that error: $risk(D) = cost(D) * P(D)$. Such decisions with a low risk are strong candidates for high-level automation. In fact, such an automation can prevent humans from being overloaded, so that instead of concentrating on simple decisions, they can carry out other, more important functions.[46]

42. Hoch and Schkade, "Psychological Approach Support to Decision Systems."
43. Parasuraman, Sheridan, and Wickens, "A model for types and levels of human interaction with automation.," p. 287.
44. Ibid., p. 291.
45. ibid., p. 287.
46. Ibid., p. 292.

| | |
|---|---|
| HIGH | 10. The computer decides everything, acts autonomously, ignoring the human. |
| | 9. informs the human only if it, the computer, decides to |
| | 8. informs the human only if asked, or |
| | 7. executes automatically, then necessarily informs the human, and |
| | 6. allows the human a restricted time to veto before automatic execution, or |
| | 5. executes that suggestion if the human approves, or |
| | 4. suggests one alternative |
| | 3. narrows the selection down to a few, or |
| | 2. The computer offers a complete set of decision/action alternatives, or |
| LOW | 1. The computer offers no assistance: human must take all decisions and actions. |

Figure 2.4: Levels of automation of decision and action selection. [45]

### 2.2.2 Decision Support in BPM Environments

Operational support in BPM consists of three dimensions: *Detection*, *prediction* and *recommendation*.[47] Its goal should be to "provide directions and guidance rather than enforcing a particular route."[48]

The requirement for operational support in those three dimensions is comprehensive process monitoring to enable data-based analysis parallel to the process execution. The first part of this section introduces the concept of process monitoring, and possible performance dimensions to be able to identify problematic cases. After that various process analysis techniques, and prediction and recommender algorithms stemming from machine learning are discussed.

Two notable architectures for real-time operational support in BPM environments have been proposed. The first is a generic implementation supporting prediction, time-based conformance checking and time-based recommendations. It is based on process mining and utilises an annotated transition system for its support functionality. A prototype based on this architecture was implemented in ProM.[49] The second is a theoretical architecture for real-time decision support, which can utilise either discrete event

47. van der Aalst, Pesic, and Song, "Beyond Process Mining: From the Past to Present and Future," p. 41.

48. van der Aalst, W. M. P., "TomTom for Business Process Management (TomTom4BPM)," in *Advanced Information Systems Engineering*, ed. Eck, P. van, Gordijn, J., and Wieringa, R., Lecture Notes in Computer Science (Springer Berlin Heidelberg, 2009), p. 3, isbn: 978-3-642-02143-5, doi:10.1007/978-3-642-02144-2_2.

49. van der Aalst, Pesic, and Song, "Beyond Process Mining: From the Past to Present and Future," p. 43.

simulation or analytic tools for its decision support.[50]

In a BPM environment decision support can be provided in regard to the process flow or a concrete decision task. The underlying process model provides a structured environment, where possible paths and actions are known for any given point. This is called the *decision space*.[51]

#### *2.2.2.1 Process Monitoring*

The requirement for any automated real-time decision making is *Process Monitoring*. It is one of the five distinct phases of BPM, and its goal is to provide real time information about the status of a business process. Generally, this enables an organisation to make well-informed business decisions. Recently it has been also known as Business Activity Monitoring, and earlier approaches in WFM used the term operative process controlling or workflow monitoring.[52]

In WFM the monitoring is responsible for generating exception reports, e.g. for potentially overdue items in a production process, and for generating status reports. Also it collects audit logs (security-relevant record) and process logs (event logs). Fig. 2.5 shows the division in technical and organisational monitoring. Technical monitoring is used for performance measurements, e.g. system response time and system load. Organisational monitoring measures efficiency, e.g. idle times or workload.[53]

To quantify those terms, measurable (key) performance indicators (KPIs) have been introduced. Based on these indicators a degree of performance for business-specific critical success factors can be identified at any given time. Modern BPM systems usually display information about the current KPIs in a graphical interface.[54]

50. Fritzsche, M. et al., "Extending BPM Environments of Your Choice with Performance Related Decision Support," in *Business Process Management*, ed. Dayal, U. et al., Lecture Notes in Computer Science (Berlin, Heidelberg: Springer Berlin Heidelberg, 2009), pp. 97 ff. isbn: 978-3-642-03847-1, doi:10.1007/978-3-642-03848-8_8.

51. Vanderfeesten, I., Reijers, H. A., and van der Aalst, W. M. P., "Product-based workflow support," *Information Systems* 36, no. 2 (April 2011): p. 529, issn: 03064379, doi:10.1016/j.is.2010.09.008.

52. zur Muehlen, M. and Rosemann, M., "Workflow-based Process Monitoring and Controlling - Technical and Organizational Issues," in *33rd Hawaii International Conference on System Sciences*, c (2000), pp. 1 f. isbn: 0769504930.

53. Ibid., p. 2.

54. Dahanayake, A., Welke, R. J., and Cavalheiro, G., "Improving the understanding of BAM technology for real-time decision support," *International Journal of Business Information Systems* 7, no. 1 (2011): p. 11, doi:10.1504/IJBIS.2011.037294.

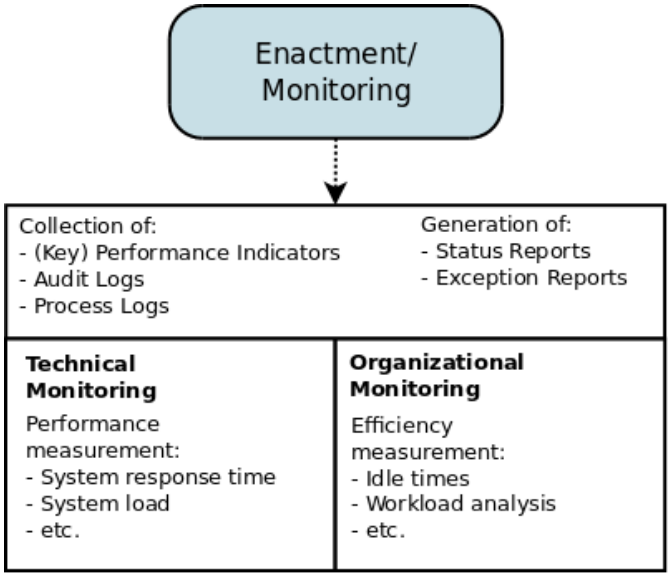



Figure 2.5: Operative process controlling in the enactment phase separated in: (a) *Technical Monitoring*, and (b) *Organizational Monitoring* (own illustration).

Typically, three dimensions of performance can be identified: time, cost and quality.[55] Table 2.1 (overleaf) lists possible KPIs for each dimension. In the time dimension, *lead time* (or flow time) denotes the total time from creation of a case to its completion. When calculating the lead time, the degree of variance may also be important. *Service time* is the time actually spent working on a case, this can be measured per activity or per case. The *waiting time* (or idle time) sums any time span the case is waiting for a resource to become available, again this can be measured per activity or per case. *Synchronization time* is relevant for cases with external triggers or concurrent activities. It counts the time span a case spends in a partially enabled state.

To derive cost performance indicators, various costing models from the application of Accounting and Controlling can be used. For example, *Activity-Based Costing (ABC)*, which assigns costs to products and services based on their consumption of each activity. *Time-Driven ABC* extends this concept into the time dimension, while *Resource Consumption Accounting (RSA)* offers a complex, modern solution to providing cost-based decision support information for managers. The key component of cost-based performance measurement often is the *average resource utilisation*.

The quality dimension focuses on the service delivered to the customer. Questionnaires can directly measure subjective customer satisfaction, while

55. VAN DER AALST, "Business Process Management: A Comprehensive Survey," p. 20 f.

indirect indicators such as *complaints per case* or *total number of defect products* also help to measure performance in this dimension.

| **Time** | **Cost** | **Quality** |
|---|---|---|
| Lead time | Activity-Based Costing (ABC) | Customer satisfaction |
| Service time | Time-Driven ABC | Avg. number of complaints (per case) |
| Waiting time | Resource Consumption Accounting (RCA) | Number of product defects |
| Synchronization time | Average resource utilisation | |

Table 2.1: Examples of performance indicator in the three dimensions of performance: (a) *time*, (b) *cost*, and (c) *quality*

#### *2.2.2.2 Business Process Simulation*

Having established performance dimensions and concrete indicators to measure in the monitoring phase, analysis is required to draw conclusions based on the collected data. Traditionally analysis was seen as a separate phase in the BPM life-cycle after the process execution, but more recent models integrate the diagnosis into the execution phase as *data-based analysis* (cf. previous sections). This has enabled real-time support systems, which utilise such data directly.

In contrast to analytic methods used in the Business Intelligence (BI) field, Business Process Simulation (BPS) is an alternative approach to analyse business processes and their data. The idea is to run a process repeatedly and to collect information about performance indicators in each run, resulting in confidence intervals for the indicators. To run a simulation, several types of information are necessary. First of all, a workflow model of the process defines the ordering of tasks, associated resources, basically the general structure of the process. To control the simulation, additionally a simulation environment is necessary. In real-life a PAIS interacts directly with the model, while in a simulation scenario such a system is not available. The behavioural characteristics of the simulation environment includes specifications of external events like the arrival of new cases, and attributes determining the service time of tasks. Those characteristics are defined in the form of probability distributions, which should reflect the

real-life distributions as good as possible. In general, simulation tries to abstract the details of the business processes, and to replace them by such stochastic distributions. Finally, the simulation needs to respect real-time constraints like number of resources and their availability. Such a *steady-state* simulation is aimed at finding long-term trends and the initial state of a system is of no importance.[56]

Traditional simulation approaches are only connected indirectly with the PAIS they originate from, and are used for helping with business process re-engineering or design by providing what-if views on a process model. Van der Aalst et al. propose the concept of *advanced simulation*, which aims to directly connect PAIS and simulation. This allows real-time simulation by taking into account the current state of a process and historic data collected for the process. This kind of *transient-state* simulation typically is aims to discover short-term developments and can be used for operational decision support by simulating the outcome of all possible decisions at a given point.[57]

A common approach to simulating business processes is to use Discrete Event Simulation (DES). It models the operation of a system as a time-ordered sequence of events. The complete sequence describes the entire experience of an entity as it flows through the simulation.[58] This is very similar to the workflow-based perspective of business processes, and thus the business process model can be directly translated into an appropriate simulation model. Each entity in the system has its own state and can carry information.

In contrast to continuous simulation – which breaks up the time into small slices and updates the system state according to all activities occurred in each slice – discrete-event simulation can run faster, since it can jump from event to event and does not have to consider the time frame in between.

56. VAN DER AALST, W. M. P. et al., "Business Process Simulation: How to get it right?," in *Handbook on Business Process Management*, ed. VOM BROCKE, J. and ROSEMANN, M., International Handbooks on Information Systems (Berlin: Springer, 2010), pp. 2–3.

57. ROZINAT, A. et al., "Workflow Simulation for Operational Decision Support," *Data & Knowledge Engineering* 68, no. 9 (2008): p. 837.

58. HLUPIC, V. and ROBINSON, S., "Business process modelling and analysis using discrete-event simulation," in *Proceedings of the 30th conference on Winter simulation* (Los Alamitos, CA, USA: IEEE Computer Society, 1998), p. 1365.

#### 2.2.2.3 Machine Learning

Both machine learning and data mining provide useful methods to process information. Machine learning can be defined as "the study of data-driven methods capable of mimicking, understanding and aiding human [...] information processing tasks."[59] Data mining is the application of machine learning methods to large and complex data sets aimed at discovering and explaining new, previously unknown insights and relations.[60]

Machine learning algorithms use example data or past experience to produce either predictive or descriptive knowledge. Three types of methods can be distinguished: supervised learning, reinforced learning and unsupervised learning. The common factor between all methods is that given a set of input variables $X$ some output value $Y$ is generated.

In supervised learning the task is to learn the mapping from $X$ to $Y$ based on the input data set. Here, the input data set contains the correct output values for all samples (*labeled data*). Supervised learning methods usually utilise regression models to approximate a good mapping.[61] In the context of supervised learning, the input vector is often called *predictor value* and the output data *target value*. An example is the prediction of sales numbers for the future based on historic data.

Reinforced learning is used for problems where an outcome depends on a sequence of actions. Here no concrete input-output pairs are provided for learning, but instead the learning problem often requires exploration mechanisms.[62] The algorithm only learns based on the quality of the outcome, for example a game of chess can be described as such a problem; each move seen independently might not be very meaningful, but the sequence of moves decides about winning or losing the game.[63]

In unsupervised learning the goal is to find regularities in the input data, and hence no concrete output data is available (*unlabeled data*). Unsupervised learning is closely related to density estimation in statistics. Cluster algorithms can be used to perform unsupervised learning, e.g. to group customers in a (possibly previously unknown) group of similar customers.[64]

59. Barber, D., *Bayesian Reasoning and Machine Learning* (Cambridge University Press, 2012), p. III, isbn: 9780521518147.
60. Alpaydin, E., *Introduction to Machine Learning*, 2nd ed., ed. Dietterich, T. et al., Adaptive Computation and Machine Learning (Cambridge, MA, USA: MIT Press, 2010), p. 2, isbn: 9780262012430.
61. Ibid., pp. 9 ff.
62. Barber, *Bayesian Reasoning and Machine Learning*, pp. 144 f.
63. Alpaydin, *Introduction to Machine Learning*, pp. 13 f.
64. Ibid., pp. 11 f.

The complexity and computation time of learning algorithms usually depends on the dimension of the input and the size of the data set used for training the algorithm. A smaller number of input features also can lead to simpler models, which should always be preferred over a more complex solution (cf. *Occam's razor*). Thus a reduction of the input dimensionality is desirable in a real-time scenario. Such a reduction can be achieved by feature selection or feature extraction. The selection of features aims at reducing the dimensionality by discarding the uninformative, and therefore for the output less important, input variables. Feature extraction aims at finding a combination of the original input features with a reduced dimensionality still containing the same level of information. A method which can be used for such a feature extraction is the Principal Component Analysis (PCA).[65]

### 2.2.3 Related Approaches

#### *2.2.3.1 Overview of Approaches*

Previous approaches to real-time decision making support can be grouped into two sorts, *Prediction* and *Recommendation* or a combination thereof. All approaches utilise partial traces, i.e. logs of process instances which are still running, and are located in the combined enactment and data-based analysis phase of the BPM life-cycle.

A naive approach to prediction would be using simple heuristics like average-based estimation, where the remaining cycle time is estimated as the average cycle time minus the already spent time.[66] Using this approach on attributes with a high variance naturally leads to bad predictions. Therefore, a naive approach is only suitable as a baseline to compare against.

Van Dongen et al. use non-parametric regression to predict the total cycle time of a case. Their approach also estimates which activities will be executed in the future flow.[67] It outperforms the naive approach and could be further improved by including case-specific data. However, deciding which type of variable to use for those data attributes cannot easily be automated

65. Alpaydin, *Introduction to Machine Learning*, p. 110.

66. van Dongen, B. F., Crooy, R. A., and van der Aalst, W. M. P., "Cycle Time Prediction: When Will This Case Finally Be Finished?," in *Proceedings of the OTM 2008 Confederated International Conferences, CoopIS, DOA, GADA, IS, and ODBASE 2008. Part I on On the Move to Meaningful Internet Systems*, Lecture Notes in Computer Science (Berlin, Heidelberg: Springer Berlin Heidelberg, 2008), p. 5.

67. Ibid.

since it typically requires insights into the process and its semantics.[68] More details on such regression-based approaches are presented by Crooy..[69]

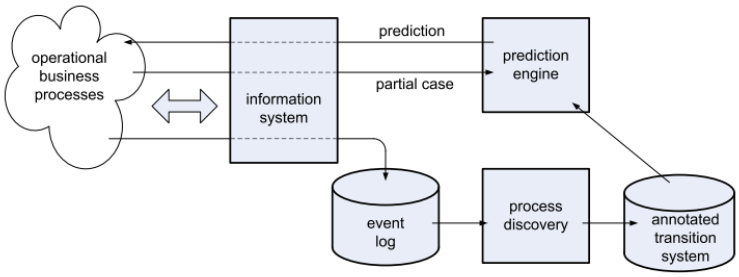



Figure 2.6: Two-step prediction approach by deriving a transition system and using a prediction engine. [70]

The same goal of predicting cycle time is pursued by van der Aalst et al., but with a different approach as depicted in Fig. 2.6. Here a two-step approach is taken by firstly constructing an explicit process model in the form of an annotated transition system with an adjustable degree of abstraction. This model is then used for predictions. This approach outperforms previous approaches based on simulation or regression, both in terms of quality and computation time.[71]

Another approach for giving predictions is *short-term simulation*. Based on design information (process model), state information (current real-time state) and historic information a simulation model in form of a colored petri net can be built. This enables the automatic creation of simulation models, allowing for a direct coupling of real process to simulation. With this model, a short-time simulation system for operational decision support as shown in Fig. 2.7 is feasible.[72]

Schonenberg et al. introduce a recommendation approach aimed at providing more process flexibility at run-time. It offers support based on historic data (earlier experiences), but does not limit the user by only allowing a static control flow like many traditional workflow systems do. The approach

68. van Dongen, Crooy, and van der Aalst, "Cycle Time Prediction: When Will This Case Finally Be Finished?," p. 18.

69. Crooy, R., "Predictions in Information Systems: a process mining perspective." (Master Thesis, Technische Universiteit Eindhoven, 2008), pp. 17-39.

70. van der Aalst, W. M. P., Schonenberg, M., and Song, M., "Time Prediction Based on Process Mining," *Information Systems* 36, no. 2 (2011): p. 2, doi:10.1016/j.is.2010.09.001.

71. Ibid., pp. 30 f.

72. Rozinat et al., "Workflow Simulation for Operational Decision Support."

73. ibid., p. 2.

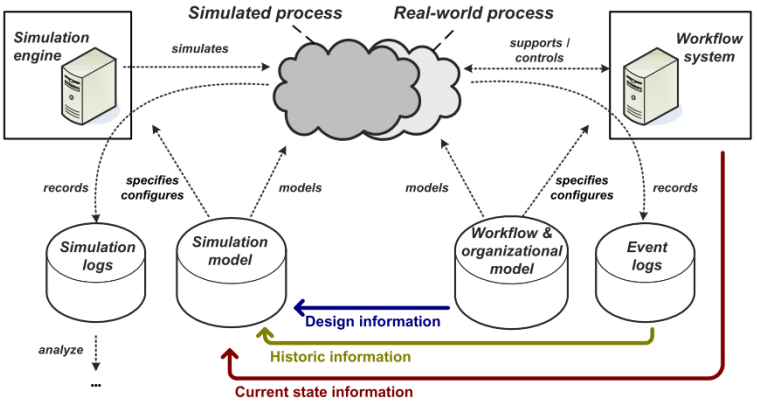



Figure 2.7: Short-term simulation service approach.[73]

is based on an abstraction mechanism to compare current partial cases with earlier executions, and a target function which is optimised. Fig. 2.8 (overleaf) shows the architecture of this approach. Their evaluation is based on the assumption that the user's goal is to minimise cycle time, and is conducted in a controlled experiment. But their approach supports other target functions as well, basically any other performance indicator can be optimised. Results indicate that traces with recommendation support often outperform traces without such support, showing the value of historic information.[74]

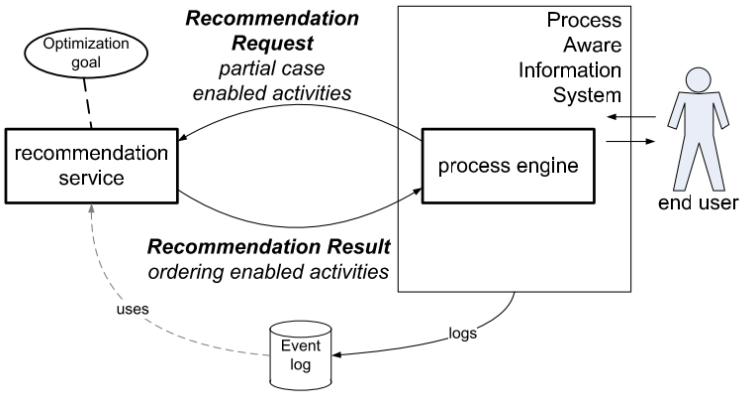



Figure 2.8: Recommendation service approach.[75]

74. Schonenberg, H. et al., "Supporting Flexible Processes Through Recommendations Based on History," in *Business Process Management*, ed. Dumas, M., Reichert, M., and Shan, M., Lecture Notes in Computer Science (Berlin, Heidelberg: Springer Berlin Heidelberg, 2008), p. 15, ISBN: 978-3-540-85757-0, doi:10.1007/978-3-540-85758-7_7.

