Current Trends in Workflow Mining

L K Ajayi¹, A A Azeta², I T Owolabi³, O O Damilola⁴, F Chidozie⁵, A E Azeta⁶, O Amosu⁷

¹²³⁴Department of Computer and Information Sciences, Covenant University, NIGERIA
⁵Department of Political Science and Int. Relations, Covenant University, NIGERIA
⁶Department of PTTIM, FIIRO, NIGERIA
⁷Department of PTTIM, FIIRO, NIGERIA

lydia4reel@gmail.com¹, ambrose.azeta@covenantuniversity.edu.ng², temiadebo@gmail.com³, osofuyeodundami@yahoo.com⁴, felix.chidozie@covenantuniversity.edu.ng⁵, aeeigbe@gmail.com⁶, funmiotesh@gmail.com⁷

Abstract. Current workflow management systems need a deep insight into the design of organization’s business process. Creating a business workflow design is sometimes complicated, inaccurate, time-consuming and non-error free. This business process design requires adequate tools and techniques needed to help organizations or companies achieve their desirable target. To aid the support of workflow, scholars have proposed the use of workflow mining which helps organizations have a deep insight into their business processes. This study gives an in-depth analysis on the current trend of workflow mining and also discusses their strengths and limitations/weaknesses. Most existing approaches produces models that are generalized, incomplete, not well represented and shows only fuzzy pictures of the workflow investigated. The challenging issues that arise as a result of inaccurately applying the right algorithm to construct a work model from workflow logs are also highlighted. There is need to make appropriate selection for a good mining algorithm to avoid any setback and understanding the organization business workflow processes for business design, modeling, resource strategy and planning control. It is essential to have a comprehensive characterization study of workflow mining algorithms. The resulting insights provides a foundation for such workflow mining algorithms necessary for improving business processes schemes as well as for the effectiveness of future services and application

Keywords: Workflow Mining, Algorithms, Visually Impaired, Applications, Industry

1. Introduction

Workflow Mining (WFM) can be described as a process mining whose aim is to determine process knowledge from users-interaction logs and actually represents them as high-level workflow models [1]. Business processes are a big part of an organization’s basic activities which shapes how organizations run their day to day activities. It is a technique for extracting workflow models from workflow logs [2]. The aim of every organization is to be able to deliver and produce in fewer amounts of time market acceptable product and services. A good way to accomplish this objective is to have a well-defined business processes. This brought about workflow management system which is needed to be able to design operational processes [3].
Workflow mining is a research agenda that helps organizations or companies to have a deep insight into their business or project processes. Workflow mining is referred in medical term as an X-ray machine that scans and give in and out information and reality that occurs and likely to occur in an organization [1]. Workflow logs captured generally previously unknown useful information on how business processes are executed in real time basis. Workflow models are then derived through this workflow mining which helps to better the development of business process of an organization. Workflow mining has the capabilities to minimize the design time, error, cost and at the same time increase reliability and quality of products. This allow organizations or companies to spend a lesser amount of time planning, scheduling, analyzing, implementing, managing and addressing their business process design. Considerable work has been done on making use of different work mining algorithm to mine event logs in order to generate a process model capable of giving support to workflow process design.

Most existing approaches create models that are generalized and incomplete. These models shows only fuzzy pictures of the workflow investigated and further explain that adaptation and analysis of the workflows are not properly represented by these models [4]. This brought about the existence of work mining algorithms. Most algorithms have been applied inappropriately having unknown limitations that later generates problems and setback for businesses on the long run. In order to avoid any generating issue in the future, people involved in the work project needs have a deep insight on the limitations/weaknesses as well as strengths of each algorithm to be used and which is the optimum one to be applied appropriately for projects.

Workflow generally refers to the process in which tasks, information, data and documents are conveyed from one individual to another. It refers to an automation of a process whereby the data involved is processed by different tasks. It supports analysis and design of a new system by reducing analysis time and cost because in most cases, log data are readily available without imposition of additional cost [5]. Most software systems like Oracle, PeopleSoft, Customer Relationship Management (CRM) has adopted the use of workflow technology and most problem encountered along the way are solved through the use of workflow knowledge which are generally applied. Workflow management is a fast growing technology that is mostly used by businesses in various companies and organizations. One of its major features involves computerization that combines both human behavior and machine-based actions. Organizations make use of workflow management systems to handle workflow that is needed for process completion and execution.

