A review of business process mining: state-of-the-art and future trends

A. Tiwari and C.J. Turner
School of Applied Sciences, Cranfield University, Cranfield, UK, and
B. Majeed
British Telecom, Computational Intelligence Group, Ipswich, UK

Abstract

Purpose – This paper seeks to examine the area of business process mining, providing an overview of state-of-the-art techniques. An outline of the main problems experienced in the practice of process mining is given along with reference to work that addresses the most challenging issues experienced in this field. This paper also aims to examine the application of soft computing techniques to process-mining problems.

Design/methodology/approach – This paper is based on a comprehensive review of literature covering more than 50 research papers. These papers are analysed to identify current trends and future research directions in the field.

Findings – Process-mining techniques are now becoming available as graphical interface-driven software tools, where flow diagram representations of processes may be manipulated as part of the mining task. A significant number of papers employ mining heuristics to aid in the task of process discovery. Soft computing algorithms are increasingly being investigated to aid the accuracy and speed of mining algorithms. Many papers exist that address common mining problems such as noise and mining loops. However, problems such as duplicate tasks, mining perspectives and delta analysis require further research.

Originality/value – The contribution of this paper is to provide a summary of the current trends in process-mining practice and point out future research directions. A review of the work in this new and expanding area has been provided in the form of illustrative graphs and tables that identify the trends in this area. This is the most comprehensive and up-to-date review of business process-mining literature.

Keywords Process management, Work flow, Computer applications

Paper type General review

1. Introduction

The need for companies to learn more about how their processes operate in the real world is a major driver behind the development and increasing use of process-mining techniques. The practice of business process mining derives from the field of data mining. Data mining refers to the extraction of knowledge from large data sets though identification of patterns within the data. Data mining practice has been developed and adapted to create the business process-mining techniques that are now being used to mine data logs containing process execution data to reconstruct actual business processes. In addition customised algorithms (and algorithms borrowed from other fields of computing) have been developed specifically to address the needs of process mining specialists. Business process-mining techniques use execution logs of business processes. These are typically hosted within business process management (BPM)

Business Process Management Journal
Vol. 14 No. 1, 2008
pp. 5-22

1463-7154
DOI 10.1108/14637150810849373

systems, though they may also be accessible though other process-related systems installed within a company.

Currently many approaches to process mining make use of heuristic algorithms ("rules of thumb" based on assumptions about business process patterns). Though, a growing number of methods make use of soft computing techniques such as genetic algorithms and neural network technologies. This paper will examine a number of process-mining approaches and outline the major technical issues process mining as a practice has to overcome.

2. Process-mining techniques

Agrawal *et al.* (1998) were early pioneers of process mining. Their algorithmic approach to process mining allowed the construction of process flow graphs from execution logs of a workflow application. The discipline of process mining also has its roots in the work of Cook and Wolf (1998a) who attempted to discover software process models from the data contained in event logs. van der Aalst (2004a) compares the method of extracting process models from data with that of distillation. In terms of business process mining, van der Aalst (2004a) states that almost any transactional information system can provide suitable data.

van der Aalst (2003) identifies two broad types of workflow meta models. These are graph- and block-orientated models; each with their own language and graphical representation. Aguilar-Saven (2004) adds net-based languages to this definition (with block-oriented models/languages being grouped under the term workflow languages). van der Aalst (2003) does not make this distinction between net and graph models describing net-based models, such as Petri nets as a form of graph-oriented model.

The most common form of graph oriented meta-model is the directed graph. Agrawal *et al.* (1998) was one of the first to use directed graphs in process mining. This author describes a number of constructs involved in the actual graph. Activities, usually enclosed in boxes or circles, are referred to as vertices and the arrows between the activities, that indicate the direction of flow, are known as edges.

Some workflow meta-models may be used to define workflow models as well as act as a language for the display of process-mining activities. An example of this is provided by Herbst and Karagiannis (2004) and their InWoLvE workflow mining system. It is often useful to be able to define an "ideal" workflow template so that mined process models may be compared against it for conformance purposes.

Cook and Wolf (1998a), while not directly relating their work to that of business process discovery, did examine the use of three statistical analysis methods for use in mining tasks:

(1) *RNet*. A statistical method that examines past behaviour to describe a process state.

(2) *Markov (Markovian approach)*. A hybrid statistical and algorithmic approach that looks at both past and future behaviour to define a potential current state.

(3) *Ktail*. This algorithmic method examines future behaviour to describe a potential current state.

It was the opinion of Cook and Wolf that the Markov and Ktail approaches promised the most in terms of process discovery, allowing a process analyst to introduce existing knowledge in order to refine the results. In later work, the authors Cook *et al.* (2004)

extend their findings to allow for the discovery of concurrent models of behaviour in workflow systems. Concurrent workflows are characterised by simultaneous threads of process execution. There are a number of techniques that may be used to perform mining of business processes:

- *Genetic algorithms*. Algorithms designed around the process of Darwinian natural selection (Alves de Medeiros *et al.* 2004a, b, c).
- *General algorithmic approach*. Custom algorithms designed for mining processes by individual authors (van der Aalst and Song, 2004).
- *Markovian approach*. An algorithm that examines past and future behaviour to define a potential current state (Cook and Wolf, 1998a).
- *Neural network*. Models the human mind in its ability to "learn" and then identify patterns in data (Cook and Wolf, 1998a).
- *Cluster analysis*. Divides a group of solutions into homogenous sub groups (Schimm, 2004).

