International Journal of Science and Engineering Applications
Volume 10-Issue 09, 116 - 121, 2021, ISSN:- 2319 - 7560

Trace Clustering: A Preprocessing Method to Improve the Performance of Process Discovery

Huiling LI
College of Computer Science and Engineering, Shandong University of Technology
Zibo, 255000, China

Shuaipeng ZHANG
College of Computer Science and Engineering, Shandong University of Technology
Zibo, 255000, China

Xuan SU
College of Computer Science and Engineering, Shandong University of Technology
Zibo, 255000, China

Abstract: The information system collects a large number of business process event logs, and process discovery aims to discover process models from the event logs. Many process discovery methods have been proposed, but most of them still have problems when processing event logs, such as low mining efficiency and poor process model quality. The trace clustering method allows to decompose original log to effectively solve these problems. There are many existing trace clustering methods, such as clustering based on *vector space approaches*, *context-aware trace clustering*, *model-based sequence clustering*, etc. The clustering effects obtained by different trace clustering methods are often different. Therefore, this paper proposes a preprocessing method to improve the performance of process discovery, called as trace clustering. Firstly, the event log is decomposed into a set of sub-logs by trace clustering method, Secondly, the sub-logs generate process models respectively by the process mining method. The experimental analysis on the datasets shows that the method proposed not only effectively improves the time performance of process discovery, but also improves the quality of the process model.

Keywords: process discovery; trace clustering; process model; log similarity; quality measure

1. INTRODUCTION

Process mining [1-3] aims to extract effective information about business processes from event logs to discover, monitor and improve actual processes. Process mining mainly includes: 1) Process discovery takes the event log as input to the automatic production process model; 2) Conformance checking can be used to check if reality, as recorded in the log, conforms to the model and vice versa; 3) Enhancement is to extend or improve an existing process model using information about the actual process recorded in some event log. In addition, process mining also includes process prediction [4-5] and business process automation [6]. Process discovery is one of the most challenging process mining tasks, aims to discover a business process model form an event log. In the past two decades, researchers have proposed various process discovery approaches, e.g, Alpha Miner [7], Heuristic Miner [8], Inductive Miner [9], etc.

However, with the advent of the era of big data, business processes produce larger and more complex event logs. For these event logs, most existing process discovery approaches unable to mine the information correctly, and usually lead to process discovery low efficiency. In the process mining manifesto [10], Professor Van der Aalst and others take that existing process mining methods are difficult to handle the massive amounts of data is generated ASML's wafer scanner. as an example, therefore, dealing with large-scale and complex event log problems is one of the important challenges of process mining.

When dealing with complex and large-scale event log, the event log is reasonably decomposed into several sub-logs, and then the sub-logs are discovered by the existing process discovery approaches, thereby improving the efficiency of process discovery and the quality of process models. An effective way to decompose the event log is to cluster the trace in the event log, so that the process model combination corresponding to the clustered sub-logs can clearly and completely express the behavior in the original event log. On the one hand, the preprocessing operation of trace clustering can effectively improve the time performance of the process discovery method, and on the other hand, it also reduces the probability of complex process models (similar to the spaghetti process model), and then more intuitively understand the process model. To this end, we propose a preprocessing method to improve the performance of process discovery, called as trace clustering. The sub-logs by the trace clustering methods are mined by the existing process discovery approaches to generate the sub-process models. Finally, Checking the conformance of the above sub-logs with the original log by measuring fitness, precision, F-Measure to verify the feasibility and efficiency of the trace clustering preprocessing operation.

2. RELATED WORK

2.1 History of Process Mining Algorithms

In 2002, Wil van der Aalst proposed the *Alpha Miner* in [7]. From the perspective of workflow, it is based on the direct follow activity relationship between logs to mine the activity associations in event logs. The disadvantage of Alpha Miner is that it unable to flexibly handle noise, incomplete event logs, and cannot identify short loops, map non-local dependencies, and handle non-free choice structures. Many researchers have devoted themselves to improving and extending the *Alpha Miner*, and different algorithms have been proposed to solve these limitations.

For this reason, Weijters & van der Aalst et al. (2003) extended the Alpha Miner in [8] and considered the frequency of directly follow activity relationship, and calculated the dependency/frequency parameter to obtain the heuristic network. The algorithm is called *Heuristic Miner*. It can handle noise and allows comparison between manually designed models and execution processes. This algorithm is the most commonly used and customized because it

guarantees good adaptability, but it cannot provide complete reliability because uncommon paths are not incorporated into the model.

Jansen-Vullers et al. (2006) created a new algorithm based on integer programming technology. It shows that it is possible to search for the best settings using objective functions and applying integer programming techniques. This method finds all the solutions of a system of equations, and implements a minimization function through integer programming techniques.

Leemans et al. (2013) proposed an extensible framework called *Inductive Miner*[9]. The purpose of the algorithm is to discover block-structured process model that is reasonable and suitable for the behavior observed on the event log. This algorithm represents the minimum information of the discovery process model. Inductive Miner provided polynomial time complexity and feasible computational cost.

In 2017, vanden Broucke and Weerdt extended the most popular Heuristic algorithm and proposed the *Fodina Miner*[11]. This method is robust to noise and can identify repetitive activities. In addition, the algorithm is flexible, allowing users to choose to adjust the discovery process.

2.2 Quality Evaluation Index

This article uses the following three indicators to evaluate the quality of the event log, where L represents the event log and M represents the process model.

Index 1-Fitness

Fitness[12] quantifies the degree to which the process model can accurately reproduce the trace recorded in the event log, and it quantifies the ability of the process model to regenerate the trace recorded in the event log. A degree of fitness of 1 means that the process model can regenerate all trace in the event log, and a low degree of fitness indicates that most of the behaviors in the event log cannot be reproduced by the process model;

Index 2-Precision

Precision[13] quantification of some behaviors that can be repeated in the process model but not seen in the event log. It measures the ability of the process model to only generate traces in the event log. A precision of 1 means that all traces generated by the process model are included in the event log, and low precision means that the process model allows more behavior than the event log.

Index 3-F-measure

The F-measure value[14] is defined as the harmonic mean value of fitness and precision, calculated as follows:

$$F\text{-}measure(L,M) = \frac{2 \times fitness(L,M) \times precision(L,M)}{fitness(L,M) + precision(L,M)}$$

Where *fitness* (*L*, *M*) is the degree of fitness of the process model found in the event log relative to the original log, and *precision* (*L*, *M*) is the precision of the process model found in the sample log relative to the original log.

3. Framework

This paper proposes a process mining algorithm based on trace clustering. On the basis of the existing process mining algorithm, the log is preprocessed for trace clustering operation, and then the clustered sub-logs are respectively applied to the existing process mining algorithm performs. Finally, evaluates the obtained process model. Fig.1 shows an overview of our approach.

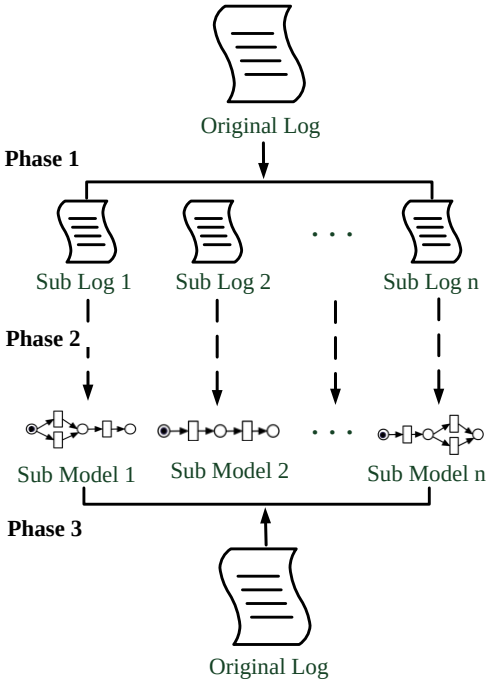


Figure 1. An Approach Overview

Phase 1: Preprocessing based on trace clustering

There are many existing trace clustering approaches, such as K-means trace clustering, active learning clustering, etc. After the original log is processed by the trace clustering method, a set of sub-logs are obtained, so that they belong to the same sub-log The traces the same sub-log of are similar, and the traces belonging to different sub-logs are different. The trace clustering method used in this process requires preset parameters, such as the number of clusters, and setting different parameters may affect the final log quality.

Phase 2: Event log process discovery

There are many existing process discovery methods, such as Alpha Miner, Heuristic Miner, and Inductive Miner. According to the event log input by the user, these mining algorithms are used to obtain the corresponding process model. It is worth noting that the parameter settings of the process mining algorithm may result in different process models, and the default parameter settings are used in this article.

Phase 3: Process model quality assessment

The feasibility and efficiency of the method proposed in this article can be evaluated from the following two perspectives.

(1) Process model quality: In order to quantify the quality of the process model, we first process the original log into a set of sub-logs from the trace clustering method by the existing process discovery methods for each sub-log to obtain the corresponding sub-process models, and separate the sub-process model from the original log checking conformance to measure fitness, precision, and F-Measure to quantify the quality of the new process model. By comparing the quality of the process model

with the original log, the feasibility of the method proposed in this article is demonstrated;

(2) Process discovery efficiency: The efficiency of process discovery can be quantified by the time it takes to obtain the process model. The less time it takes to obtain the process model, the higher the efficiency of process discovery. The efficiency of the method proposed in this paper is demonstrated by comparing the time it takes to obtain the process model.

4. TRACE CLUSTERING METHOD

4.1 Vector Space Method

Song et al. [15] proposed a method to construct a vector space model for the trace in the event log. This method is based on a set of configuration files, each of which measures multiple characteristics of each trace from a specific angle, such as activities, directly follow relation, etc., these features can form a corresponding feature matrix. Based on the feature matrix, multiple distance metrics (Euclidean distance, etc.) are used to calculate the distance between any two traces in the event log. Finally, the traditional clustering algorithms such as K-means clustering is applied in data mining to group the traces in event logs into sub-logs.

4.2 Context-aware Trace Clustering

Bose and van der Aalst described this trace clustering technique in [16,17], which extends the previous trace clustering method by improving the context awareness of control flow. The context awareness here refers to the control flow attributes of the trace in the event log, rather than context information such as organizer, case data, etc. In [16], Bose and van der Aalst proposed a general edit distance technique based on Levenshtein[18], in which editing operations include insertion, deletion or replacement. In [17], the idea of context-aware trace clustering was further developed, and the idea of generating a vector space model for the traces in the event log was reconsidered, using conservative patterns or subsequences to replace the previous activities as the basis of the vector space model. In this way, the concepts of maximum, supermax, and near-supermax repetition are defined to create a feature set that determines a certain trace vector. The corresponding trace clustering method in this article is Guide Miner Tree trace clustering.

4.3 Model-based Sequence Clustering

Ferreira et al. [19] proposed a trace clustering that is different from previous methods. Inspired by the work of Cadez et al. [20] in the field of Web usage mining, they proposed to cluster traces by learning a hybrid first-order Markov model using an expectation maximization (EM) algorithm. In [21], this model-based trace clustering technique was applied to server logs, proving its availability in real life.

De Weerdt et al. proposed in [22] the problem of finding the optimal distribution of traces on a given number of clusters, so as to maximize the combined accuracy of the associated process model. This method changes the goal of traditional trace clustering, which is based on grouping similar traces to find the optimal distribution and solves the problem of finding the optimal trace distribution. It proposes a top-down greedy algorithm and a standard for trace selection. Not because they exhibit similar behavior, but because they fit a particular process model well. The corresponding trace clustering method in this article is *ActiTrac* trace clustering.