75. ibid., p. 3.

A recommendation approach in product-based workflow design is presented by Vanderfeesten et al. They introduce recommendation strategies for *workflow products*, informational productions like decisions. Information processed in the workflow is described by data elements, for each case having a different value. Actions on data elements are modelled as operations, defining input and output elements. Additionally, operations can have a number of attributes like execution cost, processing time or failure probability.[76] The authors present a global decision strategy based on *Markov Decision Processes (MDPs)*, basically a Markov chain extended with actions (for allowing choice) and rewards. This approach is guaranteed to find an optimal solution, but like other analytic solutions it needs to calculate the full state space to find the solution – a problem also known as state space explosion problem.[77] Since this is neither feasible for big processes nor at run-time, it only acts as benchmark strategy for local recommendation strategies. Local strategies are heuristics only considering the currently available set of decisions at any point in the process. A possible naive strategy is a random selection, other strategies are lowest cost, lowest failure probability, etc. These strategies can be combined, executed sequentially or weighted. The evaluation has shown that heuristics come close to the optimal solution calculated by MDPs. A shortcoming of the approach is the exclusive consideration of cases in isolation, optimisation on process level or concurrent processes is not supported.[78]

Conforti et al. propose an approach for history-aware real-time risk detection,[79] and risk-aware prediction and recommendation.[80] They identify three fault types for a use case, and provide decision support for risk reduction by predicting the most likely fault severity for each choice. Their approach supports three performance dimensions: time, cost and reputation. Based on the prediction, they recommend the best course of action. By utilising a decision tree trained on historical process data they were able

76. Vanderfeesten, Reijers, and van der Aalst, "Product-based workflow support," pp. 4-7.

77. Fritzsche et al., "Extending BPM Environments of Your Choice with Performance Related Decision Support," p. 100.

78. Vanderfeesten, Reijers, and van der Aalst, "Product-based workflow support," pp. 32.

79. Conforti, R. et al., "History-Aware, Real-Time Risk Detection in Business Processes," in *On the Move to Meaningful Internet Systems: OTM 2011*, ed. Meersman, R. et al., Lecture Notes in Computer Science (Berlin, Heidelberg: Springer Berlin Heidelberg, 2011), pp. 100 f. isbn: 978-3-642-25108-5, doi:10.1007/978-3-642-25109-2_8.

80. Conforti, R. et al., "Supporting Risk-Informed Decisions during Business Process Execution," in *Advanced Information Systems Engineering*, ed. Salinesi, C., Norrie, M. C., and Pastor, Ó., Lecture Notes in Computer Science (Berlin, Heidelberg: Springer Berlin Heidelberg, 2013), pp. 116 f. isbn: 978-3-642-38708-1, doi:10.1007/978-3-642-38709-8_8.

to significantly reduce the overall severity and number of faults.[81]

#### *2.2.3.2 Classification of Approaches*

The approaches discussed in the last section all are based on the common ground of providing support on the operational level, i.e. in real-time. To achieve such a support, they use partial traces, which either are used to classify a case, or predict a target attribute. This target attribute may be flexible, e.g. by using a target function. However, for their evaluation all approaches concentrate on a single target attribute.

Additionally the approaches can be be classified by their type (prediction, recommendation or combination thereof) and the level of automation they provide. Determining this level is not always easy, since some of the approaches do not present an integrated system but only algorithmic basics. This means that the level of automation in most cases could still be increased. The pure recommendation approaches already use a higher level of automation, since they pre-filter possible selections for the user. Table 2.2 (overleaf) shows an overview of the approaches and their respective classification. The naive approach only describes basic algorithms and therefore has no automation level.

81. Conforti et al., "Supporting Risk-Informed Decisions during Business Process Execution," p. 14.

| **Approach** | **Model** | **Target attribute** | **Type** | **Level of Automation** |
|---|---|---|---|---|
| naive | Mean Average | Lead time | Prediction | – |
| Van Dongen et al. (2008) and Crooy (2008) | Non-parametric regression | Lead time | Prediction | 2 |
| Van der Aalst et al. (2011) | Model-based prediction | Lead time | Prediction | 2 |
| Schonenberg et al. (2008) | Top-N recommender | Lead time | Recommendation | 3 |
| Rozinat et al. (2008) | Short-term simulation | Lead time and no. of cases | Prediction | 2 |
| Vanderfeesten et al. (2011) | Markov Decision Processes and Top-N recommender | Total costs | Recommendation | 4 |
| Conforti et al. (2011) and Conforti et al. (2013) | Decision Tree | Risk severity | Prediction and Recommendation | 2 |

Table 2.2: Classification of real-time decision support approaches in BPM environments.

## 2.3 Findings

This chapter has described the requirements for operational decision support in BPM, and discussed various approaches to giving predictions and recommendations at real-time. In all approaches, the key factor is the direct communication between the enactment engine which controls and monitors the workflow, the recommendation service and the event log database which stores historical event data. While these data sources have previously been used separately or for off-line analysis, e.g. in rule-based detection engines or for process re-design, the combination of these data sources enables operational decision support. Furthermore, an underlying process model helps by structuring the decision space, so that supervised learning algorithms can be used.

All approaches lack a view on the global process level, i.e. they ignore competing cases, global resource availability and other external factors which influence the process execution. Furthermore, decision support systems have to consider multiple performance indicators, since an isolated optimisation of process criteria easily leads to a biased optimisation.[82] To avoid such a biased view, the process quality needs to be measured in terms of multiple KPIs coming from all performance dimensions, weighted by their importance for the analyst and the use case.

82. zur Muehlen and Rosemann, "Workflow-based Process Monitoring and Controlling - Technical and Organizational Issues," p. 4.

# Chapter 3

# Approach

An analysis of related approaches regarding operational decision support systems in BPM environments has shown that such a scenario is both technically feasible and can be beneficial for process quality.

Shortcomings of existing approaches include the inability to automatically chose meaningful features, and the exclusive concentration on case-related data. The application of established data mining algorithms in an integrated BPM system in an automated fashion requires the extraction of relevant features from the event logs stored by the business process execution system. The more data those traces contain, the harder it is to select relevant features. To tackle this issue, this thesis introduces an extraction strategy based on weighted performance indicators fully configurable by the user. The utilisation of real event logs instead of sample data sets helps to develop a solution which can be integrated into a real BPM environment.

Furthermore, the approach pursued in this thesis aims to evaluate whether a broader perspective, such as the management view introduced in Chapter 1 (also referred to as a global *process view*) can help to improve the recommendation quality of decision support systems in the operational business process execution phase. In this chapter a set of global features is defined, which later on are included in the data set used for the training of the learning algorithms.

The current approach is clearly placed within the enactment phase of the BPM life-cycle as it touches both execution of the process and real-time enabled data-based diagnosis. The primary goal is to *highlight* problematic process instances at run-time. Problematic instances are instances which need attention to meet some generic, configurable performance score. The secondary goal is to provide meaningful recommendations on a course of action to mitigate the arising problems for such instances.

Chapter 4 describes the implementation of the approach, from selecting a suitable real-life data set to the architecture used for the experiments. The analysis of results in Chapter 5 compares the quality of business process executions regarding some key performance indicators of recommendation-aided and unaided business executions. This shows whether the approach presented in this chapter actually improves the performance indicators, and if so how the consideration of global variables change the recommendation quality.

## 3.1 Requirements

Based on the analysis of related work conducted in the last chapter, some requirements for an operational decision support system can be identified. They mainly concern the interoperability of decision support systems with existing BPM solutions and the data sources used for training.

**Integration** By developing a solution with clearly defined interfaces, it can be used in existing BPM systems. The approach should be able to handle event logs in a standardised format like eXtensible Event Stream (XES), which allows the communication with existing systems and the analysis of the results with established tools.

**Data** The approach should utilise four data classes:

**Design data** in form of a process model.

**Case data** in form of partial log traces of running process instances.

**Historical data** in form of event logs of previous process executions.

**Contextual data** in form of global attributes like resource utilisation, active cases, etc.

## 3.2 Feature Extraction

Data can originate from a variety of sources. This section discusses various data sources such as event logs or process models, and the features which can be extracted from them.


Figure 3.1: Structure of a *Process Log*. A log contains traces, and a trace contains a sequence of events. (own illustration).

### 3.2.1 Data Sources

In a BPM system, information about its process instances is often available as a process log. Fig. 3.1 shows the structure of a process log. A log contains information about all completed and running process instances (cases). Each instance can be represented as a trace, which consists of a time-ordered sequence of events. Events describe all activities a case has encountered in its life time. All these entities can have additional information in the form of arbitrary properties. Table 3.1 (below) shows an example event log.

Several standards to represent a process log have been proposed. One generally-acknowledged standard is XES, an XML-based standard for event logs. Here some commonly used attributes have been standardised, e.g. IDs, time stamps, and life-cycle phases.[1]

| **Case ID** | **Event ID** | **Timestamp** | **Activity** | **Resource** |
|---|---|---|---|---|
| 1 | 56432 | 2014-03-10 00:38:44.54 | Submit request | Customer #124 |
| 1 | 56433 | 2014-03-10 08:31:22.12 | Decline request | John |
| 2 | 56439 | 2014-03-11 11:15:00.85 | Submit request | Customer #21 |
| 2 | 56440 | 2014-03-11 11:25:58.71 | Answer request (START) | Peter |
| 2 | 56441 | 2014-03-11 11:28:29.33 | Answer request (COMPLETE) | Peter |

Table 3.1: A part of an event log.

Formally an event can be defined as $e$, and characterised by its attributes

1. Günther, C. W. and Verbeek, E., *Extensible Event Stream (XES): Standard Definition v2.0*, 2014, accessed May 1, 2014, `http://www.xes-standard.org/`.

$\#_n(e)$. Here, $n$ denotes the attribute's name, e.g. for the standard extensions the time stamp would be defined as $\#_{time}(e)$, the activity as $\#_{activity}(e)$, the life cycle transition as $\#_{trans}(e)$ and the resource as $\#_{resource}(e)$. Cases are handled analogously to events, they are defined as $c$ and also can have attributes $\#_n(c)$. One special attribute of a case is its trace $\hat{c} = \#_{trace}(c)$, a finite sequence of events recorded for the trace. Each case and event has a unique ID. A complete log is defined as $L$. With these definitions the thesis follows the formalisation standard proposed by Van der Aalst.[2]

In the example given in Table 3.1 the trace of case 1 is defined as $\hat{1} = \#_{trace}(1) = \langle 56432, 56433 \rangle$.

A complete event log acts as a source for **historical data**, while a (partial) trace acts as a source for **case data**. Many features and performance indicators can be extracted from an event log, as discussed in detail in the next section. Such kinds of data are necessary to detect problematic instances at run time and to predict the performance of a case.

**Design data** is necessary to give meaningful recommendations. If a process model is provided, it can provide the *decision space* for a location within the model. The decision space contains all activities which can be legally performed at a given time (legally in respect to the process model). If no process model is provided, it can be mined from historical event data. This often needs manual post-processing to yield in an usable model for a recommendation scenario.

**Contextual data** or global data has to be provided by the respective BPM system or other external sources. This information can be used to improve detection, predictions and recommendations.

### 3.2.2 Performance Indicators

The features extracted from the log and the environment can be used as performance indicators either as-is or as in combination. The three performance dimensions (cost, time, and quality) presented in the last chapter are considered to be conflicting goals. While one can strive to achieve a perfect quality, usually only one other dimension can be fulfilled as well, e.g. achieving a short time-frame, consequently leading to high costs.

In project management, the *triple constraint* or *magic triangle* as visualised in Fig. 3.2, represents this conflicting relationship between the three goals

2. VAN DER AALST, "Business Process Management: A Comprehensive Survey," pp. 98–106.

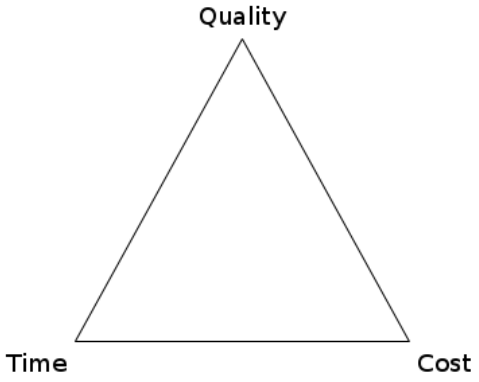



Figure 3.2: The *Magic Triangle* (adapted version).[3]

of cost, time and quality optimisation.[4] This relationship does not only hold true in project management, but is reflected in the basics of applied economics.

Even if monitoring performance indicators of all dimensions, it is often only possible to optimise a decision for one or two of the goals. Thus a trade-off between dimensions has to be made. This can be achieved by either only considering one dimension at a time, or by weighting the dimensions according to a desired outcome. Additionally a risk estimation can help to judge the importance of one dimension at any time.

Usually Key Performance Indicators (KPIs) are defined as performance indicators highly relevant to the success of an organisation. In this approach, a KPI is defined as an arbitrary combination of performance indicators. The weights of the combination can either be user-defined or automatically created, and can be adapted on-the-fly. With this approach, relevant KPIs can be constructed automatically by analysing the input features and outcomes.

3. Wrt. Broy, M. and Kuhrmann, M., *Projektorganisation und Management im Software Engineering*, Xpert.press (Springer Berlin Heidelberg, 2013), p. 5, ISBN: 9783642292897, doi:10.1007/978-3-642-29290-3

4. Hofmann, M., *Performance-orientiertes Projektmanagement: Konzeption zum Umgang mit einmaligen, komplexen Aufgaben*, Unternehmensführung & Controlling (Springer Fachmedien Wiesbaden, 2014), p. 31, ISBN: 9783658047986, doi:10.1007/978-3-658-04799-3.

## 3.3 Decision Support

Decision support in BPM has three dimensions: Detection, Prediction and Recommendation. They are partly dependant on each other, i.e. to give a good recommendation detection and prediction methods can be used.

Business processes themselves can be seen as directed graphs and are often modelled as such. They can be interpreted as Markov Decision Processes (MDPs), where outcomes are partly random and partly under the control of a decision maker. MDPs extend Markov Chains by modelling possible decisions for each state as actions and by adding rewards for the transitions. The respective state transition function depends only on the current state and the action, so that the Markov property holds true. This means it is a memoryless process, and outcomes are independent from previous decisions. By constructing a structure holding all possible states of the process, an optimal solution can be found for MDPs. However, for complex processes this is not feasible due to the state space explosion problem, which results from the exponential growth of the state space with the process size. Furthermore, it is questionable whether the Markov property holds true in reality, i.e. whether states in a real business process can be considered memoryless.

Hence, the approach taken in this thesis uses machine learning methods instead to give predictions and recommendations. Such learning algorithms can utilise the raw features, the resulting performance indicators and the KPIs as defined in the last section. As outlined in Chapter 2, there are three categories of algorithms; supervised learning, reinforced learning and unsupervised learning algorithms.

For this approach supervised learning was chosen due to the well-defined features presented in the last section. This means that a target value can be any feature or performance indicator, or even a combination thereof. The only restriction is that the target variables have to be known in the learning phase.

### 3.3.1 Predictor and target variables

In machine learning, predictor variables are used to predict a target variable. Table 3.2 (overleaf) summarises the available types of data at run-time. Based on those types predictor and target variables can be extracted.

This section discusses the extraction of generic indicators, which should be applicable to any event log. With just a plain event log, case-based data

| Case-based data | Historic data | Design data | Contextual data |
|---|---|---|---|
| Custom trace data, e.g. customer information | Lead time | Process Model | Number of open cases |
| Number of events | Service time | Decision Space | Resource utilisation |
| Number and length of loops | Wait time | | Weekday |

Table 3.2: Examples of available information at run-time in four categories of data.

and historic data can be extracted. Contextual data can be provided by the BPMS or by external sources like traffic or weather forecasts. Some of the statistics described in the next paragraph can be collected either globally for a whole process log, or in reference to attributes like a resource or activity.

Possible candidates are:

**Trace Length** is the number of events in the trace, or $|\hat{c}|$.

**Cycle Time** is the lead time of a case, or the current lead time of a partial trace. It can be calculated as $\#_{time}(\hat{c}(n)) - \#_{time}(\hat{c}(1))$ with $n = |\hat{c}|$.

**Service Time** is the time actually worked on a case. It can be defined as the time spent by activities between the `START` and `COMPLETE` life cycle transitions with the same resource handling the event. For the implementation should be noted that the extraction also should work for nested activities.

**Queue Time** is the time a case spends waiting in a queue, e.g. because no resource is available. Depending on the log, this can for example be defined as the time spent between the `SCHEDULE` and `START` life cycle transitions.

**Wait Time** The time a case spends waiting otherwise, e.g. on customer side.

**Cost** of a case or event. If no information about the cost are available, the service time can be used to estimate costs, e.g. if multiplied by an average resource cost per hour. As discussed in Chapter 1, a variety of cost-based accounting strategies exist.

**Loops** in the trace and their length. A loop is a repeating sequence of activities in a trace. These cam be further narrowed down to sequentially

repeating patterns. A high number of loops might be an indicator for problems in the case.

**Competing cases** are cases which run concurrently to a given case and need to go through the same activity, or be handled by the same resource.

**Outcome** of a case. This definition is highly specific to the use case, but as a nominal value very suitable for a classification task. It is part of the quality dimension.

**Current weekday** In many scenarios, the current day can give an indication about expected cycle time or costs by considering weekends or pubic holidays, and the resource availability on those days.

**Weather** might be an important contextual factor for certain organisations.

**Risk** of a decision. The risk tries to quantify the chance and cost of an error.

While all of the variables are suitable as predictors, it does not make sense to use all of them as target variables as well. Predicting the weather based on historic case data will usually not help in operational decision making scenarios. Therefore the target variables are usually not considered to be contextual or design data, but rather reside in one of the performance dimensions quality, cost and time.