Table I highlights the main references concerned with the application of these techniques to business process mining. In the compilation of this table, the authors found that journals in the subject areas of BPM, data management, knowledge engineering and software engineering provided the best source of papers in the area for business process-mining techniques. The subject area of manufacturing process management was excluded from the review since it lies outside the focus of this paper on business processes.

Figure 1 shows the number of papers that detail process-mining techniques. These papers are listed in Table I. It is interesting to note that most authors in this area have devised custom (other) algorithmic approaches to process mining, 39 per cent of the papers in total. Only 10 per cent of papers detail data mining-based approaches and only 6 per cent of papers detail the use of soft computing techniques when applied to process-mining tasks. Petri net modelling seems to be the most popular notation method with 29 per cent of papers detailing its use in the modelling of business processes.

Much of the work reported in the area of process mining deals with the problem of discovering a process model from a set of process instances. In this vein, such discovery, using a directed graph, a finite state machine or a Petri net for representing process instances, aims at discovering a process model that best describes the set of process instances (van der Aalst, 2003). However, it may not always be possible to assume the existence of a single process model to which all process instances comply; here the goal must be to discover a set of frequently occurring temporal patterns. An example of the work in this area is the paper of Hwang *et al.* (2004) who model frequently occurring temporal patterns from event-based data. Dong and Li (1999) describe a method of identifying emerging patterns from data. While their work does not consider business processes it may nevertheless be applied to this domain, identifying changes in the way processes are being carried out in a live environment.

Other notable papers in this area include Hwang (2002) who present an algorithm that they claim can produce process flow graphs in a more detailed and efficient manner; an algorithm which builds upon the process-mining activities of Agrawal *et al.* (1998).

Table I.
Main techniques used for process mining

Papers	Process-mining techniques: Data mining techniques	Genetic algorithms	Markovian approach	Cluster analysis	Neural networks	Other algorithmic approaches	Modelling technique used: Event-driven process chains	Petri nets
van der Aalst *et al.* (2005a)		X					X	X
van der Aalst *et al.* (2005b)						X		X
Zhang *et al.* (2003)						X		
Herbst and Karagiannis (2004)	X							
van der Aalst and Alves de Medeiros (2005)						X		
van der Aalst (2004b)						X		X
Schimm (2004)	X			X				
Schimm (2003)	X							
Greco *et al.* (2005)	X							
Weijters and van der Aalst (2001)						X		X
Alves de Medeiros *et al.* (2004a)		X						X
Agrawal *et al.* (1998)	X							
Adams *et al.* (2001)				X				
Dongen *et al.* (2005a, b)						X		
Dustdar *et al.* (2004)						X		X
Dongen and van der Aalst (2005b)						X	X	X
Dongen *et al.* (2005a)						X	X	X
Maruster *et al.* (2002b)						X		X
Weijters and van der Aalst (2003)						X		X
van der Aalst and Song (2004)						X		X
Alves de Medeiros *et al.* (2005)		X						X
Dongen and van der Aalst (2004a)						X		X
Maruster *et al.* (2002a)						X		X
Dongen and van der Aalst (2004b)						X	X	X
Alves de Medeiros *et al.* (2003)						X		X
Gaaloul and Godart (2005)						X		

(*continued*)

Table I.

Papers	Process-mining techniques: Data mining techniques	Genetic algorithms	Markovian approach	Cluster analysis	Neural networks	Other algorithmic approaches	Modelling technique used: Event-driven process chains	Petri nets
Hammori *et al.* (2004)	X							
Golani and Pinter (2003)						X		
Cook and Wolf (1998a)			X		X			
Hwang (2002)	X							
Cook *et al.* (2004)						X		X
Hwang *et al.* (2004)						X		
Alves de Medeiros *et al.* (2004c)						X		X
Chen and Yun (2003)						X		
Greco *et al.* (2004)				X				
van der Aalst *et al.* (2002)						X		X
Dongen *et al.* (2005b)						X		
Mannila and Rusakov (2001)			X					
Herbst and Karagiannis (1998)			X					
Cook and Wolf (1998b)						X		
van der Aalst (2005)						X		X
Alves de Medeiros *et al.* (2004b)						X		X
Dongen and van der Aalst (2005a)						X		

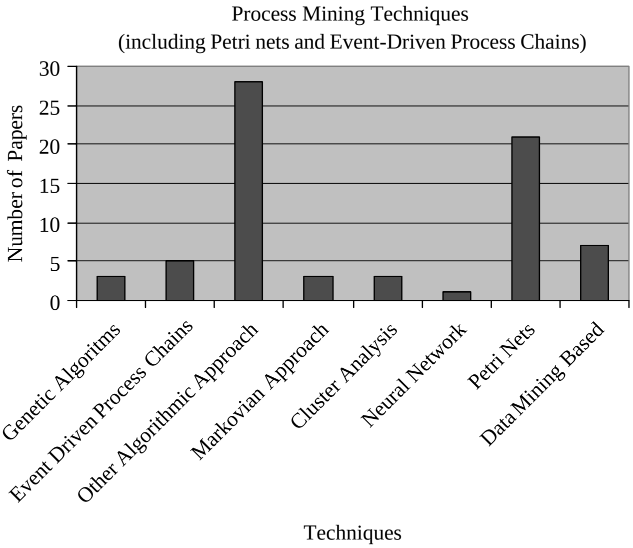

Figure 1.
Papers detailing process-mining techniques

A number of software tools exist for process mining. Herbst and Karagiannis (2004) describe their InWoLvE workflow mining tool and detail a number of process-mining algorithms incorporated in the software. The tool deals with four classes of workflow models:

(1) injective sequential;

(2) injective parallel;

(3) non-injective sequential; and

(4) non-injective parallel.