Workflow management systems are explicitly driven by process models which means a completely specified design for workflows is needed in order to establish a given workflow process [6]. Installation and establishment of WFM needs a workflow model that will describe deeply how the business processes should be managed. The main focus of workflow mining is to collect data at run-time in order to support workflow analysis and design. It concerns itself with the origin of a graphical representation of workflow model in a recorded event data which involves analyzing event logs and constructing a high-level illustration of a particular system in a form of workflow models [7]. It basically involves reversing the conventional method by gathering data at run-time to form workflow logs and then automatically derives a model from the logs containing data generated. These models can either be Petri-Net models or Hidden Markov Models (HMM) but the most commonly used model is the Petri-Net models which shares several similarities with models that are already established. These models captures the real behavior of a subsystem and entails how users communicate directly or indirectly with the system which can basically be used for as an input data for different assignments such as error detection, system diagnosis, policy performance checks and prediction [1]. As a result of these benefits, workflow mining has engrossed the curiosity of the research community and enjoyed global applications domain over the last few decades and its relevance has been widely acknowledged. Workflow mining is a type of process mining or business process mining that relies on existing workflow.

The remaining part of this paper is arranged as follows: Section two discusses the literature review. Section three highlights statement of the problem. Current research issues in workflow
mining in relation to constructing and using workflow models and algorithms are described in section four. Section five highlights the comparative analysis of workflow mining algorithms. The application of workflow mining is presented in section six, while section seven concludes the paper.

2. Literature review
A typical case study of workflow mining is a research conducted in [8] that entails surgery work mining which uses ten process logs each describing a Laparoscopic Cholecystectomy. The ten process logs are then used to develop Hidden Markov Model (HMM), instead of petri-like model with a surgical simulator. Design of a workflow analysis and a graphical user interface is also introduced that make use of a graph representation of the Hidden Markov Model for visualization of surgery operations. The surgery aspect involves removal of gallbladder which is usually performed because of symptomatic gallstone diseases [8]. Another case study includes a workflow e-commerce place, which provides a platform to allow people buy and sell their products. There are typically four different types of actors involving Administrators, the Merchant, the Regular buyers and the Window shoppers. The role of the Administrators involves maintenance and management like checking, updating catalogs, removing catalogs, etc. The Merchants role is to check orders, arrange shipment of products and query products. The role of the Regular buyers is to check for desirable products based on their wants and place order and the window shoppers also query and purchase products [1].

Envisaging that there are different work models that exist in the system, such as customer information management, catalogs management, products return management, and product purchase management, etc. After logging into the system, users can effectively carry out any workflow as long as there is access to the system. Every single individual of an organization knows how to get his/her work done but in view of business processes and workflow, a global view is required. Every participant knowledge in an organization can be gathered together to form a global process view. Most research work done concentrates more on workflow diagnosis and performance analysis rather than the traditional approach which is workflow modeling, simulation and implementation.

Workflow mining mainly supports diagnosis/design and trains people on how to interact and work with the new system. It uses several machine learning approaches to generate a workflow model from workflow logs. One of the problems faced today is that a very good designer has to create a well-detailed model that represents an accurate routing of the system. The main output of workflow mining is workflow logs. Workflow logs are set of events that suit a particular workflow. These events consist of workflow attributes which are workflow, task, case, activity, instance, event type, timestamp, resource, container, etc. [4]. They are basically used to construct a process specification that represents registered behavior adequately. Workflow mining can also be applied in the area of e-learning for the visually impaired, application and industry.

3. Statement of the Problem
The problem of workflow mining is to develop a viable workflow model that represents the workflow system at hand with little information. A workflow log may have challenging issues such as noise, incomplete (i.e. short of workflow instances like activity, process, etc.), incorrect or refer to exceptions. For example, a hospital workflow system, two casually unrelated events like blood sample and death may occur intermittently without involving a casual relation [7].