### 3.3.2 Detection of problematic process instances

To detect and highlight problematic process instances at run-time, two kinds of information can be utilised. The prediction of KPIs can offer valuable clues about the condition of a business process. If the projected performance in one or more dimensions lies under a certain threshold, some corrective action needs to be taken to mitigate the bad performance. In previous approaches this threshold often is user-defined and the exception detection triggers an alert in the monitoring system. The current approach tries to set the threshold dynamically by considering historic data, so that, for example, the worst-performing quarter can be used as a limit.

Furthermore, in case of detecting a problematic instance, not just an alert is triggered, but the risk of a decision can be estimated and an appropriate recommendation for a course of action can be given. A naive approach of risk estimation is used for the evaluation. Risk is estimated by the cost of an

erroneous decision multiplied by the probability of an error. Additionally, the confidence of the recommendation can be calculated.

Based on the estimated risk and the recommendation confidence it can be determined whether an automation would be beneficial to the situation. Based on these two factors there are four possibilities. If the risk is high and the confidence low, the system will not take automated action, but rather notify a human operator (automation level 2, the computer offers a set of alternatives). This case supports the decision maker by detecting a problematic instance which requires human attention. If both the risk and confidence are high, other factors have to be taken into consideration as well. Those factors could include a pre-defined threshold for risk, or the number of risky decisions taken in recent time. In any case, a decision maker should be notified, possible together with a recommendation for further action (automation level 3, selection narrowed down to a few).

With low-risk decisions the system has more room to operate autonomously. When there is low risk and high confidence, the system may act automatically and inform the human decision maker if asked for a summary (automation level 8). A decision task combining low risk and low confidence is another border case like described in the high risk and high confidence scenario. It boils down to the willingness of the process owner to take risks, and depends on the concrete use case.

### 3.3.3 Prediction of performance indicators

For the prediction of performance indicators both linear regression and a multilayer perceptron (feedforward neural network) are evaluated. Both types are supervised learning algorithms and can be used to establish a mapping between a vector of input features $\vec{X}$ and a result $Y$, which aims to be valid for a given training set. Such algorithms assume that the prediction of future variables depends on their historical behaviour. Such an assumption may not be valid for long-term developments, but is sufficient for short-term predictions.

### 3.3.4 Classification of process instances

The classification of cases can be done in regard to a nominal performance indicator. This indicator may be constructed artificially, e.g. the outcome may depend on the existence of certain activities in the case. The current approach compares the decision tree implementation J48 (based on C4.5) and multinominal logistic regression (One-vs-All) for such classification.

## 3.4 Business Process Simulation

The learning algorithms introduced in the last chapter need a training data set to learn either the regression coefficients, the decision rules or to construct the underlying neural network. To generate such a training data set, BPS can be utilised. As described in Chapter 2, a simulation consists of a process model and a simulation environment. To integrate recommendations, the environment has not only to provide probabilities for certain paths, but also a mechanism to influence the decisions made in the process flow by issuing recommendations.

On the other hand a simulation model should aim to reflect the reality as good as possible. To evaluate the current approach based on a real life setting, a real event log is chosen, analysed and eventually used as a simulation model. In a first step, this model is then used to generate event logs which serve as training data and as an evaluation benchmark. In a second step, the simulation is conducted with a recommendation-enabled environment, and the event logs are used to compare the performance indicators with the original simulation.

To create a simulation model from real life event logs, some preparations have to be undertaken. The event log needs to be of sufficient quality to be able to extract a process model, e.g. by means of process mining. Based on this model the work-flow of the simulation can be designed and the transition probabilities can be determined. An simulation model should abstract the internal details of a business process, and solely rely on stochastic distributions. Due to those challenges, real processes often can only be modelled in an approximate fashion. The key characteristics of the original process should be reflected in the simulation model and output to be of any use for further analysis.

For the current approach an event-based simulation is used. Possible entities are business cases and resources. A case holds all information associated to this case, much like the structure of a case in an event log; and a resource holds information about its name, group and organisation. A business event implements an activity from the event log, and it is always associated with a case and optionally with a resource. Due to these similarities to the event log format introduced at the beginning of this chapter, such a simulation can easily produce valid event logs, which can be stored for further analysis with external tools like process mining suites.

## 3.5 Integrated Architecture

This section illustrates the integration of the three components: The feature extraction discussed in Section 3.2, the decision support system described in Section 3.3 and the simulation described in Section 3.4.

Fig. 3.3 (below) shows the initial data collection phase. The simulation model is used with a simulation environment to conduct simulation runs. These runs are based on probabilities as defined in the model, and produce event logs. The monitoring component in the environment registers these events, calculates the performance indicators at each step, and stores the collected event logs. The data collected here acts as benchmark for the evaluation of the decision support system.

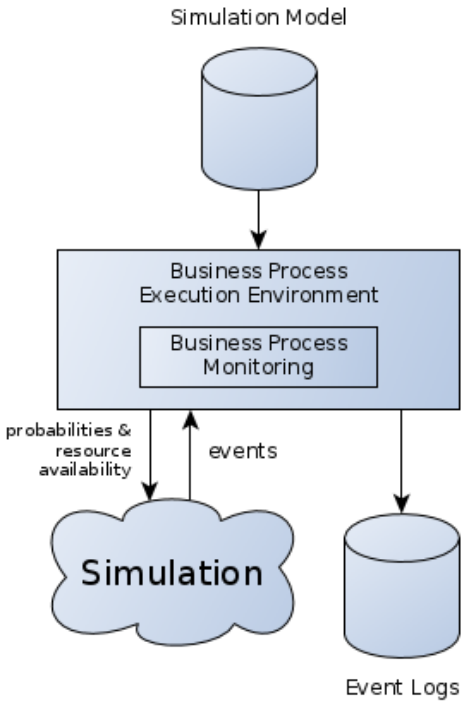



Figure 3.3: Architecture of the benchmark simulation and initial data collection (own illustration).

Fig. 3.4 (overleaf) shows the integration of a decision support system into this architecture. While the internal mechanisms of the execution environment are not changed, the simulation is not based solely on probabilities anymore, instead the control flow can be actively controlled by the given recommendations. To enable recommendations the architecture is extended by a decision support component. This decision support system receives

global data, design data (decision space), and contextual data from the environment. At the same time the support system has access to the historical event logs stored from the first phase, which are used to train the learning algorithms.

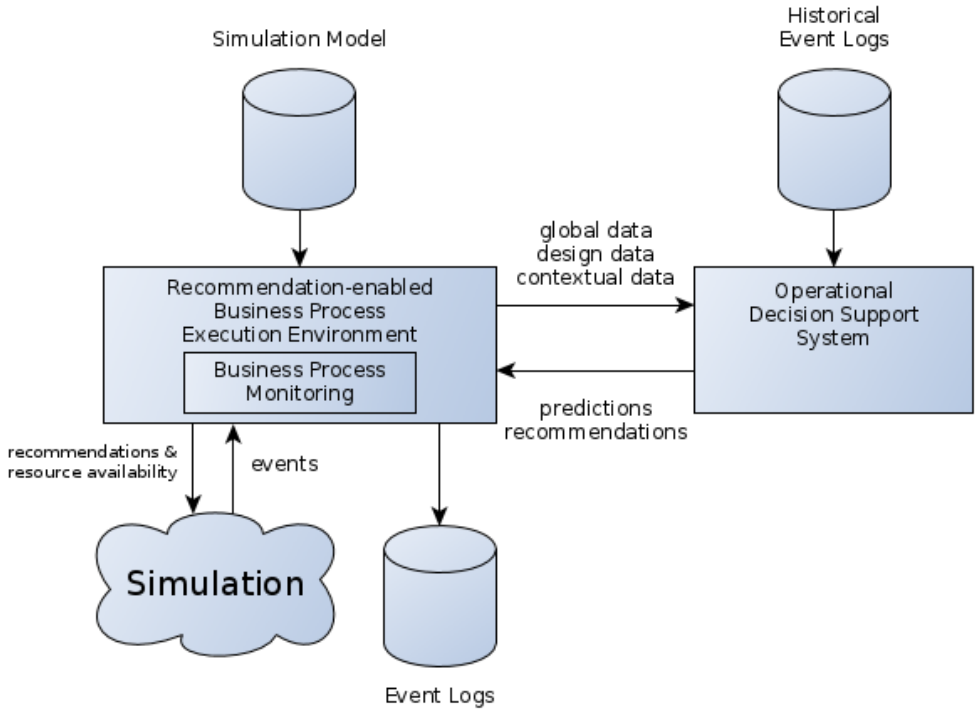



Figure 3.4: Architecture of the recommendation-enabled simulation (own illustration).

## 3.6 Shortcomings and Assumptions

Possible shortcomings of this approach include the manual feature definition, which is use-case specific. With the generic features discussed in this chapter alone, it might not be possible to make good recommendations for specific use cases. Furthermore, the weighting of the factors and performance dimensionalities has to be determined manually as well. This could be improved by using an automated optimisation algorithm, which is able to find optimal parameters for a given problem. The risk and cost estimation is based on naive approaches, but can easily be exchanged with more sophisticated alternatives.

In regards to the simulation, the proposed model will not accurately reflect reality, but rather approximate it roughly. Also it is assumed that a given

process log can be transformed in a simulation model. This might not always hold true, especially if the existing event logs are of bad quality. To solve this problem, the automated extraction of a simulation model could be utilised in future approaches.

# Chapter 4

# Implementation

This chapter describes the implementation of the real-time decision support approach in BPM environments as discussed in the last chapter.

It starts with a survey of event logs in the public domain, with the goal of selecting a log suitable for the evaluation of the decision support approach. An event log (or process log) contains traces of one or more cases (or process instances). A trace consists of a sequence of events usually collected by an information system.

After a suitable log is selected, an analysis of the log is conducted to gain a deeper understanding of the data set. Based on this analysis a simulation model is creates which then is implemented in a Java-based simulation framework. To apply the recommendation approach to this simulation, a small business process execution and monitoring system is implemented, which coordinates the experiment and collects data for the evaluation in the next chapter.

## 4.1 Data Survey

In the following paragraphs three types of sources for finding a suitable event log are discussed, and in the next sections concrete examples of those are described in detail.

A starting point to find event logs are the homepages of process mining tools such as *ProM*[1] or *Disco*.[2] ProM is an open-source framework for pro-

1. Cf. *ProM 6.3: Description and Example Logs*, 2013, accessed April 19, 2014, `http://www.promtools.org/prom6`.

2. Cf. Rozinat, A., *Disco User's Guide*, 2012, accessed April 19, 2014, `http://`

cess mining developed primarily by the Process Mining Group of the Eindhoven Technical University. Many other authors contributed plug-ins (currently about 500 available plug-ins[3]) and other improvements to the project. Disco is a proprietary process mining tool developed by Fluxicon Process Laboratories, a company started by two PhDs graduates of the Eindhoven Technical University. Under their Academic Initiative, Fluxicon offers free academic licenses to certain tertiary students. While ProM only provides the user with simple example logs to aid them in learning the basics of the tool, Disco offers two more sophisticated process logs for the same reason. One set to guide the user through the learning process, and a larger one derived from real process logs to explain the advanced features of their tool.

A second source of event logs are books about Process Mining. The book *Process Mining: Discovery, Conformance and Enhancement of Business Processes* written by Wil van der Aalst[4] describes some logs, most of which are available online.[5] The majority of these logs serve an educational purpose, e.g. they are only adequate to explain and show process mining algorithms and their shortcomings. One promising artificial data set introduced by van der Aalst (a peer-review process) is also found in the next source and is discussed in the following section.

Another great source of data is the data repository of the 3TU.Federation,[6] a network of three technical universities in the Netherlands: Delft University of Technology (TU Delft), Eindhoven Technical University (TU/e) and the University of Twente. This repository stores data sets originating from technical and scientific research in the Netherlands. The data is publicly available and can be used, among other things, for scientific research. 3TU hosts a collection of event logs published by the IEEE Taskforce on Process Mining, with both synthetic and real data sets.[7]

`fluxicon.com/disco/files/Disco-User-Guide.pdf`.

3. *ProM 6.3: Description and Example Logs*.

4. van der Aalst, *Process Mining: Discovery, Conformance and Enhancement of Business Processes*.

5. van der Aalst, W. M., *Event logs and models used in Process Mining book*, 2011, accessed October 12, 2013, `http://www.processmining.org/event%5C_logs%5C_and%5C_models%5C_used%5C_in%5C_book`.

6. *3TU Datacentrum*, 2014, accessed April 19, 2014, `http://data.3tu.nl/repository`.

7. IEEE Task Force on Process Mining, *IEEE Task Force on Process Mining - Event Logs*, accessed April 19, 2014, `http://data.3tu.nl/repository/collection:event%5C_logs`.

### 4.1.1 Real Life Event Logs

The data discussed in this and the next section comes from the 3TU repository. This section introduces three event log sets published for the yearly Business Process Intelligence Challenge (BPIC). This challenge is held in conjunction with the International Conference on Business Process Management and invites participants to analyse a set of real-life event logs focusing on one or more questions provided by the original process owner.

#### *4.1.1.1 BPIC'13: Volvo IT Support (VINST)*

**Description** The set of event logs was provided by Volvo IT for the BPIC 2013. It is divided into three subsets, two coming from their problem management system[8] and one from their incident management system.[9] This combined information system is called VINST. Within the process, two separate organisational units exist.

**Content** The incident management event log contains 7,554 cases and a total of 65,533 distinct events, and the problem management system contains 2,306 cases and 9,011 events over a time span of roughly 2 years, although some outlier cases are active for a much longer period (4-6 years). Cases start with an automated or manual incident/problem report, and they end with a resolution or closure of the case – except for the open problems set, which contains events without resolution.

**Questions** The process owner was interested in four questions:[10]

1. Push to Front (incidents only): Is there evidence that cases are pushed to the 2nd and 3rd line too often or too soon?
2. Ping Pong Behavior: How often do cases ping pong between teams and which teams are more or less involved in ping-ponging?
3. Wait User abuse: Is the *wait user* substatus abused to hide problems with the total resolution time?

8. Steeman, W., *BPI Challenge 2013, closed problems*, 2013, doi:10.4121/uuid:c2c3b154-ab26-4b31-a0e8-8f2350ddac11; Steeman, W., *BPI Challenge 2013, open problems*, 2013, doi:10.4121/uuid:3537c19d-6c64-4b1d-815d-915ab0e479da.

9. Steeman, W., *BPI Challenge 2013, incidents*, 2013, doi:10.4121/uuid:500573e6-accc-4b0c-9576-aa5468b10cee.

10. Cf. original data set description Volvo IT, *VINST data set*, 2012, accessed April 19, 2014, http://www.win.tue.nl/bpi/%5C_media/2013/vinst%5C_data%5C_set.pdf.

4. Process Conformity per Organisation: Where do the two organisational units differ and why?

**Performance Indicators** One general indicator is the total flow time (resolution time). Other indicators tailored to the questions are the push to front policy conformity (cases handled by the 1st level support), number of handovers between support teams (ping-pong behaviour) per product or team, the correlation between number of handovers and resolution time or life time so far, and finally the time spent in the *Wait-user* status as indicator for possible abuse cases.

**Findings** In respect to the original questions, participants of the challenge came to the following conclusions. The overall Push-to-back ratio is normal, but prominent with certain teams and products. More contextual data is needed to determine causes of this behaviour (e.g. busy periods, incident types). The log exposed different types of ping-pong behaviour, *linear* and *circular* ping-pongs. Finally, there are differences in the process flows of the two organisational units.[11]

**Problems** The case description offers only little information about strategy, goals or expectations. This makes a qualitative judgement impossible, the only way of judging a team is based on another teams performance. Also there is no information about the process or product, which makes it hard to determine causes of the observed behaviours.

**Suitability for this thesis** Promising – The process model is not overly complex. The original questions do not directly relate to decision support scenarios, but the identified performance indicators allow a broad monitoring. Previous findings are mostly irrelevant for such a scenario, however the missing context could prove to be a challenge.

Fig. 4.1 and Fig. 4.2 depict process models of the open and completed process logs in a rather coarse resolution. Uncommon activities and paths have been removed, and the numbers are absolute frequencies.

#### *4.1.1.2 BPIC'12: Dutch Financial Institute*

**Description** This set of event logs was provided by an anonymous Dutch financial institution for the BPIC 2012. Each case represents a per-

11. Cf. Bautista, A. D. et al., "Process Mining in Information Technology Incident Management: A Case Study at Volvo Belgium" (2013); Kang, C. J. et al., "Process Mining-based Understanding and Analysis of Volvo IT's Incident and Problem Management Processes" (2013).

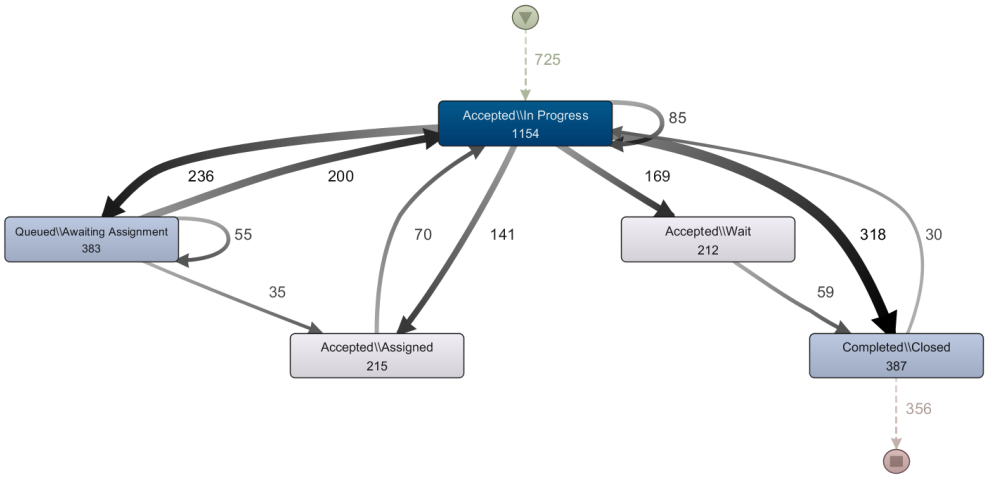



Figure 4.1: Process model extracted from the open VINST logs (own illustration).

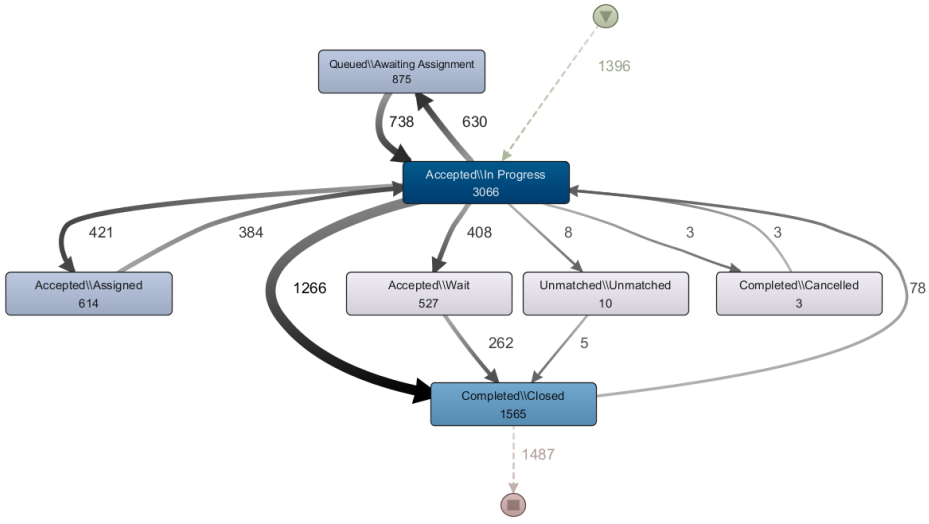



Figure 4.2: Process model extracted from the completed VINST logs (own illustration).

sonal customer loan or overdraft approval process.[12]

**Content** The event log contains 13,087 cases (applications) and some 262,200 events. Three sub processes (Application, Offer, Work) were merged to create this log. A trace always starts with the customer submitting an application through a webpage. This event contains the amount of

12. VAN DONGEN, B., *BPI Challenge 2012*, 2012, doi:10.4121/uuid:3926db30-f712-4394-aebc-75976070e91f.

money requested by the customer. Traces end with a decision regarding the application (Accept/Reject/Cancel).