5. EXPERIMENT ANALYSIS

5.1 Experimental Environment Settings

The open source process mining tool platform ProM (see http://www.promtools.org/) provides a fully pluggable experimental environment for process mining. It can be extended by adding plug-ins, currently contains more than 1,600 plug-ins, the tool and all plug-ins are open source.

The experiments in this article are all based on PC Intel Core i5-4210M 2.60GHz CPU, 12GB RAM environment, using Java language.

5.2 Simulation Data Structure

This article uses WoPed simulation tool (see as https://woped.dhbw-karlsruhe.de/) to construct a Petri net model. The model is constructed as follows, and then a jar package is generated from the log to generate a simulation Log. It contains 206 traces, 3228 events and 20 activities. The process model is shown in Figure 2.

The Petri net in Figure 2 is designed to generate different event logs with three types of behavior: a trace with activity X, a trace with activity Y, and a trace with neither activity X nor activity Y. Please note that it has chosen to include a large number of parallel and circular behaviors to approximate the complexity of a real event log.

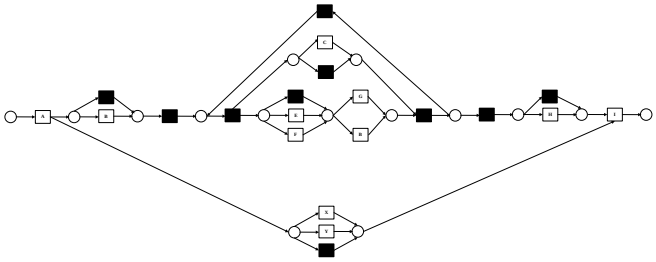


Figure 2. Example Petri net used for generating different classes of behavior according to the presence/absence of activities X and Y.

5.3 Simulation Log Analysis

We use four classic process discovery methods (Alpha Miner, Heuristic Miner, Inductive Miner,Fodina Miner) and three representative trace clustering methods(ArciTrac Trace Clustering,Guide Tree Miner,K-Means) to analyze the feasibility and accuracy of the method proposed in this article from the following two aspects.

5.3.1 Time Performance Analysis

By comparing the time taken by the original log to obtain the process model using each process mining method and the total time used by the process model obtained by the clustered sub-logs through each process mining method, it shows that the trajectory clustering method improves the process mining to a certain extent. The time of the method, the results obtained are shown in Table 1.

It can be seen from Table 1 that for most mining algorithms, the sum of the sub-logs after trajectory clustering is mostly less than the time to mine the original log directly through the process discovery method, which shows that the trace clustering process Later, the time performance of the process

Table 1. Process discovery algorithm time performance comparison(ms)

Trace Clustering Method	Number of clusters	Process discovery algorithm			
		Alpha Miner	Heuristic Miner	Inductive Miner	Fodina Miner
OriginalLog		31	90	153	35
Arctrac	3	32	24	55	30
	4	19	41	31	18
	5	26	49	39	10
Guide Tree Miner	3	19	24	133	23
	4	26	49	39	28
	5	29	63	26	19
K-means	3	19	44	29	30
	4	25	55	57	12
	5	29	43	42	20

discovery method has been further improved. In fact, if the processed sub-logs are processed on a distributed platform, the time performance will be further improved.

It is worth noting that this article does not compare the processing time of trace clustering. The reason is that this article only discusses whether the effect of clustering has further improved the process discovery method. In addition, this article also compares the trace clustering time statistics are performed, as shown in Table 2. From Table 2, it can be seen that the processing time of trace clustering is to a certain extent far longer than the time used for process discovery. The time of different trace clustering is different, which is due to the different operations in different trace clustering method caused.

Table 2. Trace clustering preprocessing time of log (ms)

Trace Clustering Method	Number of clusters	Trace clustering time
ArciTrac	3	6215
	4	5411
	5	4849
Guide Tree Miner	3	9024
	4	4415
	5	7451
K-Means	3	2189
	4	2163
	5	2160

5.3.2 Process Model Quality

By comparing the quality of the process model generated by the above process discovery method with the quality of the process model generated by the newly proposed method, the quality of the traditional process model is to compare the fitness, precision, and F-Measure of the process model generated by the original log and the original log. The Measure value is quantified; the new method proposed in this paper is to obtain the corresponding process model through the existing process discovery method through the several sub-logs generated by trace clustering, and then respectively do the fitness degree and the original log of the respective process model and the original log. The accuracy and F-Measure index values are quantified, and then the weighted average is used to obtain the final evaluation value. The results obtained are shown in Table 3, Among them, *F* represents fitness, *P* represents precision, *F1* represents *F-Measure*.

It can be seen from Table 3 that, except for the Alpha algorithm, which cannot obtain the relevant results, the final evaluation quality values obtained by the other process mining algorithms are all greater than the quality of the logs directly evaluated. This shows that the new method proposed in this paper has improved the process to a certain extent. Accuracy of discovery. Take the clustering method *ArciTrac*, when the number of clusters is 4 as an example, it is found that the fitness value of the method is reduced, but the accuracy value is increased, and the harmonic average value of the two is F-Measure value is increased, which shows that the quality of the process model has been improved.

6. CONCLUSIONS

This paper proposes a preprocessing method, called as trace clustering to improve the performance of the process discovery methods. The analysis on the simulation experiment data set shows that the method proposed in this paper can not only effectively improve the time performance of the process discovery method, but also improve the quality of the process model.

Table 3. Comparison of evaluation indicators

Trace Clustering Method	Number of clusters	Process Mining Algorithms											
		Alpha Miner			Heuristic Miner			Inductive Miner			Fodina Miner		
		F	P	F1	F	P	F1	F	P	F1	F	P	F1
Original Log		-	-	-	0.8557	0.795	0.8242	0.9527	0.516	0.67	0.847	0.768	0.8057
ArciTrac	3	-	-	-	0.7492	0.987	0.852	0.7909	0.9326	0.8546	0.7501	0.993	0.8547
	4	-	-	-	0.77	0.9379	0.8409	0.8207	0.8045	0.7988	0.7756	0.9378	0.8423
	5	-	-	-	0.78	0.9338	0.8427	0.8206	0.7637	0.7778	0.7667	0.9358	0.8373
Guide Tree Miner	3				0.8436	0.8353	0.839	0.925	0.5824	0.703	-	-	-
	4				0.836	0.841	0.8379	0.9185	0.5403	0.763	-	-	-
	5				0.8379	0.8511	0.8438	0.8791	0.616	0.7056	-	-	-
K-Means	3				0.8446	0.8546	0.8496	0.9263	0.5161	0.6587	0.8329	0.7294	0.777
	4				0.8333	0.8667	0.85	0.9165	0.3907	0.5471	0.8475	0.76	0.8008
	5				0.8441	0.8704	0.8567	0.9309	0.5606	0.6977	0.827	0.777	0.8007

7. REFERENCES

[1] W. M. P. VAN DER AALST. Data science in action[M]//Process mining. Springer, Berlin, Heidelberg, 2016: 3-23.

[2] LIU C, DUAN H, ZENG Q T, et al. Towards comprehensive support for privacy preservation cross-organization business process mining[J]. IEEE Transactions on Services Computing,2019,12(4): 639-653.

[3] ZENG Q, SUN S X, DUAN H, et al. Cross-organizational collaborative workflow mining from a multi-source log[J]. Decision support systems, 2013, 54(3): 1280-1301. Tavel, P. 2007 Modeling and Simulation Design. AK Peters Ltd.

[4] POURBAFRANI M, VAN ZELST S J, VAN DER AALST W M P. Scenario-based prediction of business processes using system dynamics[C]//OTM Confederated International Conferences" On the Move to Meaningful Internet Systems". Berlin, Germany: Springer-Verlag, 2019: 422-439.

[5] QAFARI M S, VAN DER AALST W. Fairness-Aware Process Mining[C]//OTM Confederated International Conferences" On the Move to Meaningful Internet Systems". Berlin, Germany: Springer-Verlag, 2019: 182-192.

[6] GAO J, VAN ZELST S J, LU X, et al. Automated robotic process automation: A self-learning approach[C]//OTM Confederated International Conferences" On the Move to Meaningful Internet Systems". Berlin, Germany: Springer-Verlag, 2019: 95-112.

[7] VAN DER AALST W, WEIJTERS T, MARUSTER L. Workflow mining: Discovering process models from event logs[J]. IEEE transactions on knowledge and data engineering, 2004, 16(9): 1128-1142.

[8] WEIJTERS A, RIBEIRO J T S. Flexible heuristics miner (FHM)[C]//2011 IEEE symposium on computational intelligence and data mining (CIDM). Washington, D. C., USA: IEEE, 2011: 310-317.

[9] LEEMANS S J J, FAHLAND D, VAN DER AALST W M P. Discovering block-structured process models from event logs-a constructive approach[C]//International conference on applications and theory of Petri nets and concurrency. Berlin, Germany: Springer-Verlag, 2013: 311-329.Forman, G. 2003. An extensive empirical study of feature selection metrics for text classification. J. Mach. Learn. Res. 3 (Mar. 2003), 1289-1305.

[10] VAN DER AALST W, et al. Process mining manifesto[C]//Business Process Management Workshops: BPM 2011 International Workshops, Clermont-Ferrand, France, August 29, 2011, Revised Selected Papers, Part I. Berlin, Germany: Springer-Verlag, 2011, 99: 169-194.

[11] VAN DER AALST W, et al. Process mining manifesto[C]//Business Process Management Workshops: BPM 2011 International Workshops, Clermont-Ferrand, France, August 29, 2011, Revised Selected Papers, Part I. Berlin, Germany: Springer-Verlag, 2011, 99: 169-194.

[12] VANDEN BROUCKE S K L M, De Weerdt J. Fodina: a robust and flexible heuristic process discovery technique[J]. decision support systems, 2017, 100: 109-118.

[13] ADRIANSYAH A, MUNOZ-GAMA J, CARMONA J, et al. Alignment based precision checking[C]//International conference on business process management. Berlin, Germany: Springer-Verlag, 2012: 137-149.

[14] SONG M, GÜNTHER C W, VAN DER AALST W M P. Trace clustering in process mining[C]//International conference on business process management. Berlin, Germany: Springer-Heidelberg, 2008: 109-120.

[15] BOSE R P J C, VAN DER AALST W M P. Context aware trace clustering: Towards improving process mining results[C]//proceedings of the 2009 SIAM International Conference on Data Mining. Society for Industrial and Applied Mathematics, 2009: 401-412.

[16] BOSE R P J C, VAN DER AALST W M P. Trace clustering based on conserved patterns: Towards achieving better process models[C]//International Conference on Business Process Management. Berlin, Germany: Springer-Heidelberg, 2009: 170-181.

[17] LEVENSHTEIN V I. Binary codes capable of correcting deletions, insertions, and reversals[C]//Soviet physics doklady. 1966, 10(8): 707-710.

[18] FERREIRA D, ZACARIAS M, MALHEIROS M, et al. Approaching process mining with sequence clustering: Experiments and findings[C]//International conference on business process management. Berlin, Germany: Springer-Heidelberg, 2007: 360-374.

[19] CADEZ I, HECKERMAN D, Meek C, et al. Model-based clustering and visualization of navigation patterns on a web site[J]. Data mining and knowledge discovery, 2003, 7(4): 399-424.

[20] VEIGA G M, FERREIRA D R. Understanding spaghetti models with sequence clustering for ProM[C]//International conference on business process management. Berlin, Germany: Springer-Heidelberg, 2009: 92-103.