The term injective relates to the uniqueness of the activity nodes of the process model. An injective problem class means that each node of the process is unique (there are no repetitions of nodes in the process model). A sequential problem class (process model) does not contain any splits of joins, unlike a parallel problem class. The ability of this tool to mine process models with non-unique nodes is an advantage. The mining approach taken by this tool is inductive, using rules and success criteria to enable the identification of processes from data.

In terms of soft computing techniques authors such as Alves de Medeiros *et al.* (2005) and van der Aalst *et al.* (2005) have explored the use of evolutionary computing techniques, in the form of genetic algorithms. In terms of neural network techniques only Cook and Wolf (1998a) have explored this area in relation to process discovery.

It can be seen from Figure 2 that the number of papers published in the area of process mining has increased significantly in the past few years (peaking in 2004). Of particular interest is the recent publication of papers concerning the use of soft computing techniques in process mining. Many of the process-mining techniques employed by authors involve the inclusion of heuristics at some point in order to initiate and "fine tune" the identification of processes within data. Alves de Medeiros *et al.* (2003) argue that there are problems in using heuristics in that very complex

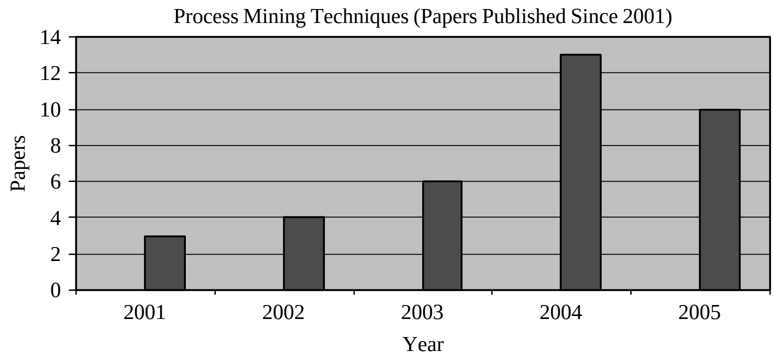

Figure 2.
Process-mining technique (papers published since 2001)

process constructs cannot be handled correctly by algorithms employing heuristic techniques. Instead these authors encourage the examination and use of the "α" algorithm which has been designed to work, without heuristics, for a defined set of conditions and process constructs. The algorithm has, more recently, been modified (Alves de Medeiros *et al.*, 2004b) to cope with mining problems such as short loops though both versions of the algorithm have difficulties with noisy data. Mining problems such as these are discussed in more detail in Section 3.

At present, both genetic algorithms and neural networks have been used to perform process-mining tasks. Cook and Wolf (1998a) conducted some work with a neural network-based method called RNet. Such a method "learns" how to recognise process patterns in data through a feedback mechanism, learned errors are passed back in the system to correct the method. Cook and Wolf (1998a) state that at the time of their research this method was still in the early stages of development.

Genetic algorithms have been used more recently in process-mining activities. Alves de Medeiros *et al.* (2004a) have investigated the use of genetic algorithms in the area of process mining, discovering their applicability in the mining of noisy event logs. This technique allows for process patterns to be represented as chromosome strings. van der Aalst *et al.* (2005a) claim that process-mining problems such as hidden activities and non-free choice constructs can also be addressed by genetic algorithms. The technique used by the authors makes use of causal matrices to represent the relations between activities of a process. The matrix is designed to show if there is a causal relationship between an activity and the other activities of a process and display the expected output for each activity.

Alves de Medeiros *et al.* (2005) point out that more research in the genetic algorithm approach is necessary to ensure that the process models mined by using this approach always reflect the behaviour found in process logs, as current algorithms tend to allow extra behaviour not found in logs. Table I details a number of papers that address process-mining problems using soft computing techniques. As can be seen from this table a number of papers employ a genetic algorithm approach to partially or fully solve current process-mining problems. However, many process-mining problems are still solved using methods derived from data mining.

3. Process-mining problems

The main issues still encountered in the area of business process mining are outlined by van der Aalst (2004a) as:

- *Noise.* Logged data may be incorrect or incomplete creating problems when data is being mined.
- *Hidden tasks.* Tasks that exist but cannot be found in the data.
- *Duplicate tasks.* Two process nodes may refer to the same process model.
- *Non-free choice constructs.* Are controlled choices that depend on choices made in other part of the process model.
- *Mining loops.* A process may be executed several times, loops may be simple involving one or more events or more complex.
- *Different perspectives.* Process events may be appended with additional information for mining purposes.
- *Delta analysis.* Comparison of process model and reference model to check for similarity/disparity.
- *Visualising results.* The results of process mining may be presented in graphical form in terms of a management panel.
- *Heterogeneous results.* Access to information systems based on different platforms.
- *Concurrent processes.* Mining of processes occurring at the same time.
- *Local/global search.* Local strategies restrict the search space and are less complex, global strategies are complicated but have a better chance of finding the optimal solution.
- *Process re-discovery.* The selection of a mining algorithm which can rediscover a class of process models from a complete workflow log.