**Questions** The process owner was interested in the following questions:

1. Are there estimators for the total lead time?
2. Which resources generate the highest activation rate of applications?
3. What does the process model look like?
4. Which decisions have a significant influence on the process flow?

BPIC participants could choose to analyse the process as a whole, or only one of the three sub processes.

**Performance Indicators** Applicable performance indicators are lead time (total flow time), resource utilisation in terms of wait time (i.e. waiting for a customer response) vs. work time, and resource efficiency (time).

**Findings** Four main conclusions can be drawn. The first finding is that automated application cancellations can occur earlier than the established default of 30 days. Secondly, through event level data, insights into resource performance can be gained. The resource deployment in the organisation as captured by the event logs is not optimal, so called *specialists* can work more efficiently then all-rounders. And lastly, the use of decision/classification trees can help with work prioritisation early in the process.[13]

**Problems** Extensive pre-processing is necessary to reduce the overall complexity of cases (4000+ variants on a process with just 6-7 key steps). This includes the removal of redundant business events and concurrent events. The pre-processing requires business judgement to further reduce the complexity. Additionally, missing context information makes an interpretation of the event log regarding to the original questions hard. For example, the data set contains no (additional) information about customer demographics, or a customer history.

**Suitability for this thesis** Good – The process model can be simplified by omitting redundant events and by concentrating on one of the three

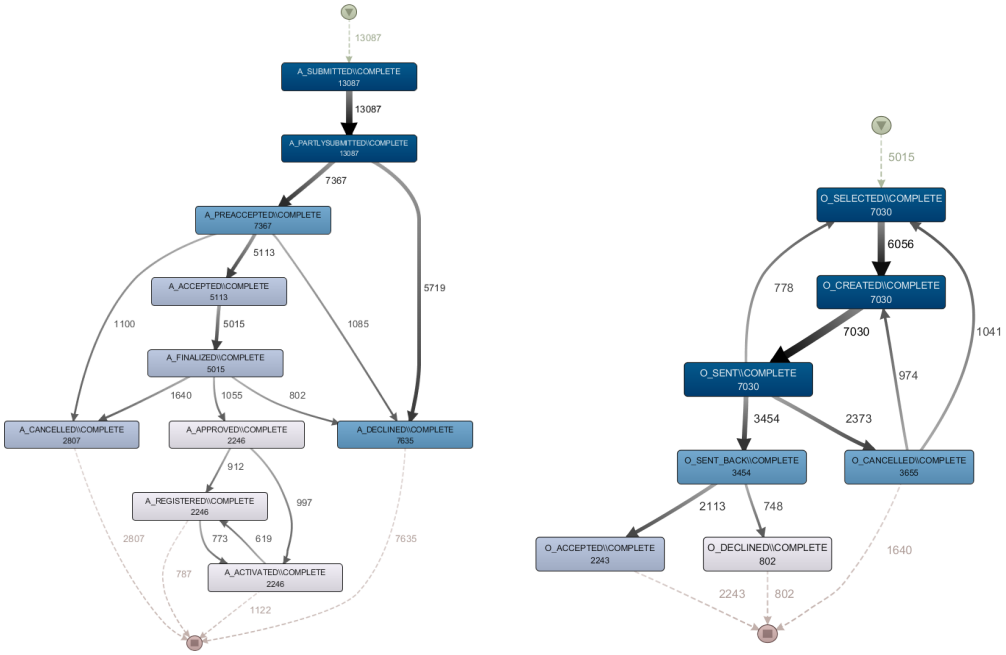
13. Cf. Bautista, A. D., Wangikar, L., and Akbar, S. M. K., "Process Mining-Driven Optimization of a Consumer Loan Approvals Process" (2012), 1–26; Molka, T., Gilani, W., and Zeng, X.-J., "Dotted Chart and Control-Flow Analysis for a Loan Application Process" (2012).

sub processes. Related approaches have shown that a decision support scenario is feasible, e.g. for prioritisation early in the process.

Fig. 4.3a shows the process model for the application sub process. While all cases are represented in this sub process, it only contains 23% of the total events.

Fig. 4.3b shows the model for the offer sub process. Since not all loan applications receive get an offer, only 38% of the cases are represented in this sub process, and it contains 11% of the events.

The work item sub process is not pictured here, since it is more complex than the previously mentioned sub processes and needs simplification to be displayed graphically. Roughly three quarters (73%) of the cases are represented in this sub process, and it contains 64% of the events.




(a) *Application* sub process (b) *Offer* sub process

Figure 4.3: Process models for two sub processes of a loan application process at a Dutch financial institution. (own illustration).

### *4.1.1.3 BPIC'11: Dutch Academic Hospital*

**Description** The last real life set of event logs was provided by a Dutch academic hospital for the BPIC 2011. Each case represents a patient of

a Gynaecology department and contains information about activities and their organisational units.[14]

**Content** The event log contains 1,143 cases (care-flow of patients) and a total of 150,291 distinct events. Cases do not have a common start or end event.

**Questions** The original process owner did not state any specific questions, the participants were encouraged to focus either on a specific aspect of interest in detail or a broader analysis considering more aspects.

**Performance Indicators** Possible indicators are total flow time, or number of cycles in a case. Due to the highly personalised nature of each treatment, however such indicators are not widely applicable.

**Findings** Obtaining a streamlined flow model is hard, but possible if the analysis is reduced to the organisational dimensions and by pre-processing the raw event data. Furthermore, the trace alignment technique has proven to be suitable for mining a process model, and segmenting the patients based on the urgency of their illness.[15]

**Problems** The obtained process model is a *Spaghetti-like process model* (Bose 2013) with many nodes and possible flows, which is is not easily comprehensible for humans. This is illustrated in Fig. 4.4 (overleaf).

**Suitability for this thesis** Poor – Due to the complex structure it is hard to obtain a streamlined process model, and consequently a complete simulation model.

### 4.1.2 Synthetic Event Logs

This section introduces synthetic event logs available from the 3TU data repository. Unlike the real life event logs discussed before, the description of questions has been omitted here.

14. van Dongen, B., *Real-life event logs - Hospital log*, 2011, doi:10.4121/uuid:d9769f3d-0ab0-4fb8-803b-0d1120ffcf54.

15. Cf. Bose, R. P. J. C. and van der Aalst, W. M., "Analysis of Patient Treatment Procedures" (2011), doi:10.4121/uuid; Caron, F. et al., "Beyond X-Raying a Care-Flow: Adopting Different Focuses on Care-Flow Mining" ().

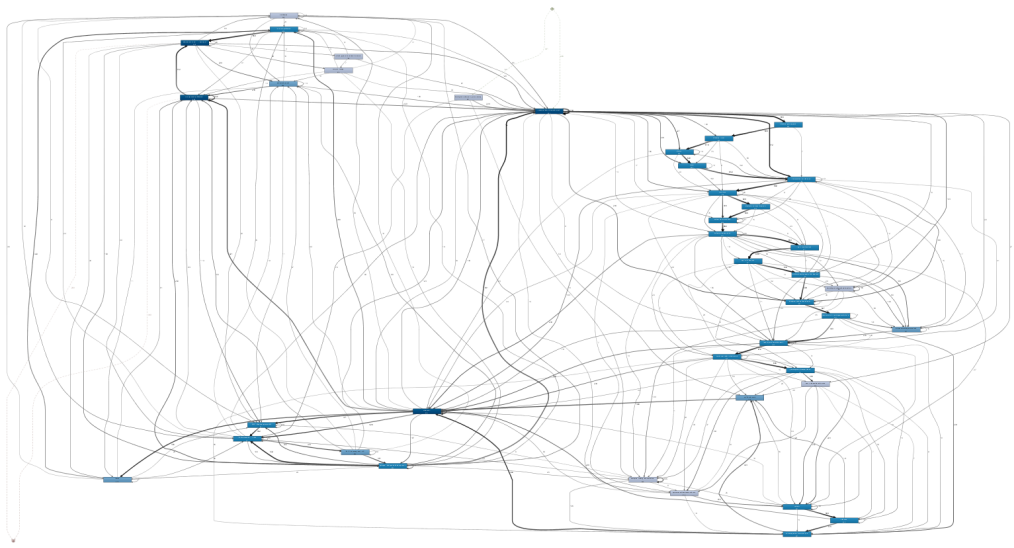


Figure 4.4: Spaghetti-like process model of a Dutch Academic Hospital created from non-pre-processed event data. Despite the complexity, it contains only the 6% most-frequent activities and 20% of the paths (own illustration).

#### *4.1.2.1 Artificial Digital Copier*

**Description** This artificial event log was created to test a new two-step approach in data mining utilising a *fuzzy miner*.[16] The log describes the workflow of a simple digital photo copier supporting various operations; photocopying, scanning, and printing. Additionally the scanned documents can be sent via email or FTP upload. The event logs were generated by simulation and contain very detailed internal logs.[17]

**Content** The event log contains 100 cases and a total of 35,000 distinct events. Cases always start and end with a common event.

**Performance Indicators** Although not the focus of the original research, a possible indicator is the total flow time.

**Findings** To remove the low level details an analyst might not be interested in, a two-step approach is suitable. In the *Pattern Abstraction* plugin developed for ProM, firstly patterns are discovered, then filtered, which leads to a selection of abstractions. Based on these abstractions, the log can be transformed to a representation containing only

16. See Bose, R. P. J. C., Verbeek, E. H. M. W., and van der Aalst, W. M. P., "Discovering Hierarchical Process Models Using ProM," in *CAiSE Forum 2011* (London, UK, 2012), pp. 33-35.

17. Bose, R. P. J. C., *Artificial Digital Photo Copier Event Log*, 2011, doi:10.4121/uuid:f5ea9bc6-536f-4744-9c6f-9eb45a907178.

interesting parts.

**Problems** The log is very detailed and mostly consists of internal nodes (decisions made by the printer's software), which are unambiguous and do not require decision support.

**Suitability** Poor – There are no clear decision support tasks.

#### *4.1.2.2 Loan Application Example*

**Description** This event log was created to evaluate new process mining techniques.[18] There are four different configurations of the process, however they only contain dummy events without business context.

**Content** The event log contains 100 cases and a total of 590 distinct events. Cases always start and end with a common event.

**Performance Indicators** Without any contextual information, only generic indicators such as lead time can be considered.

**Findings** The log enabled new insights into process mining techniques.

**Problems** None – The original research was not examined in detail since the logs are not suitable for this thesis and new process mining techniques are not in the scope of this thesis.

**Suitability** Poor – The logs do not contain enough contextual information to be of any use for decision support scenarios. Since the logs were created to evaluate new process mining techniques, and are therefore tailored to model creation problems, this does not come as a surprise.

Figure 4.5 (overleaf) shows a process model generated from this data set. The missing business context can be clearly seen by the anonymous labels. The numbers depict the absolute frequency of occurrences.

#### *4.1.2.3 Review Example*

**Description** This event log set contains traces of a peer review process.[19] It was used to evaluate a new approach for predicting lead times of

18. Buijs, J., *Loan application example*, 2013, doi:10.4121/uuid:bd8fcc48-5bf3-480e-8775-d79d6c700e90.

19. van der Aalst, W. M. P., *Synthetic event logs - review example large.xes.gz*, 2010, doi:10.4121/uuid:da6aafef-5a86-4769-acf3-04e8ae5ab4fe.

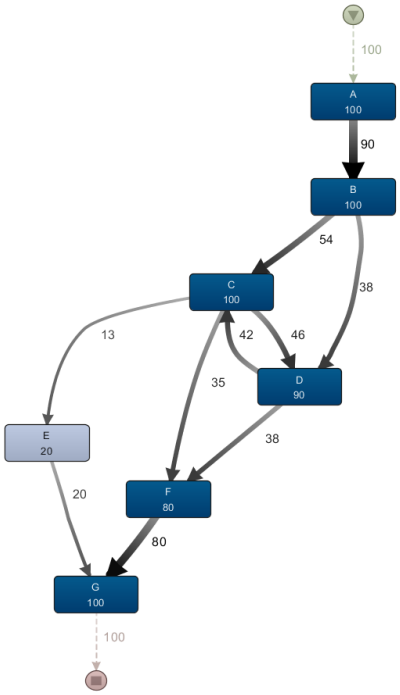



Figure 4.5: Process model extracted from configuration 1 of the artificial loan application logs (own illustration).

processes.[20]

**Content** The event log contains 10,000 cases and about 150,000 distinct events. A case (a review process) always starts with inviting reviewers. Those reviewers can either return their review, or decide not to answer. When all reviews are collected, a decision is made whether to accept or reject the paper.

**Performance Indicators** The original research focused on the total flow time, since the goal of the approach was to predict the total flow time. Alternatively, the number of exception (i.e. review time-outs) could be used on a resource basis.

**Findings** By obtaining a *transition model* of the abstracted process logs, good predictions are possible.

**Problems** The original research does not state any problems associated with this event log.

**Suitability** Promising – A possible decision support scenario could include

20. van der Aalst, Schonenberg, and Song, "Time Prediction Based on Process Mining," p. 20.

which reviewer to invite in the first place, e.g. to filter out those who are not likely to answer, or who would take too long to complete their review.

For this thesis, this data set presents the most promising set of all synthetic logs introduced in this section. Figure 4.6 shows the process model generated from this data set. As in the previous example, the numbers depict the absolute frequency of occurrences.

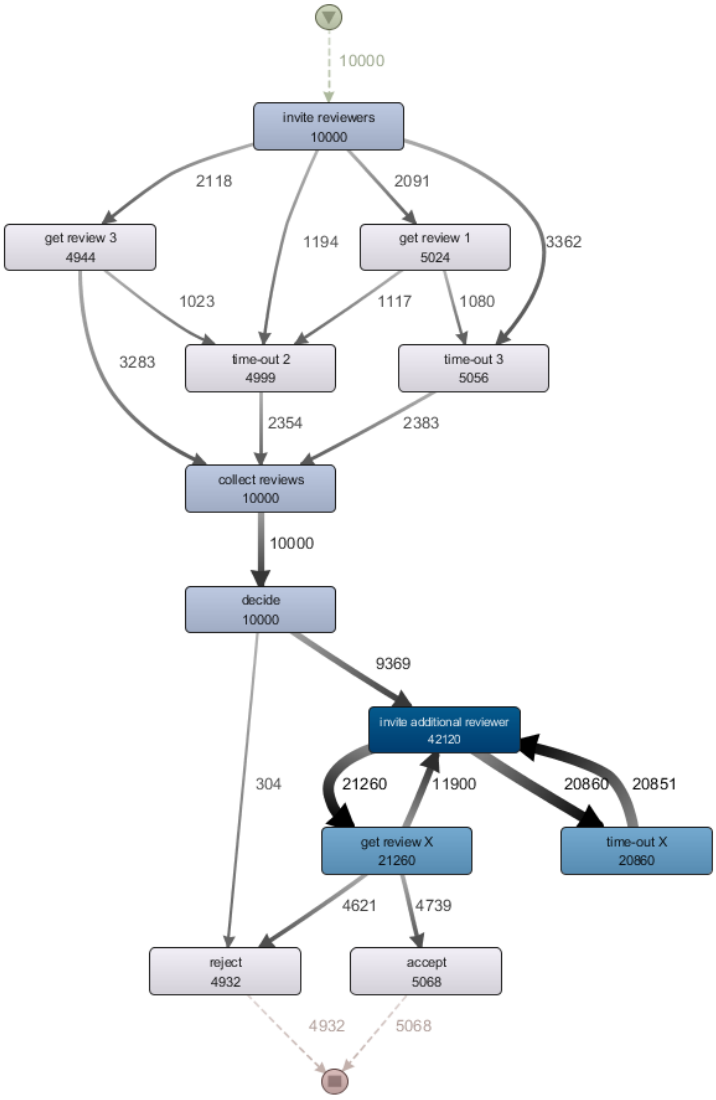



Figure 4.6: Process model extracted from the artificial review example logs (own illustration).

### 4.1.3 Conclusion

All event logs discussed are only available in raw form, so before they can be used to design a simulation model they need to be converted to a process model by means of process mining.

The real-life event logs need at least some, and in certain cases extensive, pre-processing in order to conduct a traditional analysis. This poses a challenge for abstract run-time analyses. Furthermore no real-life logs explicitly define exceptional cases, they are only detectable by observation of behaviour and deduction of a standard process model. Once the logs have been processed however, they offer potential to be used in a real-time test scenario.

The BPIC 2013 event log is promising due to its simple structure. However, missing contextual information make it hard to utilise for a decision support system, which could only concentrate on generic performance indicators like lead time. This event log would profit more from an organizational analysis.

The BPIC 2011 event log is not suitable, since it produces spaghetti-like process models which are extremely hard to comprehend.

The most promising real-life data set is the BPIC 2012 event log described in Section 4.1.1.2, containing traces of personal loan applications. Seen as a complete model, it is still very complex, but has the advantage of being composed of three sub processes. When focusing at one of the sub processes, the model gets considerably more manageable.

Most of the synthetic event logs have been developed to test new process mining techniques. This and the lack of business context make them less usable in a real-time scenario. The only promising event log is the review example, which can be simplified to a clearly arranged model with a couple of decision tasks. However, missing contextual information make it hard to utilise.

## 4.2 Process Model

The data from the BPIC 2012 (logs of financial institution) is the best candidate for the evaluation of a decision support system. The event logs

are available in Mining eXtensible Markup Language (MXML)[21] and XES[22] formats, which all modern process mining suites can read.

To analyse and visualise the event logs the process mining tools *Disco*[23] and *ProM* have been used. The mathematical language *GNU Octave* and the Java library *OpenXES* were also helpful to extract and analyse the necessary data.

### 4.2.1 Analysis

The event log is from a Dutch financial institute, containing events related to an application process for a personal loan or overdraft. The workflow of a successful application is described as follows:

> An application is submitted through a webpage. Then, some automatic checks are performed, after which the application is complemented with additional information. This information is obtained through contacting the customer by phone. If an applicant is eligible, an offer is sent to the client by mail. After this offer is received back, it is assessed. When it is incomplete, missing information is added by again contacting the customer. Then a final assessment is done, after which the application is approved and activated.[24]

Each trace has the global attributes `AMOUNT_REQ`, the amount of money requested by the customer, and `REG_DATE`, the date of application. Three separate sub processes can be identified in the data. The log contains 24 distinct event classes, which are prefixed according to their respective sub process with `A_` for states of the application (10 states), `O_` for states of the offer belonging to an application (7 states) and `W_` for states of work items (7 states). Work states support three life cycle phases: `SCHEDULE` when the work item is created in the queue, `START` when it is obtained by a resource and `COMPLETE` when it is released by the resource.