[21] DE WEERDT J, VANDEN BROUCKE S, VANTHIENEN J, et al. Active trace clustering for improved process discovery[J]. IEEE Transactions on Knowledge and Data Engineering, 2013, 25(12): 2708-2720.

[22] PAGE L, BRIN S, MOTWANI R, et al. The PageRank citation ranking: Bringing order to the web[R]. Stanford InfoLab, 1999.

International Journal of Science and Engineering Applications
Volume 10-Issue 09, 122 - 124, 2021, ISSN:- 2319 - 7560

Formation of Cavitation and Supercavitation in a Rectangular Shape Nozzle

Espanta Ferdowsian
Department of Mechanical Engineering
University of Maryland
USA

Abstract: Flow inside a rectangular shape nozzle is simulated in this study. Finite volume scheme is utilized as the main solver for the current study. Second order scheme is utilized to discretize pressure. Second order upwind scheme is utilized for solving momentum equation. Then the momentum equation is coupled with the continuity equation to obtain the pressure and velocity at each cell. Cavitation inception and super cavitation is also found and discussed in this study and the results were also verified with previous Winklhofer et al. experiments.

Keywords: Cavitation; Multiphase; Turbulent; Flow; Mesh

1. INTRODUCTION

Spray atomization is affected enormously by turbulence and cavitation in a diesel injector nozzle and then combustion performance is affected as well [1-5]. Cavitation has been investigated widely in real-size diesel injector nozzle even though there are many ambiguous questions remained unanswered. High fuel injection pressure and small-scale size of the nozzles makes it very hard to visualize the flow inside nozzles and therefore many simulations are necessary to be done in order to understand the process of cavitation. Experimental studies of cavitation due to complex nature of the mentioned phenomena are very limited to transparent models under a certain boundary conditions [6-12]. With development of CFD (Computational Fluid Dynamic) numerical simulation has been of interest in the recent years as it is faster and less expensive.

2. NUMERICAL SCHEME

Continuity and momentum equations for an incompressible fluid flow can be written as following [1, 4]:

$$\frac{\partial \bar{u}_i}{\partial x_i} = 0$$

$$\frac{\partial \bar{u}_i}{\partial t} + \bar{u}_j \frac{\partial \bar{u}_i}{\partial x_j} = -\frac{1}{\rho}\frac{\partial \bar{P}}{\partial x_i} + \upsilon \frac{\partial^2 \bar{u}_i}{\partial x_j \partial x_j} - \frac{\partial}{\partial x_j} R_{ij}$$

In which $\bar{u}_i$ is the mean velocity, t the time, x_i is the position, ρ is the constant density, $\bar{P}$ is the mean pressure, $R_{ij} = \overline{u'_\iota u'_j}$ is the Reynolds stress tensor and υ is the kinematic viscosity and finally $u'_i = u_i - \bar{u}_\iota$ is the fluctuating components of the velocity.

The RSTM evaluates differential transport equations in order to obtain turbulence stress components:

$$\frac{\partial}{\partial t} R_{ij} + \bar{u}_k \frac{\partial}{\partial x_k} R_{ij} = \frac{\partial}{\partial x_k}\left(\frac{\upsilon_t}{\sigma^k} R_{ij}\right) - \left[R_{jk}\frac{\partial \bar{u}_j}{\partial x_k}\right] - C_1 \frac{\epsilon}{k}\left[R_{ij} - \frac{2}{3}\delta_{ij} k\right] - C_2\left[P_{ij} - \frac{2}{3}\delta_{ij}\epsilon_1\right]$$

In which the turbulent production can found from the following equations:

$$P_{ij} = -R_{jk}\frac{\partial \bar{u}_j}{\partial x_k} - R_{jk}\frac{\partial \bar{u}_i}{\partial x_k}$$

The turbulent dissipation rate is also calculated from the following equation:

$$\frac{\partial \epsilon}{\partial t} + \bar{u}_j \frac{\partial \epsilon}{\partial x_i} = \frac{\partial}{\partial x_i}\left[\left(\upsilon + \frac{\upsilon_t}{\sigma^\epsilon}\right)\frac{\partial \epsilon}{\partial x_i}\right] - C^{\epsilon 1}\frac{\epsilon}{k} R_{ij}\frac{\partial u_i}{\partial x_j} - C^{\epsilon 2}\frac{\epsilon^2}{k}$$

Instantaneous fluid velocity was also obtained from the following equation:

$$\frac{du_i}{dt} = -\frac{u_i - \bar{u}_\iota}{T_I} + \left(\frac{2\overline{u'_\iota}^2}{T_I}\right)\xi_i(t)$$

The average time in which the turbulent eddies spent time in a particle track is obtained as following:

$$T_I = \int_0^\infty \frac{u'_p(t)u'_p(t+s)}{u'_q u'_q} ds$$

Bubble trajectory has also been obtained from the following equations:

$$R\ddot{R} + \frac{3}{2}\dot{R}^2 = \frac{1}{\rho}[(P_g - P_0 - P_s) - 4\mu\frac{\dot{R}}{R} - 2\frac{\sigma}{R} + \frac{R}{c}\frac{d}{dt}(P_g)]$$

$$P_g(t) = (P_{g0})(\frac{R_0}{R})^{3\gamma}$$

$$P_{g0} = P_{inf} + 2\frac{\sigma}{R_0} - P_{cr}$$

Bubble motion equation has also been used in order to obtain location of each cell during the calculation period:

$$\frac{dU_b}{dt} = 2g\vec{j} - \frac{3}{\rho}\nabla\vec{p} + \frac{3}{4}\frac{C_D}{R}(\vec{U} - \overrightarrow{U_b})|\vec{U} - \overrightarrow{U_b}| + \frac{3}{R}(\vec{U} - \overrightarrow{U_b})\dot{R} + \frac{1.542}{R}\nu^{0.5}(\vec{U} - \overrightarrow{U_b})\left|\frac{d\vec{U}}{dy}\right|^{0.5} sgn(\frac{d\vec{U}}{dy})\vec{j}$$

$$C_D = \frac{24}{Re}(1 - 0.197Re^{0.63}) + \frac{24}{Re}(2.6 \times 10^{-4}Re^{1.38})$$

3. RESULT AND DISCUSSION

At the beginning a structured dominant mesh is used in this study which is following the measurements introduced by Winklhofer et al. [13]. Figure 1 shows structured dominant mesh used in this study in which a fully structured mesh is used in the orifice area of the nozzle. Moreover, figure 2 shows close up view of the mesh near the orifice inlet.

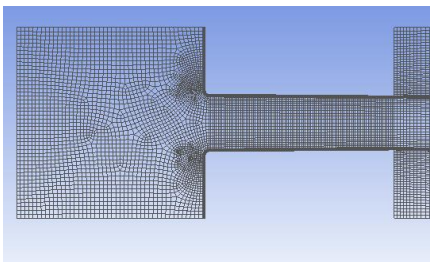


Figure 1. Mesh topology of the domain.

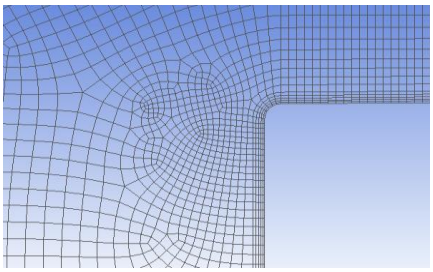


Figure 2. Close up view of the mesh near the orifice inlet.

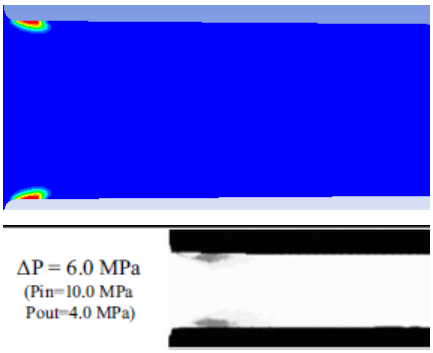


Figure 3. Contour of vapor volume fraction when cavitation inception occurs.

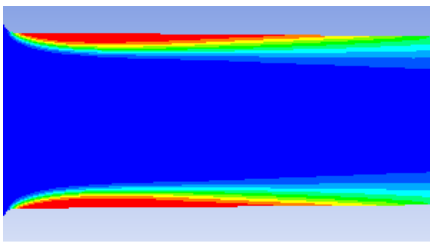


Figure 4. Contour of vapor volume fraction when super cavitation occurs.

As mentioned earlier the flow is solved using finite volume scheme where the two equations of continuity and momentum were solved together in order to obtain the velocity and pressure profile. Then using a structured dominant mesh as shown in figure 1 and 2 the whole domain is divided in to smaller domains to make it feasible for utilization of the current numerical scheme.

Figure 3 and 4 shows formation of cavitation at two stages which are namely inception and final. During

the cavitation inception in which the inlet pressure is fixed to 10 MPa and the outlet pressure is fixed to 4.0 MPa, only a smaller region of cavitating region forms which leads to affect the inlet radius region. It can also be seen from figure 3 that the inception region predicted from the simulation is matching with previous experimental result shown in black and white.

Finally, figure 4 shows formation of super cavitation when the inlet pressure is fixed to 10 MPa and the outlet pressure is fixed to 2.5 MPa. It can be seen that the whole orifice area is covered by the vapor volume fraction in this case and any further increase in the pressure difference would make the flow choke and the whole physical properties would change.

4. CONCLUSION

In this study the flow is simulated in a rectangular shape nozzle using finite volume scheme. In order to make the simulation process feasible the domain is divided in to smaller regions using structured dominant mesh.

1- Cavitation inception was found and verified using previous experimental contour.

2- Super cavitation was found and verified using previous experimental contour.

5. REFERENCES

[1] SMJ Zeidi, M.M., Numerical investigation of the effect of different parameters on emitted shockwave from bubble collapse in a nozzle. Journal of Particle Science & Technology, 2021. 6(2): p. 13.

[2] Zeidi, S. and M. Mahdi, Investigation the effects of injection pressure and compressibility and nozzle entry in diesel injector nozzle's flow. Journal of Applied and Computational Mechanics, 2015. 1(2): p. 83-94.

[3] Zeidi, S. and M. Mahdi. Investigation of viscosity effect on velocity profile and cavitation formation in Diesel injector nozzle. in Proceedings of the 8th international conference on internal combustion engines. 2014.

[4] Zeidi, S.M.J. and M. Mahdi, Evaluation of the physical forces exerted on a spherical bubble inside the nozzle in a cavitating flow with an Eulerian/Lagrangian approach. European Journal of Physics, 2015. 36(6).

[5] SMJ Zeidi, M.M., Effects of nozzle geometry and fuel characteristics on cavitation phenomena in injection nozzles. Proceedings of the 22st Annual International Conference on Mechanical Engineering-ISME, 2015.

[6] Azadeh Yazdi, M.N., Sepideh Amirahmadian, Nasim Sabetpour, Amirmasoud Hamedi, Utilization of Schnerr-Sauer Cavitation Model for Simulation of Cavitation Inception and Super Cavitation. International Journal of Aerospace and Mechanical Engineering, 2021. 15(7).

[7] Mohammadreza Nezamirad , S.A., Nasim Sabetpour ,Amirmasoud Hamedi ,Azadeh Yazdi, Effect of Needle Height on Formation of Cavitation in a Six-Hole Diesel Injector Nozzle. 6th national conference on Mechanical and Aerospace Engineering, 2021.