In order to counter these problems, workflow algorithms need to be more studied to have a deep insight on their strengths and weaknesses/limitations in order to select the optimum best to be used to derive a workflow model for practical application. Most research work only gave a broad survey on the current trend of workflow mining but didn’t properly align the limitation or strength of each algorithm used and how and when it is proper to be applied to future work in order to derive a realistic workflow model for a given business process. There is therefore need to present a qualitative and comparative analysis of different current workflow algorithms showcasing their strengths and weaknesses/limitations.
Therefore, the objectives of this paper are as follows:

- To review the concepts of workflow mining with its different mining algorithms.
- To discuss the application areas of workflow mining.
- To make a quantitative and comparative analysis of workflow mining algorithms detailing their strengths and limitations/weaknesses.

4. Current research issues in workflow mining in relation to constructing and using workflow models

There are several problems that are peculiar in workflows and calling for attention to allow practical application and for a proper workflow model to be derived. They include the following [9]:

Minning Hidden Tasks: Hidden task are also called invisible task. Under this research issue, workflow log may contain invisible or hidden tasks and if a model contains invisible tasks, then the consequential model is not a precise representation for workflow model.

Duplicate Tasks Mining: Duplication of tasks is also another current research issue in workflow log Mining.

Loops Mining: A workflow log may contain several types of loops. There is the one-length loop or the two-length loops and they are also referred to as short loops.

Mining non-free-choice constructs: Another underlying issue is mining non-free choice constructs that combines both choice and synchronization.

Noise and incompleteness: Noise are called incorrect logged data and they are the most common problem in work mining. It can generally appear if tasks are wrongly logged or event log represents special cases.

Visualizing Results: the overall results of workflow mining needs to be presented in a graphical form.

Concurrence: it is also a research issue that involves mining different processes at the same time.

Delta Analysis: this involves comparing the reference model and mined model to see if the mining model fits for future application.

Others are mining sequence, splits and joins. In order to avoid problematic constructs from occurring or been solved, there is need to have a deep insight into work mining algorithms and techniques highlighting their strength and weaknesses to enable practical application.

4.1. Algorithms for workflow mining

Workflow algorithms are used for mining workflow models using event logs and then the models derived can then be tested for fitness using delta analysis, conformance checking, etc, and helps to extract meaningful information from event logs [10]. The general use of workflow algorithms is for mining purposes in many practical applications, such as transaction logs of a resource planning system of an organization can be used to determine process models, product models, etc. However, workflow algorithms have different properties and specialism i.e. some algorithms are seen to counter the issue of problematic constructs aligned above over other algorithms and been able to select the optimum best algorithm for accuracy and confidence which are needed for future application to represent an actual model when compared.

However, in terms of mined workflow model imitating an actual real model, literature study shows that workflow algorithms have specific limitations/weaknesses and at the same time strengths. These brought about a comparative analysis of these algorithms for future applications as outlined below:

Alpha (α++) algorithm

The study in [23] introduced the alpha (α++) algorithm been the first algorithm that discovers concurrency. The steps in alpha includes read an event log, collate the tasks, infer the relations been ordered which is the main step, build the net based on the inferred relation and give the resulting output net. It has different limitations and strengths as shown below;
**Strength**

a. Alpha algorithm has the ability to rediscover a large set of workflow model.
b. It is also the first algorithm to acknowledge the context of concurrency.
c. It describes sequences of tasks.
d. It has been proven to mine length two loops and length one loop, tasks that are invisible tasks.
e. It tackles the issue of duplicate tasks.
f. It mines timed workflow logs and calculates all given performance metrics.
g. It exhibits simplicity and reveals basic work mining concepts, terms and interesting ideas and theoretical limits but has its limitations which have been proven not to be the best starting point for real life logs as it will only produce results but not the best one.

**Limitations/weaknesses**

a. It exhibits the inability to deal with noise.
b. It cannot be used as a benchmark.
c. It is also simple to be applied to real-life logs.
d. It has been proven to have the inability to mine short loops.
e. Inability to also mine non-free choice constructs.
f. It doesn’t consider redesigning problem even though it tackles mining duplicate tasks.
g. Inability to detect certain multiple kind of task.