An application ends with one of three possible outcomes. As described in the successful scenario above the first possible outcome is the *approval* of a

21. Verbeek, H., *Mining eXtensible Markup Language (MXML): Definition*, 2011, accessed April 22, 2014, `http://www.processmining.org/logs/mxml`.

22. Günther and Verbeek, *Extensible Event Stream (XES): Standard Definition v2.0*.

23. *Disco Process Mining Tool*, accessed April 14, 2014, `http://www.fluxicon.com/disco`.

24. *BPIC 2012 Event Log Description*, 2012, accessed April 26, 2014, `http://www.win.tue.nl/bpi/2012/challenge`.

loan application, which only happens after the initial application has been completed, an offer sent to the customer and the application passed a final assessment. Alternatively an application can be *declined* at any point in the process. The last alternative is the *cancellation* of an application, which again can happen at any time. The data indicates that this is often triggered automatically after a certain time-out period, e.g. 30 days without activity after sending an offer to the customer or requesting more information from the customer. Since the most recent event is not necessarily indicative of the outcome, determining the actual outcome of a case requires analysing the whole trace. The order of events may differ depending on the resource working on the case or due to possible concurrency in the workflow. For the remainder of this analysis it is assumed that a case containing the `A_ACTIVATED` event was accepted, a case containing `A_DECLINED` was declined and a trace containing `A_CANCELLED` was cancelled.

About 60% of all applications are declined in the process. Of those, 45% are declined immediately by automated checks, and another 28% are declined immediately after an employee has looked into the case. 22% of all applications are cancelled, the majority of these cancellations occurs 31 days after the initial submission. This is probably due to aforementioned automated time-out of applications in the system. Successful applications represent the remaining 18% of cases.

It is not easy to obtain a fitting and comprehensible process model from the raw data. The event logs need extensive pre-processing and manual enhancement. Some of the problems (and additional insights into the data) have been described in the original submissions of the BPIC 2012.[25] Some of the submissions concentrated on the mining aspect and were able to extract a model with a very good fit to the data.[26] In this thesis however the goal is to extract a model suitable for further simulation, and not necessarily a perfect process model. As such the event log seems to be suitable for an evaluation.

### 4.2.2 Simplification

While the analysis in 4.2.1 gives a first insight into the process, it is still necessary to extract a streamlined process model suitable for simulation. The

25. Bautista, Wangikar, and Akbar, "Process Mining-Driven Optimization of a Consumer Loan Approvals Process"; Molka, Gilani, and Zeng, "Dotted Chart and Control-Flow Analysis for a Loan Application Process."

26. Adriansyah, A. and Buijs, J., "Mining Process Performance from Event Logs" (2012).

structure is complex and it is hard to extract concrete probability distributions. The event logs itself still hold some more potential for simplifications, but this makes it necessary to manually pre-process the event logs to obtain a simpler model. In this sections further simplifications applied to the event log are described.

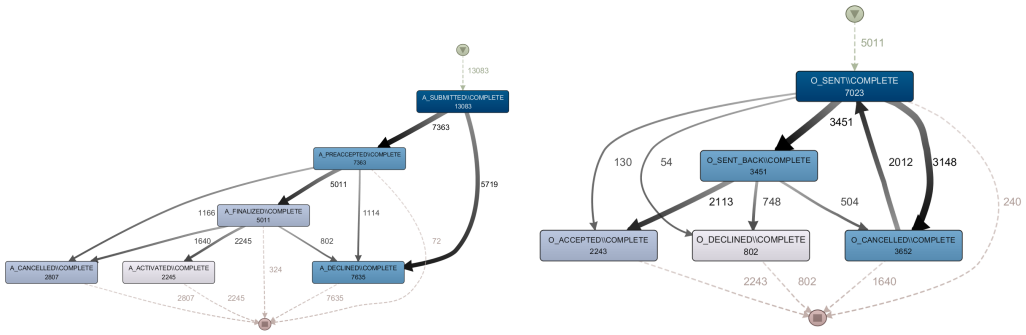
To start with some of the cases in the log are not completed yet and can be removed. Any case with an unknown outcome – i.e. all cases which do not contain at least one of the events `A_ACTIVATED`, `A_DECLINED` or `A_CANCELLED` – have been discarded for further analysis. This leaves 12,688 cases and 249,451 events (roughly 95% of the original data set).

| **Event (Lifecycle)** | **Description** |
|---|---|
| A-Events | |
| `A_PARTLYSUBMITTED` (complete) | Always follows `A_SUBMITTED`. |
| `A_ACCEPTED` (complete) | Usually precedes `A_FINALIZED`. |
| `A_APPROVED` (complete) | Usually precedes `A_REGISTERED`. |
| `A_REGISTERED` (complete) | Usually precedes `A_ACTIVATED`. |
| O-Events | |
| `O_SELECTED` (complete), `O_CREATED` (complete) | Usually precedes `O_SENT`. |
| W-Events | |
| `W_*` (schedule) | Schedule events are automatically created. While important for calculating assignment times or the likes, they do not represent a direct decision task. |

Table 4.1: Pre-processing of the original event logs: Event reduction.

Furthermore it can be seen that some activities are redundant, e.g. `A_PARTLYSUBMITTED` always follows `A_SUBMITTED`. Additionally all events associated with the `SCHEDULE` life-cycle transition can be ignored, because those are created automatically when a work item has been released by a resource. Table 4.1 summarises all events which have been removed from the original log. Those redundant events total 60,184 events, and after removing them from the log 189,267 events remain (72% of the original events in the data set). Of the 24 distinct original states only 17 states remain. This allows a

clearer view into sub processes A and O as shown in Fig. 4.7 (overleaf).




(a) Simplified *Application* sub process (b) Simplified *Offer* sub process

Figure 4.7: Simplified sub process models (own illustration).

The complete model still cannot easily be represented comprehensibly. To gain a better understanding of the process structure, separate streamlined process models for the three outcomes can be extracted. These models exclude rarely activated activities and exceptional cases.

### *4.2.2.1 Successful application*


Figure 4.8: Timeline of the accepted applications (own illustration).

This stream represents the normal execution of a loan application process an has been roughly outlined by the process owner (cf. last section). For a successful application three phases can be identified. The description of the phases contains more details than the original workflow description, e.g. fraud checks, calls to the customer, and other details omitted in the process owner's description. Every contact attempt made with the customer is usually repeated until contact was made eventually (e.g. by calling at different times of the day). The description contains the name of the respective work item, while application and offer states are taken straightforwardly from Fig. 4.7.

**Initial phase.** The submission is processed automatically and either pre-approved or declined. Pre-approved applications are checked by an employee (`W_Afhandelen leads`). An exception is made if the organization suspects a fraudulent intent, in this case the application

is assessed for fraud by a specialist (`W_Beoordelen fraude`). After completing an application, it is either declined or an offer is created. If an offer was created, the *offer phase* begins.

**Offer phase.** This offer is then sent to the customer, and the staff tries to contact the customer via phone (`W_Nabellen offertes`). Sometimes the offer is cancelled after the conversation and a new one is created and sent to the customer, presumably due to negotiations on the phone. Eventually the offer is returned by the customer, which completes the application (`W_Completeren aanvraag`) and the *validation phase* starts.

**Validation phase.** The goal of this phase is to finalise the application. Firstly the existing information is validated (`W_Valideren aanvraag`), and if some data is missing the customer is contacted (`W_Nabellen incomplete dossiers`). When all necessary information is collected, the application is activated.

The *fraud check* is only performed on 107 cases (less than 1% of the data set), and will be neglected in the simulation. The work item *Completion of application* (`W_Completeren aanvraag`) seems to be a composition of several sub-tasks, basically covering the first two phases: Initial contact (`W_Afhandelen leads`), fraud assessment (`W_Beoordelen fraude`), and customer contact (`W_Nabellen offertes`). By absolute occurrences, contacting the customer is the most performed task in the process.

An illustration of the streamlined model for successful applications can be found in the appendix (Fig. B.2).

#### *4.2.2.2 Declined application*

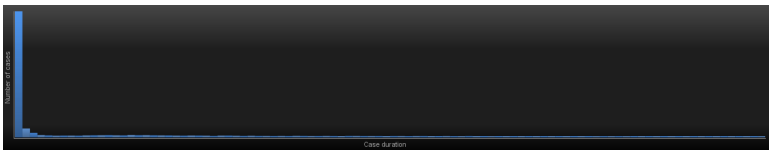


Figure 4.9: Timeline of the declined applications. Most applications are declined early in the process (own illustration).

Declined applications generally follow the same three-phase schema as successful applications, but the application can be declined by an employee at any point of time. Nonetheless, 60% of the declined applications are declined in the first phase.

An illustration of the streamlined model for declined applications can be found in the appendix (Fig. B.3).

#### *4.2.2.3 Cancelled application*

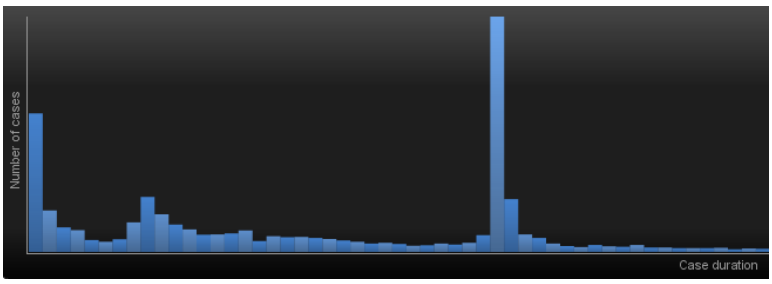



Figure 4.10: Timeline of the cancelled applications. The spike represents day 31 where applications presumably are cancelled automatically due to lack of feedback (own illustration).

The same three-phase schema holds true for cancelled applications. The three phases are visible in the streamlined model, but the majority of cancellations happen in the offer and validation phases. This is most likely due to the time-outs discussed previously.

An illustration of the streamlined model for cancelled applications can be found in the appendix (Fig. B.4).

### 4.2.3 Conclusion

The event log on hand represents a complex business process and cannot easily be transformed into a single unified and comprehensible process model by means of simple process mining. To tackle this problem some potentials for further manual simplification have been identified, and by removing redundant events the overall complexity could be reduced considerably. The simplified data set is 28% smaller on the event side, while only losing 5% of the cases, eachof which is fully attributable to incomplete process instances.

This simplified model was further decomposed into the three application types *accepted*, *declined* and *cancelled* to allow better insights into the underlying process structure. With this method a clear three-phase structure was identified. Additionally, Fig. 4.8, Fig. 4.9 and Fig. 4.10 show unique timeline patterns for each of the application types.

## 4.3 Simulation

Based on the understanding of the business process gained so far, a simulation model can be built. The model designed in this section does not cover the complete business process, rather only some key elements for the purpose of showing the feasibility of the approach and to enable a first evaluation. However, in this simplified version of the process the likelihood and probability distribution of occurring events is based on the real data set.

For the concrete implementation the Java-based framework *DESMO-J*[27] was chosen. It has been developed mainly by the modelling and simulation group of the computer science department at the University of Hamburg and allows the creation of object-oriented simulation models. It supports both discrete-event simulation and continuous simulation, and even the combination of both approaches in one simulation model, and can be integrated into existing BPM systems.[28]

### 4.3.1 Simulation Model

The simplified process model used for the simulation model is shown in Fig. B.5 (appendix). It contains eleven discrete activities, and two additional artificial start and end events (`START` and `END`). The decision space is outlined in Table 4.2. The ordering of activity has been adjusted, so that a work item always completed before a state change in the offer or application happens. This simplifies the model significantly, and does not change the overall workflow. The simulation model contains a possible loop when contacting the customer, since this activity was repeated quite often in the original logs. This is the only part in the model where an application can be cancelled, which was usually triggered by a time-out in the original process. Additionally, the original activity descriptions have been translated to English.

27. University of Hamburg: Department of Computer Science, *Discrete Event Simulation Modeling in Java (DESMO-J)*, 2014, accessed April 27, 2014, `http://desmoj.sourceforge.net`.

28. Gehlsen, B. and Page, B., "A Framework For Distributed Simulation Optimization," in *Proceedings of the 2001 Winter Simulation Conference* (Arlington, VA, USA: ACM, 2001); Rücker, B., "Building an open source Business Process Simulation tool with JBoss jBPM" (Master Thesis, Stuttgart University of Applied Science, 2008); Göbel, J. et al., "The discrete event simulation framework DESMO-J: Review, comparision to other frameworks and latest development," in *Proceedings of the 27th European Conference on Modelling and Simulation*, vol. 4 (Aalesund, Norway, 2013), 100–109, ISBN: 9780956494467.

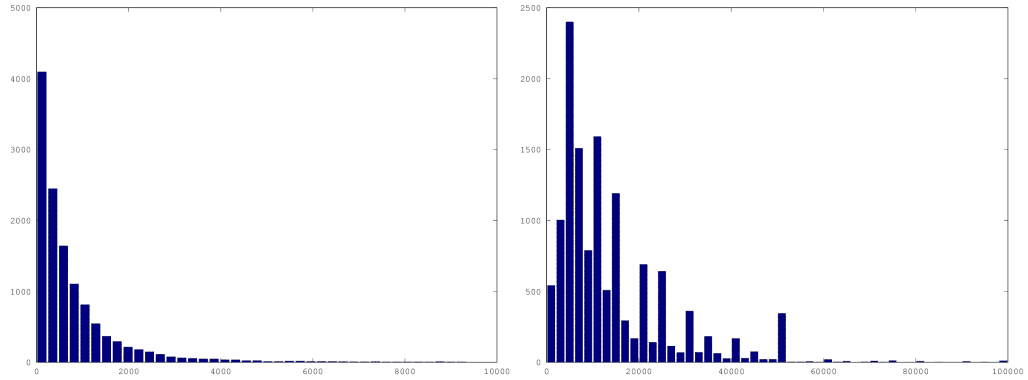
| Activity | Decision Space |
|---|---|
| A_SUBMITTED | A_DECLINED, A_PREACCEPTED |
| A_PREACCEPTED | W_AssessApplication |
| W_AssessApplication | A_DECLINED, O_CREATED |
| O_CREATED | O_SENT |
| O_SENT | W_ContactCustomer |
| W_ContactCustomer | W_ContactCustomer, O_CREATED, O_SENT_BACK, A_CANCELLED |
| O_SENT_BACK | W_ValidateApplication |
| W_ValidateApplication | A_ACTIVATED, A_DECLINED |

Table 4.2: Decision space of the simulation model.

The probability distributions were gathered by analysing historical data, i.e. the simplified event logs of the original process. Fig. 4.11 illustrates this based on real data extracted from the original event log. The histograms 4.11a and 4.11b respectively, show the arrival interval of loan applications in seconds, and the amount of money requested per application. While the arrival rate can be approximated by a $\chi^2$ or Gamma distribution, the amounts requested per case do not relate to one of the common distributions. While it clearly exhibits some exponential behaviour, the preference of customers to chose even numbers (as 500, 750, or 5000) is clearly visible and distorts the distribution. In such cases, instead of using a probability distribution, a random sample was drawn from the original population to approximate the correct distribution.

The process was modelled as a directed graph with edge annotations. The annotations define the transition probability, and the transition duration of the respective edge. This allows for a clean, semi-automatic integration with DESMO-J by separating the process model from the simulation model. In this way simulation environment provides the framework around the process model. Tab. B.1 (appendix) shows the used edge annotations (transition probabilities and distributions used for duration sampling) in detail.

Events can be distinguished into two categories, simple events which only reference a case, and resource-dependent events, which need a case and a resource to trigger. Simple events are for example the submission or initial rejection of an application, i.e. events with an automated resource. On the other hand, work items reserve an available resource for a time period




(a) Arrival rate of applications. (b) Amount requested per case.

Figure 4.11: Histograms of (a) the arrival rate of new loan applications in seconds (showing only 40 of 100 bins), and (b) the amounts requested by customers (own illustration).

determined by the underlying distribution for the given activity.

The only constraints imposed on the simulation are workflow-based restrictions, as defined by the decision space. The original event log does not indicate any resource bottlenecks, therefore resources are assumed to be available all the time. Additionally, the probability distributions defined for the time spans already incorporate small variations.

### 4.3.2 Simulation State

At any given time, the simulation state is known. It contains information about the control flow (partial traces), case data (contextual data, e.g. amount of money requested in loan application), resource data (e.g. availability and utilisation), and statistics of each events and entities.

This state is kept synchronised with the Business Process Execution and Monitoring (BPEM) environment, which acts as simulation environment and controlling instance, and is described in more detail later on. This is achieved by sending event logs to the BPEM whenever the state changes, i.e. an activity is executed.

## 4.4 Recommendation Service

The original goal of this thesis is to explore decision support in BPM environments. A decision support system aims at relieving the human operator,

e.g. by decreasing their work load. A recommendation service enables such a decision support on operational business processes.

The information provided by the recommendation service can be utilised on various levels of automation as described in Chapter 2. One possibility is to merely support the human decision maker by providing information. This can be done by retrieving the predicted performance of an active trace or a possible choice, consequently allowing a comprehensive view of the current and future performance indicators. Such information allows the human decision maker to better assess the decision task at hand. On the scale of automation levels this would be regarded as low automation, e.g. level 2 or 3.

Another type of support for a human decision maker is the (semi-)automatic assessment of a decision task. By estimating the risk of a decision, which can for example be defined as probability of a wrong decision multiplied by the cost of an error, and the confidence of the prediction made by the learning algorithm, a recommendation for a course of action can be made. When confronted with a low-risk decision and a highly confident prediction the system can autonomously take action, while it might fall back to support through providing information in other cases. This allows the human to just concentrate on important tasks by filtering trivial decisions, and helps them to make a good decision, if an intervention is necessary.

### 4.4.1 Architecture

The recommendation service offers an interface for the process engine (or the simulation, respectively). It serves three major purposes: the prediction of a performance indicator, the classification of a (partial) trace, and the assessment of a situation as well as the recommendation of an action based on this assessment.

The recommendation service is initialised with an event log object, which resembles the historical data. Based on this event log, the learning algorithms are trained. After the initialisation, the service accepts partial traces and responds with either a prediction or a classification for the given trace. Internally, the service converts the logs and traces to data sets usable by the machine learning libraries.

### 4.4.2 Prediction and Classification

At the core of the recommendation service are the predictors and classifiers. As outlined in Chapter 3, supervised learning algorithms have been chosen, specifically linear regression and a multilayer perceptron for prediction, and multinominal logistic regression and a C4.5 decision tree for classification. The implementations are provided by the data mining software *Weka*[29] in form of a Java library.

The target value for the prediction is cycle time, while the classification uses the outcome (accepted, declined, cancelled). Additional to the generic features discussed in Chapter 3, some use case specific input features have been added. This includes the case attribute `AMOUNT_REQ`, the outcome of an application as discussed previously, and the length of the longest loop in the trace. Additionally the cost of a case was estimated as the total service time minus penalties for cancelled and declined applications. Accepted applications are assumed to be economically profitable for the financial institution and therefore more desirable, while cancelled and declined applications represent less of a desirable outcome. While this is still a fairly naive approach, the alternative of estimating the cost by service time multiplied with a hourly wage is useless in this case, since the cost would then fully correlate to the service time and provide no additional value to the input feature vector.