A number of references exist that attempt to address many of the above process-mining problems. A cross section of some of these papers is shown in Table II. In the compilation of this table the authors, again, found that journals in the subject areas of BPM, data management, knowledge engineering and software engineering provided the best source of papers in the area for business process-mining problems. The subject area of manufacturing process management was excluded from the review since it lies outside the focus of this paper on business processes.

Figure 3 shows the papers available that address process-mining problems. Table II provides full details about these papers. The most commonly tackled problem is that of noise, with 20 per cent of papers detailing its mitigation. Mining loops and concurrent processes are also dealt with by a substantial number of papers (17 and 15 per cent, respectively). However, problems of heterogeneous data sources and mining perspectives are not so well researched with only 2 per cent of papers each.

It can be seen from Figure 4 that there has been an increase in the number of papers addressing process-mining problems in the past couple of years. Many of these papers involve the use of heuristic techniques to solve mining loop and noise problems.

The problem of noise in process data has been described by van der Aalst (2004a) as the presence of "incorrectly logged information" within a process log. van der Aalst

	Noise	Hidden tasks	Duplicate tasks	Non-free choice constructs	Mining loops	Different perspectives	Delta analysis	Visualising results	Heterogenous data sources	Concurrent processes	Local global search	Process rediscovery
van der Aalst (2005)							X					
Greco *et al.* (2004)											X	
Cook and Wolf (1998b)										X		
van der Aalst *et al.* (2002)				X								X
Dongen *et al.* (2005b)								X				
Weijters and van der Aalst (2003)	X				X							X
Golani and Pinter (2003)										X		
Cook and Wolf (1998a)	X							X				
Alves de Medeiros *et al.* (2004c)					X			X		X		
Zhang *et al.* (2003)										X		
Alves de Medeiros *et al.* (2004b)	X				X							
Gaaloul and Godart (2005)	X									X		
Alves de Medeiros *et al.* (2003)					X							
Dongen and van der Aalst (2004b)	X											
Dongen and van der Aalst (2004a)								X				
Alves de Medeiros *et al.* (2004a)					X							
Dongen and van der Aalst (2005b)			X									
Agrawal *et al.* (1998)	X		X									
Alves de Medeiros *et al.* (2005)	X	X		X								

(*continued*)

Table II.
Papers dealing with process-mining problems

Table II.

	Noise	Hidden tasks	Duplicate tasks	Non-free choice constructs	Mining loops	Different perspectives	Delta analysis	Visualising results	Heterogenous data sources	Concurrent processes	Local global search	Process rediscovery
Weijters and van der Aalst (2001)	X				X					X		
Schimm (2003)					X							
Dongen and van der Aalst (2005a)									X			
Hammori *et al.* (2004)								X				
Cook *et al.* (2004)	X									X		
Hwang (2002)	X											
van der Aalst *et al.* (2005a)	X	X	X	X	X			X		X	X	
van der Aalst and Alves de Medeiros (2005)	X	X	X		X							
Herbst and Karagiannis (2004)			X		X			X				
Dustdar *et al.* (2004)								X				
van der Aalst and Song (2004)						X						
Cook and Wolf, 1998b)										X		
Schimm (2004)										X		