This alpha algorithm led to the discovery of heuristic mining algorithm.

**Heuristic mining algorithm**

Heuristic algorithm is also the second algorithm that was developed by Dr. Ton Weijters to address the problem of alpha algorithm by considering the frequency in event log. The steps in this algorithm is to read an event log, collate the tasks, infer the relations been ordered which is based on their frequencies, build the net based on the inferred relation and give the resulting output net.

**Strength**

a. The strongest point in heuristic algorithm is that it is used for dealing with noise unlike alpha algorithm to express the main behavior in an event log.
b. It helps to remove infrequent paths.
c. It also deals with incompleteness.

**Limitations/weaknesses**

a) It doesn’t have the ability to mine duplicate tasks.
b) It doesn’t have the ability to mine non-free choice constructs.

**Genetic mining algorithm**

The steps in genetic algorithm includes reading the event log, collating and building the initial population, calculating the fitness of each individual in the population, stopping and returning each individual that is fit and creating next population using elitism and genetic operator.

**Strength**

a) It has the ability to handle noise and infrequent behavior.
b) It majorly can handle in visible tasks.
c) It has the ability to handle duplicate tasks.
d) It also has the ability to handle non-free choice constructs.
e) It is able to handle short loops.
f) It allows combination with other mining algorithm like heuristics for incremental improvement.
Limitations/weaknesses
a) It exhibits slow computational time period.
b) It requires a lot of computing power.

Fuzzy mining algorithm
The fuzzy mining algorithm was developed by Fluxicon co-founder with the name Christian W. Gunther. The steps involved includes fusing similar behaving attributes, generating meta rules, generating same fuzzy item sets and making fuzzy association rules.

Strength
a) It addresses the problem of large number of activities/tasks
b) It also addresses the problem of highly unstructured behavior.
c) It is useful on large, unstructured and noisy event logs.
d) It makes good use of correlation or significance metrics to simplify work model at a desired level of abstraction.
e) It helps in removing not so important edges, isolated node clusters and clustering highly correlated nodes into a single node clustering.

Limitations/weaknesses
a) There is total necessity to produce fuzzy rules when making use of fuzzy mining algorithm.
b) The model discovered by fuzzy mining algorithm doesn’t contain semantics and can only be used to indicate relationships between events.

Inductive mining algorithm
The inductive-based mining algorithm is used for creating a workflow tree that can be transformed into a Petri net model for a certain event log discovered by Mark Gold. The steps involved are splitting event logs repeatedly, finding the most prominent splits in the event logs, then detecting the operator and then continuing on both sub-logs.

Strength
a) It has the ability to handle incompleteness and noise.

Limitations/weaknesses
a) It doesn’t deal with concurrency
b) It doesn’t have the ability to mine some problematic construct like loops, duplicate task, hidden tasks and non-free choice construct.

Region based mining algorithm
Region mining algorithms are based on the concept of regions. They are either state-based or language-based regions. The state-based steps include discovering a transition system and converting the transition system into a Petri –net.

Strengths
a) It is used to discover more complex control flow-like structure.
b) It has the ability to deal with problematic constructs (label splitting, non-free choice, etc.).

Limitations/weaknesses
a) Region based mining exhibits over fitting.
b) It doesn’t have the ability to mine other problematic constructs like loops, hidden tasks, non-free constructs, noise, incompleteness, etc.
c) In order to apply this algorithm, the event log must be complete.

5. Comparative analysis of workflow mining algorithms

The appropriateness of an algorithm needs to be evaluated for future applications. This led to a comparative analysis of workflow algorithms which is based on the following criteria.