The results can be used as-is or combined by a weighted ranking function in the following form:

$$decision(x_p) = w_c * cost(x_p) + w_t * time(x_p) + w_q * quality(x)$$

This recommendation has been integrated into the simulation, and can give recommendations for a given trace at each decision point.

## 4.5 Experiment

The experiment combines the simulation and the recommendation service, with the goal of extracting metrics to evaluate the approach in the next chapter. To be able to utilise the recommendation service in the simulation a small business process execution and monitoring environment has been designed and implemented. The execution environment is able to collect

29. Machine Learning Group at the University of Waikato, *Weka 3: Data Mining Software*, 2014, accessed April 28, 2014, `http://www.cs.waikato.ac.nz/~ml/weka/`.

event logs from the simulation, provide historic data to the recommendation service for learning purpose, and keep track of the KPIs and their changes throughout the simulation.

### 4.5.1 Design

The experiment is designed according to the integrated architecture presented in Sec. 3.5 and illustrated in Fig. 3.3 (benchmark and Fig. 3.3 (recommendation-enabled). Initially event logs have to be generated and stored to serve as historic data for the supervised learning algorithms. These logs are created by simulating the business process and storing the resulting logs. At the same time KPIs of the original simulation are collected for evaluation purposes.

Selecting features and training of the learning algorithms marks the second step. For this, all features discussed in Chapter 3 (performance indicators) are extracted from the log. Depending on the recommendation approach (prediction vs. classification), a numerical or categorical target value is selected and the remaining predictor variables are analysed with a principal component analysis to reduce the dimensionality to a minimum. This helps to reduce the noise and prevents regressions from over-fitting the data.The resulting data set is used as a base data set for 10-fold cross-validation. This means the data set is partitioned into 10 sets, of which nine are used as training set and one as test set. This partition is cycled 10 times, so that each sub-set eventually has acted as both training and test set. Based on the findings only the most promising feature set is retained.

The last phase is a simulation run utilising the recommendation service for the decision tasks instead of the pre-defined decision probabilities. This approach still takes constraints into account, e.g. a valid decision space and resource availability. As in the first phase, KPIs and event logs are collected for further evaluation.

### 4.5.2 Execution Environment

To conduct the experiment all parts have to be integrated. The simulation and the simulation model have already been described in this chapter. Subsequently the recommendation service was discussed in detail.

The last component necessary to conduct the experiment is the business process execution and monitoring environment. Fig. 4.12 shows the architecture of a simple execution environment. It offers an interface to add

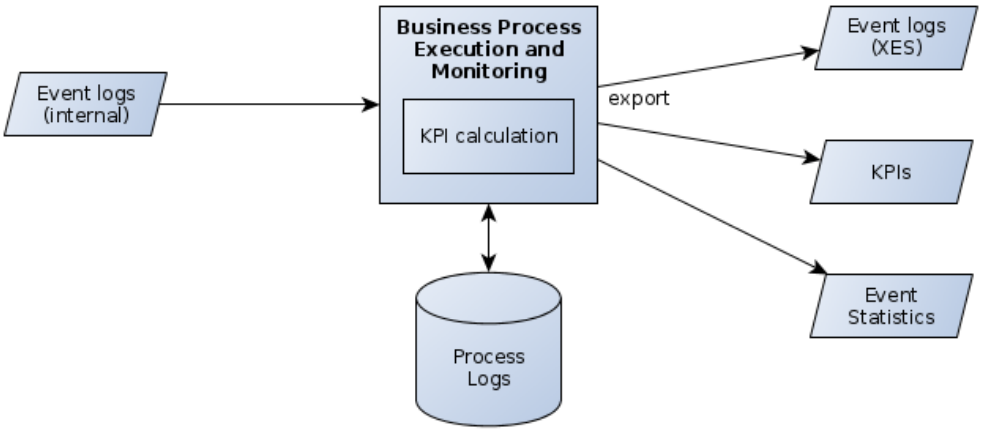



Figure 4.12: Architecture of the Business Process Execution and Monitoring Environment (own illustration).

event logs (e.g. generated by the simulation), to export the stored event logs in a standardised format (XES), and to extract the KPIs calculated in the course of the execution phase. The environment also holds statistical information about the events in the log, which can be retrieved. Internally it uses a storage to keep track of all active and completed cases.

# Chapter 5

# Analysis of Results

To start with, this chapter will introduce appropriate evaluation metrics for prediction and classification models. After that two prediction models and two classification models are evaluated, and compared against each other based on an event log generated via simulation. In the next step, the performance development of the decision tree over time is described. Finally a conclusion is drawn from the findings and problems are briefly discussed.

## 5.1 Evaluation Metrics

The first metric used for this evaluation is the *root-mean-square error* (RMSE), a measure of difference between predicted values $\hat{y}_t$ and actually observed values $y_t$. It is used to measure the quality of a prediction model. It is defined as:

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{n}(\hat{y}_t - y_t)^2}{n}} \tag{5.1}$$

Here, $n$ stand for the degrees of freedom, i.e. the size of the input feature vector. Further, the RMSE can be normalised:

$$\text{NRMSE} = \frac{RMSE}{x_{max} - x_{min}} \tag{5.2}$$

To measure the performance of a rule-based classification model, the *support*, *confidence* and *lift* can be calculated. Given a set of antecedents $X$ and a set of consequences $Y$, the support *supp* of a rule $X \Rightarrow Y$ is defined

as the proportion of items in the data set for which the rule is applicable, i.e. items with matching antecedents and consequences. The confidence is defined as:

$$\text{conf}(X \Rightarrow Y) = \frac{supp(X \cup Y)}{supp(X)} \tag{5.3}$$

It measures the ratio of items for which the rule is applicable and items which have the same antecedents. Finally, *lift* basically describes whether following a classification rule is better than a random choice of the target variable. It is defined as:

$$\text{lift}(X \Rightarrow Y) = \frac{supp(X \cup Y)}{supp(X) \times supp(Y)} \tag{5.4}$$

Furthermore for classification tasks the *precison*, *recall* and *f-measure* can be calculated. The precision for a class is the ratio of true positives, i.e. the number of items correctly classified, and total number of elements labelled as positives. Recall is the number of true positives divided by the total number of elements that actually belong to the positive class. The f-measure combines precision and recall, and weighs them evenly:

$$f = 2 \times \frac{precision \times recall}{precision + recall} \tag{5.5}$$

## 5.2 Performance Evaluation

To obtain a data set the process described in the last chapter was simulated over four weeks. This produced an event log with 1,224 cases and 11,053 events. To obtain a training set, all incomplete cases have been removed, which leaves a usable training set with 1,148 cases and 10,347 events. From these cases the attributes `TraceLength`, `AmountRequested`, `LongestLoopLength`, `CycleTime`, `WorkTime` and `Outcome` were extracted and used in the evaluation. In each case the overall prediction and classification quality was tested by a 10-fold cross validation.

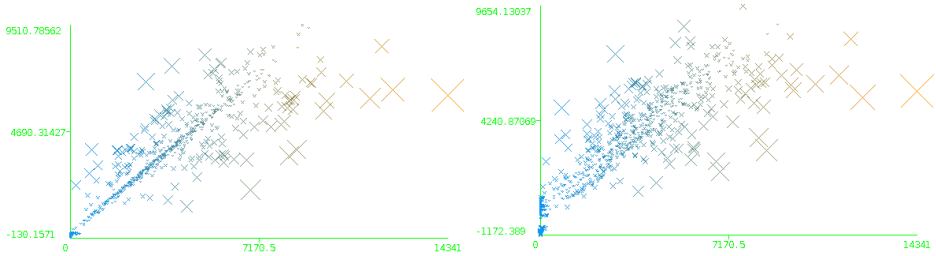
### 5.2.1 Prediction Quality

The cycle time was used as target value for both prediction models. It was measured in minutes, and the minimum value observed was 0, the maximum

14,341 and the arithmetic mean 2380.

#### 5.2.1.1 Linear Regression

Fig. 5.1a shows the results for linear regression. The observed RMSE was 795.15, resulting in a normalised RMSE of approximately 5.54 percent. The illustration shows that linear regression provides acceptable results, except for some outliers with high cycle times.




(a) Linear Regression (b) Multilayer Perceptron

Figure 5.1: Cross-validation results for cycle time prediction. Actual cycle time is denoted on the x-axis, and predicted cycle time on the y-axis (own illustration).

#### 5.2.1.2 Multilayer Perceptron

The observed RMSE for a multilayer perceptron was 941.41, resulting in a normalised RMSE of approximately 6.56 percent. This is reflected in Fig. 5.1b, which depicts the results and shows that this multilayer perceptron is less accurate than linear regression.

### 5.2.2 Classification Quality

To measure the classification performance, the cases have been classified into their predicted outcome. Of 1,148 total cases, 554 ended with the outcome `ACCEPTED`, 567 with `DECLINED` and 27 with `CANCELLED`.

#### 5.2.2.1 Logistic Regression

Logistic regression was able to correctly classify 90.06 percent of the instances. Details about the accuracy by class are listed in Table 5.1a.

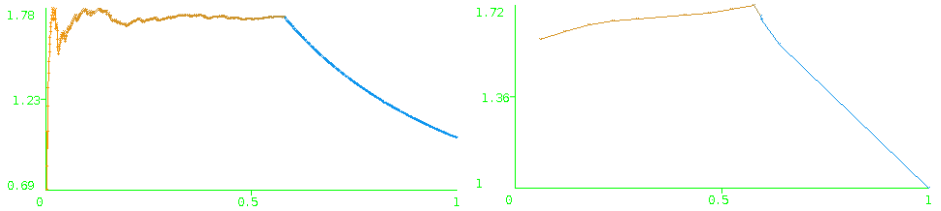
| Class | Precision | Recall | F-Measure |
|---|---|---|---|
| ACCEPTED | 0.830 | 0.998 | 0.907 |
| DECLINED | 1.000 | 0.802 | 0.890 |
| CANCELLED | 0.963 | 0.963 | 0.963 |
| weighted avg. | 0.917 | 0.901 | 0.900 |

(a) Logistic Regression

| Class | Precision | Recall | F-Measure |
|---|---|---|---|
| ACCEPTED | 0.830 | 0.995 | 0.905 |
| DECLINED | 0.996 | 0.802 | 0.889 |
| CANCELLED | 0.963 | 0.963 | 0.963 |
| weighted avg. | 0.915 | 0.899 | 0.898 |

(b) Decision Tree

Table 5.1: Detailed accuracy of classification methods by class.




(a) Logistic Regression (b) Decision Tree

Figure 5.2: Lift for the outcome ACCEPTED. The number of instances is denoted on the x-axis, and the lift on the y-axis (own illustration).

#### *5.2.2.2 C4.5 Decision Tree*

The decision tree was able to correctly classify 89.89 percent of the instances. Details about the accuracy by class are listed in Table 5.1b. In contrast to the clear difference in prediction performance described before, the classification methods evaluated here do not differ that much. This is reflected in Fig 5.2 which compares the lift of both approaches for the outcome ACCEPTED.

### 5.2.3 Performance development over time

However, these results only hold true for a complete event log, since the performance is only measured based on complete traces. These results are helpful to initially assess the suitability of an algorithm, but to judge the overall quality, the development of the classification performance has to be examined over the complete life-time of a case.

Fig. 5.3 (overleaf) outlines the average confidence of correct classifications over time (blue line). Additionally the ratio between correct and erroneous classifications is shown (orange line). The graph shows clearly that the classification success within the first seven steps is rather low, less than 50 percent of the classifications is correct. On the other hand, the remaining

correct classifications have a perfect confidence of 1.0. Towards the end a higher ratio of traces is classified correctly, and the confidence rises steadily.

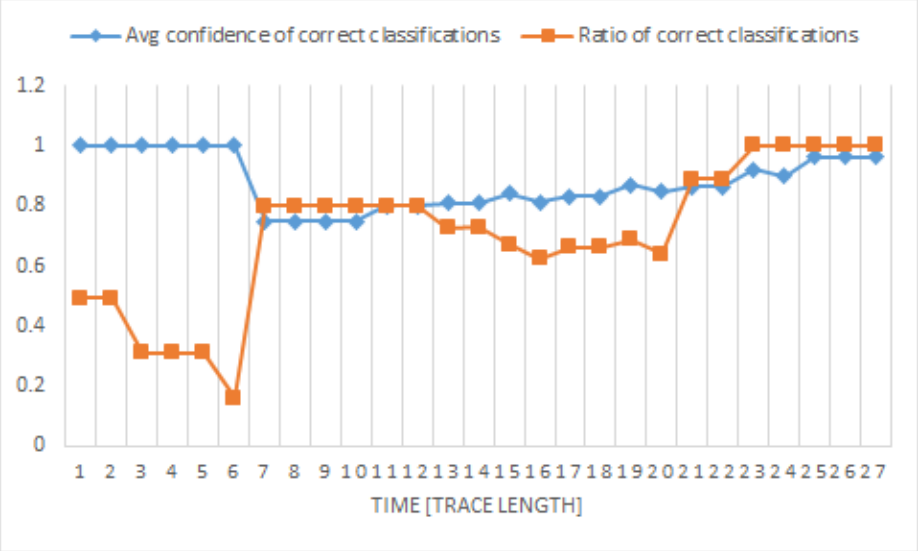



Figure 5.3: Classification distribution and confidence over time. The time is denoted on the x-axis, and the confidence of the correct classifications (blue) as well as the ratio of correct classifications (orange) on the y-axis. (own illustration).

## 5.3 Interpretation

The evaluation has shown that a rather simple machine learning algorithm utilising only a few features from an event log can deliver good results. However, in a BPM environment predictions and classifications typically are not needed at the end of a business process. Instead the process needs to be correctly assessed as early as possible in its life-time.

When focusing on time-dependent performance quality, it is clear that process instances often can not be classified early on with such traditional techniques. In the beginning of their life span cases often have not yet developed strong characteristics. In the model used for the current approach, the only unique attribute is the amount requested in a loan application, which is clearly not enough information to be of any help. In a real-life scenario usually more contextual information would be available, e.g. historical data regarding the customer from the organisation's database.

# Chapter 6

# Conclusion

## 6.1 Summary

Research in the area of operational decision support aims to improve business processes, their underlying models and in the end the actual process execution. Decision support systems specifically target human decision makers, and aim to help them in their decision tasks at all levels of BPM.

The goal of this thesis was to provide an insight into existing approaches of applying decision support technology to BPM environments, and furthermore to identify requirements for integrated support systems.

A survey of related approaches has shown that there are three types of operational support: detection, prediction and recommendation. While detection acts merely as an informative means, it can be used to develop further support based on the detection of a problematic process instance at run-time. The prediction of performance indicators can support both the detection and recommendation approaches. However, most approaches take a theoretical view onto operational decision support systems. Only some concrete solutions have been proposed, and of these only one is a fully integrated solution.

Furthermore, the survey has helped to identify certain requirements which were then incorporated into the approach developed in the course of this thesis. These requirements include the integration of a decision support system with process execution and monitoring phases, and the use of a variety of data sources. These sources provide case-based, design-based, contextual and historical data, which all play an important role for operational support systems.

This thesis introduced an approach to providing operational decision support by utilising supervised machine learning algorithms. It aims to combine detection, prediction and recommendation into a configurable environment. Historical event data is used for predictive analysis, the training of a decision tree and logistic regression, while a process monitoring engine acts as data aggregator.

A survey of publicly available real-life and synthetic event logs was conducted, to eventually select a suitable log which would benefit from operational decision support. The event log of a personal loan application process by financial institution was chosen and subsequently analysed. The analysis revealed that it is hard to construct a comprehensive process model from the raw data. The difficulty stems from concurrent activities, differences in the workflow of the various resources involved in the process and its number of activities. Subsequently, the log was significantly simplified so that a streamlined process model could be created.

This simplified model acted as a guide for a simulation model through the incorporation of properties like work flow structure, flow times and decision probabilities into the simulation model. This model then was used to obtain a benchmark data set of event logs, which also acted as training set for the decision support system. Even though the simulation model was an abstraction of the process model discovered in the original event log, it proved itself valuable for the evaluation of the approach.

The proposed approach was implemented as a proof-of-concept prototype, and evaluated by performing simulation runs and collecting the accumulated data. A small Business Process Execution and Monitoring environment connected simulation, recommendation service and process monitoring and allowed the exportation of the collected event logs. Additionally it stored prediction and classification results, which then were analysed together with the event logs with third party software.

The evaluation has shown that it is possible to achieve a good recommendation performance in regards to metrics such as root-mean squared error, recall and precision, even with a few input features. However, it has also shown that these findings only apply for process instances which have reached the last part of their life-time. When focusing on the development of the performance indicators over time, it is obvious that young cases do not have enough data associated to provide confident recommendations early on.

In summary, the thesis introduced the groundwork necessary for operational decision support in BPM and demonstrated that the proposed approach is both feasible and can improve the process quality.

## 6.2 Outlook

Some technical remarks can be made regarding the work presented in this thesis. In the current approach, simulation was used to mitigate the cold start problem of recommender systems utilising supervised learning algorithms. While the data produced by such simulation is not perfect in regards to the observed reality, it is still representative enough to draw conclusions applicable to such reality.

Building a simulation model which accurately reflects the reality and offers a sufficient level of abstraction is a big challenge. When starting with real life event logs, the underlying process model is often not available or existent. This makes accurate modelling of the process flow hard. Amongst other things, process mining aims to automatically discover such process models. The second problem is the time-consuming and laborious manual task of fitting probability distributions. This could be improved by utilising automatic density estimation, which itself is an unsupervised learning task. In the future a stronger focus on the automatic creation of simulation models is necessary to make hem suitable for research in operational decision support and eventually in real-life systems. A first approach to solve this problem is given by Rozinat et al.[1] and future approaches could advance the automated creation of simulation models even further.

Furthermore, in this thesis resource-based constraints were ignored, and only process flow and duration information based on historical data were utilised. In the presented use case resource bottlenecks were not an issue, but in future approaches resource modelling should be considered.

While the basic prediction and classification mechanisms were evaluated, the advanced recommendations were not easily applied to the model used in this thesis. Future approaches need to incorporate the findings of the evaluation and work around the restrictions imposed, especially by the early life-time phases of process instances, where not enough information is available to provide meaningful decision support.

This thesis focused mainly on operational decision support in BPM, however decision support systems also have potential to be used in other life-cycle phases of BPM. Process improvement as a key corncern of BPM, lacks automated decision support systems, e.g. for helping process designers with process re-engineering. While separated tools and approaches for conformance checking, process discovery and the simulation of process model alternatives exist, the combination of those approaches is widely neglected by

1. Cf. Rozinat et al., "Workflow Simulation for Operational Decision Support."

BPM researchers.[2]

The work conducted in this thesis provides an up-to-date overview of recent approaches for process-oriented operational decision support, and a starting point for future research in both operational decision support and off-line decision support systems in this area.

2. Cf. van der Aalst, "Business Process Management: A Comprehensive Survey," pp. 28-30.

# Appendix A

# Glossary

**Decision Support System (DSS)**
A decision support system is a computer-based interactive system, that can be used to support decision makers in complex decision making and problem solving, instead of replacing them. It utilises data and models and solves problems with varying degrees of structures.[1]

**Recommender System (RS)**
The goal of a RS is to "generate meaningful recommendations to a collection of users for items or products that might interest them".[2]

**Expert System (ES)**
An expert system consists of a knowledge base and an inference engine. The knowledge base is expressed as rules (If $A$ then $B$), and the inference engine is used for reasoning. An expert system is "a computer system that emulates the decision-making ability of a human expert".[3]

**Automated Advising System (AAS)** An automated advising system is usually considered an expert system,[4] ans is sometimes referred to as

1. See Eom et al., "A Survey of Decision Support System Applications (1988-1994)"; Shim, J. et al., "Past, present, and future of decision support technology," *Decision Support Systems* 33, no. 2 (2002): 111–126.