[8] Mohammadreza Nezamirad, N.S., Azadeh Yazdi, Amirmasoud Hamedi, Investigation the Effect of Velocity Inlet and Carrying Fluid on the Flow inside Coronary Artery. International Journal of Aerospace and Mechanical Engineering, 2021. 15(7).

[9] Nasim Sabetpour, A.Y., Sepideh Amirahmadian, Mohammadreza Nezamirad, Amirmasoud Hamedi, Formation of Vapor Volume Fraction in a real size nozzle using Schnerr and Sauer approach. Forth Conference on Technology Development in Mechanical and Aerospace Engineering, 2021.

[10] Mohammadreza Nezamirad, S.A., Nasim Sabetpour, Azadeh Yazdi, Amirmasoud Hamedi, Effect of Needle Height on Discharge Coefficient and Cavitation Number. International Journal of Aerospace and Mechanical Engineering, 2021. 15(7).

[11] Mohammadreza Nezamirad, S.A., Nasim Sabetpour, Azadeh Yazdi, Amirmasoud Hamedi, Effect of Needle Height on Discharge Coefficient and Cavitation Number. International Journal of Aerospace and Mechanical Engineering, 2021. 15(7).

[12] Mohammadreza Nezamirad, S.A., Nasim Sabetpour, Azadeh Yazdi,Amirmasoud Hamedi, Effect of Different Diesel Fuels on Formation of the Cavitation Phenomena. International Journal of Aerospace and Mechanical Engineering, 2021. 15(7).

[13] Winklhofer, E., et al. Comprehensive hydraulic and flow field documentation in model throttle experiments under cavitation conditions. in Proceedings of the ILASS-Europe conference, Zurich. 2001

International Journal of Science and Engineering Applications
Volume 10-Issue 09, 125 - 131, 2021, ISSN:- 2319 - 7560

Selected Soft Computing Algorithms for Job Shop Problem (JSP)

Adedeji Oluyinka Titilayo[*1]
Department of Information System Science, Ladoke Akintola University of Technology, Ogbomoso, Nigeria
otadedeji@lautech.edu.ng

Alade Oluwaseun Modupe [2*]
Department of Cyber Security Science, Ladoke Akintola University of Technology, Ogbomoso, Nigeria
oalade75@lautech.edu.ng

Makinde Bukola Oyeladun [3*]
Department of Computer Science, Osun State College of Technology, Esa-Oke. Nigeria
bukolamakinde22@gmail.com

OYELEYE Taye E
Department of Computer Science and Engineering, Ladoke Akintola University of Technology, Ogbomoso, Nigeria

(*Corresponding author's e-mail: otadedeji@lautech.edu.ng, oalade75@lautech.edu.ng, bukolamakinde22@gmail.com)

Abstract: Job Shop Problem (JSP) is an optimization problem in computer science and operations research in which jobs are assigned to resources at particular times. Each operation has a specific machine that it needs to be processed on and only one operation in a job can be processed at a given time. This problem is one of the best known combinatorial optimization problems. The aim of this project is to adapt Bat, Bee, Firefly, and Flower pollination algorithms, implement and evaluate the developed algorithms for solving Job Shop Problem.

Keywords: scheduling. Optimization, job shop problem, BAT, BEE)

1. INTRODUCTION

Scheduling is the allocation of shared resources over time to competing activities. It has been the subject of a significant amount of literature in the operations research field. Emphasis has been on investigating machine scheduling problems where jobs represent activities and machines represent resources; each machine can process at most one job at a time practical and varied. They arise in diverse areas such as flexible manufacturing systems, production planning, computer design, logistics, communication, etc. A scheduling problem is to find sequences of jobs on given machines with the objective of minimizing some function of the job completion times.

In a simpler version of this problem, flow shop scheduling, all jobs pass through all machines in the same order. A more complex case is represented by a job shop scheduling problem where machine orderings can be different for each job. Job shop problem (JSP) is one of the hardest combinatorial optimization problems. It belongs to the class of Non-deterministic Polynomial (NP) hard problems, consequently there are no known algorithms guaranteed to give an optimal solution and run in polynomial time. That means, classical optimization methods (branch and bound method, dynamic programming) can be used only for small scale tasks. (Binato, Hery, Loewenstern and Resende, 2002).

2. THEORECTICAL BACKGROUND

Job Shop Problem (JSP)

JSP is a static optimization problem, since all information about the production program is known in advance. General job shop problem is the probably most studied one by academic research during the last three decades and is notoriously difficult problem to solve. The JSP is a Non-deterministic Polynomial (NP) hard problem and among those optimization problems, it is one of the least tractable known problem (Garey and Johnson, 1979). It is purely deterministic, since processing time and constraints are fixed and no stochastic events occur.

JSP also illustrates some of the demands required by a wide array of real-world problems. In a shop floor, machines process jobs and each job contain a certain number of operations. Each operation has its own processing time and has to be processed on a dedicated machine. Each job has its own machine order and no relation exists between machine orders of any two jobs. For each job, the machine order of operations is prescribed and is known as technological production recipe or technological constrain, which are static to a problem instance. (Garey and Johnson, 1979).

Operations to be processed on one machine form an operation sequence for this machine. For a given problem, an operation sequence for each machine is called a schedule. Since each operation sequence can be permuted independently of operation sequences of other machines, the problem with n jobs and m machines can have a maximum of different solutions. The completion time of all jobs is known as makespan. The objective is to find a feasible schedule with minimum makespan. Feasible schedules are obtained by permuting the processing order of operations on machines without violating the technological constraints. (El-Bouri, Azizi and Zolfaghari, 2007)

Makoto and Hiroshi considered the JSP problem to minimize the total weighted tardiness with job-specific due dates and delay penalties, and a heuristic algorithm based on the tree search procedure was developed for solving the problem. Gomes and Barbosa presented an integer linear programming model to schedule flexible job shop, which considered job re-circulation and parallel homogeneous machines Loukit and Jacques dealt with a production scheduling problem in a flexible job shop with particular constraints-batch production.

An example of two jobs to be performed three machines (2×3) job shop problem is illustrated in Table 1 In this problem, each job requires three operations to be processed on a pre-defined machine sequence. The first job (J_1) need to be initially operated on the machine M_1 for 5time units and then sequentially processed on $M_2$and M_3 for 4 and 9 time units, respectively. Likewise, the second job (J2) has to be initially performed on M_3 for 5 time units and sequentially followed by M_1 and M_2 for 6 and 7 time units, respectively. The design task for solving JSP is to search for the best schedule(s) for operating all pre-defined jobs in order to optimize either single or multiple scheduling objectives, which is used for identifying a goodness of schedule such as the minimization of the makespan (C_{max}) (Ge HW, Sun, Liang, Qian, 2009).

Table1 An example of 2-jobs 3-machines scheduling problem with processing times.

(Ge HW, Sun , Liang, Qian, 2009)

Job (J_j)	Operation (O_{jk})	Time (t_{jk})	Machine (M_k)		
			M_1	M_2	M_3
J_1	O_{11}	5	5	-	-
	O_{12}	4	-	4	-
	O_{13}	9	-	-	9
J_2	O_{23}	5	-	-	5
	O_{21}	6	6	-	-
	O_{22}	7	-	7	-

I. BAT ALGORITHM

Bat algorithm was developed by Xin-She Yang in 2010. The algorithm exploits the so-called echolocation of the bats. The bats use sonar echoes to detect and avoid obstacles. It is generally known that sound pulses are transformed into a frequency which reflects from obstacles. The bats navigate by using the time delay from emission to reflection. They typically emit short, loud sound impulses.

II. BEE ALGORITHM

The BCO was inspired by bees behavior in the nature. The basic idea behind the BCO is to create the multi agent system (colony of artificial bees) capable to successfully solve difficult combinatorial optimization problems. The Artificial Bee Colonies (ABC) is another novel optimization algorithm that comes under Swarm Intelligence. ABC algorithm is inspired by social behavior of natural bees. (Karaboga and Basturk, 2007).

III. FIREFLY ALGORITHM

Firefly algorithm is inspired by the social behavior of fireflies. Most of the fireflies produce short and rhythmic flashes and have different flashing behavior. Fireflies use these flashes for communication and attracting the potential prey. The swarm of fireflies will move to brighter and more attractive locations by the flashing light intensity that is associated with the objective function of problems considered, in order to obtain efficient optimal solutions. One major improvement is the firefly algorithm (FA) which was based on the flashing characteristics of tropical fireflies. The attraction behavior, light intensity coding, and distance dependence provides a surprising capability to enable firefly algorithm to handle nonlinear, multimodal optimization problems efficiently (Xin She Yang, 2008).

IV. FLOWER POLLINATION ALGORITHM

In nature, the main purpose of the flowers is reproduction via pollination. Flower pollination is related to the transfer of pollen, which is done by pollinators such as insects, birds, bats, other animals or wind. Some flower types have special pollinators for successful pollination. The four rules of pollination have been formulated based on the inspiration from flowering plants and they form the main updating equations of the flower pollination algorithm, the main actors of performing such transfer are birds, bats, insects, and other animals. There exist some flowers and insects that have made what we can call a flower-pollinator partnership. These flowers can only attract the birds that are involved in that partnership, and these insects are considered the main pollinators for these flowers (Glover, 2007).

3. RELATED WORKS

Several works have been done in recent years to solve the Job Shop Problem. Xueni and Henry (Xueni Qiu and Henry Lau, 2012) proposed a new hybrid algorithm based on Particle Swarm Optimization (PSO) and Artificial Immune Systems (AIS) theories of clonal selection and immune network to solve the job shop scheduling problem, which is a classical combinational optimization problem. Experimental results demonstrated that the algorithm was competitive among other methods and optimal solutions were obtained within a reasonable computation time, especially for small size problems. However, there were occurrence of unexpected events and disturbances in the scheduling process.

Chaudry (Chaudry, 2012) presented a modified permutation chromosome representation for schedules with alternative machines for given operations. Although all the alternatives were included in the chromosome, the second and subsequent appearances of an operation in the chromosome

did not contribute to the overall calculation of the fitness value. However, all occurrences were included in the crossover and mutation operations, allowing for alternatives.

This chromosome representation did not represent a customization of the Genetic Algorithm (GA) method to accommodate a particular problem. Instead, it was a generalization of the existing representation, which in turn became a reduced form of this generalized form. However, the second chromosome representation also produced the same results as the modified permutation representation.

Beck and Wilson (2007) addressed job shop scheduling when the durations of the activities were independent random variables. A theoretical framework was created to formally define this problem and to prove the soundness of two algorithm components: Monte Carlo simulation to find upper bounds on the probabilistic makespan of a solution and a partial solution; and a carefully defined deterministic JSP whose optimal makespan was a lower bound on the probabilistic makespan of the corresponding probabilistic JSP. Then used these two components together with either constraint programming or Tabu search to define a number of algorithms to solve probabilistic JSPs, introduced three solution approaches: a branch-and-bound technique using Monte Carlo simulation to evaluate partial solutions; an iterative deterministic search using Monte Carlo simulation to evaluate the solutions from a series of increasingly less constrained problems based on a parameterizable lower bound; and a number of deterministic filtering algorithms which generate a sequence of solutions to a deterministic JSP, each of which was then simulated using Monte Carlo simulation. Empirical evaluation demonstrated that the branch-and-bound technique was only able to find approximately optimal solutions for very small problem instances. The iterative deterministic search performed as well as, or better than, the branch-and-bound approach for all problem sizes. However, for medium and large instances, the deterministic filtering techniques performed much more strongly while providing no optimality guarantees.