Table 1: A Comparative Analysis of Workflow Mining Algorithm

| Criteria                        | Alpha(++) mining algorithm | Heuristic mining algorithm | Genetic mining algorithm | Fuzzy mining algorithm | Inductive-based mining algorithm | Region-based mining algorithm |
|---------------------------------|----------------------------|-----------------------------|--------------------------|------------------------|----------------------------------|------------------------------|
| Mining(short) one-length loop   | Yes [2,11]                 | Yes [12]                    | Yes [2,13]               | Yes [9]                | No [17]                          | No [15]                      |
| Mining(short) two-length loops  | Yes [11,16]                | Yes [9]                     | Yes [13]                 | Yes [16]               | No [14]                          | No [17]                      |
| Mining invisible/hidden tasks   | Yes [2,9,16], No [12]      | No [12]                     | Yes [13]                 | No [16]                | No [15]                          | No [15]                      |
| Mining duplicate tasks         | No [13,16], No [23]        | Yes [13]                    | No [16]                  | No [17, 21]            | Yes [38]                         | No [38]                      |
| Noise-dealing/incompleteness    | No [25,35]                 | Yes [5]                     | Yes [5]                  | Yes [25]               | Yes [17]                         | No [13]                      |
| Mining non-free choice constructs | Yes [13]               | No [25]                     | Yes [2,14]               | No [9]                 | No [14]                          | No [15]                      |
| Concurrency                    | Yes [13,17]                | Yes [13]                    | Yes [13]                 | Yes [13,19]            | No [13]                          | Yes [13]                     |
| Dependency                     | Yes [6]                    | Yes [6]                     | Yes [6]                  | Yes [6]                | Yes [10]                         | No [6]                       |
| Simplicity                     | Yes [10,20]                | Yes [2]                     | Yes [21]                 | Yes [10,20]            | No, It Is Complex [20]           | No [10,20]                   |
| Time efficiency                | Yes [2, 20]                | Yes [2, 20]                 | No [2, 20]               | No [6]                 | No [6]                           | No [6]                       |
| Linear in size of log          | Yes [22]                   | Yes [20]                    | Yes [20]                 | Yes [20]               | Yes [20]                         | No, its polynomial [22]      |

In Table 1, the criterion used are based on general problematic constructs that workflow mining faces since most algorithm contains constructs such as noise. The concurrency criteria mean ability to resolve conflicts between two event classes. The dependency criteria mean ability to work well with other algorithms. The simplicity criterion means how easy and simple the algorithms are for application, and linear in size of log criteria means is the performance ratio linear with the size of the event log. In order to get the optimum best between the six workflow mining algorithms outlined above, chart-graph based was applied on different criteria as shown in Figure 1.
Figure 1: A Chart Graph showing a retailed comparative analysis of workflow mining algorithms.

From the chart analysis in Figure 5, Alpha algorithm gives 82% yes and 18% no, Heuristic algorithm gives 73% yes and 27% no, Genetic algorithm gives 91% yes and 9% no, Fuzzy algorithm gives 64% yes and 36% no, Inductive algorithm gives 7% yes and 73% no and Region-based algorithm gives 18% yes and 82% no. From this analysis, Genetic algorithm shows a percentage of 91% yes and 9% no making it the optimum best compared to other algorithm outlined.

6. Application of workflow mining
The application of workflow mining is useful from microscopic, macroscopic and mesoscopic perspectives and its application are given by the following:
- Transportation engineering, planning, design and operations, problems and solutions.
- Healthcare system where flow of multi-disciplinary patients is analyzed, surgery flow system, etc.
- Banking and financial domain to improve organization process, accounts auditing, customer attraction, etc.
- Educational organizations.
- E-commerce, logistics, E-government, insurance, etc.
- Construction process system.
- Simulation of human behavior and decision-making.

7. Discussion and conclusion
Workflow mining has been reviewed from earliest inception till date involving the techniques, tools, and algorithms. These further led to series of different workflow mining algorithms highlighting their limitations/weaknesses and at the same time their strengths. As of today, workflow mining models derived from event logs should prove to be more accurate and fit which is critically needed for any future applications. As was presented in [24, 25], this current study did not include hypothesis in its model formulation and testing, rather a comparative analysis of workflow mining was presented.