2. Cf. Melville, P. and Sindhwani, V., *Recommender Systems*, 2010, p. 829.

3. Jackson, P., *Introduction To Expert Systems*, 3 ed. (Addison-Wesley, 1998), pp. 1-14, isbn: 9780201876864.

4. Cf. Harlan, R. M., "The Automated Student Advisor: a large project for expert systems courses," in *Proceedings of the twenty-fifth SIGCSE symposium on Computer science education - SIGCSE '94* (New York, New York, USA: ACM Press, 1994), 31–35, isbn: 0897916468, doi:10.1145/191029.191046, http://portal.acm.org/citation.cfm?doid=191029.191046; Siegfried, R. M., Wittenstein, A. M., and Sharma, T., "An automated advising system for course selection and scheduling," *Journal of Com-*

a decision support system.[5]

*puting Sciences in Colleges* 18, no. 3 (2003): 17–25.

5. Wagner, J. J., "Support Services for the Net Generation: The Penn State Approach," *College and University Journal* 81, no. 1 (2005): pp. 5-10.

# Appendix B

# Detailed Process Models

| | |
|---|---|
| Events | 189,267 |
| Cases | 12,687 |
| Activities | 23 |
| Median case duration | 13.7 hrs |
| Mean case duration | 8.3 d |
| Start | 01.10.2011 00:38:44 |
| End | 14.03.2012 15:34:10 |

(a) All event logs.

| | |
|---|---|
| Events | 72,915 |
| Cases | 2,245 |
| Activities | 20 |
| Median case duration | 14.4 d |
| Mean case duration | 16.7 d |
| Start | 01.10.2011 00:38:44 |
| End | 14.03.2012 15:34:10 |

(b) Only accepted applications.

| | |
|---|---|
| Events | 52,388 |
| Cases | 7,635 |
| Activities | 20 |
| Median case duration | 7.3 mins |
| Mean case duration | 49.2 hrs |
| Start | 01.10.2011 08:11:08 |
| End | 14.03.2012 15:20:26 |

(c) Only declined applications.

| | |
|---|---|
| Events | 63,964 |
| Cases | 2,807 |
| Activities | 19 |
| Median case duration | 17.4 d |
| Mean case duration | 18.6 d |
| Start | 01.10.2011 09:45:25 |
| End | 14.03.2012 15:30:49 |

(d) Only cancelled applications.

Figure B.1: Key statistics for financial event log (own illustration).

| **Transition** $t$ | $P(t)$ | **Duration** ($\mu$ **in min**) |
|---|---|---|
| START → A_SUBMITTED | 100% | Gamma distribution: $\mu = 17, k = 1, \beta = 2$ |
| A_SUBMITTED → A_DECLINED | 26% | Poisson distribution: $\mu = 2, \lambda = 1$ |
| A_SUBMITTED → A_PREACCEPTED | 74% | Poisson distribution: $\mu = 2, \lambda = 1$ |
| A_PREACCEPTED → W_AssessApplication | 100% | Exponential distribution: $\mu = 60, \lambda = 1$ |
| W_AssessApplication (START) → W_AssessApplication (COMPLETE) | 100% | Normal distribution: $\mu = 20, \sigma = 10$ |
| W_AssessApplication (COMPLETE) → A_DECLINED | 17% | Poisson distribution: $\mu = 2, \lambda = 1$ |
| W_AssessApplication (COMPLETE) → O_CREATED | 83% | Normal distribution: $\mu = 10, \sigma = 5$ |
| O_CREATED → O_SENT | 100% | Normal distribution: $\mu = 1, \sigma = 0.2$ |
| O_SENT → W_ContactCustomer (START) | 100% | Normal distribution: $\mu = 30, \sigma = 15$ |
| W_ContactCustomer (START) → W_ContactCustomer(COMPLETE) | 100% | Exponential distribution: $\mu = 11, \lambda = 1$ |
| W_ContactCustomer (COMPLETE) → W_ContactCustomer (START) | 40% | Exponential distribution: $\mu = 1200, \lambda = 1$ |
| W_ContactCustomer (COMPLETE) → O_CREATED | 10% | Normal distribution: $\mu = 5, \sigma = 1$ |
| W_ContactCustomer (COMPLETE) → O_SENT_BACK | 40% | Normal distribution: $\mu = 12, \sigma = 2$ |
| W_ContactCustomer (COMPLETE) → A_CANCELLED | 10% | Normal distribution: $\mu = 3, \sigma = 0.5$ |
| O_SENT_BACK → W_ValidateApplication (START) | 100% | Exponential distribution: $\mu = 30, \lambda = 1$ |
| W_ValidateApplication (START) → W_ValidateApplication (COMPLETE) | 100% | Normal distribution: $\mu = 50, \sigma = 20$ |
| W_ValidateApplication (COMPLETE) → A_ACTIVATED | 83% | Normal distribution: $\mu = 12, \sigma = 2$ |
| W_ValidateApplication (COMPLETE) → A_DECLINED | 17% | Normal distribution: $\mu = 1, \sigma = 0.2$ |

Table B.1: Edge annotations of the directed graph simulation model.

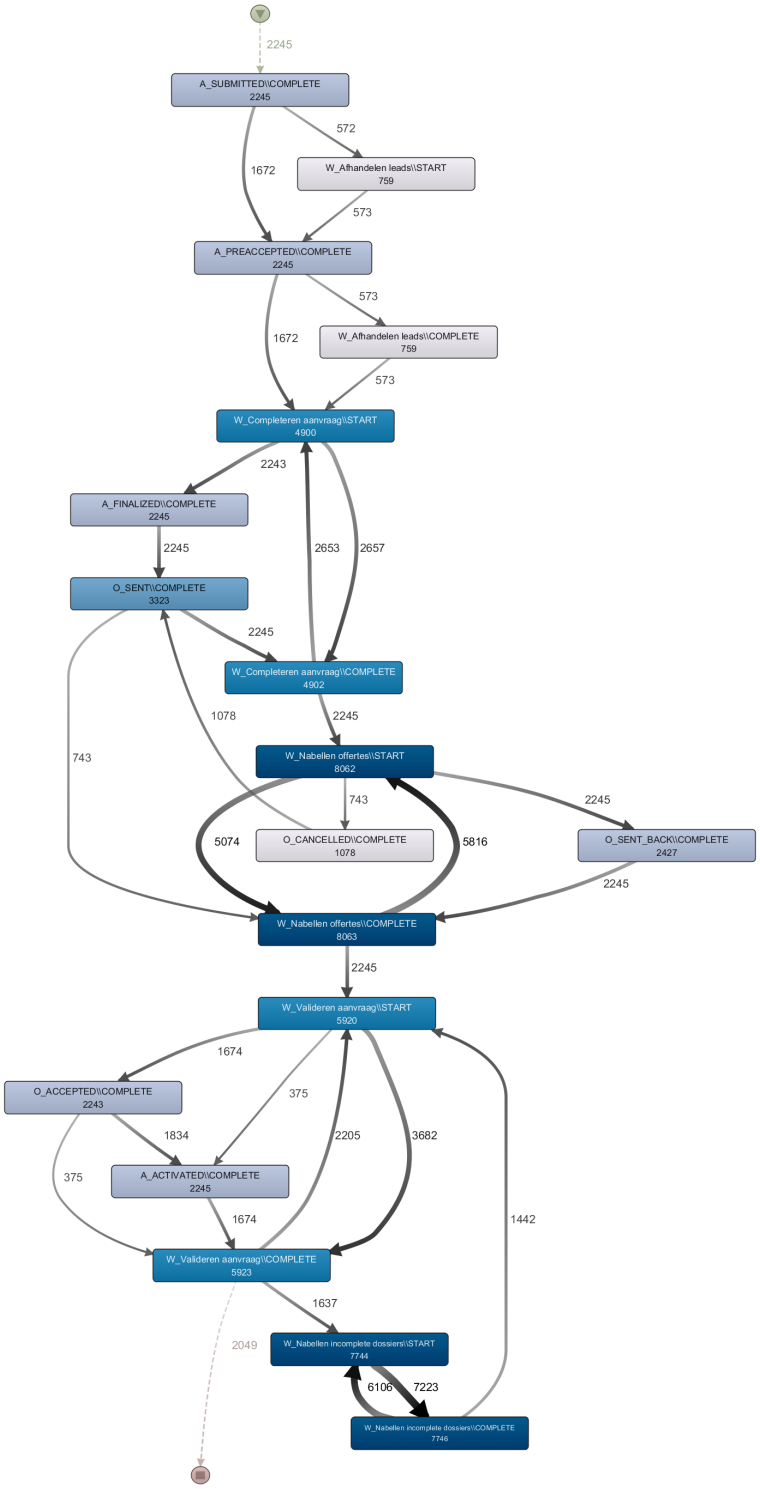



Figure B.2: Streamlined process model for successful applications (own illustration).

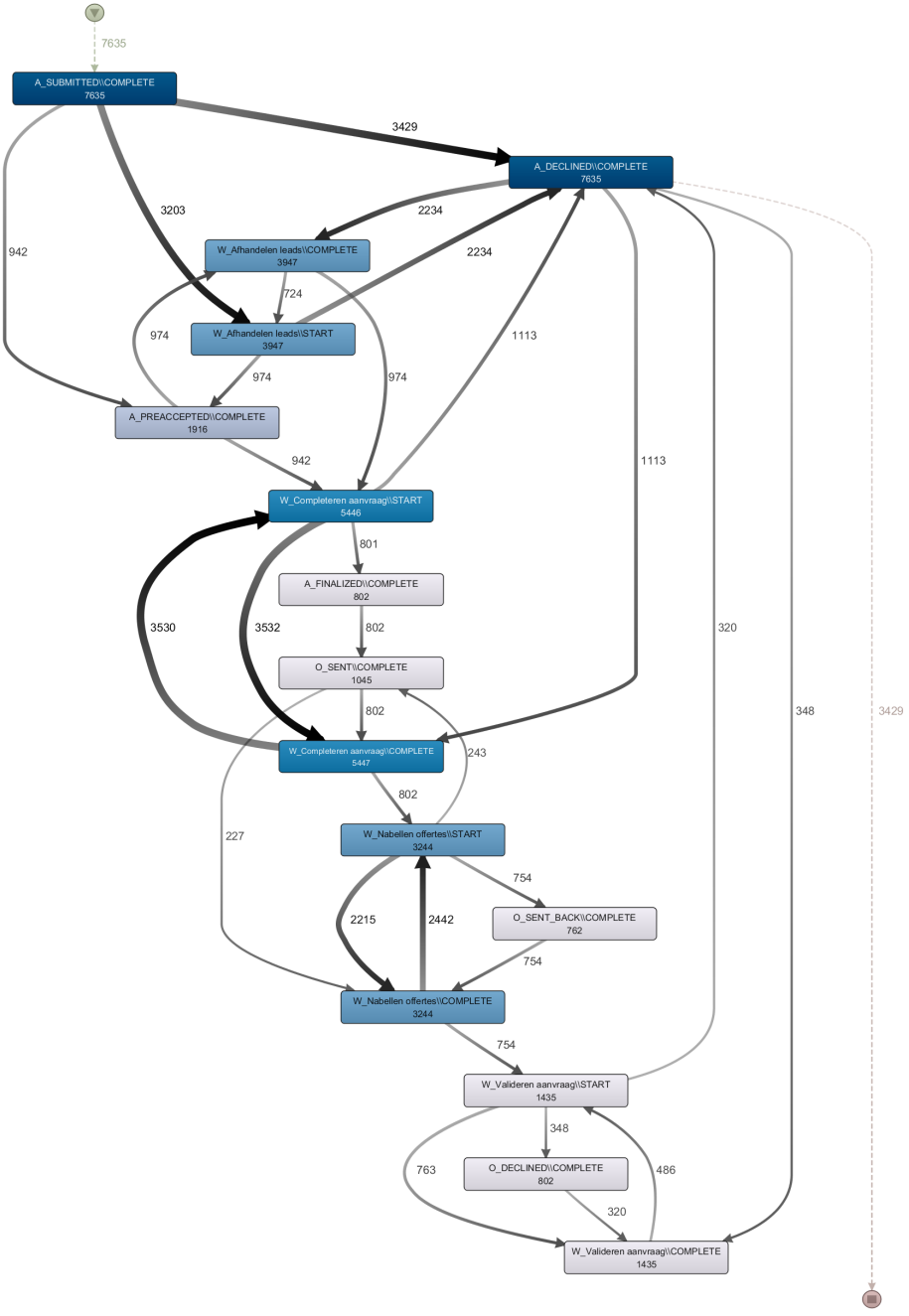



Figure B.3: Streamlined process model for declined applications (own illustration).

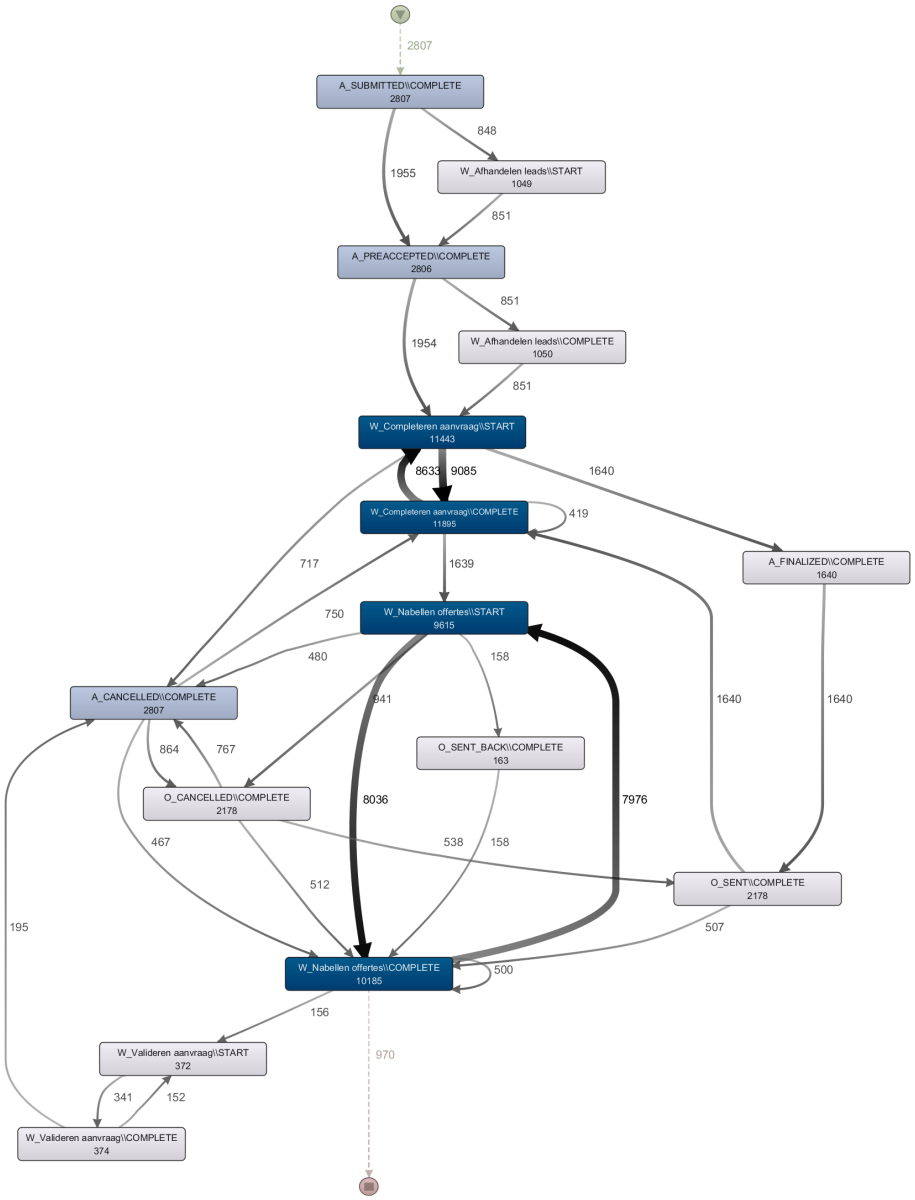



Figure B.4: Streamlined process model for cancelled applications (own illustration).

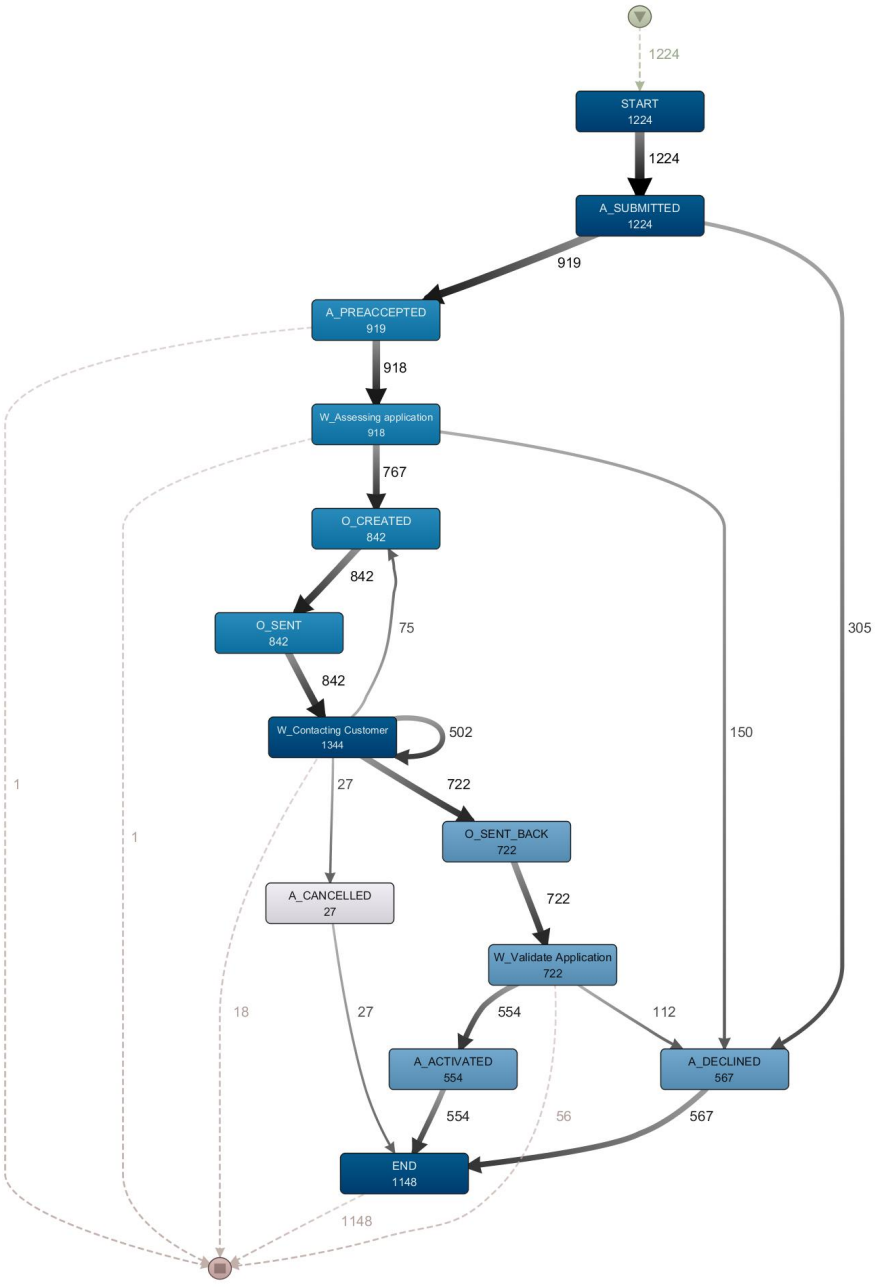



Figure B.5: Simplified process model used for simulation (own illustration).

