Enterprise Information Systems
Vol. 6, No. 2, May 2012, 139–163

Industrial application of semantic process mining

Jon Espen Ingvaldsen* and Jon Atle Gulla

Department of Computer and Information Science, Norwegian University of Science and Technology, Trondheim, Norway

(*Received 3 August 2010; final version received 29 May 2011*)

Process mining relates to the extraction of non-trivial and useful information from information system event logs. It is a new research discipline that has evolved significantly since the early work on idealistic process logs. Over the last years, process mining prototypes have incorporated elements from semantics and data mining and targeted visualisation techniques that are more user-friendly to business experts and process owners. In this article, we present a framework for evaluating different aspects of enterprise process flows and address practical challenges of state-of-the-art industrial process mining. We also explore the inherent strengths of the technology for more efficient process optimisation.

Keywords: process mining; business process intelligence; process models; business process management; data mining; process analysis; semantics; ontologies

1. Introduction

Process mining is an emerging technology that complements traditional business intelligence and business process monitoring environments. While traditional business intelligence and business process monitoring systems focus on the extraction of known performance indicators, process mining assumes a more volatile environment and emphasises the construction of process models to explain the real and partly hidden process flows of companies (van der Aalst and Weijters 2004). Process mining is a growing research discipline, as businesses grow ever more complex and the business environments change more rapidly. According to Gartner Group, process discovery now accounts for about 40% of the costs of implementing business process management (StereoLOGIC 2010).

The potential benefits of industrial process mining have been demonstrated in some recent case studies (Ingvaldsen and Gulla 2006, van der Aalst *et al*. 2007, Funk *et al*. 2008, Rozinat *et al*. 2009). As an automatic technique working on in-house data, it provides a cost-effective analysis that can be used for continuous process monitoring. The analysis does not rest on subjective manual contribution, and its statistical orientation helps us deal with process anomalies and exceptions without losing the big picture.

Early work on process mining relied on assumptions about event log noise and completeness that enabled a clean mapping to business process formalisms (Folino *et al*. 2009). Whereas the initial process mining techniques were not directly applicable to industrial enterprise systems, current research is addressing noisy

*Corresponding author. Email: jonespi@idi.ntnu.no

ISSN 1751-7575 print/ISSN 1751-7583 online

http://dx.doi.org/10.1080/17517575.2011.593103
http://www.tandfonline.com

environments and the full complexity of real-world business processes. This involves dealing with noise (Maruster *et al*. 2006), allowing for behaviours not directly described in event logs (van der Aalst *et al*. 2008), identifying process instances in unlabelled event logs (Ferreira and Gillblad 2009) and resolving the complexities and 'spaghettiness' of discovered process mining models (Weijters and Medeiros 2006, Günther and van der Aalst 2007, Song *et al*. 2009). In spite of the success of several recent case studies and the introduction of more robust and scalable techniques for discovering accurate business flow models, though, we have still not seen any widespread industrial adoption of the technology.

There are inherent limitations and challenges in process mining, and the technique is intended for particular circumstances and particular purposes. The technology's reliance on available, but often spurious log data poses a challenge, especially if these data come from heterogeneous systems with different structures and formats. Process mining also tends to produce results that are hard to interpret in business terms and requires special mining expertise.

In this article, we present a framework for evaluating process mining in general and discuss the industrial benefits and challenges of our semantic process mining approach. Our approach, which makes use of ontologies and semantically driven data mining, has been implemented as a full-fledged process mining tool called EVS (Enterprise Visualisation Suite) and applied to real production data from ERP systems like SAP. While ontologies are employed to harmonise heterogeneous log data and run process analyses in terms understandable to business users, search and data mining features support a more exploratory and comparative process analysis than what is traditionally offered by process mining applications.

The structure of this article is as follows: section 2 introduces the concept of process mining, while section 3 describes our framework for understanding the real process flows of enterprises. In section 4, we demonstrate our extraction of process models and also discuss the inherent challenges of log-supported process model extraction. Section 5 explains how four different viewpoints help us uncover important process aspects that are not exposed in standard process models in EPC or similar notations. A comparison of our semantic approach to standard process mining is given in section 6, followed by the conclusions in section 7.

2. Process mining

According to van der Aalst *et al*. (2008) and Rozinat *et al*. (2009), process mining techniques attempt *to extract non-trivial and useful information from event logs*. The technology automate activities that would otherwise be done manually by process experts and depend heavily on the availability and quality of process-related data, Cook and Wolf (1998) were early pioneers in the field of process mining. They developed algorithmic approaches to process mining that allowed the construction of process flow graphs for software development projects from event logs. Further contributions to the discipline of *business* process mining are found in the work of van der (van der Aalst and Weijters 2004), among others.

Examples of questions that can be addressed by process mining include the following (Aalst 2009):

- Process discovery: 'What is really happening?'
- Conformance checking: 'Do we do what was agreed upon?'

- Performance analysis: ‘Where are the bottlenecks?’
- Process prediction: ‘Will this process instance be late?’
- Process improvement: ‘How to redesign this process?’

In general, process mining is geared towards dynamic and chaotic business environments, in which processes are both unstable and not very well understood. Like process modelling initiatives, the technique is applied to uncover unknown aspects of process flows and help us understand how processes are executed in real life. In highly dynamic environments, we cannot assume that tasks are carried out according to pre-defined policies or standards. Since it is based on automatic tools and transaction logs, the technique does not suffer from the subjective or fragmentary knowledge of the people involved in process modelling. On the other hand, since extensive logs from ERP systems and other enterprise systems are necessary, the quality of these logs also constrains the effectiveness of the technique.

Process mining is particularly useful in volatile business sectors and flexible work environments, in which processes are planned and modified on the fly and involve complex structures of enterprises, products, services and consumers (Günther 2009). The low costs of running automatic process mining analyses help us keep the real business processes under close scrutiny almost on a day-to-day basis. Business activity monitoring (BAM) solutions offer similar automatic features, but require that more of the process knowledge is known beforehand and incorporated into the BAM solution.

Figure 1 shows how process mining relates to other well-known process analysis techniques. It shares the cost-effectiveness and objectivity with BAM solutions and the flexibility and exploratory nature with conceptual modelling approaches.

The basic building blocks for process mining algorithms are event traces from transaction logs. A trace describes the execution of sequentially related events. The definitions of which events are related to each other depend on the nature of the respective process mining projects. Examples of traces include process instances in a workflow management system, operations on related documents in ERP systems, or paths of tagged items through RFID-based sensor networks. As these examples show, the applicability of process mining is broad and process mining techniques are not restrained to environments that are necessarily process-aware. The only requirement is that the systems produce traceable event logs that document parts of its actual behaviour (Mans *et al.* 2009). Most workflow management systems,

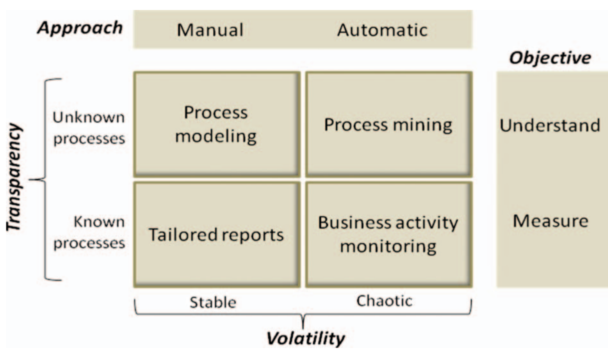

Figure 1. Techniques for analysing process quality.

ERP systems, case handling systems, CRM systems and middleware frameworks provide the necessary log information. Recent work also shows how embedded systems like X-ray machines, mobile phones, car entertainment systems and production systems increasingly log events suitable for process mining (van der Aalst 2008).

An event contains information about the operation (activity name or identifier), timestamp information about when it was executed, as well as a set of other context attributes, like references to involved resources. Context attributes are important for describing different process perspectives and the environment around alternative process paths or process characteristics.

Over the last years, process mining has also adopted data representations and techniques from disciplines like semantics and data mining. Semantic process mining has been introduced as a research area combining traditional process mining and ontologies. Traditional process mining is purely based on labels found in the event logs and they do not benefit from the actual semantics behind these labels. Semantic process mining takes advantages of the rich knowledge expressed in ontologies and associated event log entries and extracts semantic models that enable reasoning and describe discovered process flows at conceptual levels. Semantic process mining allows also for machine reasoning and usage of known business terms. A typical event log is very system technical, and by involving ontologies we can lift the analysis from a system technical level to a conceptual and more human readable level. It is fundamental for semantic process mining that the event log elements are annotated with relations to ontological concepts.