The scheduling of jobs to machines is a very challenging problem with several constraints that have to be explored in different ways. The use of certain heuristics and exact methods have been used to solve JSP such as Tabu Search. Although Tabu Search has gained popularity in recent years in terms of space i.e. space complexity, however, it is hard to understand the concept of k-insertion and implementing the algorithms. Other problems such as premature convergence and getting stuck in local optima would be given proper attention.

Therefore, Bat, Bee, Firefly and Flower Pollination algorithms were adapted and implemented for solving Job Shop Problem and to carry out performance evaluation of selected algorithms for solving Job Shop Problem.

4. METHODOLOGY

Finding a solution to the Job Shop Problem requires that we set up the Bat, Bee, Firefly and Flower pollination algorithms in a specialized way. The following parameters were used to carry out this project.

BAT

Bat algorithm was guided by the following parameters:

i. population size: 100
ii. loudness: 0.25
iii. pulse rate: 0.1
iv. search variable dimension: 50
v. maximum iteration:100

BEE

Bee algorithm was guided by the following parameters:

i. number of scout bees: 30
ii. maximum iteration: 100

FIREFLY

Firefly algorithm was guided by the following parameters:

i. maximum iteration: 100
ii. number of fireflies: 100
iii. attraction co-efficient: 0.2
iv. mutation co-efficient: 0.2
v. light coefficient: 1

FLOWER POLLINATION

Flower pollination algorithm was guided by the following parameters:

i. switch probability: 0.8
ii. population size: 100
iii. maximum iteration: 100

In this paper, the sequences for each job are stored in a symmetric matrix, as shown in the figure 2. Here in this paper are 6 jobs to be carried out on 6 machines.

```
T=[1  3  6  7  3  6
   8  5  10  10  10  4
   5  4  8  9  1  7
   5  5  5  3  8  9
   9  3  5  4  3  1
   3  3  9  10  4  1];


js=[3  1  2  4  6  5
    2  3  5  6  1  4
    3  4  6  1  2  5
    2  1  3  4  5  6
    3   2  5  6  1  4
    2   4  6  1  5  3];
```

Figure 2: Data for 6 jobs using 6 machines

5. RESULTS AND DISCUSSION

In other to check that the proposed algorithms give optimal result, each of the algorithms (Bat, Bee, Firefly and Flower Pollination) were applied to 6 numbers of jobs using 6 machines respectively, setting iteration limit to 100 for 5 successive runs.

After using Firefly algorithm on 6 jobs and 6 machines for the first run, the best processing time = 55s as shown in Figure 3.1 and Figure 3.2.

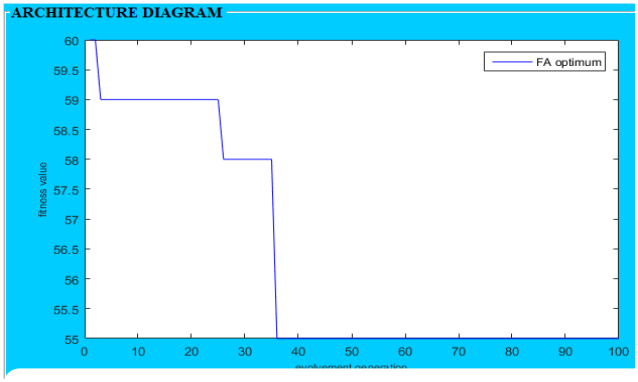


Figure3.1: Architecture Diagram showing fitness value using Firefly algorithm

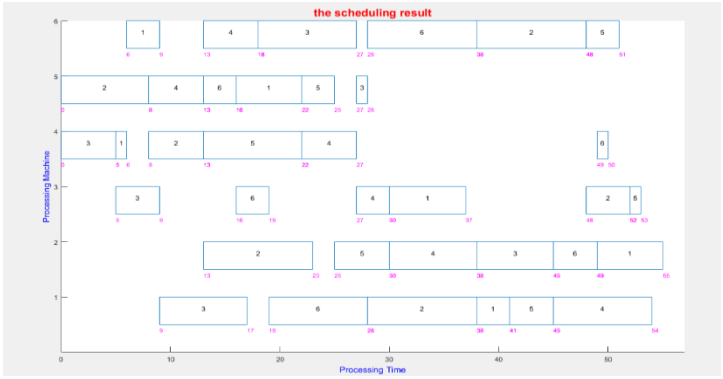


Figure3.2: Scheduling Result Showing Time Sequence using Firefly algorithm

After using Bee algorithm on 6 jobs and 6 machines for the first run, the best processing time = 55s as shown in Figure 3.3 and Figure 3.4.

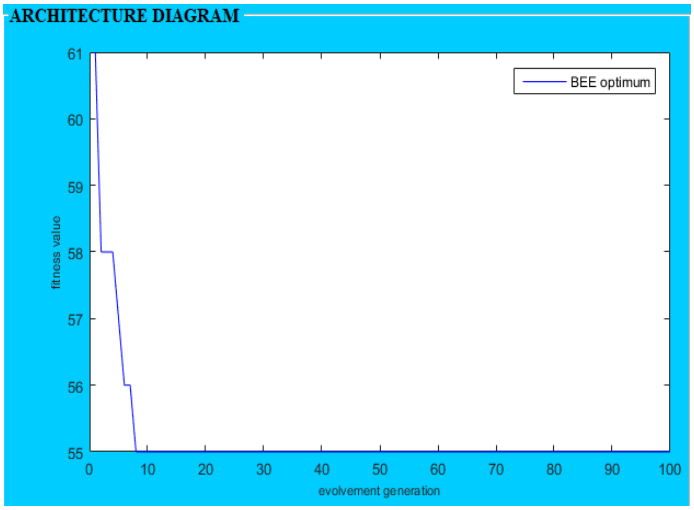


Figure3.3: Architecture Diagram showing fitness value using Bee algorithm

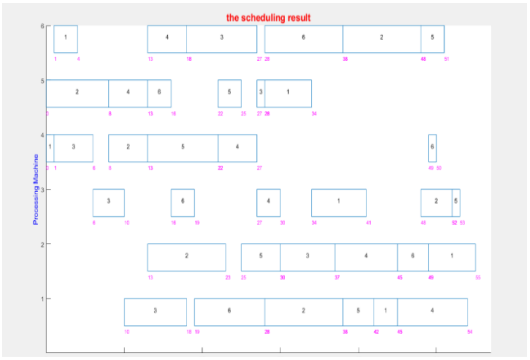


Figure3.4: Scheduling Result Showing Time Sequence using Bee algorithm

After using Bat algorithm on 6 jobs and 6 machines for the first run, the best processing time = 55s as shown in Figure 3.5 and Figure 3.6.

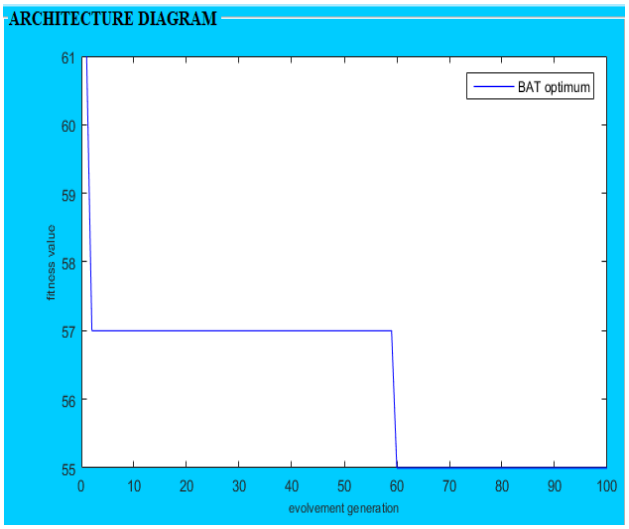


Figure3.5: Architecture Diagram showing fitness value using Bat algorithm

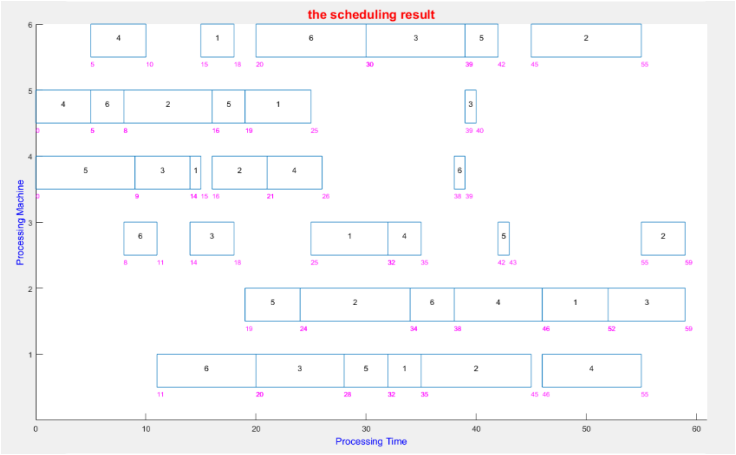


Figure3.6: Scheduling Result Showing Time Sequence using Bat algorithm

After using Flower Pollination algorithm on 6 jobs and 6 machines for the first run, the best processing time = 59s as shown in Figure 3.7 and Figure 3.8.

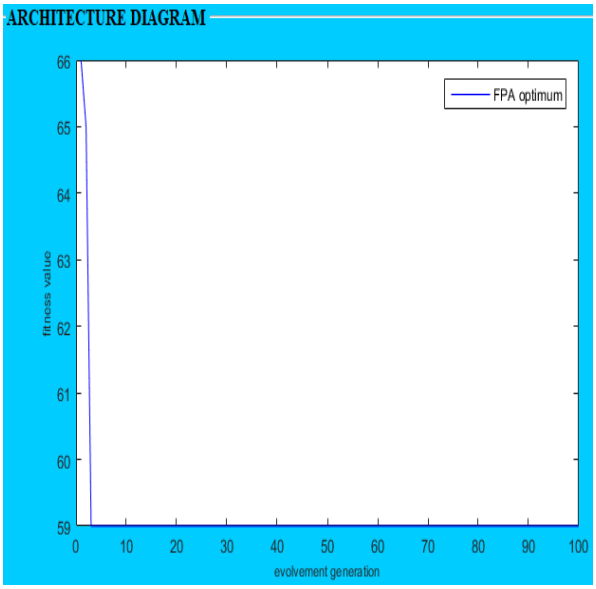


Figure3.7: Architecture Diagram showing fitness value using Flower Pollination algorithm

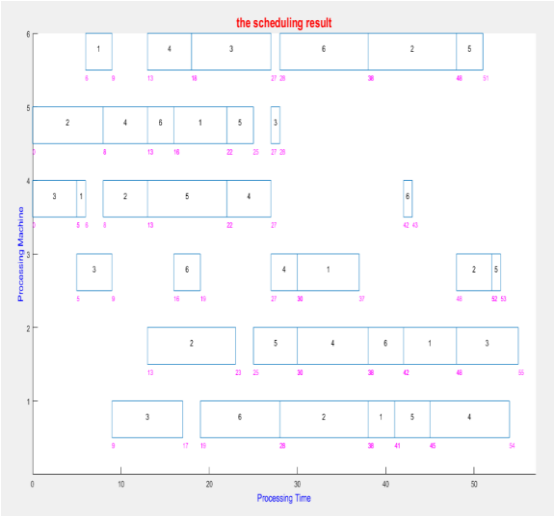


Figure3.8: Scheduling Result Showing Time Sequence using Flower Pollination algorithm

Table 2: Comparisons of Algorithms Based on Processing Time

Experimental Run	FIREF	BEE	BAT	FLOWER POLLINATION
	Best Processing Time(s)	Best Processing Time(s)	Best Processing Time(s)	Best Processing Time(s)
FIRST	55	55	55	59
SECOND	57	57	55	55
THIRD	55	55	58	55
FOURTH	55	57	55	55
FIFTH	57	55	55	55
AVERAGE	55.8	55.8	55.6	55.8

Based on the above comparison table for test data, Bat Algorithm yielded comparatively the best processing time of 55.6 seconds when compared to other selected algorithms.