# Bibliography

*3TU Datacentrum*, 2014. Accessed April 19, 2014. `http://data.3tu.nl/repository`.

Adriansyah, A. and Buijs, J. "Mining Process Performance from Event Logs." 2012.

Alpaydin, E. *Introduction to Machine Learning*. 2nd ed. Edited by Dietterich, T., Bishop, C., Heckerman, D., Jordan, M., and Kearns, M. Adaptive Computation and Machine Learning. Cambridge, MA, USA: MIT Press, 2010. isbn: 9780262012430.

Anderson, J. R. *Cognitive Psychology and Its Implications*. 6th ed. Worth Publishers, 2004. isbn: 978-0716701101.

Barber, D. *Bayesian Reasoning and Machine Learning*. Cambridge University Press, 2012. isbn: 9780521518147.

Bautista, A. D., Akbar, S. M. K., Alvarez, A., Metzger, T., and Reaves, M. L. "Process Mining in Information Technology Incident Management: A Case Study at Volvo Belgium." 2013.

Bautista, A. D., Wangikar, L., and Akbar, S. M. K. "Process Mining-Driven Optimization of a Consumer Loan Approvals Process." 2012.

Becker, J., Kugeler, M., and Rosemann, M., eds. *Prozessmanagement: Ein Leitfaden zur prozessorientierten Organisationsgestaltung*. 7th ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012. isbn: 978-3-642-33843-4. doi:`10.1007/978-3-642-33844-1`.

Bose, R. P. J. C. *Artificial Digital Photo Copier Event Log*, 2011. doi:`10.4121/uuid:f5ea9bc6-536f-4744-9c6f-9eb45a907178`.

Bose, R. P. J. C. and van der Aalst, W. M. "Analysis of Patient Treatment Procedures." 2011. doi:`10.4121/uuid`.

Bose, R. P. J. C., Verbeek, E. H. M. W., and van der Aalst, W. M. P. "Discovering Hierarchical Process Models Using ProM." In *CAiSE Forum 2011*, 33–48. London, UK, 2012.

*BPIC 2012 Event Log Description*, 2012. Accessed April 26, 2014. `http://www.win.tue.nl/bpi/2012/challenge`.

Broy, M. and Kuhrmann, M. *Projektorganisation und Management im Software Engineering*. Xpert.press. Springer Berlin Heidelberg, 2013. isbn: 9783642292897. doi:`10.1007/978-3-642-29290-3`.

Buijs, J. *Loan application example*, 2013. doi:`10.4121/uuid:bd8fcc48-5bf3-480e-8775-d79d6c700e90`.

Caron, F., Vanthienen, J., De Weerdt, J., and Baesens, B. "Beyond X-Raying a Care-Flow: Adopting Different Focuses on Care-Flow Mining."

Conforti, R., Fortino, G., Rosa, M. L., and ter Hofstede, A. H. M. "History-Aware, Real-Time Risk Detection in Business Processes." In *On the Move to Meaningful Internet Systems: OTM 2011*, edited by Meersman, R., Dillon, T., Herrero, P., Kumar, A., Reichert, M., Qing, L., Ooi, B.-C., et al., 100–118. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011. isbn: 978-3-642-25108-5. doi:`10.1007/978-3-642-25109-2\_8`.

Conforti, R., Leoni, M. de, La Rosa, M., and van der Aalst, W. M. P. "Supporting Risk-Informed Decisions during Business Process Execution." In *Advanced Information Systems Engineering*, edited by Salinesi, C., Norrie, M. C., and Pastor, Ó., 116–132. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013. isbn: 978-3-642-38708-1. doi:`10.1007/978-3-642-38709-8\_8`.

Crooy, R. "Predictions in Information Systems: a process mining perspective." Master Thesis, Technische Universiteit Eindhoven, 2008.

Dahanayake, A., Welke, R. J., and Cavalheiro, G. "Improving the understanding of BAM technology for real-time decision support." *International Journal of Business Information Systems* 7, no. 1 (2011): 1–26. doi:`10.1504/IJBIS.2011.037294`.

Davenport, T. H. and Beck, J. C. *The attention economy: understanding the new currency of business*. Boston, MA: Harvard Business School Press, 2001. isbn: 1578518717.

DAVENPORT, T. H. and SHORT, J. E. "The New Industrial Engineering: Information Technology and Business Process Redesign." *MIT Sloan Management Review* (Cambridge, MA, USA) 31, no. 4 (1990). `http://sloanreview.mit.edu/article/the-new-industrial-engineering-information-technology-and-business-process-redesign/`.

*Disco Process Mining Tool*. Accessed April 14, 2014. `http://www.fluxicon.com/disco`.

DUMAS, M., VAN DER AALST, W. M., and TER HOFSTEDE, A. H. *Process-Aware Information Systems*. Hoboken, NJ, 2005. ISBN: 9780471663065.

EOM, S. B., LEE, S. M., KIM, E., and SOMARAJAN, C. "A Survey of Decision Support System Applications (1988-1994)." *The Journal of Operational Research Society* 49, no. 2 (1998): 109–120.

FRAME, J. D. *The New Project Management: Tools for an Age of Rapid Change, Complexity, and Other Business Realities*. 2nd ed. Jossey-Bass, 2002. ISBN: 978-0-7879-5892-3.

FRITZSCHE, M., PICHT, M., GILANI, W., SPENCE, I., BROWN, J., and KILPATRICK, P. "Extending BPM Environments of Your Choice with Performance Related Decision Support." In *Business Process Management*, edited by DAYAL, U., EDER, J., KOEHLER, J., and REIJERS, H. A., 97–112. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009. ISBN: 978-3-642-03847-1. doi:`10.1007/978-3-642-03848-8\_8`.

GEHLSEN, B. and PAGE, B. "A Framework For Distributed Simulation Optimization." In *Proceedings of the 2001 Winter Simulation Conference*. Arlington, VA, USA: ACM, 2001.

GÖBEL, J., JOSCHKO, P., KOORS, A., and PAGE, B. "The discrete event simulation framework DESMO-J: Review, comparision to other frameworks and latest development." In *Proceedings of the 27th European Conference on Modelling and Simulation*, 4:100–109. Aalesund, Norway, 2013. ISBN: 9780956494467.

GÜNTHER, C. W. and VERBEEK, E. *Extensible Event Stream (XES): Standard Definition v2.0*, 2014. Accessed May 1, 2014. `http://www.xes-standard.org/`.

Harlan, R. M. "The Automated Student Advisor: a large project for expert systems courses." In *Proceedings of the twenty-fifth SIGCSE symposium on Computer science education - SIGCSE '94*, 31–35. New York, New York, USA: ACM Press, 1994. ISBN: 0897916468. doi:10.1145/191029.191046. http://portal.acm.org/citation.cfm?doid=191029.191046.

Hevner, A. R., March, S. T., Park, J., and Ram, S. "Design science in information systems research." *MIS Quarterly* 28, no. 1 (2004): 75–105. ISSN: 0276-7783. http://www.hec.unil.ch/ypigneur/HCI/articles/hevner04.pdf.

Hlupic, V. and Robinson, S. "Business process modelling and analysis using discrete-event simulation." In *Proceedings of the 30th conference on Winter simulation*, 1363–1370. Los Alamitos, CA, USA: IEEE Computer Society, 1998.

Hoch, S. J. and Schkade, D. A. "Psychological Approach Support to Decision Systems." *Management Science* 42, no. 1 (1996): 51–64.

Hofmann, M. *Performance-orientiertes Projektmanagement: Konzeption zum Umgang mit einmaligen, komplexen Aufgaben*. Unternehmensführung & Controlling. Springer Fachmedien Wiesbaden, 2014. ISBN: 9783658047986. doi:10.1007/978-3-658-04799-3.

IEEE Task Force on Process Mining. *IEEE Task Force on Process Mining - Event Logs*. Accessed April 19, 2014. http://data.3tu.nl/repository/collection:event%5C_logs.

———. "Process Mining Manifesto." In *BPM 2011 International Workshops*, edited by Daniel, F., Barkaoui, K., and Dustdar, S. 2012. ISBN: 978-3-642-28115-0.

Jablonski, S. and Bussler, C. *Workflow Management: Modeling Concepts, Architecture and Implementation*. London, UK: International Thomson Computer Press, 1996. ISBN: 9781850322221.

Jackson, P. *Introduction To Expert Systems*. 3 ed. Addison-Wesley, 1998. ISBN: 9780201876864.

Kang, C. J., Kang, Y. S., Lee, Y. S., Noh, S., Kim, H. C., Lim, W. C., and Hong, R. "Process Mining-based Understanding and Analysis of Volvo IT's Incident and Problem Management Processes." 2013.

Machine Learning Group at the University of Waikato. *Weka 3: Data Mining Software*, 2014. Accessed April 28, 2014. http://www.cs.waikato.ac.nz/~ml/weka/.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., and Byers, A. H. *Big data: The next frontier for innovation, competition, and productivity.* Technical report June. McKinsey Global Institute, 2011.

Medina-Mora, R., Winograd, T., Flores, R., and Flores, F. "The action workflow approach to workflow management technology." In *Proceedings of the Conference on Computer-Supported Cooperative Work (CSCW '92)*, 281–288. 1992. isbn: 0897915437.

Melville, P. and Sindhwani, V. *Recommender Systems*, 2010.

Meyer, S., Sperner, K., Magerkurth, C., and Pasquier, J. "Towards Modeling Real-World Aware Business Processes." In *Proceedings of the Second International Workshop on Web of Things*, 1–6. June. New York, NY, USA: ACM, 2011. isbn: 9781450306249. doi:10.1145/1993966.1993978.

Moen, R. and Norman, C. *Evolution of the PDCA Cycle.* Technical report. 2006.

Molka, T., Gilani, W., and Zeng, X.-J. "Dotted Chart and Control-Flow Analysis for a Loan Application Process." 2012.

North, M. J. and Macal, C. M. *Managing Business Complexity: Discovering Strategic Solutions with Agent-Based Modeling and Simulation.* New York, NY, USA: Oxford University Press, 2007. isbn: 978-0195172119.

Parasuraman, R., Sheridan, T. B., and Wickens, C. D. "A model for types and levels of human interaction with automation." *IEEE transactions on systems, man, and cybernetics. Part A, Systems and humans* 30, no. 3 (May 2000): 286–97. issn: 1083-4427.

*ProM 6.3: Description and Example Logs*, 2013. Accessed April 19, 2014. http://www.promtools.org/prom6.

Rozinat, A., Wynn, M. T., van der Aalst, W. M. P., ter Hofstede, A. H. M., and Fidge, C. J. "Workflow Simulation for Operational Decision Support." *Data & Knowledge Engineering* 68, no. 9 (2008): 834–850.

Rozinat, A. *Disco User's Guide*, 2012. Accessed April 19, 2014. http://fluxicon.com/disco/files/Disco-User-Guide.pdf.

Rücker, B. "Building an open source Business Process Simulation tool with JBoss jBPM." Master Thesis, Stuttgart University of Applied Science, 2008.

Schonenberg, H., Weber, B., van Dongen, B., and van der Aalst, W. M. P. "Supporting Flexible Processes Through Recommendations Based on History." In *Business Process Management*, edited by Dumas, M., Reichert, M., and Shan, M., 51–66. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008. isbn: 978-3-540-85757-0. doi:10.1007/978-3-540-85758-7_7.

Shim, J., Warkentin, M., Courtney, J. F., Power, D. J., Sharda, R., and Carlssone, C. "Past, present, and future of decision support technology." *Decision Support Systems* 33, no. 2 (2002): 111–126.

Siegfried, R. M., Wittenstein, A. M., and Sharma, T. "An automated advising system for course selection and scheduling." *Journal of Computing Sciences in Colleges* 18, no. 3 (2003): 17–25.

Steeman, W. *BPI Challenge 2013, closed problems*, 2013. doi:10.4121/uuid:c2c3b154-ab26-4b31-a0e8-8f2350ddac11.

———. *BPI Challenge 2013, incidents*, 2013. doi:10.4121/uuid:500573e6-accc-4b0c-9576-aa5468b10cee.

———. *BPI Challenge 2013, open problems*, 2013. doi:10.4121/uuid:3537c19d-6c64-4b1d-815d-915ab0e479da.

Sternberg, R. J. and Sternberg, K. *Cognitive Psychology*. 6th ed. Cengage Learning, 2011. isbn: 978-1133313915.

University of Hamburg: Department of Computer Science. *Discrete Event Simulation Modeling in Java (DESMO-J)*, 2014. Accessed April 27, 2014. http://desmoj.sourceforge.net.

van der Aalst, W. M. P. "Business Process Management: A Comprehensive Survey." *ISRN Software Engineering* 2013 (2013): 1–37. issn: 2090-7680. doi:10.1155/2013/507984.

———. *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. Heidelberg: Springer Berlin / Heidelberg, 2011. isbn: 9783642193446. doi:10.1007/978-3-642-19345-3.

———. *Synthetic event logs - review example large.xes.gz*, 2010. doi:10.4121/uuid:da6aafef-5a86-4769-acf3-04e8ae5ab4fe.

———. "TomTom for Business Process Management (TomTom4BPM)." In *Advanced Information Systems Engineering*, edited by Eck, P. van, Gordijn, J., and Wieringa, R., 2–5. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2009. isbn: 978-3-642-02143-5. doi:10.1007/978-3-642-02144-2_2.

VAN DER AALST, W. M. P. and JABLONSKI, S. "Dealing with workflow change: identification of issues and solutions." *Computer systems science and engineering* 15, no. 5 (2000): 267–276.

VAN DER AALST, W. M. P., NAKATUMBA, J., ROZINAT, A., and RUSSELL, N. "Business Process Simulation: How to get it right?" In *Handbook on Business Process Management*, edited by VOM BROCKE, J. and ROSEMANN, M. International Handbooks on Information Systems. Berlin: Springer, 2010.

VAN DER AALST, W. M. P., PESIC, M., and SONG, M. "Beyond Process Mining: From the Past to Present and Future." In *22nd International Conference on Advanced Information Systems Engineering*, 38–52. Hammamet, Tunisia: Springer Berlin Heidelberg, 2010. doi:10.1007/978-3-642-13094-6_5.

VAN DER AALST, W. M. P., SCHONENBERG, M., and SONG, M. "Time Prediction Based on Process Mining." *Information Systems* 36, no. 2 (2011). doi:10.1016/j.is.2010.09.001.

VAN DER AALST, W. M. P., WEIJTERS, T., and MARUSTER, L. "Workflow mining: discovering process models from event logs." *IEEE Transactions on Knowledge and Data Engineering* 16, no. 9 (September 2004): 1128–1142. ISSN: 1041-4347. doi:10.1109/TKDE.2004.47.

VAN DER AALST, W. M. *Event logs and models used in Process Mining book*, 2011. Accessed October 12, 2013. http://www.processmining.org/event%5C_logs%5C_and%5C_models%5C_used%5C_in%5C_book.

VAN DONGEN, B. F., CROOY, R. A., and VAN DER AALST, W. M. P. "Cycle Time Prediction: When Will This Case Finally Be Finished?" In *Proceedings of the OTM 2008 Confederated International Conferences, CoopIS, DOA, GADA, IS, and ODBASE 2008. Part I on On the Move to Meaningful Internet Systems*, 319–336. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008.

VAN DONGEN, B. *BPI Challenge 2012*, 2012. doi:10.4121/uuid:3926db30-f712-4394-aebc-75976070e91f.

———. *Real-life event logs - Hospital log*, 2011. doi:10.4121/uuid:d9769f3d-0ab0-4fb8-803b-0d1120ffcf54.

VANDERFEESTEN, I., REIJERS, H. A., and VAN DER AALST, W. M. P. "Product-based workflow support." *Information Systems* 36, no. 2 (April 2011): 517–535. ISSN: 03064379. doi:10.1016/j.is.2010.09.008.

VERBEEK, H. *Mining eXtensible Markup Language (MXML): Definition*, 2011. Accessed April 22, 2014. `http://www.processmining.org/logs/mxml`.

VOLVO IT. *VINST data set*, 2012. Accessed April 19, 2014. `http://www.win.tue.nl/bpi/%5C_media/2013/vinst%5C_data%5C_set.pdf`.

WAGNER, J. J. "Support Services for the Net Generation: The Penn State Approach." *College and University Journal* 81, no. 1 (2005): 3–10.

WESKE, M. *Business process management: concepts, languages, architectures.* 2nd ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012. ISBN: 978-3-642-28616-2. doi:`10.1007/978-3-642-28616-2`.

WILDE, T. and HESS, T. "Forschungsmethoden der Wirtschaftsinformatik - Eine empirische Untersuchung." *Wirtschaftsinformatik* 49, no. 4 (2007): 280–287.

———. *Methodenspektrum der Wirtschaftsinformatik: Überblick und Portfoliobildung.* Technical report 2. München: Institut für Wirtschaftsinformatik und Neue Medien, 2006.

ZUR MUEHLEN, M. and HANSMANN, H. "Workflowmanagement." In *Prozessmanagement: Ein Leitfaden zur prozessorientierten Organisationsgestaltung*, 7th ed., edited by BECKER, J., KUGELER, M., and ROSEMANN, M., 367–400. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012. ISBN: 978-3-642-33843-4. doi:`10.1007/978-3-642-33844-1\_11`.

ZUR MUEHLEN, M., RECKER, J., and INDULSKA, M. "Sometimes Less is More: Are Process Modeling Languages Overly Complex?" In *The 3rd International Workshop on Vocabularies, Ontologies and Rules for The Enterprise.* IEEE Publishers, 2007.

ZUR MUEHLEN, M. and ROSEMANN, M. "Workflow-based Process Monitoring and Controlling - Technical and Organizational Issues." In *33rd Hawaii International Conference on System Sciences.* c. 2000. ISBN: 0769504930.

# Eidesstattliche Erklärung / Declaration of Originality

Ich versichere, dass ich die Arbeit selbstständig angefertigt, keine anderen als die angegebenen Hilfsmittel benutzt und alle wörtlichen oder sinngemäßen Entlehnungen deutlich als solche gekennzeichnet habe.
Ich erkläre weiterhin, dass die vorliegende Arbeit noch nicht im Rahmen eines anderen Prüfungsverfahrens eingereicht wurde.

I hereby declare that this thesis represents my original work and that I have used no other sources except as noted by citations. All data, tables, figures and text citations which have been reproduced from any other source, including the internet, have been explicitly acknowledged as such.
Furthermore I declare that this work has not been submitted within the scope of another examination procedure.

Saarbrücken, Monday 5$^{\text{th}}$ May, 2014

(Place, Date) (Signature)