Use of traditional data mining algorithms has also shown its applicability within process mining. Trace clustering is a research field within process mining that applies clustering algorithms (like k-Means, Self Organizing Maps (SOM) and sequence clustering) on event log structures to extract homogeneous subsets that form more harmonised and understandable process mining models (Ferreira *et al*. 2007, Song *et al*. 2009). Rule extraction algorithms (like decision trees and decision rules) have shown its ability to identify unknown patterns and relations that can complement the understanding of discovered flow models (Rozinat and van der Aalst 2006). Data mining algorithms like non-parametric regression have also been applied in process mining projects to construct predictive models that can estimate the most likely outcome of initiated process instances (van Dongen *et al*. 2008).

3. Understanding enterprise process flows

The overall objective of process mining is to understand all log-related aspects of the current business process situation that are relevant to future process improvement. The focus is on current structures and activities, and the intention is to provide a set of dimensions that together form a comprehensive understanding of the processes.

As opposed to the situation in business process modelling, process mining considers real business processes to form both a *logical* and an *empirical* level of analysis, as shown in Figure 2. From a logical point of view, a process consists of activities that are carried out according to certain temporal or logical constraints. The logical process model gives us the procedures for how tasks can be performed and corresponds to business process models like EPC and BPMN. It may, for example, indicate that purchase orders are created on the basis of purchase requisitions, or that requisitions need to undergo a specific approval process before

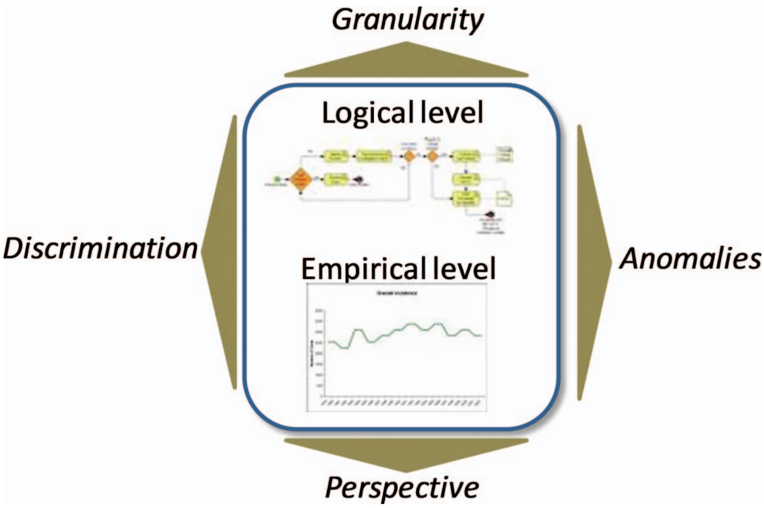

Figure 2. Dimensions for understanding real process flows.

purchase orders are created. When evaluating actual processes, the logical level indicates to what extent the real processes are consistent with the policies or standards of the enterprise.

However, the logical level cannot tell us whether resources are used efficiently or processes are performed satisfactorily. If the duration of one process varies significantly, for example, it is important to understand what characterises slow and fast process executions. The empirical level captures aggregated statistics on transaction times and frequencies, load, resources used, etc. For the business process analysis, the empirical data give us objective clues into the quality of processes and the utilisation of resources.

An underlying assumption in process mining is that both the logical and empirical level of actual business processes need to be uncovered and understood. Moreover, an in-depth analysis of business processes normally requires that the following four types of process viewpoints be supported:

- *Granularity*: Transaction logs are specified at the level of transaction codes, completion times and user IDs. At higher abstraction levels, these transactions turn into events and processes at various levels of detail. Similarly, user information in logs may be aggregated into roles, departments and companies for more high-level analyses of processes.
- *Perspective*: Even though a business process is usually assessed from a process perspective, it may be useful to restructure the model according to social networks, document flows, geographical dependencies, etc. These are all different perspectives that emphasise different aspects of processes.
- *Discrimination*: The discrimination viewpoint refers to the ability to run comparative analyses of selected process instance sets and expose the discriminating features of the sets. For example, a comparison of purchasing processes run by different departments may show that the total process duration is more dependent on the department than the ordered products or the respective vendors.

- *Anomalies*: For any selected set of process instances, it is useful to identify, characterise and present exceptional process executions. The anomalies typically describe cases with undesirable or surprising process properties and call for a closer analysis of how the enterprise should deal with these cases.

These four viewpoints stem from research paper studies and industrial feedback from several process mining case studies. In the following, we will use the framework above to discuss some important challenges in process mining and demonstrate how semantic approaches help us both with identifying logical and empirical process flow and present interesting views of the models.

4. Extraction of process models

EVS is a comprehensive process mining environment that has been developed over the last 5 years. It includes functionality for harmonising process-relevant data from ERP systems using ontological process models, extracting and presenting various representations of business processes and analysing process data with data mining methods. It is one of the few process mining tools that can deal with large-scale industrial data with incomplete transaction logs and inconsistent log structures. We will, in the following, use EVS to illustrate some of the challenges and opportunities of process mining environments in general. For more information about the implementation of EVS, see Ingvaldsen and Gulla (2006), Ingvaldsen and Gulla (2008) and Ingvaldsen (2011).

Process mining environments extract business process models that merge the logical structure of processes with empirical data describing the execution of these processes. Figure 3, for example, shows the delivery process that has been extracted by EVS from the SAP transaction log of a medium-sized Norwegian food company. The data spans 1 month of operations and includes a total 2092 SAP transactions. The real process forms an intricate web of activity dependencies that can be difficult to read and is far more complex than most employees' perception of the process.

The rounded boxes in the diagram are processes or process steps, and the arrows show temporal relationships between dependent activities. The size of the rounded boxes indicates the relative frequency of these process steps, i.e. frequently executed steps like *Create Invoice* have larger boxes than rare process steps. Similarly, the arrows are thicker the more frequently the work flow follows these paths. The process steps *Change Customer* and *Release Orders for Billing* are not connected because they carry no dependencies to other process steps. As shown in the figure, the delivery process is often initiated by changed sales orders that trigger the picking and packing of deliveries. The next step would normally be the outbound delivery, though it may also be followed by the cancellation of transfer orders that may necessitate the creation of a delivery recipient before the outbound delivery is completed. The last step in the process is the creation of invoices for the actual deliveries. When packing and preparing the outbound delivery or setting up the invoice, modifications of the original sales orders may often occur.

What complicates the model is that the users seem to go back and forth between many of the process steps. Even though this is not an efficient way of carrying out the process, there might be good reasons for these process cycles that compensate for the prolonged duration of the process or the duplication of work. There are also exceptional process steps like *Canceling transfer order* that tend to be ignored in

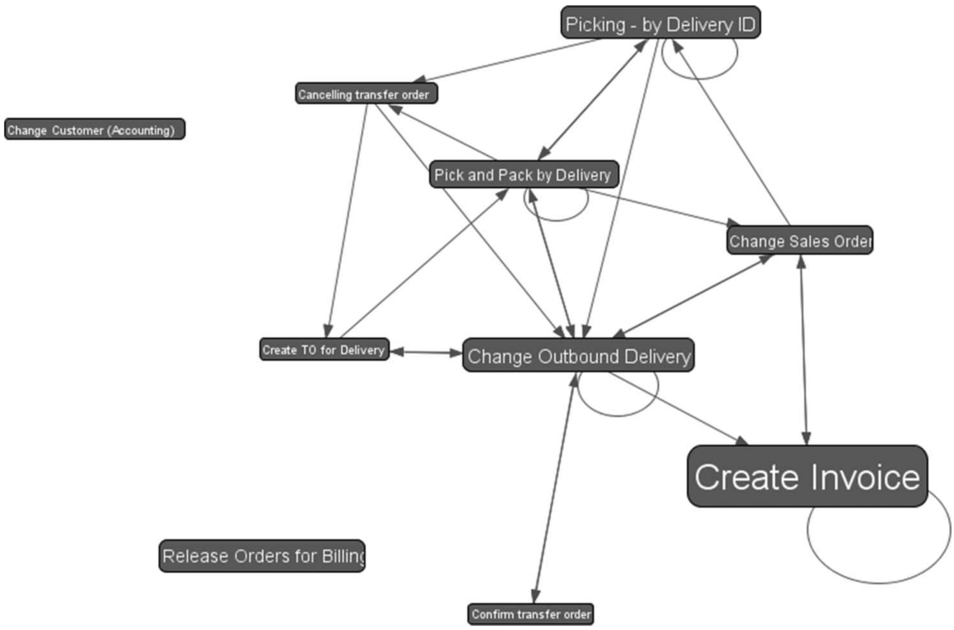

Figure 3. Constructed delivery process.

manually made process models, but may have a deep effect on the efficiency of the process as a whole.