Table 3: Software Complexity Metrics for Each Algorithm

Performance Metrics Result

The result of the software complexity metrics is given in Table 3:

Metrics	BAT	BEE	FIREFLY	FLOWER P A
Vocabulary(n)	82	89	102	148
Lines of Code (LOC)	63	89	92	91
Cyclomatic Complexity(G)	11	11	9	8
Calculated Program Length (N_h)	480.3865	540.9707	640.1314	1018.0621
Maintainability Index (M.I)	78.4369	72.4140	71.6290	70.0986

From the above table, Bat algorithm had the best Vocabulary(n) with a unit of 82 and the best Calculated Program Length (N_h) with a unit of 480.3865. However, Flower Pollination algorithm had a Cyclomatic Complexity of 8, Maintainability Index of 70.0986. It can be stated that Bat and Flower Pollination algorithms performed well in terms of software complexity, however Bat algorithm had lesser lines of code compare to Flower Pollination algorithm.

Conclusion

In this paper, Bat, Bee, Firefly and Flower Pollination algorithms were adapted and implemented for solving Job Shop Problem. The experimental result obtained on standard JSP showed that Bat Algorithm (BA) provides better results than Bee, Firefly, Flower Pollination Algorithms in most of the instances. Future works can be tailored towards hybridization of two algorithms to solve Job Shop Problem and observe whether there is any improvement to the results presented to this work.

References

[1] Abdel-Raouf O, Abdel-Baset M, El-Henawy I (2014a) An improved flower pollination algorithm with chaos. Int J Educ Managt Eng 4(2):1–8.

[2] Binato, Silvio & J. Hery, W & M. Loewenstern, D. (2000). A Grasp for Job Shop Scheduling. Essays and Surveys on Metaheuristics. 15. 10.1007/978-1-4615-1507-4_3.

[3] BJ, Glover. (2008). Understanding Flowers and Flowering: An integrated approach.Understanding Flowers and Flowering: An integrated approach. 1-256.10.1093/acprof:oso/9780198565970.001.0001.

[4] Bruker, P., Schlie, R. (1990): Job Shop Scheduling with Multi-Purpose Machine, Computing,Vol. 45, 1990.

[5] Chaudhry, Imran Ali. (2012). Job Shop Scheduling Problem with Alternative Machines UsingGenetic Algorithms. Journal of Central South University of Technology. 19.10.1007/s11771-012-1145-8.

[6] Dervis Karaboga, Bahriye Basturk (2007): A powerful and efficient algorithm for numerical Function Optimization: Artificial Bee Colony (ABC) Algorithm J Glob Optim, Volume 39, Issue 3, pp 459-471.

[7] Dubey HM, Pandit M, Panigrahi BK (2015b) A biologically inspired modified flower pollinationalgorithm for solving economic dispatch problems in modern power systems. Cognit Comput 7(5):594–608.

[8] El-Bouri A., Azizi N., and Zolfaghari, S. (2007). A comparative study of a new heuristic based on adaptive memory programming and simulated annealing: The case of job shopscheduling. *European Journal of Operational Research, 177(3), 1894-1910.*

[9] Fogel D.B., Fogel L.J. (1996) An introduction to evolutionary programming. In: Alliot JM., Lutton E., Ronald E., Schoenauer M., Snyers D. (eds) Artificial Evolution. AE 1995. Lecture Notes in Computer Science, vol 1063. Springer, Berlin, Heidelberg

[10] Garey, M. R., & Johnson, D. S. (1979). Computers and Intractability: A Guide to the Theory of NP-Completeness. W.H. Freeman and Company, New York.

[11] Ge HW, Sun L, Liang YC, Qian F (2008). An Effective PSO and AISBased Hybrid IntelligentAlgorithm for Job-Shop Scheduling. IEEE Trans. Syst., Man, Cybern. A, Syst. Humans.38: 358-368.

[12] H. P. Schwefel. (1981). “Numerical Optimization of Computer Models,” John Wiley & Sons, Chichester. Volume 3, Issue 1

[13] Hayes-Roth, Frederick. (1975). Review of "Adaptation in Natural and Artificial Systems by John H. Holland", The U. of Michigan Press, 1975. Intelligence/sigart Bulletin - SIGART. 1515. 10.1145/1216504.1216510.

[14] J. Christopher Beck and Nic Wilson (2007): Proactive Algorithms for Job Shop Scheduling with Probabilistic Durations. *Journal of Artificial Intelligence Research 28 (2007) 183–232.*

[15] John McCall (2004): Genetic algorithms for modelling and optimization. *Journal ofComputational and Applied Mathematics 184 (2005) 205 – 222.*

[16] Koza, J. R. (1992) Genetic Programming: On the Programming of Computers by Means of Natural Selection. Cambridge, MA: The MIT Press.

[17] Lars Chittka, James D.Thomson, Nickolas M. Waser (1999): Flower Constancy, Insect Psychology, and Plant Evolution.

[18] Liangshan Shao, Yuan Bai and Yunfei Qiu, Zhanwei Du (2012): Particle swarm optimizationalgorithm based on semantic relations and its engineering applications.

[19] Lukasik S, Zak S (2009) Firefly algorithm for continuous constrained optimization task, ICCCI2009. In: Nguyen NT, Kowalczyk R, Chen SM (eds) Lecture notes in artificial intelligence, vol 5796. Springer, Berlin, pp 97–100.

[20] Marta Castilho Gomes, Ana Paula Barbosa-Povoa and Augusto Q Novais (2005): Optimal scheduling for flexible job shop operation.

[21] Mohamed Abdel-Basset and Laila Shawky (2008): Flower pollination algorithm: a comprehensive review.

[22] Nishant Pathak, Sudhanshu Prakash Tiwari (2012): Travelling Salesman Problem Using Bee Colony With SPV. *International Journal of Soft Computing and Engineering (IJSCE)*

[23] Pavlyukevich. I, (2007): Levy flights, non-local search and simulated annealing, *Journal of Computational Physics*, vol 226, no2, pp 1830-1844.

[24] Rechenberg I. (1965) Cybernetic solution path of an experimental problem. Royal Aircraft Establishment, Farnborough p. Library Translation 1122.

[25] Sana Jawarneh and Salwani Abdullah (2015): Sequential Insertion Heuristic With Adaptive Bee Colony Optimisation Algorithm for Vehicle Routing Problem With Time Windows.

[26] Sato T and Hagiwara M (1997): Bee system Finding solution by a concentrated search. Proc. IEEE International Conference on Systems, Man and Cybernetics, 12-15, Hyatto Rlandoo, Rlndof, Lordia, USA. 3954-3959.

[27] Teodorovic, D. & Dell'Orco, M. (2005): Bee colony optimization: A cooperative learning approach to complex transportation problems. Advanced OR and AI Methods in Transportation, pp. 51-60.

[28] Xueni Qiu, Henry Y.K. Lau (2012): An AIS-based Hybrid Algorithm with PSO for Job Shop Scheduling Problem.

[29] Yang X-S (2008) Nature-inspired metaheuristic algorithm. Luniver Press, Beckington.

[30] Yang X-S (2009) Firefly algorithms for multimodal optimization, In: Stochastic algorithms: foundations and applications, SAGA, Lecture Notes in Computer Sciences, 5792, 169 178.

[31] Yang XS (2010) Firefly algorithm, stochastic test functions and design optimisation. Int J BioInspired Comput 2(2):78–84.

[32] Yang, X. S., (2010). A New Metaheuristic Bat-Inspired Algorithm, in: Nature Inspired Cooperative Strategies for Optimization (NISCO 2010) (Eds. Cruz, C.; Gonz´alez, J. R.; Pelta,

[32] Zaid Abdi Alkareem Alyasseri, Mohammed Azmi Al-Betar, Ahamad Tajudin Khader, Mohammed A. Awadallah (2018): Variants of the Flower Pollination Algorithm: A Review.

International Journal of Science and Engineering Applications
Volume 10-Issue 09, 132 - 139, 2021, ISSN:- 2319 - 7560

Voice Recognition System for Door Access Control Using Mobile Phone

Falohun Adeleye Samuel[1]*
Department of Computer Engineering, Ladoke Akintola University of Technology, Ogbomoso. Nigeria.
asfalohun@lautech.edu.ng

Adedeji Oluyinka Titilayo[2*]
Department of Information System Science, Ladoke Akintola University of Technology, Ogbomoso. Nigeria
otadedeji@lautech.edu.ng

Adegbola Oluwole Abiodun [3*]
Department of Electronic and Electrical Engineering, Ladoke Akintola University of Technology, Ogbomoso, Nigeria.
oaadegbola@lautech.edu.ng

Alade Oluwaseun Modupe[4*]
Department of Cyber Security Science, Ladoke Akintola University of Technology, Ogbomoso, Nigeria. Email
oalade75@lautech.edu.ng

Makinde Bukola Oyeladun[5*]
Department of Computer Science, Osun State College of Technology, Esa-Oke. Nigeria
bukolamakinde22@gmail.com

Isamot Rokeeb Mayowa
and
Ajadi Mayowa Samuel
Department of Computer Engineering, Ladoke Akintola University of Technology, Ogbomoso. Nigeria.

Abstract: Security is one of the most important issues for any individual or organization, and as technology has advanced, numerous techniques to protecting lives and property have been deployed through door access control systems. The typical method of unlocking a door is to open it with a real key or by twisting the door knob. Physical keys that are used to open doors are subject to duplication and can be misplaced. Furthermore, typical biometric technologies and other technologies are vulnerable to a variety of failures, such as a person's finger being cut off to produce a fingerprint scan, a pin being hacked using various methods or permutations, and a person's photo being used for facial recognition. Furthermore, it is more difficult for people with physical disabilities to unlock a door system without the assistance or support of others. For example, it is difficult for a person in a wheelchair to open a door system without the assistance or support of another person. As a result, a speech recognition access control system that can accommodate both able-bodied and impaired people is unavoidable. This paper demonstrates how voice recognition may be used to access door systems by creating a door access control system that employs speech recognition to simplify the work of providing access to door systems via a mobile phone connected via Bluetooth. The system's performance is perfectly in line with its design.

Keywords-Door; Voice Recognition; Bluetooth; Arduino; Microcontroller.

1. INTRODUCTION

Voice recognition is just one of several biometrics applications that can be used in access control systems. Voice recognition (32%) is the most popular biometric measure, followed by fingerprints (27%), facial scan (20%), hand geometry (12%), and iris scan (12%), according to a Unisys poll (10 percent). Consumers prefer speech recognition biometric systems, according to this survey. By reading in the unlock phrase or password, voice recognition would be utilized to unlock many sorts of doors such as office doors, garage doors, and gates. The speech recognition system must first go through a training phase in order to understand and learn different voices from different people, and then go through a testing step to confirm that the system recognizes voices correctly. Different people would read in phrases to open a door system in this arrangement.