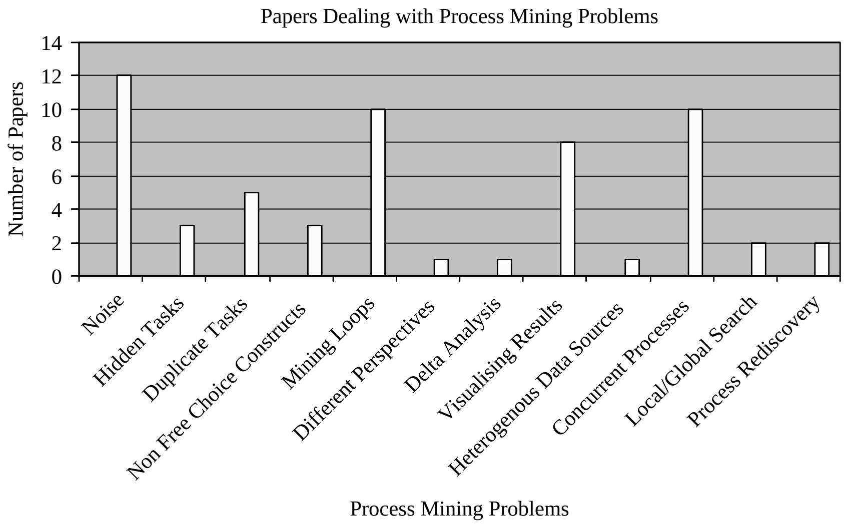

Figure 3.
Papers dealing with process-mining problems

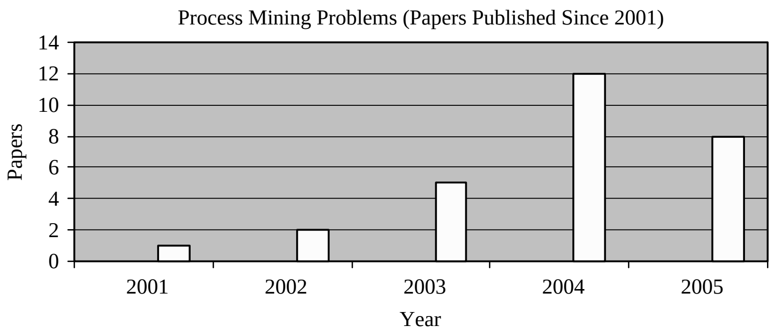

Figure 4.
Process-mining papers published since 1998

goes on to mention the value of algorithms that may discern exception from normal process execution traces. Agrawal *et al.* (1998) were one of the first authors to address the problem of noise in event logs; applying an algorithmic approach to this issue. Cook and Wolf (1998a) also examine a range of software process discovery algorithms that can deal with noise. Later work by Alves de Medeiros *et al.* (2005) employs genetic algorithms to address mining problems involving noise and other issues.

The visualisation of process-mining results is another issue that has been addressed by several authors. The issue concerns the display of mining results to the user, especially the graphical representation of mined processes by models. Cook and Wolf (1998a) are among the first to address process mining visualisation issues through the use of a graphical interface. They provide a method of displaying process models derived from mined process data. One of the most recent developments in the visualisation of business processes and the practice of process mining is outlined by van Dongen *et al.* (2005b). The ProM framework is a tool to support process-mining activities. The framework is a "pluggable" software tool able to support many sub tools ranging from graphical process display tools to mining algorithms and analysis tools. The framework encourages a unified approach to be adopted in the design of process mining software so users are presented with a generic interface capable of hosting a range of process mining and modelling activities.

Work on mining loops includes the contributions of Alves de Medeiros *et al.* (2004b) who addressed the mining of short loops from event logs. This work is complemented by a paper on the mining of mobile systems by Alves de Medeiros *et al.* (2004c) which also deals with mining loops.

The mining of concurrent processes has been explored by Schimm (2004) who concentrates on the task of mining exact workflows from event-based data. Cook *et al.* (2004) examine the problem of mining concurrent workflows in greater depth. These authors have developed algorithms to detect the presence of concurrent behaviour in event logs. Four metrics are used to detect concurrency in process logs in the work of Cook and Wolf (1998b):

- entropy;
- event type counts;
- deciding causality; and
- periodicity.

Entropy gives a measure of the amount of information contained in an event sequence and the number of occurrences of events within a process. It is possible to look for particular entropy values that may suggest a concurrent branching of a process in the event data. Event type counts are used to determine the likelihood of a fork occurring joining branches of a process together. The count of event types following a fork should be the same as the count at the fork occurrence. This metric is often used alongside the entropy measure for process branching. Causality of events must be decided when examining a concurrent process in order to decide if certain events are sequentially causally related in the presence of multiple threads of execution within a process. A summation of the event frequencies is used as part of this metric. The fourth metric, periodicity, examines the probability of an event sequence occurring, by examining the period of occurrence of event sequences within the event log. The issue of process rediscovery is fundamental to process mining and is addressed by van der Aalst *et al.* (2002). These authors explore the class of business processes, which in their view can be rediscovered from workflow logs. Process event sequences can be identified from a process log by looking for sequences of events and measuring the frequency of these event sequences in a log. The mining algorithm outlined by the authors cannot model transitions that share an input place. The algorithm is also not able to detect implicit events in a process.

One area in particular that has scope for further investigation is that of process mining perspectives. van der Aalst and Song (2004) have explored the mining of business processes from a social perspective, where staff interactions in business processes can be mapped out. This involves appending extra information to process event logs. It is possible that qualitative information relating to certain parts of a business process (or set of processes) may be captured in event logs for analysis.

Few papers have been published detailing the comparison of mined business process models with "ideal" models. van der Aalst (2005) outlines a node mapping technique for such a delta analysis task. In such methods syntactical differences between an ideal process model and a model of an actual process are highlighted though van der Aalst also describes a way of taking into account behavioural differences. The method put forward examines two models from the point of view of behaviour inheritance, and is able to construct a model based on all available executions of a process showing only the activities that are common to both processes.

The problem of duplicate activity instances within process data could also benefit from further investigation. Herbst and Karagiannis (2004) describe a method for removal of repeated nodes using counters. Their technique differs from Agrawal's (1998) approach in that mining parameters can be set to force a high degree of specialisation in mined nodes. However, the authors concede that further techniques need to be considered in order to totally eliminate this particular problem.

4. Discussion and conclusions

This paper provides an analysis and comparison of key research efforts relating to business process-mining problems and techniques. In general, a large number of papers, over the past few years, are employing techniques that incorporate heuristics for process mining. This can relate to the initialisation of a process to provide an algorithm with some guidelines as to the process models being sought or in the "tuning" of the algorithm as it conducts its search. Additionally process mining software tools are becoming available to use with enterprise process management systems (or on a stand alone basis). Such tools allow for a dynamic visual representation of mined processes along with the modelling of new processes. A significant number of papers also detail process-mining techniques based on modified data mining approaches, with most employing data mining search algorithms. However, many other papers introduce novel process mining specific techniques not used in data mining practices such as the work of Hwang *et al.* (2004), who introduce algorithms designed to detect temporal patterns in process data.