On the empirical side, we notice that invoices are created far more frequently than outbound deliveries. In this particular test period, 1197 invoices were created on the basis of only 234 outbound deliveries, which may indicate that invoice processing should be analysed in more detail for possible optimisation. We also notice the relative high number of changed sales orders. Since changed sales orders may necessitate new invoices, the 144 changed orders also add to the worrying number of invoices produced. Interestingly, only one order for transferring goods internally was confirmed in this period.

The model shown in Figure 3 is the result of a log analysis process, in which a number of assumptions about the interpretation of logs have been made. The frequency counts for each activity in Figure 3 is shown in Table 1. The most important of these concern the following four structural challenges in industrial process mining:

- Operational ambiguity
- Temporal ambiguity
- Aggregation dependencies
- Manual process uncertainties

In the following, we discuss these challenges with respect to typical cases from real-world ERP systems.

4.1. Operational ambiguity

Fundamental to process mining is the mapping of events in transaction logs to the corresponding process steps and processes in the inferred business process models.

Often the logs are not detailed enough to enable an unambiguous mapping of events to processes, or they contain erroneous or noisy data that clutters the analysis.

Take for example the *Change Sales Order* transaction in the delivery process in Figure 3. This transaction may also be part of sales processes that address the negotiations with the customers, or it may be part of service processes that do not include any packed delivery at all. The identification of the correct process was possible in this case because the events in the transaction log linked back to preceding transactions, which were found to be part of the delivery process. If these links between events had not been recorded, which is often the case in enterprise systems, we would not be able to associate all events with the correct processes.

The excerpt from the SAP reference model in Figure 4 illustrates the complexity of recognising processes from transactions. As shown in the Figure 4, SAP's R/3 sales module is functionally described at five different abstraction levels, of which the two lowest ones are transactions and business processes. Each process typically links to a number of transactions, which means that these transactions need to be carried out to complete the process. However, a particular transaction in SAP may also be part of several processes. For example, the transaction *V-01: Create sales call* in Figure 4 can be part of either *Sales activity processing* or *Subsequent debt for empties and returnable packaging*. If both of these processes are running simultaneously, it may be hard to decide to which process a particular V-01 transaction belongs.

Business processes often cross organisational boundaries and may involve multiple IT systems. To describe end-to-end processes completely, event log structures from multiple sources need to be gathered and aligned. Many enterprise systems today produce high quality event log structures that describe much of the context of events and how the events relate to each other. There are, however, other systems that produce fragmentary event logs that do not easily lend itself to process mining.

Process aware middleware frameworks can help us extract event data across multiple systems, but without an alignment of events from multiple IT systems, process mining tools cannot make full use of the data. Unfortunately, there is still no accepted unified process log ontology available in industry that can relate events

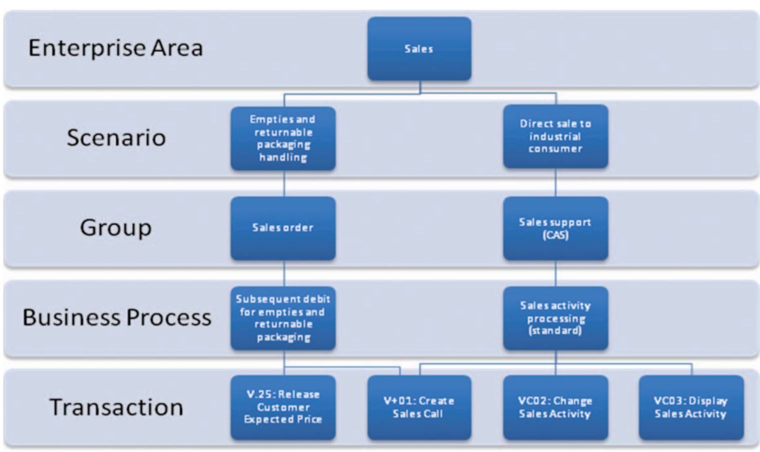

Figure 4. Ambiguous abstraction from transactions in SAP reference model.

from different enterprise systems. SAP's reference model, for example, describes ontological structures for their ERP system, though it is not compatible with other ERP systems' log structures and is too restrictive to capture all possible process behaviours. Process mining frameworks like EVS and ProM harmonise event logs on the basis of custom ontologies (Dongen *et al.* 2005, Ingvaldsen 2011). These ontologies have a minimal and extendable core, where industry and customer specific details are specified in related domain ontologies. Since there is no standardised log ontology, though, domain ontologies are often engineered or customised when new enterprise system solutions are to be analysed. This lack of standardised log structures prevents us from easily analysing inter-system processes or applying process mining environments on new ERP solutions.

4.2. *Temporal ambiguity*

Another challenge relates to enterprise systems' ability to log operations through multiple stages. Figure 5(a) shows a trace from an enterprise system that is capable of logging both operations' start times and completion times. Such a level of detail allows process mining algorithms to describe concurrency in the business flow (Schimm 2003, Wen *et al.* 2009). We can describe how and where the overall process time is spent and uncover any waiting periods between operations. In process improvement projects, we would like to identify such waiting periods and analyse to what extent they may be shortened or eliminated all together.

However, many enterprise systems log only the completion times of events. Figure 5(b) illustrates the same trace as in (a), but this time we only have information about events as they are completed. As we do not know when the operations started, we cannot distinguish between the time spent on the actual work and the time spent waiting for the operation to start. Although the preciseness of information has decreased, we can, however, describe overall processing times and the amount of time consumed between the completion of any two operations.

4.3. *Aggregation dependencies*

Aggregation dependencies refer to the tendency of aggregating items in operations differently in different parts of the process. The multiple items in one operation may be dealt with by multiple operations in the next stage, where each of these operations also includes any number of items from other preceding operations. Operations handling multiple items in the same operation are common in many business sectors and are often vital for efficient process completion. For process mining, however, aggregation dependencies pose some serious challenges in terms of trace identification and statistical calculations.

Most operations in enterprise systems are document-based, i.e. they create or modify documents that refer to the content of other documents. We say that the operations consume sets of documents (their input) and produce other documents or new versions of existing ones (their output). A typical document consists of a document header and a set of items that each shares the header information. Take, for example, the *purchase order* document in SAP. A purchase order is a sales contract between a buyer and a single vendor detailing the exact goods or services to be rendered. The document header contains common information like shipping address and payment terms, while the goods or services at the item level come with

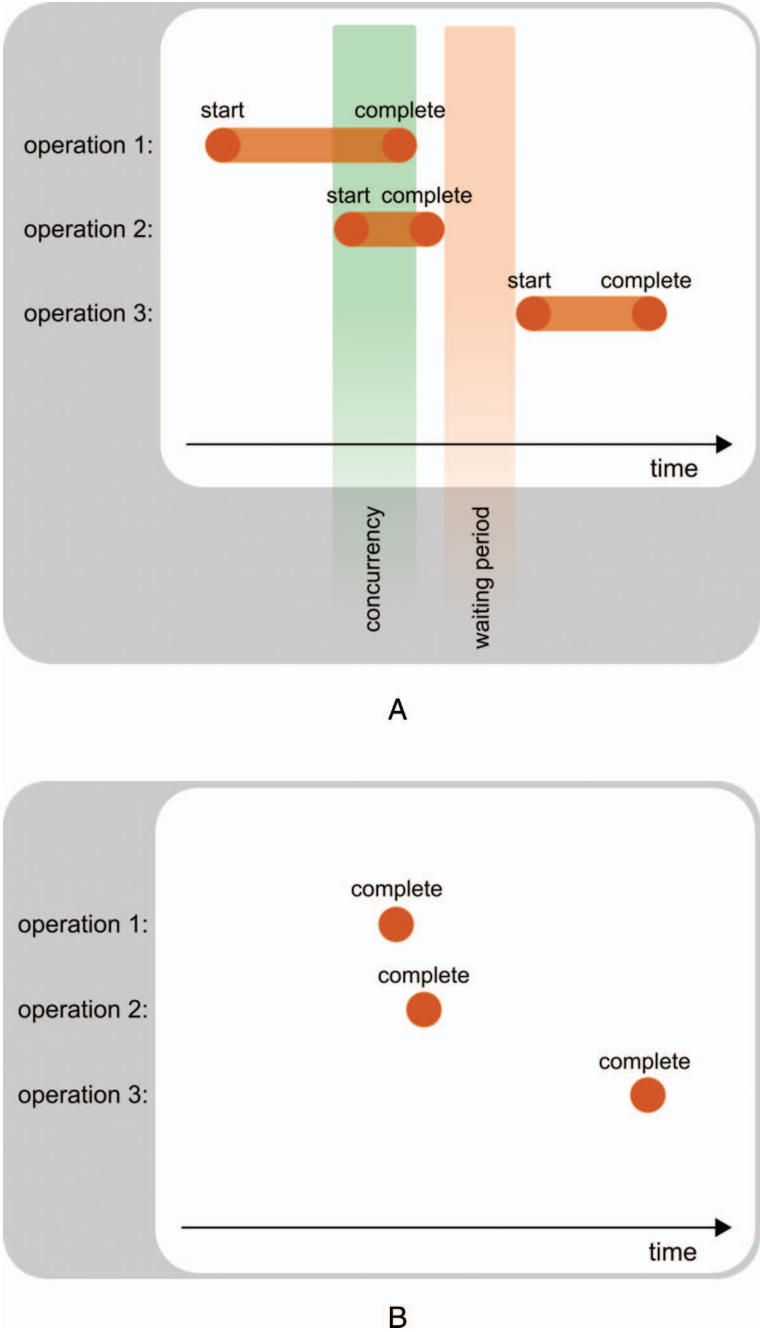

Figure 5. Detailed log in (a) specifies both start time and completion time of logs. In (b) only the completion time is recorded.