The future trend of work mining should look into improving the limitation of each algorithm for future applications. The limitation of genetic algorithm which is time efficiency and low computation period needs to be improved, since it has been proved to be the optimum best algorithm so far for future application.
Acknowledgement
In this study, we would like to express our deep appreciation for the support and sponsorship provided by Covenant University Centre for Research, Innovation and Discovery (CUCRID).

References
[1] Liu, X., Alshangiti, M., Ding, C., Yu, Q., Engineering, K., Alshangiti, M., Yu, Q. (2018). Accepted Manuscript. https://doi.org/10.1016/j.datak.2018.04.002.
[2] Gupta, E. (2015). Process mining algorithms, (March).
[3] Medeiros, A. K. A. De, Aalst, W. M. P. Van Der, & Weijters, A. J. M. M. (2003). Workflow Mining : Current Status and Future, (i), 389–406.
[4] Schimm, G. (2004). Mining exact models of concurrent workflows, 53, 265–281. https://doi.org/10.1016/j.compind.2003.10.003
[5] Singh, G. N., & Aggarwal, S. (2011). A Process Model For Workflow Mining, 4(2), 719–722.
[6] Aalst, W. M. P. Van Der, Weijters, A. J. M. M., & Maruster, L. (2016). Workflow Mining :
[7] Aalst, W. Van Der. (2010). Process Mining : Making Knowledge Discovery Process Centric, 13(2), 45–49.
[8] Blum, T., Padoy, N., & Feußner, H. (2008). Workflow mining for visualization and analysis of surgeries. https://doi.org/10.1007/s11548-008-0239-0
[9] Aleem, S., Capretz, L. F., & Ahmed, F. (2015). Business Process Mining Approaches : A Relative Comparison, (4), 1557–1564.
[10] Wang, J., Wong, R. K., Ding, J., Guo, Q., & Wen, L. (2010). Efficient Selection of Process Mining Algorithms, 1–14.
[11] Weber, P., Bordbar, B., & Ti", P. (2015). A framework for comparing process mining algorithms, 1–4.
[12] Cervavolo, P. (2016). Focus On Process Mining.
[13] Nascimento, F., M. P. Van Der, Weijters, A. J. M. M., & Marusten, L. (2016). Workflow Mining :
[14] Ghawi, R. (2016). Process Discovery using Inductive Miner and Decomposition . A Submission to the Process Discovery Contest @ BPM2016. Technical Report, (October).
[15] Pourmasoumi, A., & Bagheri, E. (2016). Business Process Mining, 1(1), 1–8.
[16] Karla, A., & Medeiros, A. De. (2016). Mining Algorithms, 1–53.
[17] Turner, C. J., Tiwari, A., & Olaiya, R. A. (2012). Business Process Mining : From Theory to Practice, 18, 493–512.
[18] Ghawi, R. (2016). Process Discovery using Inductive Miner and Decomposition . A Submission to the Process Discovery Contest @ BPM2016. Technical Report, (October).
[19] Lou, J., Yang, S., & Li, J. (2010). Mining Program Workflow from Interleaved Traces.
[20] Rozinat, A., & Medeiros, A. K. A. De. (2016). Towards an Evaluation Framework for Process Mining Algorithms.
[21] Dongen, B. F. Van, Medeiros, A. K. A. De, Verbeek, H. M. W., & Weijters, A. J. M. M. (2013). The ProM framework : A new era in process mining tool support, (i).
[22] Liu, T., Cheng, Y., & Ni, Z. (2012). Knowledge-Based Systems Mining event logs to support workflow resource allocation. Knowledge-Based Systems, 35, 320–331. https://doi.org/10.1016/j.knosys.2012.05.010
[23] Aalst, W. M. P. Van Der, Bolt, A., & Zelst, S. J. Van. (2017). RapidProM : Mine Your Processes and Not Just Your Data, 1–25.
[24] Nicholas-Omoregbe O S Azeta A A Chiazor I A and Omoregbe N 2017 Predicting the adoption of e-learning management system: A case of selected private universities in Nigeria. Turkish Online Journal of Distance Education-TOJDE 18(2) 106-121.
[25] Azeta A A Misra S Azeta V I Osamor V C 2019 Determining suitability of speech Enabled examination result management system. Wireless Networks 1-8.