A number problems encountered by process mining researchers have been partly or fully addressed over recent years. One of the most frequently addressed problems encountered in process mining is noise, closely followed by mining loops. Another area that has seen much research is that of the mining of concurrent processes. However, research for this paper has highlighted a number of process mining problems that could be investigated further. Mining perspectives is a relatively new area. The identification of information within a process and its migration and transformation though that process is potentially of great interest. The comparison of processes is still a fairly *ad hoc* task. More research is required to enable the production of a generic framework for the quantified comparison of processes. The use of Petri nets as a modelling technique is fairly common, 29 per cent of papers employ this technique. In general, most process-mining papers have been published since 2001, the years 2004 and 2005 account for 65 per cent of all process mining papers. The visualisation of business processes, especially though process-mining software has received more attention over the past few years. This trend is likely to continue as more attention is given to the manipulation of mined processes and ease of use by users, especially those users that are not expert in process mining.

Table III displays a list of common mining problems and shows which techniques can be used to address them. As can be seen from the table genetic algorithms can be

Table III. Mapping process-mining challenges to techniques in current research

	Data mining based	Neural networks	Genetic algorithms	Markovian approach	Other algorithmic approaches
Noise	Agrawal *et al.* (1998) and Hwang (2002)	Cook and Wolf (1998a)	Alves de Medeiros *et al.* (2005) and van der Aalst *et al.* (2005a)	Cook and Wolf (1998a)	Gaaloul and Godart (2005), Cook *et al.* (2004), Weijters and van der Aalst (2003, 2001), Alves de Medeiros *et al.* (2004b) and van der Aalst and Alves de Medeiros (2005)
Hidden tasks			Alves de Medeiros *et al.* (2005) and van der Aalst *et al.* (2005a)		van der Aalst and Alves de Medeiros (2005)
Duplicate tasks	Herbst and Karagiannis (2004) and Agrawal *et al.* (1998)		van der Aalst *et al.* (2005a)		Dongen and van der Aalst (2005b) and van der Aalst and Alves de Medeiros (2005)
Non-free choice constructs			Alves de Medeiros *et al.* (2005) and van der Aalst *et al.* (2005a)		van der Aalst *et al.* (2002) and Weijters and van der Aalst (2001)
Mining loops	Herbst and Karagiannis (2004) and Schimm (2003)		van der Aalst *et al.* (2005a) and Alves de Medeiros *et al.* (2004a)		Weijters and van der Aalst (2001, 2003), Alves de Medeiros *et al.* (2004a, b, 2003) and van der Aalst and Alves de Medeiros (2005)
Different perspectives					van der Aalst and Song (2004)
Delta analysis					van der Aalst (2005)
Visualising results	Herbst and Karagiannis (2004) and Hammori *et al.* (2004)		van der Aalst *et al.* (2005a)		Dustdar *et al.* (2004), Alves de Medeiros *et al.* (2004c), Dongen and van der Aalst (2004a) and Dongen, *et al.* (2005a, b)
Heterogeneous data sources					Dongen and van der Aalst (2005a)
Concurrent processes	Schimm (2004)		van der Aalst *et al.* (2005a)		Golani and Pinter (2003), Alves de Medeiros *et al.* (2004c) and Weijters and van der Aalst (2001)
Local/global search			van der Aalst *et al.* (2005a)		Greco *et al.* (2004)
Process re-discovery					van der Aalst *et al.* (2002) and Weijters and van der Aalst (2003)

used to address many mining problems, though data mining-based approaches are still relevant for the most common problems such as noise and mining loops. Neural network and Markovian-based techniques have been used to solve problems of noise in mining data though their wider application has not been explored in great depth. Many problems such as delta analysis and mining perspectives require a custom built algorithm.

While many of the process mining problems can be addressed by a combination of modified data mining approaches and custom built algorithms there is no single approach that can address all of the problems encountered in process mining; many custom process mining algorithms exist though they tend to solve only one or two problems. The approach with the widest application to mining problems is that of genetic algorithms. This area is currently receiving a significant amount of attention from the research community. Over the last few years a number of authors have started to develop the use of genetic algorithms in the pursuit of better process mining results. This approach has shown promise in the areas of noise reduction and the mining of hidden tasks. Attempts to incorporate neural network techniques to undertake process-mining tasks have been fewer, possibly due to the complexity involved in the approach.

In the opinion of the author the development of a soft computing-based framework for the automated mining of processes would be of great benefit. Important aspects of such a framework would include techniques to address the mining of temporal patterns, the comparison of processes and the mining of different perspectives. The use of soft computing techniques may be relevant in the identification of temporal patterns and their distillation into a single representative process model, for example, neural networks may be used to identify patterns and genetic algorithms used for the process distillation task.

Future research may also focus on more sophisticated methods of automated process mining without the need for human input. Work in the area of fuzzy logic may provide "human" like decision making capabilities to be used as part of the pattern identification stage. Further, work on identifying common process patterns encountered in organisations may be required along with standardisation of the way enterprise information systems record process data (and what data elements they record about a process).

References

Adams, N.M., Hand, D.J. and Till, R.J. (2001), "Mining for classes and patterns in behavioural data", *Journal of the Operational Research Society*, Vol. 52, pp. 1017-24.

Agrawal, R., Gunopulos, D. and Leymann, F. (1998), "Mining process models from workflow logs", in Schek, H.J. (Ed.), *Proceedings of the 6th International Conference on Extending Database Technology: Advances in Database Technology*, Springer Verlag, Heidelberg.