prices, quantities and other item-specific conditions. The items of a purchase order can refer to several items of previously made purchase requisitions, and they can themselves be referred to in one or more invoices and good receipts. These document references are traced in process mining to discover and describe real work practices, though the construction of process instances from document references is not straightforward.

Figure 6 illustrates the problem of analysing document flows in purchasing processes. Whereas the items from two purchase requisitions were gathered to construct a single purchase order, the vendor split up the delivery into two separate deliveries. The vendor also sent two separate invoices for the sold items. Each document creation in this flow is annotated with timestamp information (t1 to t7) that is used to describe the temporal order of the actual transactions. Transaction t1 is executed before t2, which again is executed before t3 and so on.

Figure 6 shows seven events that in principle could be merged and defined as one single trace. However, the strategy of combining all events that involve related documents tends to result in traces that span too large document flow networks and are too cluttered. It may be the case that only a subset of the items specified in *PR2* is included in *PO1*, and the remaining items in *PR2* may be included in other purchase orders that potentially involve other purchase requisitions, goods receipts and invoices, as well as other vendors.

Another strategy would be to create one trace for each end output, i.e. for each document that no other operation consumes. Since the invoice processing and goods receipt operations in Figure 6 do not produce any output that is consumed by other

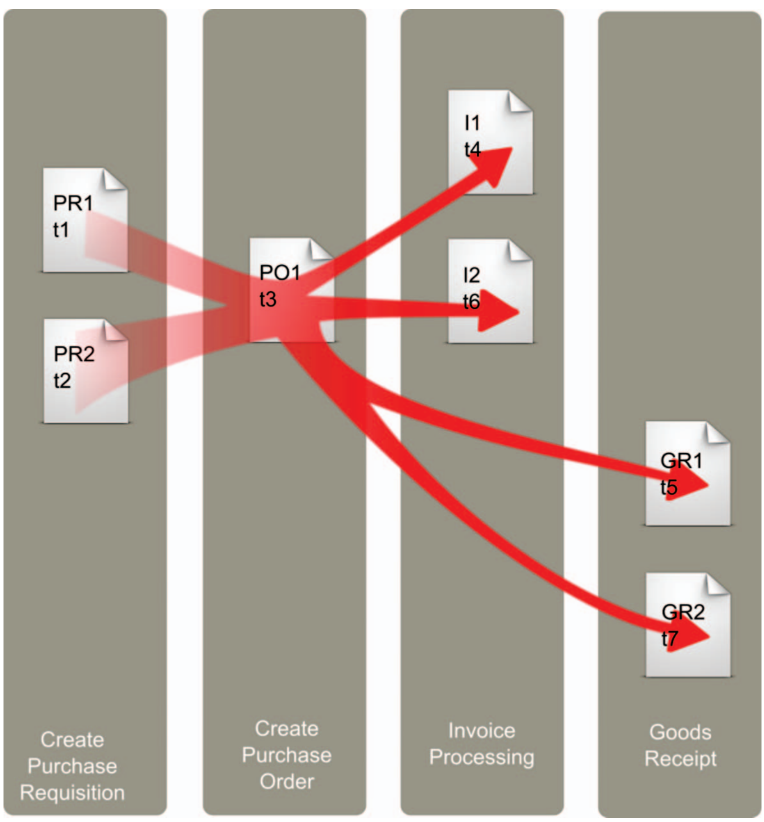

Figure 6. Aggregation complexity in SAP purchasing.

operations, their outputs are considered end outputs. By tracing the document consumption relations, we then get four traces that all start with the same purchase requisition operations (see Figure 7). This strategy is inspired by the emphasis on output in commonly cited business process definitions. Davenport defines a business process '*as a structured, measured set of activities designed to produce a specific output for a particular customer or market*' (Davenport 1993). Hammer and Champy have a definition that can be considered a subset of Davenport's, describing a process as '*a collection of activities that takes one or more kinds of input and creates an output that is of value to the customer*' (Hammer and Champy 1993). Although this strategy evades the problem of large and nested document networks and creates well defined trace structures, we easily end up with traces that refer to the same events. For example, the two purchase requisition operations and the purchase order operation are all included as events in all of the four traces. This means that certain events are repeated in several places, which may affect the reliability of statistical load calculations.

Figure 8 shows a model that could be extracted from the four traces in Figure 7. The business process model contains four steps, and each of the steps and the flow relationships between them are enriched with numbers stating how often the step or relation has occurred in the past. The numbers in front of the parentheses show occurrence counts based on the number of involved traces, while the numbers inside the parentheses are occurrence counts based on the number of involved events (independently of the number of traces). As we can see, the occurrence numbers for

Figure 7. Four traces deduced from transactions in Figure 6.

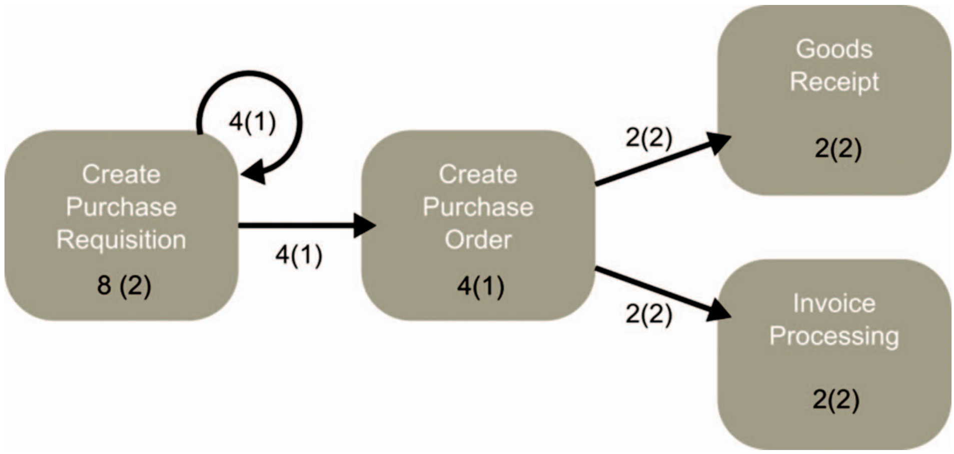

Figure 8. Naive merging of traces into process models.

the two first steps are misleading when the calculations are based on the number of involved traces.

Another typical example of aggregation dependencies comes from the shipping industry. Items sent from a vendor to a customer are processed in one package, but is often moved and handled in containers with other package on its way to the customer. Even though aggregation dependencies exist in many real-world processes, most process mining techniques still assume that events are only members of a single trace and do not deal with these cases. This simplification in many process mining tools has hampered their value in many real process improvement projects.

4.4. Manual process uncertainties

Even for IT-supported businesses, there are typically a significant number of manual operations that do not leave any traces in the underlying enterprise systems. Negotiations, human communication and decision making are all manual operations that are present in many business processes, but rarely recorded in transaction logs. As they do not leave traces in the transaction logs they are neither seen by the process mining algorithms nor included as elements of the discovered models.

This limitation of the log data poses a serious challenge to current process mining approaches. The manual operations are often among the most time-consuming parts of the process, and they can explain many of the problems encountered in the analysis of the recorded operations in the log. So far there is no good solution to this problem, and the users need manually to add information about manual operations and relate them to the operations supported by the enterprise system.

5. Analysis along multiple process viewpoints

Early process mining tools emphasised the mapping of log data onto a single process notation with no abstraction levels. The challenge was to construct complete process models that could be verified against process policy documents and allowed logical tests for consistency, completeness and deadlocks to be carried out. Formal modelling languages like Petri Nets were used for this analysis, and experts were normally needed to explain these models to the relevant process owners.