The voice is analyzed by the system, which extracts various features and intents from it. These qualities are recorded in a database so that the next time the system hears that voice, the properties in the database are matched and appropriate feedback is supplied. (2017, Cho et al.)

Security is one of the most important issues for any individual or organization, and as technology has advanced, numerous techniques to protecting lives and property have been deployed through door access control systems. The typical method of unlocking a door is to open it with a real key or by twisting the door knob. Physical keys that are used to open doors are subject to duplication and can be misplaced. Furthermore, typical biometric technologies and other technologies are vulnerable to a variety of failures, such as a person's finger being cut off to produce a fingerprint scan, a pin being hacked using various methods or permutations, and a person's photo being used for facial recognition. Furthermore, it is more difficult for people with physical disabilities to unlock a door system without the assistance or support of others. For example, it is difficult for a person in a wheelchair to open a door system without the assistance or support of another person. Hence, the need for a voice recognition access control system that can serve both abled-bodied and disabled individuals is inevitable.

The purpose of this study is to break the limitations of conventional door access control systems in ensuring safety of lives and properties. One limitation includes the inability of the current system to provide the best security due to different hacks on the system such as key duplication. It is also believed that it would serve as a stepping stone to the development of better voice recognition systems that would be aimed at providing better security and making life easier.

2. THEORECTICAL BACKGROUND

A system that restricts access to a location or resource selectively is known as an access control system. A door can be used as a physical means of preventing entry to specific people who do not have the appropriate access credentials, such as a key, keycard, fingerprint, voice password, RFID card, security token, or coin. In the remains of Nineveh, ancient Assyria's capital, the earliest known key and lock devices were uncovered. Since then, technology has progressed with the arrival of computers, which provide access control through the use of computer programs and software. One of the newest forms of access control is voice recognition, which entails decoding human speech and identifying the speaker.

There are two types of recognition: speaker recognition and speech recognition. Speaker recognition is the process of recognizing a person based on the features or characteristics of their speech. Speech recognition is the process of recognizing what the speaker has spoken. The act of confirming a speaker's identity in a system is known as speaker authentication. Enrollment (training phase) and verification are the two phases of a speech recognition system (testing phase). The speaker's voice is recorded as input signals, then features or qualities are extracted to create a template or model during the enrolment phase. A sample speech utterance is compared to models already stored in the system in the verification step to determine the best match (es) (De Vries *et al*, 1992).

3. REVIEW OF RELATED WORK

Both the industry and academics are making remarkable advancements in voice recognition systems. The structural and functional designs of speech recognition door access control systems have evolved over time, and research is ongoing to develop voice recognition algorithms that are flexible and capable of accurately detecting the voices of various individuals. Some major advancements in speech recognition door access control systems include:

Intelligent Voice-Based Door Access Control System using Adaptive-Network-Based Fuzzy Inference Systems (ANFIS) for Building Security. Wahyudi *et al* (2007). This study looked at how a number of technologies, such as PIN pads, keys (both traditional and electronic), identity cards, cryptography and dual control processes, are used to protect secure facilities from illegal access. Speaker verification, or the capacity to authenticate a speaker's identification by analyzing speech, is an appealing and generally inconspicuous method of providing security for access into a sensitive or secure location. The research paper went on to say that a person's voice cannot be accurately stolen, lost, forgotten, guessed, or impersonated. The study presented the design and development of a voice-based door access control system for building security because of these benefits.

Real Time Recognition based Building Automation System. G. Muthuselvi *et al* (2014). This research examined how technology was used in houses to respond to the demands and instructions of the occupants in order to improve daily life at home. The embedded system was designed to detect and understand human voice instructions, which were then utilized to toggle various workloads. A speech recognition system, as well as an 8051 microcontroller kit and relays, were used to create the design. The results of the processing were then presented on a Liquid Crystal Display (LCD), which was primarily used to show system states.

Door Automation System using Bluetooth-Based Android for Mobile Phone. L. Kamelia *et al*. (2014). This study examined at how Bluetooth on an Android phone was utilized to automate the process of opening a door. Home controllers that are connected to a Windows-based PC are the most popular. This research used Bluetooth, a component of smart home technology, on a mobile device to make the procedure easier and more efficient. The hardware for the door-lock system consists of an android smart phone acting as the task master, a Bluetooth module acting as the command agent, an Arduino microcontroller acting as the controller center / data processing center, and a solenoid acting as the door lock output.

Agbo David O et al., (2017) Designed and implemented a door locking system using

android app. The application was created with the help of an Android app that produces a password that is recognized by Bluetooth to control the opening and closing of a door that is located a long distance away from the user. The Bluetooth module put on the door receives commands from Android phones and sends them to the microcontroller, which controls the door's opening and closing. The hardware was created on experimental boards after the design was modelled in the Proteus integrated development environment. The system's performance is perfectly in line with its design. The method can be utilized in a variety of situations when access to a container must be restricted.

Kamoru et al., (2018) designed motion detector alarm and security system. The model was developed utilizing an embedded microcontroller system capable of detecting intruder movements in a restricted area and then triggering an alarm system, motion detector system. However, a passive infrared sensor was used to identify the person's mobility based on their body heat. The passive infrared (PIR) sensor, which was utilized as an alternative detector in this project, was connected to a microcontroller, which activated the alarm

system and any other output devices linked to warn the house owner. Given the amount of time and resources saved, the project's future development was excellent. This system can be used as a model for larger projects that include audio-visual cameras and send the collected image to an email in real time.

Zaid A. Mundher et al., (2019) build a Real-Time Home Security Alarm System Using a Kinect Sensor. Using the Kinect sensor and the Kinect SDK, this project aimed to build and execute a low-cost, smart, and small real-time monitoring home security system. The results reveal that using the Kinect device to develop the proposed system is both efficient and computationally simple.

However, in view of the different design done by previous researchers, they can be improved upon by incorporating the use of voice recognition through mobile phone application to access door control through Bluetooth connection, this will provide better and faster access.

4. METHODOLOGY

4.1 Principle of Operation

This microcontroller-based door access control system uses Google's open-source speech-to-text on the android application through the mobile phone to operate a door using voice input. The android application and the android mobile phone are connected via a wireless link using an HC-05 Bluetooth module. Serial communication is used to communicate between the microcontroller and other system components.

The android application connected through the Bluetooth module sends commands to the microcontroller (ATMEGA328P). It then makes a decision based on the command it has received. The device has been developed to work with a speech to text software, which will use the microcontroller to send an electrical signal to the door latch. The activity of the door latch is determined by the input command from the speech to text app on the Android mobile app: OPEN voice input will activate the door to open, and CLOSE voice input will activate the door to close, as programmed into the Bluetooth module.

LEDs are also used as visual feedback; the red LED indicates data exchange between the android and the Bluetooth module, while the yellow LED indicates the status of the door. A buzzer was utilized as a AUDIO feed back, producing a buzzing sound when the door was open or closed.

4.2 Materials Used for the Project

The following are the list of the materials used in this project;

1. Bluetooth module
2. Door latch
3. wire
4. Wood work
5. Buzzer
6. Microcontroller
7. Led
8. Mobile phone

4.3 Bluetooth Stage

The HC-05 Bluetooth module used in this project is a simple to use Bluetooth SPP (serial port protocol) module that allows for seamless wireless serial connection setup. In this project, the HC-05 Bluetooth module functions as a wireless bridge between the microcontroller and the mobile phone app, allowing serial communication between the two.

4.4 The Microcontroller Specifications

Various factors are considered in the choice of microcontroller to use for a particular purpose. These include:

1. The number of digital inputs, analogue inputs the system concerned requires; a factor which helps to determine the minimum number of inputs and outputs (I/O) that the chosen microcontroller must have and the extent of need of an internal analogue to digital converter module.
2. The size of program memory storage required
3. The magnitude of clock frequency; a factor which determines the execution rate of tasks by the microcontroller
4. The number of interrupts and timer circuits required.

In a project of this kind where the number of task that can be handled is largely dependent on the amount of memory available, a microcontroller with a large memory sufficient input/output ports and analogue/digital channels such as the ATMega328P is quite acceptable for use. The choice as to which pin will be used in a particular application is controlled by programming the various special functions registers. Figure 1 presented the block diagram of voice recognition for door access control while Figure 2 showed the circuit diagram for the voice recognition for door access control using mobile phone through Bluetooth connection.

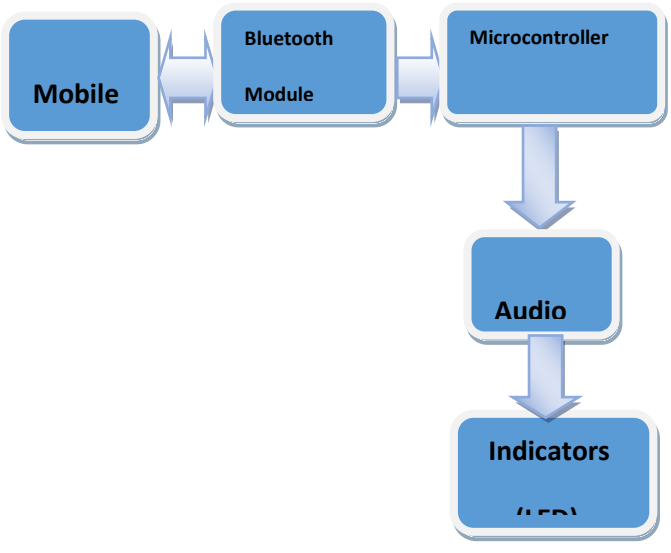


Figure 1: block diagram of voice recognition for door access control

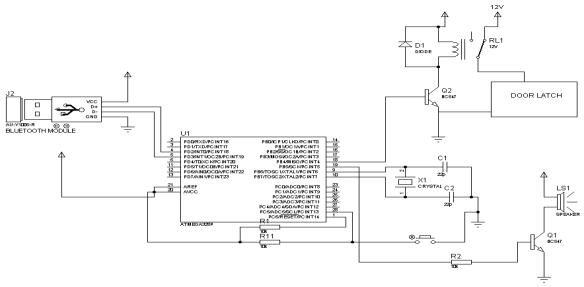


Figure 2: circuit diagram for the voice recognition for door access control using mobile phone through Bluetooth connection.

5. IMPLEMENTATION, TESTING AND RESULT

Construction

The physical realization of the project is very vital. Here the paper work is transformed into a finished hardware. After carrying out all the paper design and analysis, the project was implemented, constructed and tested to ensure its working ability. The construction of this project was done in three different stages.

1. The implementation of the whole project on a solder-less experiment board.
2. The soldering of the circuits on printed circuit boards.
3. The coupling of the entire project to the casing.

Figure 3 displayed the PCB artwork for the design of the android based voice controlled door.

Implementation

The implementation of this project was done on the breadboard. The power supply was first derived from a bench power supply in the school electronics lab. To confirm the workability of the circuits before the power supply stage was soldered. The implementation of the project on bread board was successful and it met the desired design aims with each stage performing as designed.

Soldering

The various circuits and stages of this project were soldered in tandem to meet desired workability of the project. The microcontroller stage was first soldered before the led indicator stages were done. The soldering of the project was done on a printed circuit board.