Aguilar-Saven, R.S. (2004), "Business process modelling: review and framework", *International Journal of Production Economics*, Vol. 90, pp. 129-49.

Alves de Medeiros, A.K., van der Aalst, W.M.P. and Weijters, A.J.M.M. (2003), "Workflow mining: current status and future directions", in Meersman, R. *et al.* (Eds), *CoopIS/DOA/ODBASE 2003*, Springer Verlag, Heidelberg, pp. 389-406.

Alves de Medeiros, A.K., Weijters, A.J.M.M. and van der Aalst, W.M.P. (2004a), "Using genetic algorithms to mine process models: representation, operators and results", Beta Working Paper Series, WP 124, Eindhoven University of Technology, Eindhoven.

Alves de Medeiros, A.K., Weijters, A.J.M.M. and van der Aalst, W.M.P. (2005), "Genetic process mining: a basic approach and its challenges", in Bussler, C. and Haller, A. (Eds), *Business Process Management Workshops: BPM 2005*, Springer Verlag, Heidelberg.

Alves de Medeiros, A.K., van Dongen, B.F., van der Aalst, W.M.P. and Weijters, A.J.M.M. (2004b), "Process mining: extending the α-algorithm to mine short loops", Beta Working Paper Series, WP 113, Eindhoven University of Technology, Eindhoven.

Alves de Medeiros, A.K., van Dongen, B.F., van der Aalst, W.M.P. and Weijters, A.J.M.M. (2004c), "Process mining for ubiquitous mobile systems: an overview and a concrete algorithm", in Baresi, L., Dustdar, S., Gall, H. and Materaseries, M. (Eds), *Ubiquitous Mobile Information and Collaboration Systems (UMICS 2004)*, Springer Verlag, Heidelberg.

Chen, K.C.W. and Yun, D.Y.Y. (2003), "Discovering process models from execution history by graph matching", in Liu, J. *et al.* (Eds), *IDEAL 2003*, Springer Verlag, Heidelberg, pp. 887-92.

Cook, J.E. and Wolf, A.L. (1998a), "Discovering models of software processes from event-based data", *ACM Transactions on Software Engineering and Methodology*, Vol. 7 No. 3, pp. 215-49.

Cook, J.E. and Wolf, A.L. (1998b), "Event-based detection of concurrency", *Proceedings of the 6th International Symposium on the Foundations of Software Engineering*, ACM Press, New York, NY.

Cook, J.E., Du, Z., Liu, C. and Wolf, A.L. (2004), "Discovering models of behaviour for concurrent workflows", *Computers in Industry*, Vol. 53, pp. 297-319.

Dong, G. and Li, J. (1999), "Efficient mining of emerging patterns: discovering trends and differences", *International Conference on Knowledge Discovery and Data Mining*, ACM Press, New York, NY.

Dongen, B.F. and van der Aalst, W.M.P. (2004a), "EMIT: a process mining tool", in Cortadelle, J. and Reisig, W. (Eds), *25th International Conference on Applications and Theory of Petri Nets (ICATPN 2004)*, Springer Verlag, Heidelberg.

Dongen, B.F. and van der Aalst, W.M.P. (2004b), "Multi-phase process mining: building instance graphs", in Atzeni, P., Chu, W., Lu, H., Zhou, S. and Ling, T.W. (Eds), *Conceptual Modelling – ER 2004*, Springer Verlag, Heidelberg.

Dongen, B.F. and van der Aalst, W.M.P. (2005a), "A meta model for process mining data", *Proceedings of the CAiSE'05 Workshops*, Vol. 2, pp. 309-20.

Dongen, B.F. and van der Aalst, W.M.P. (2005b), "Multi-phase process mining: aggregating instance graphs into EPCs and Petri nets", paper presented at 2nd International Workshop on Applications of Petri Nets to Coordination, Workflow and Business Process Management, at the ICATPN, Miami, FL, June.

Dongen, B.F., van der Aalst, W.M.P. and Verbeek, H.M.W. (2005a), "Verification of EPCs: using reduction rules and Petri nets", in Pastor, O. and Cunha, J.F. (Eds), *17th International Conference on Advanced Information Systems Engineering (CAiSE 2005)*, Springer Verlag, Heidelberg, pp. 372-86.

Dongen, B.F., Alves de Medeiros, A.K., Verbeek, H.M.W., Weijters, A.J.M.M. and van der Aalst, W.M.P. (2005b), "The ProM framework: a new era in process mining tool support", in Ciardo, G. and Darondeau, P. (Eds), *26th International Conference on Applications and Theory of Petri Nets (ICATPN 2005)*, Springer Verlag, Heidelberg.

Dustdar, S., Hoffmann, T. and van der Aalst, W. (2004), "Mining of *ad-hoc* business processes with TeamLog", Technical Report TUV-1841-2004-07, Technical University of Vienna, Vienna.

Gaaloul, W. and Godart, C. (2005), "Mining workflow recovery from event based logs", in van der Aalst, W.M.P. *et al.* (Eds), *BPM 2005*, Springer Verlag, Heidelberg.

Golani, M. and Pinter, S. (2003), "Generating a process model from a process audit log", in van der Aalst, W.M.P. (Ed.), *Business Process Management*, Springer Verlag, Heidelberg.