This single notation approach has proven insufficient for the complexity of most real-world business processes. The quality of processes is not only given by logical conditions on process execution, but depends heavily on the efficiency and effectiveness of the processes. It does not help much that processes conform to policy standards if they suffer from unacceptable delays, waste resources and incur unnecessary costs. The complex nature of many business sectors calls for deeper multi-faceted approaches to describing processes, in which different properties may be isolated, analysed and described at several levels of detail.

There is a growing realisation that process mining's goal of providing insight into business processes requires that the constructed models be easy to comprehend and tailored to the people responsible for the processes. Generating multiple process viewpoints has been suggested to expose the relevant aspects of the processes without distracting the users with their full complexity. Modern process mining approaches, found in ProM plug-ins and EVS, make use of ontologies and data mining to provide a rich set of viewpoints into the extracted process models:

- *Ontologies* form business-oriented abstraction hierarchies that are used to present the models in business terms at various level of detail.
- *Data mining* techniques like decision trees and clustering are used to split up process execution sets into interesting subsets or explain the characteristic features of selected executions.

Incorporated into a process mining tool, ontologies and data mining enable a more flexible generation of process model views.

5.1. Granularity

EVS stores an ontology that provides a process hierarchy from bottom-level transactions all the way to top level process areas. It also contains hierarchical information about users and other perspectives relevant for model presentation. For enterprise systems like SAP R/3 that come with reference models, this ontology is created from selected parts of these models. Figure 4, for example, shows the hierarchical levels built into the SAP reference model that is used by EVS to structure the extracted model hierarchically. A detailed account of how the EVS ontology is built up can be found in Ingvaldsen (2011).

Consider the delivery process model generated by EVS in Figure 3. It contains 10 transaction types that together describe the way sold products are packed, delivered and invoiced. Even though this model is not very complex, it may be useful to abstract from individual transactions and focus on the way processes interact. The ontology tells us that these 10 transaction types are parts of three different process activities. There is a sales order process that is responsible for creating, changing and submitting sales orders in the system. Then there is a complex delivery process that describes all available strategies for picking, packing and delivering goods internally and externally. Finally, there is a billing process that deals with invoices and release orders for billing.

If the user prefers a high-level model of the transaction model in Figure 3, EVS generates and presents the model in Figure 9. This model clearly shows that

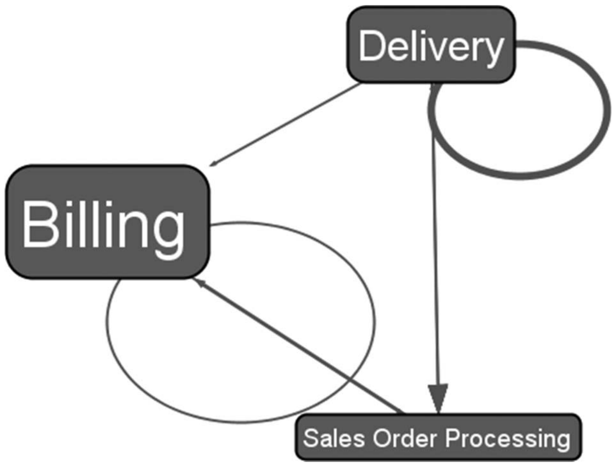

Figure 9. Abstracted view of model in Figure 3.

deliveries involve a substantial number of repetitive steps (thick circular arrow) that may slow down the whole process. When billing takes place after completing deliveries, the different sizes of the process boxes tell us that the company runs far more billing transactions than deliveries. This may indicate that they for some reason split up delivery documents when invoices are sent out, though we would need to go down to the transaction level to verify the real details.

For enterprise systems equipped with unambiguous reference models, it is fairly straightforward to build ontologies that provide good abstract views of the processes. Abstracting away from details or drilling down to verify data have proven very useful when business processes are inspected. It helps the users understand the overall process picture, though the desired level of detail depends on both the people involved and the type of analysis.

5.2. Perspectives

Business process models are normally structured around process activities that are temporally and logically dependent on each other and characterised by performance indicators like duration and frequency. The process view plays an important part in business process improvement projects, though there are other ways of analysing processes that also give important clues into process quality. It may, for example, be relevant to assess how employees collaborate to carry out processes or how goods travel across geographical distances. These alternative views of the business process model constitute different perspectives of the processes and emphasise aspects that are often difficult to address with process-structured models.

The document flow perspective shows how documents are created, changed and transferred as part of business processes. Since many enterprise systems are document-driven, this perspective is often more familiar to business process owners. Documents are concrete objects, and the document flow is visible to anyone in the business. Figure 10 shows the document flow perspective of the business process model from Figure 3. It tells us that the company handles substantially more delivery documents and invoices than sales orders. This means that they tend to split up sales

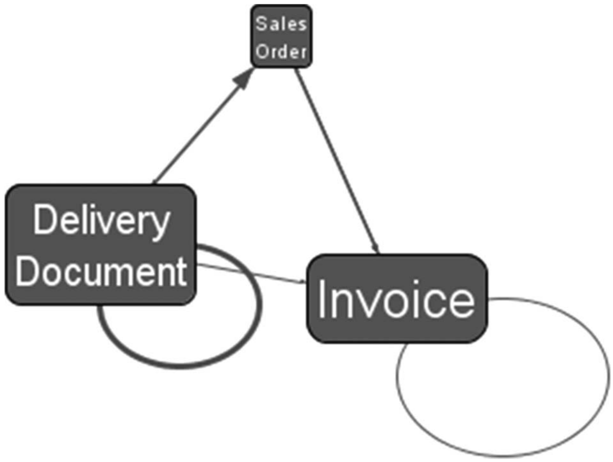

Figure 10. Document view of process.

orders into smaller units that are delivered and invoiced separately. The thick circular arrow for delivery documents reveals that the delivery process is rather complex with many operational loops updating the delivery documents.

The social network perspective is a useful perspective when organisational issues are analysed. This perspective displays the users involved and the way users collaborate or hand over work in processes. In Figure 11, the EVS-generated social network corresponding to the process model in Figure 3 and the document flow model in Figure 10 is shown. The size of the user boxes corresponds to the number of times they have taken part in the processes analysed. User *JASIV* hands over work to the users *NI-RLE, NI-LMR, NI-TM* and *PEHAN* in the delivery process. *JASIV* is mostly coordinating his work with *NI-LMR* (35 collaborations), who is also among the most central users in the whole delivery process. The delivery process in general seems very dependent on the availability of *NI-LMR, NI-RLE* and *BEALS*. The thick arrow linking *BEALS* to himself may either mean that he is responsible for several subsequent steps in the process or that he is frequently updating his own documents before sending them off.

As with the process perspective, we may present the social network at higher levels of abstraction. Rather than dealing with users, the network will then show how groups, departments or company codes interact to carry out processes. Technically, this change of perspective is feasible because the transaction logs are mapped onto a unified ontology. The model extraction module constructs graphical models that are structured around any major class in the ontology.

5.3. Discrimination

Discriminatory presentations divide process instances into groups or clusters based on certain behavioural similarities. Since absolute values of performance or quality are problematic for business processes, process mining tools tend to use comparative analyses to uncover potential weaknesses and deficiencies instead. A discriminatory comparative analysis describes performance-related or quality-related features using two overall strategies:

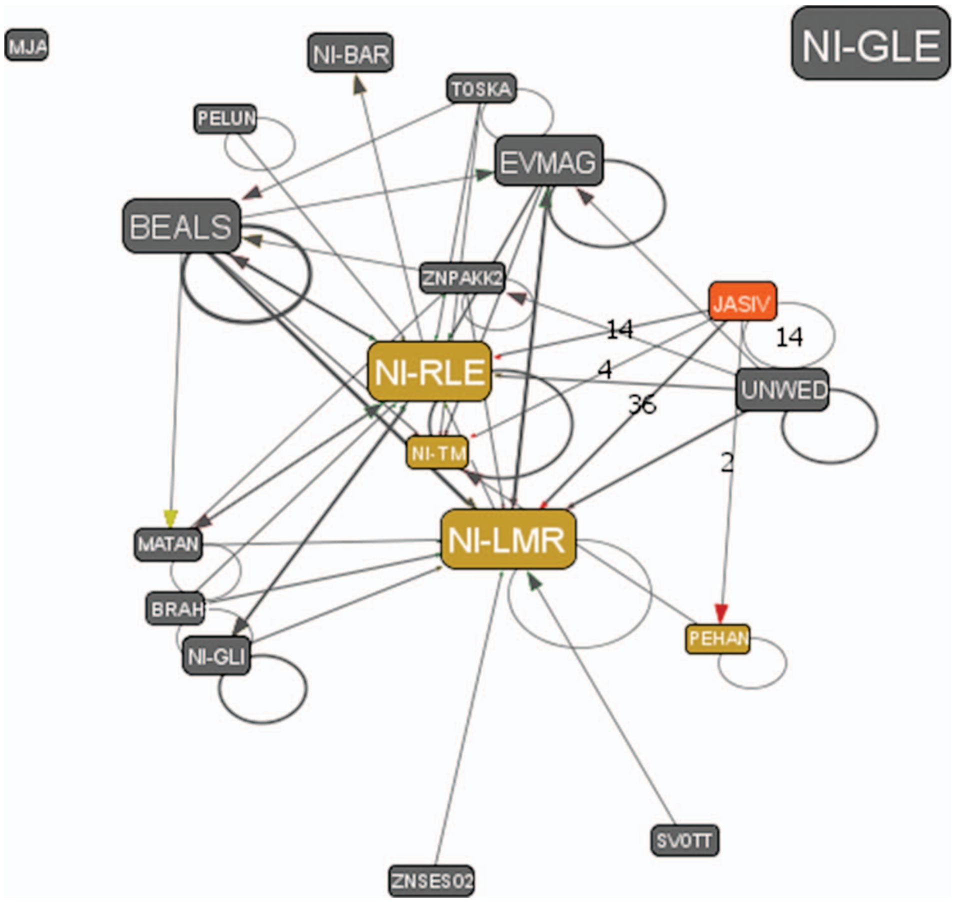

Figure 11. Social network for handover of work.

- *Discriminatory characterisation of selected clusters*. The intention is to understand the differences between specified groups of process instances. For example, we may want to know how two departments differ with respect to process duration for certain products.
- *Clustering from selected measurement variables*. The idea here is to specify a variable that may be used to evaluate qualitative aspects of processes and group process executions into clusters that can be further analysed. Process duration, resources involved, or sales values are often used as variables, and the process mining tool can use any number of variables to produce interesting process clusters.

In Figure 12, the system has clustered the delivery process from Figure 3 based on users involved and presented the delivery processes for user *JASIV*. Comparing Figures 3 and 12, we may see how JASIV contributes to the overall processes of the company.

The notion of *process quality* is problematic to define, and it has proven very difficult for process mining tools to evaluate quality in absolute terms. However, it often makes sense to compare process executions for different departments, different

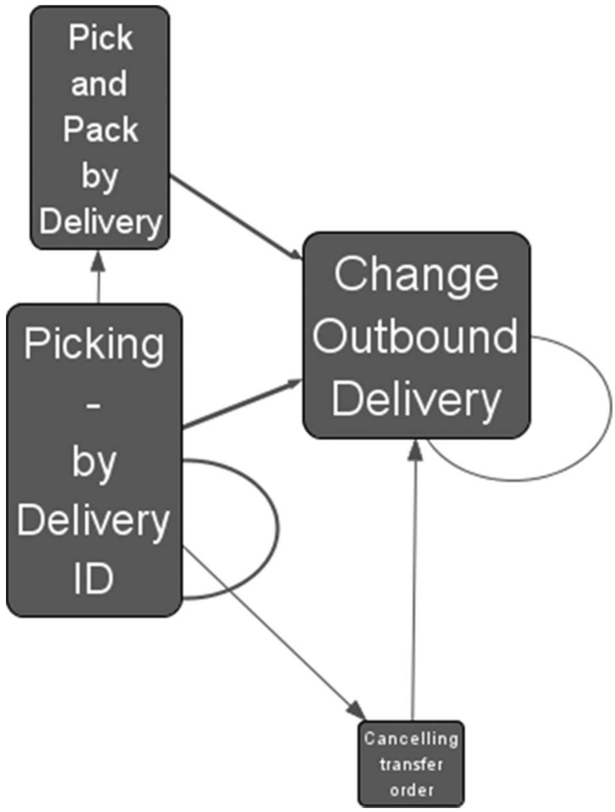

Figure 12. JASIV's part of the delivery process.

products or different vendors. Maybe more interesting is the way these tools may cluster processes instances along time lines, filtering out slow executions from fast ones and explaining what are the typical differences between the two. Even better, modern data mining techniques enable the system itself to suggest which features are the most suited for splitting up the process instances into meaningful and characteristic groups.

5.4. Anomalies

An anomaly, in the context of process mining, is any process execution that is strange, unusual, or for some reasons do not conform to an established normal behaviour. Many process oriented organisations have a defined strategy to streamline many of their processes. Examples of processes that can benefit from streamlined processes include product creation, order handling and service assurance (Vantrappen 1992). Process mining makes anomalies visible and quantifiable and allow for deeper investigations of their characteristics. Anomalies can be elements or steps in the process with a seldom occurrence, or processes that deviate with respect to measured performance indicators like duration values.

In many cases, exceptional process executions are rooted in exceptions in the surrounding environment that had to be handled with an improvised and practical solution. However, there are also cases where groups of deviating process executions share common event log characteristics. Examples of such anomalies are groups of processes that have significantly longer duration values than the norm, or processes where produced products do not meet given expectations. In such cases, process mining can be used in combination with data mining techniques to investigate underlying factors for their existence. Examples of such data mining techniques are decision trees/rules and regression analysis. By identifying likely root causes behind anomalies, we can address the underlying reasons directly and make actions for streamlining future process executions.

6. Discussion

Process mining allows organisations to gather explicit knowledge about how their processes are executed and how the true executions deviate from defined policies. It is a cost efficient technique that supports continuous process analysis and diagnosis.

In the pre-processing phase of process mining projects, event logs are extracted from various data sources like flat files and different database tables, and transformed to fit event log formats suitable for process mining. Such pre-processing work can be substantial (Ingvaldsen and Gulla 2008). In semantic process mining projects, the targeted event log elements are annotated with relations to ontological structures. For projects where such relations are not available or where the ontologies must be engineered from scratch, the costs of the pre-processing phase can limit and even exceed overall project gains.

Another challenge for process mining is to present process models discovered from event log observations in a readable way. Complex and spaghetti-like process models are typical for process mining discoveries in organisations with flexible process environments.

Although process mining can precisely spot improvement potentials, it is important to interpret process mining discoveries with caution and contextual awareness. It is especially important to see local optimisation potentials in an overall business context. We must also be aware that process mining discoveries are based on a subset of observations, constrained by time period and extent of respective event logs. To evaluate longer trends and seasonal variations, we need event logs covering longer time spans. Another challenge with event logs covering short time frames is that a relative large number of traces are incomplete and only contain a fraction of their events. If such incomplete traces are not handled properly, process mining discoveries can be corrupted and unable to represent the true business flows. An approach for eliminating the consequences of incomplete traces is to apply filters on the event logs stating that acceptable traces must pass through certain waypoints.

BAM tools share the promising aspects of cost efficiency and ability for continuous process diagnosis. In difference from process mining, BAM tools enrich presumed process structures with performance indicators such as frequency counts, throughputs and execution times. This way, organisations can monitor their process environment, see which process that consumes most of the time, and reveal processes that are frequently or seldom used. The performance measures are typically presented in aggregated drilldown structures. However, the disadvantage of BAM tools is that they depend on a correct a priori process model and do not discover process structures for executions that deviate from the presumed model.

Out in the industry, the term '*Automatic Process Discovery (ADP)*' is used as a synonym and industrial version of process mining. *Futura Reflect* by Futura Process Intelligence, *BPMne* by Pallas Athena, *Comprehend* by Open Connect, *Interstage Auto- mated Business Process Discovery and Visualization* by Fujitsu, and *Process Discovery Focus* by Iontas are some examples of commercial tools that offer some form of process discovery (Aalst 2009). Grigori (2004) describes a set of tools called the *BPI Tool Suite*. It is built on top of Hewlett-Packard's *Process Manager* and provides several features that offer various levels of automation for the management of process quality. As a front-end, they propose a Business Process Cockpit, which uses multi-coloured graphs to illustrate time-based process attributes, such as the processing times of individual activities (Mühlen 2004). The *HPPM intelligent Process Data Warehouse (PDW)* is another process analysis tool developed at Hewlett-Packard. The tool is described in Casati and Shan (2002), and they make extensive use of taxonomies to describe process quality and performance. Mühlen (2004) and Muehlen and Rosemann (2000) describe the *Process Information System based on Access (PISA) tool*, a BAM application targeting process analysis. By incorporating process models and event logs, it offers a client-server based architecture for evaluating process performance. Similar diagnostics are provided by the *ARIS Process Performance Manager (PPM)* (van der Aalst *et al.* 2008).

ProM is the most known process mining framework. It is developed at the Eindhoven University of Technology and it is a driving source behind much of the academic process mining innovations. It is built up of several plug-ins covering most of the dimensions for process mining and event log analysis. Many of these plug-ins are described in process mining publications. Although most plug-ins to ProM target the logical level, there are also plug-ins that target semantic analysis (Medeiros and van der Aalst 2009), charting and statistics specifically. Statistics are also used in many plug-ins to describe weights and load distributions in discovered process models (Dongen *et al.* 2005, van der Aalst *et al.* 2007).

ProM is an open source platform that includes an excess of 230 plug-ins. New plug-ins covering new areas of analysis functionality are constantly being developed by different research groups. Semantical analyses and parsing of event logs enriched with ontological annotations are found in separate ProM plug-ins, including:

- *Semantic organisational miner* – discovers groups of users that work together based on task similarity. Tasks are considered to be similar whenever they are instances of same concepts.
- *Semantic LTL checker* – verifies properties defined in terms of Linear Temporal Logic (LTL). By working on event logs with annotated relations to ontological structures, this plug-in allows users to formulate questions at a conceptual level (Medeiros *et al.* 2008).
- *Performance metrics in ontologies* – shows information about the throughput times of process instances and processing times of tasks and relates these measures to ontological concepts.
- *Ontology abstraction filter* – allows for manipulation of event log labels and discovery of process models of different abstraction levels. With this plug-in, the user can swap the labels in the raw event log with labels from related ontological concepts. For instance, changing the activity label of an event with a more general definition found in the ontologies. This label manipulation

allows ProM user to involve ontological structures and information in other traditional process mining plug-ins that are not aware of semantic annotations.

There are plug-ins that enable construction of formal structures like Petri Nets and event process chains (EPC) and plug-ins like the Fuzzy Miner that make use of interactive process models where the user can select appropriate complexity levels. Also other plug-ins target the granularity dimension. The ontology abstraction filter is one such plug-in that supports discovery of process models of different abstraction levels. Plug-ins to ProM cover both the activity and organisational process perspectives. Other plug-ins target the anomalies dimension. The performance metrics in ontologies plug-in provide feedback about frequencies and processing times. The discrimination dimension in ProM is handled by filtering and trace clustering functionality. The filtering functionality allows users to restrain the analysis to only incorporate traces that involve a start or end activity. The goal of trace clustering is to identify natural and homogeneous clusters of traces in the event logs. The amount of relations and elements observed in event logs are often so extensive that process models describing the observed flows become very complex. With trace clustering we eliminate much of the overall complexity that can describe the process flows within each of the identified and homogeneous clusters. Different plug-ins for trace clusters are developed for ProM and they are distinguished by their data pre-processing and selection of clustering algorithms.

Table 2 highlights differences between ProM and EVS with respect to the four process viewpoints described in section 3. In this article, we have highlighted industrial challenges such as operational ambiguity, temporal ambiguity, aggregation dependencies and manual process uncertainties. Event logs with annotations to ontological structures can enable process mining to describe discovered process models with business terms, streamline different event log sources and hide system technicalities from the raw event logs. However, noisy or erroneous data are still a challenge and the quality of the process mining results is never better than the quality of the input data.

Both temporal ambiguity and manual process uncertainties relate to the quality of the raw event log data. Discovered process mining models can only incorporate information related to the underlying event logs, and sufficient time information and logging is necessary for distinguishing between time spent on the actual work and the

Table 1. Operation frequencies for each model element in Figure 3.

Operation	Frequency
Create Invoice	1197
Change Sales Order	144
Release Orders for Billing	228
Change Customer (Accounting)	19
Change Outbound Delivery	234
Cancelling Transfer Order	5
Create TO for Delivery	5
Confirm Transfer Order	1
Pick and Pack by Delivery	108
Picking – by Delivery ID	151

Table 2. Viewpoints supported by ProM and EVS.

Viewpoints	*ProM*	*EVS*
Granularity	Ability to set level of detail of discovered process models (Fuzzy Miner plug-in) and extract models at different conceptual levels (ontology abstraction filter plug-in)	Drill down functionality that follows abstraction levels of ontology concept annotations.
Perspective	Activity and organization perspectives	All ontological annotations on events serve as potential process perspectives.
Discrimination	Event log filtering and trace clustering	Ontology-driven search for selecting analysis data.
Anomalies	Combination of manual identification and process mining approaches used	Decision trees for characterizing trace subsets. Domain ontologies for describing anomalies with familiar business terms.

time spent on waiting. For readers of discovered process mining models, it is important to be aware of manual operations and not to view discovered process mining models as complete process knowledge.

EVS handles aggregation dependencies by creating one trace for each end output, and allowing events to occur in multiple traces. With such an approach, we must be careful when statistics and performance measures are calculated and make sure that each of the involved events are only counted once. In ProM and MXML, each event is assumed to only be a member of a single trace.

There are different approaches to allow users to visualise process models at different levels of granularity. EVS and the ontology abstraction filter plug-in to ProM allow users to create process models according to different generalisation levels defined in event log-related ontologies. While this approach handles granularity in terms of generalisation of involved concepts, the Fuzzy Miner plug-in to ProM handles granularity in terms of logical complexity and readability.

In traditional event, log data formats to ProM (MXML), the process activity name and the user executing the operation are elementary. They are central elements for discovering models describing the control flow, and handover of work and social networks. In EVS, anything that is related to the context around an event is treated as equally important, and type of concept can be used to present a process perspective.

For the discrimination viewpoint, ProM and EVS contain techniques that complement each other, allowing users to define the set of events and traces that are interesting for further process mining analyses. ProM allows users to define known properties for the subset of traces we want to analyse. For instance, defining that all relevant traces should start or end with a certain activity. In this system, the user can also use trace clustering to automatically separate the whole event log into groups with minimal overlap and reduced complexity ('spaghettiness'). In EVS, free text search is used to give the user a simple, yet powerful, way of retrieving traces relevant for analyses.

Process mining can show all details of the event log-related business flows and enrich discovered models with statistical measures. All together, discovered process mining models are a good basis for identifying anomalies in the business flows. EVS allows users to apply data mining techniques to reveal hidden patterns, common characteristics for a labelled group of traces.

Gartner Group has investigated the influence of BPM tools, and they predict that by 2012, 20% of customer-facing processes will be knowledge-adaptable and assembled just in time to meet the demands and preferences of each customer, assisted by BPM technologies. Pressure to reduce the latency of change in business processes is driving a need for more dynamic and systemised measures. Gartner Group has further predicted that by 2013, dynamic BPM will be an imperative for companies seeking process efficiencies in increasingly chaotic environments (Gartner Group 2010). Process mining targets the diagnosis needs in BPM and can highlight unexploited efficiency potentials in such dynamic and rapidly changing environments.

7. Conclusion

Although process mining is a well established research discipline, it has not gained wide industrial adoption. Traditionally, information systems provide event logs for system technical reasons rather than the analysis of business process quality. With information-rich high-quality event logs, organisations are able to monitor their process performance and investigate to what extent they work in accordance with defined policies. It remains to be seen if information system vendors will improve the quality of log information structures as the capabilities of process mining increase and its importance is acknowledged.

Traditional process mining has proven itself useful for extracting flow models that describe how activities and organisational units depend on each other in dynamic business process environments. Although current process mining techniques have matured, the analysis they support is limited to describing process characteristics based on syntactic labels in logs. Semantic process mining takes advantages of the rich knowledge expressed in ontologies and associated event log entries and extracts semantic models that enable reasoning and describe phenomena at different conceptual levels and from different perspectives (De Medeiros *et al.* 2007, Pedrinaci *et al.* 2008). Ontologies are also instrumental in the integration and harmonisation of separated event log fragments.

Real-world processes are often complex. Process mining can eliminate noise from process relevant data and simplify the comprehension of complex process environments. Process mining is a complement to other manual techniques to gather business process knowledge and the main strengths of process mining discoveries are accuracy and objectivity.

Datasets where process mining can show its importance are increasing. IT supports an increasing amount of business operations. Many logistics flows involve assets tagged with RFID chips or barcodes that are read and logged through sensor networks, and people are now surrounded with mobile agents that can track gps locations and other online actions. Process mining opens for out of-the-box thinking for how to visualise business processes backed up with empirical data. The ability to use interactivity and temporal animations allow for new process visualisations that goes beyond traditional and static process modelling.

References

Aalst, W.M., 2009. Using process mining to generate accurate and interactive business process maps. *Business Information Systems Workshops* [online], 37. Available from: http://dx.doi.org/10.1007/978–3–642–03424–4_1 [Accessed 11 June 2011].