Greco, G., Guzzo, A., Manco, G. and Sacca, D. (2005), "Mining and reasoning on workflows", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 17 No. 4, pp. 519-34.

Greco, G., Guzzo, A., Pontieri, L. and Sacca, D. (2004), "Mining expressive process models by clustering workflow traces", in Dai, H., Srikant, R. and Zhang, C. (Eds), *PAKDD 2004*, Springer Verlag, Heidelberg, pp. 52-62.

Hammori, M., Herbst, J. and Kleiner, N. (2004), "Interactive workflow mining", in Desel, B.P. and Weske, M. (Eds), *Business Process Management*, Springer Verlag, Heidelberg, pp. 211-66.

Herbst, J. and Karagiannis, D. (1998), "Integrating machine learning and workflow management to support acquisition and adaptation of workflow models", *Proceedings of the 9th International Workshop on Database and Expert Systems Applications*, IEEE Computer Society, Washington, DC.

Herbst, J. and Karagiannis, D. (2004), "Workflow mining with InWoLvE", *Computers in Industry*, Vol. 53, pp. 245-64.

Hwang, S.Y. (2002), "On the discovery of process models from their instances", *Decision Support Systems*, Vol. 34, pp. 41-57.

Hwang, S.Y., Wei, C.P. and Yang, W.S. (2004), "Discovery of temporal patterns from process instances", *Computers in Industry*, Vol. 53, pp. 345-64.

Mannila, H. and Rusakov, D. (2001), "Decomposition of event sequences into independent components", *Proceedings of the First SIAM International Conference on Data Mining, Chicago, 5-7 April.*

Maruster, L., Weijters, A.J.M.M., van der Aalst, W.M.P. and van den Bosch, A. (2002a), "Process mining: discovering direct successors in process logs", *Proceedings of the 5th International Conference on Discovery Science (Discovery Science 2002)*, Springer Verlag, Berlin.

Maruster, L., van der Aalst, W.M.P., Weijters, T., van den Bosch, A. and Daelemans, W. (2002b), "Automated discovery of workflow logs from hospital data", in Dousson, C., Höppner, F. and Quiniou, R. (Eds), *Proceedings of the ECAI Workshop on Knowledge Discovery and Spatial Data*, pp. 78-84.

Schimm, G. (2003), "Mining most specific workflow models from event-based data", in van der Aalst, W.M.P. (Ed.), *Business Process Management*, Springer Verlag, Heidelberg.

Schimm, G. (2004), "Mining exact models of concurrent workflows", *Computers in Industry*, Vol. 53, pp. 265-81.

van der Aalst, W.M.P. (2003), "Workflow mining: a survey of issues and approaches", *Data & Knowledge Engineering*, Vol. 47, pp. 237-67.

van der Aalst, W.M.P. (2004a), "Process mining: a research agenda", *Computers in Industry*, Vol. 53, pp. 231-44.

van der Aalst, W.M.P. (2004b), "Workflow mining: discovering process models from event logs", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 16 No. 9, pp. 1128-42.

van der Aalst, W.M.P. (2005), "Business alignment: using process mining as a tool for delta analysis and conformance testing", *Requirements Engineering Journal*, Vol. 10 No. 3, pp. 198-211.

van der Aalst, W.M.P. and Alves de Medeiros, A.K. (2005), "Process mining and security: detecting process executions and checking process conformance", *Electronic Notes in Theoretical Computer Science*, Vol. 121, pp. 3-21.

van der Aalst, W.M.P. and Song, M. (2004), "Mining social networks: uncovering interaction patterns in business processes", in Desel, J., Pernici, B. and Weske, M. (Eds), *International Conference on Business Process Management (BPM 2004)*, Springer Verlag, Berlin.

van der Aalst, W.M.P., Alves de Medeiros, A.K. and Weijters, A.J.M.M. (2005a), "Genetic process mining", in Ciardo, G. (Ed.), *Applications and Theory of Petri Nets*, Springer Verlag, Heidelberg.

van der Aalst, W.M.P., de Beer, H.T. and van Dongen, B.F. (2005b), "Process mining and verification of properties: an approach based on temporal logic", in Meersman, R. and Tari, Z. (Eds), *On the Move to Meaningful Internet Systems 2005*, Springer Verlag, Heidelberg.

van der Aalst, W.M.P., Weijters, A.J.M.M. and Maruster, L. (2002), "Workflow mining: which processes can be rediscovered?", Beta Working Paper Series, WP 75, Eindhoven University of Technology, Eindhoven.

Weijters, A.J.M.M. and van der Aalst, W.M.P. (2001), "Process mining: discovering workflow models from event based data", in Kröse, B., Rijke, M., Schreiber, G. and Someren, M. (Eds), *Proceedings of the 13th Belgium-Netherlands Conference on Artificial Intelligence, Amsterdam, October*.

Weijters, A.J.M.M. and van der Aalst, W.M.P. (2003), "Rediscovering workflow models from event-based data using little thumb", *Integral Computer-aided Engineering*, Vol. 10 No. 2, pp. 151-62.

Zhang, S.H., Gu, N., Lian, J.X. and Li, S.H. (2003), "Workflow process mining based on machine learning", *Proceedings of the Second International Conference on Machine Learning and Cybernetics*, IEEE Computer Society, Washington, DC.

Corresponding author
C.J. Turner can be contacted at: c.j.turner@cranfield.ac.uk