Casati, F. and Shan, M., 2002. *Semantic analysis of business process executions* [online]. Berlin, Heidelberg: Springer Berlin Heidelberg. Available from: http://www.springerlink.com/content/79hr2ey4fpeteyyu [Accessed 11 June 2011].

Cook, J.E. and Wolf, A.L., 1998. Discovering models of software processes from event-based data. *ACM Transactions on Software Engineering and Methodology*, 7 (3), 215–249.

Davenport, T.H., 1993. *Process innovation: reengineering work through information technology*. Boston, Massachusetts: Harvard Business Press.

De Medeiros, A., *et al*., 2007. An outlook on semantic business process mining and monitoring. *Lecture Notes in Computer Science* [online], 4806. Available from: http://www.springerlink.com/index/g4u4543u17272t57.pdf [Accessed 11 June 2011].

Dongen, B.F., *et al*., 2005. The ProM framework: a new era in process mining tool support. *In*: G. Ciardo and P. Darondeau, eds. *Applications and Theory of Petri Nets* [online]. Berlin: Heidelberg. Available from: http://dx.doi.org/10.1007/11494744_25 [Accessed 11 June 2011].

Ferreira, D. and Gillblad, D., 2009. *Discovering process models from unlabelled event logs* [online]. Available from: http://dx.doi.org/10.1007/978–3–642–03848–8_11 [Accessed 11 June 2011].

Ferreira, D.M., *et al*., 2007. Approaching process mining with sequence clustering: experiments and findings. *Lecture Notes in Computer Science* [online], 4714 (1). Available from: http://www.springerlink.com/index/744g104721811151.pdf [Accessed 11 June 2011].

Folino, F., *et al*., 2009. Discovering expressive process models from noised log data. *In*: *Proceedings of the 2009 international database engineering & applications symposium*. ACM, 162–172. Available from: http://portal.acm.org/citation.cfm?id=1620432.1620449 [Accessed 11 June 2011].

Funk, M., *et al*., 2008. Semantic concepts in product usage monitoring and analysis. (ES Reports, No. ESR-2008-10). Eindhoven: Technische Universiteit Eindhoven, 20.

Gartner Group, 2010. *Predicts 2010: business process management will expand beyond traditional boundaries*. Stamford, Connecticut: Gartner Inc.

Grigori, D., 2004. Business process intelligence. *Computers in Industry* [online], 53 (3). Available from: http://linkinghub.elsevier.com/retrieve/pii/S0166361503001994 [Accessed 11 June 2011].

Günther, C.W. and van der Aalst, W.M.P., 2007. Fuzzy mining – adaptive process simplification based on multi-perspective metrics. *In*: *BPM*. Berlin Heidelberg: Springer-Verlag, 328–343.

Hammer, M. and Champy, J., 1993. Reengineering the corporation. *New York* [online], 18 (3). Available from: http://0-proquest.umi.com.novacat.nova.edu/pqdweb?did=793913&Fmt=7&clientId=17038&RQT=309&VName=PQD [Accessed 11 June 2011].

Ingvaldsen, J. and Gulla, J., 2008. *Business process management workshops* [online]. Berlin, Heidelberg: Springer Berlin Heidelberg. Available from: http://www.springerlink.com/content/4w3213725n4686hr [Accessed 11 June 2011].

Ingvaldsen, J.E. and Gulla, J.A., 2006. Model-based business process mining. *Information Systems Management*, 23 (1), 19–31.

Ingvaldsen, J.E., 2011. Ontology driven business process intelligence. *In*: V. Sugumaran and J.A. Gulla, eds. *Applied semantic web technologies*. UK: Taylor and Francis. ISBN: 1439801568, 9781439801567.

Mans, R., *et al*., 2009. Application of process mining in healthcare – a case study in a Dutch hospital. *In*: *Biomedical engineering systems and technologies, communications in computer and information science* [online]. Vol. 25. Springer Berlin: Heidelberg, 425. Available from: http://www.springerlink.com/index/xk8212082163l0r4.pdf [Accessed 11 June 2011].

Maruster, L., *et al*., 2006. A rule-base approach for process discovery. Dealing with noise and imbalance in process logs. *Data Mining and Knowledge Discovery*, 13, 67–87.

Medeiros, A.D., *et al*., 2008. Semantic process mining tools: core building blocks. *In*: W. Golden, *et al*., eds. *Proceedings of the 16th European Conference in Information Systems (ECIS)*, 9–11 June 2008, Galway, Ireland. 1953–1964.

Medeiros, A.D. and van der Aalst, W., 2009. *Process mining towards semantics* [online]. Available from: http://dx.doi.org/10.1007/978–3–540–89784–2_3 [Accessed 11 June 2011].

Muehlen, M.Z. and Rosemann, M., 2000. *Workflow-based process monitoring and controlling – technical and organizational issues* [online]. HICSS. Available from: http://portal.acm.org/citation.cfm?id=820360 [Accessed 11 June 2011].

Mühlen, M.Z., 2004. *Workflow-based process controlling: foundation, design, and application of workflow driven process information systems (advances in information systems and management science)* [online]. Logos. Available from: http://www.amazon.co.uk/Workflow-Based-Process-Controlling-Application-Information/dp/3832503889 [Accessed 11 June 2011].

Pedrinaci, C., *et al*., 2008. Sentinel: a semantic business process monitoring tool. *In*: *Proceedings of the first international workshop on ontology-supported business intelligence* [online]. New York, USA: ACM. Available from: http://portal.acm.org/citation.cfm?id=1452568 [Accessed 11 June 2011].

Rozinat, A. and van der Aalst, W., 2006. *Decision mining in business processes* [online]. BPM Center Report BPM0610 BPM center org. Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.115.2763&rep=rep1&type=pdf [Accessed 11 June 2011].

Rozinat, A., *et al*., 2009. Process mining applied to the test process of wafer scanners in ASML. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews* [online], 39 (4). Available from: http://portal.acm.org/citation.cfm?id=1656835 [Accessed 11 June 2011].

Schimm, G., 2003. Mining most specific workflow models from event-based data. *In*: W.M.P. van der Aalst, ed. *Business process management*. Heidelberg: Springer Verlag, 265–281.

Song, M., Günther, C.W., and Aalst, W.M., 2009. *Trace clustering in process mining* [online]. Available from: http://dx.doi.org/10.1007/978–3–642–00328–8_11 [Accessed 11 June 2011].

StereoLOGIC, 2010. *Press release: IAG partners with StereoLOGIC* [online]. Press Release. Available from: http://www.stereologic.com/news.htm [Accessed 11 June 2011].

van der Aalst, W., 2008. Challenges in business process analysis. *In*: *Enterprise information systems: 9th international conference* [online], 12–16 June 2007, Funchal, Madeira: Iceis. Revised selected papers. Available from: http://www.springerlink.com/index/l7vm0414644515j7.pdf [Accessed 11 June 2011].

van der Aalst, W. and Weijters, A., 2004. Process mining: a research agenda. *Computers in Industry* [online], 53 (3). Available from: http://linkinghub.elsevier.com/retrieve/pii/S0166361503001945 [Accessed 11 June 2011].

van der Aalst, W., *et al*., 2008. Process mining: a two-step approach to balance between underfitting and overfitting. *Software & Systems Modeling* [online], 9 (1). Available from: http://www.springerlink.com/content/u43v780550278h4l [Accessed 11 June 2011].

van der Aalst, W.M., *et al*., 2007. ProM 4.0: comprehensive support for real process analysis. *In*: J. Kleijn and A. Yakovlev, eds. *ICATPN*. *Lecture Notes in Computer Science*. New York: Springer, 484–494.

van Dongen, B., Crooy, R., and van der Aalst, W., 2008. Cycle time prediction: when will this case finally be finished. *In*: *Proceedings of the 16th international conference on cooperative information systems* [online], Coop IS, Springer. Available from: http://www.springerlink.com/index/8217374727126740.pdf [Accessed 11 June 2011].

Vantrappen, H., 1992. Creating customer value by streamlining business processes. *Long Range Planning* [online], 25 (1). Available from: http://dx.doi.org/10.1016/0024–6301(92)90310-X [Accessed 11 June 2011].

Weijters, A.J. and Medeiros, A.K., 2006. *Process mining with the HeuristicsMiner algorithm* [online]. Eindhoven: BETA Working Paper Series, WP 166, Eindhoven University of Technology. Available from: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.118.8288 [Accessed 11 June 2011].

Wen, L., *et al*., 2009. A novel approach for process mining based on event types. *Journal of Intelligent Information Systems* [online], 32 (2). Available from: http://portal.acm.org/citation.cfm?id=1519347 [Accessed 11 June 2011].