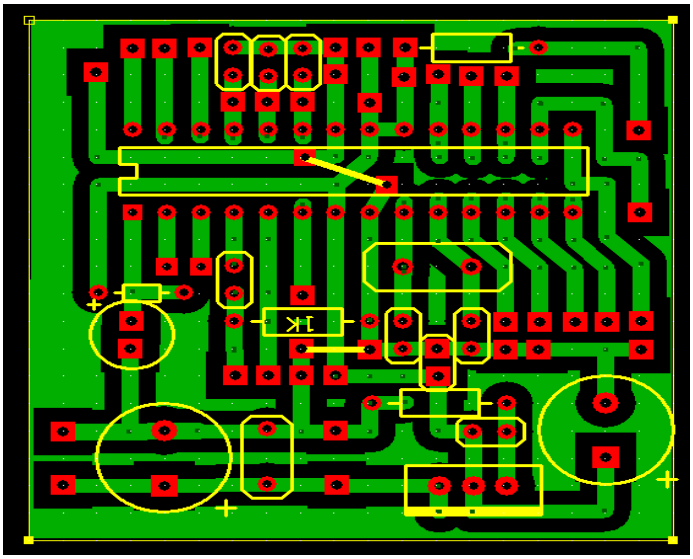


Figure 3: PCB artwork for the design of the android based voice controlled door.

Casing and boxing.

The third phase of the project construction is the casing of the project. This project was coupled and fixed to a model door for ease of demonstration.

Testing

Stage by stage testing was done according to the block representation on the breadboard, before soldering of circuit commenced on printed circuit board.

The process of testing and implementation involved the use of some test and measuring equipments stated below.

1. **Bench Power Supply**: This was used to supply voltage (5VDC) to the various stages of the circuit during the breadboard test before the power supply in the project was soldered. Also during the soldering of the project the power supply was still used to test various stages before they were finally soldered.
2. **Oscilloscope**: The oscilloscope was used to observe both the trigger and the echo signal waveforms and to ensure that all waveforms were correct and their frequencies accurate. The waveform of the oscillation of the crystal oscillator used was monitor to ensure proper oscillation at 16MHz.
3. **Digital Multi-meter**: The digital multi-meter basically measures voltage, resistance, continuity, current, frequency, temperature and transistor h_{fe}. The process of implementation of the design on the board required the measurement of parameters like, voltage, continuity, current and resistance values of the components and in some cases frequency measurement. The digital multimeter was used to check the output of the voltage regulators used in this project.

Results

The result from the research work was presented as displayed in Figure 4, 5, 6, 7, 8, 9 and 10.

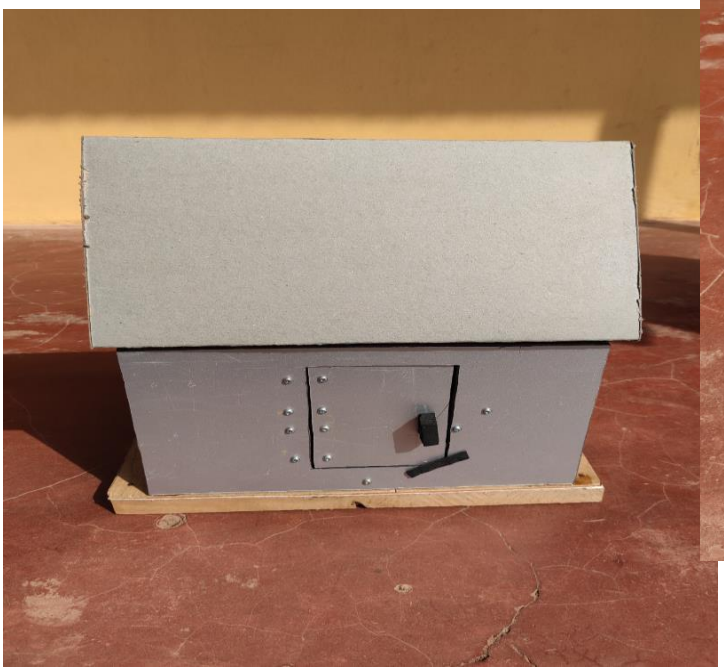


Figure 4: front view of the prototype

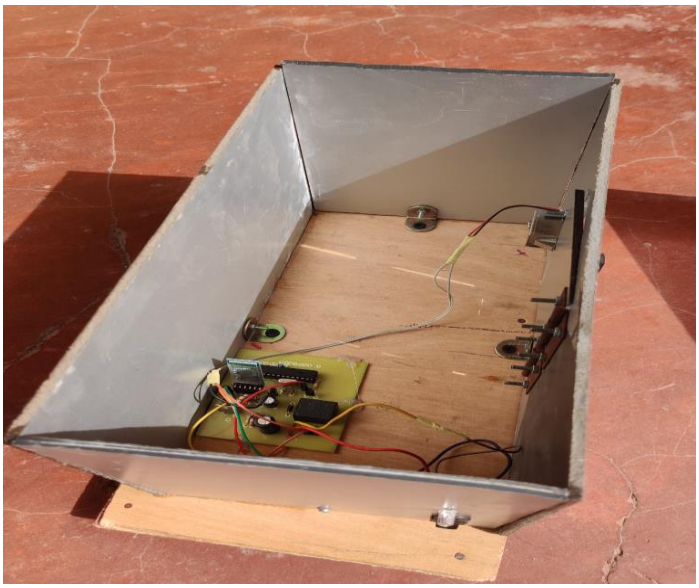


Figure 5: upward view

Figure 6: Side view of the prototype

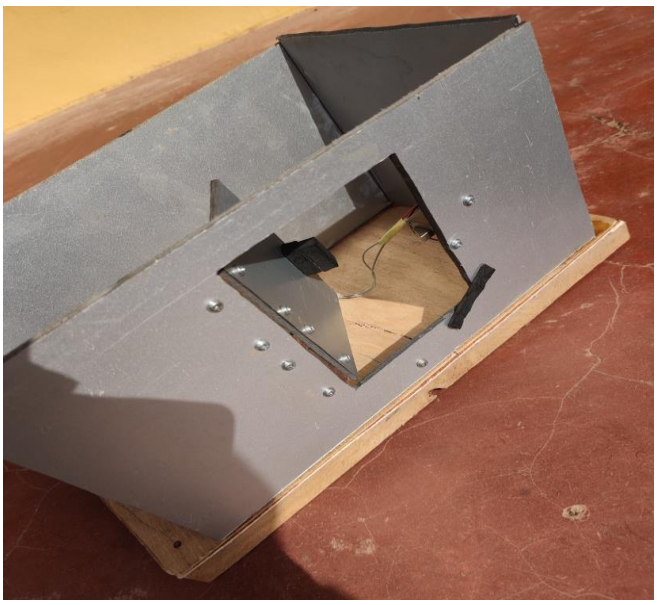


Figure 7: side of view of the prototype when opened

Figure 8: Shows interface of the Android application

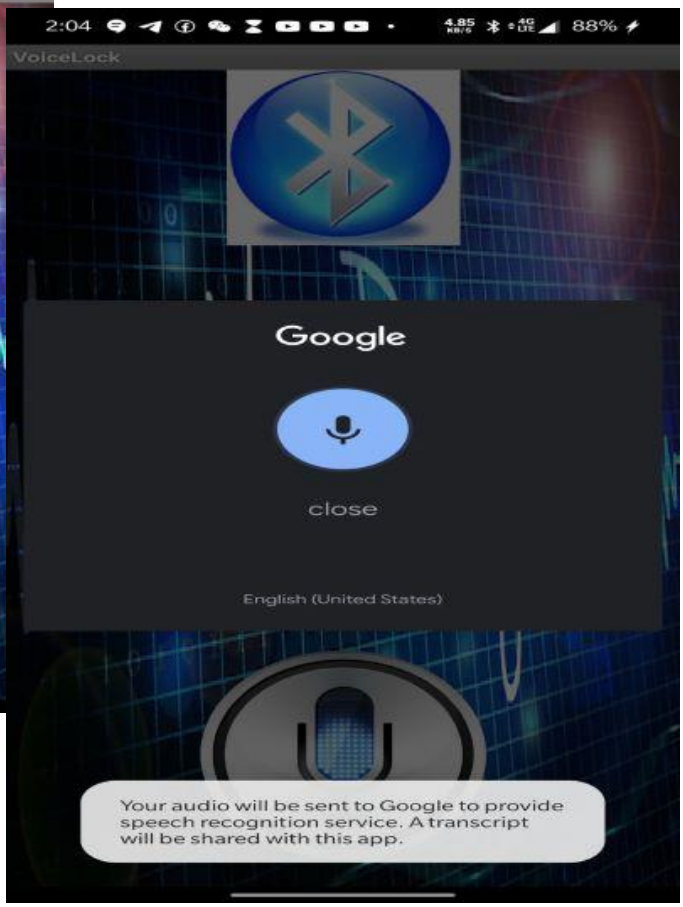


Figure 10: Show when the voice input is CLOSE

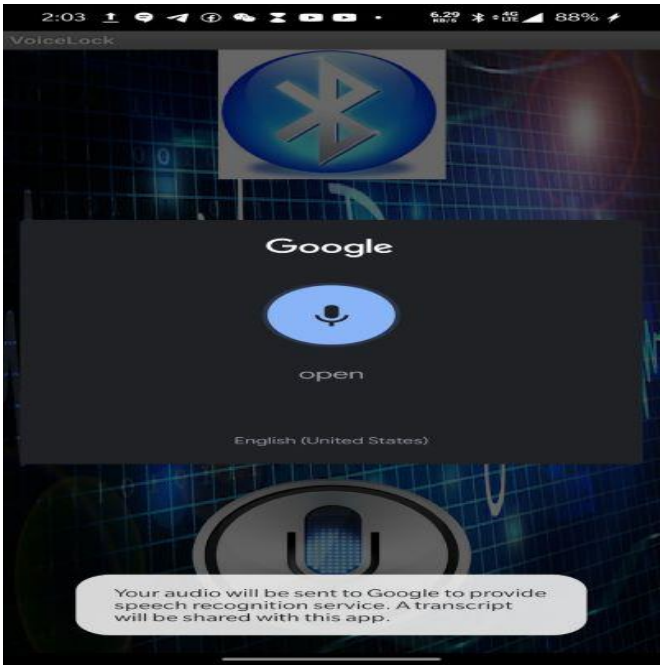


Figure 9: Shows when the voice input is OPEN

Problems Encountered

Like every research and practical engineering work, diverse kinds of problems are often encountered. The problems encountered in this project and how they were solved and maneuvered are listed below.

1. At the implementation stage of this project, the communication between the controller and the mobile used in this project was found failing. The problem was traced to both items not operating at the same frequency as designed. The oscillator was changed.
2. Network variation might cause delay in the delivery of the text message. However, the time delay was adjusted to balance the irregularities in the delivery of the message by network providers.

6. CONCLUSION AND RECOMMENDATIONS

The project which is the design and construction of the android based voice controlled door was designed considering some factors such as economic application, design economy, availability of components and research materials, efficiency, compatibility and portability and also durability. The performance of the project after test met design specifications. However, the general operation of the project and performance is dependent on the user who is prone to human error such as entering wrong voice input.

Also the operation is dependent on how well the soldering is done, and the positioning of the components on the printed circuit board. If poor soldering lead is used the circuit might form dry joint early and in that case the project might fail. Also if logic elements are soldered near components that radiate heat, overheating might occur and affect the performance of the entire system. Other factors that might affect performance include transportation, packaging, ventilation, quality of components, handling and usage.

The construction was done in such a way that it makes maintenance and repairs an easy task and affordable for the user should there be any system breakdown. The project really gave a good exposure to digital and practical electronics generally which is one of the major challenges in this field now and in future. The design of the android based voice controlled door involved research in both digital and microelectronics. The project was quite challenging and tedious but eventually was a success. However, like every aspect of engineering there is still a room for improvement and further research on the project as suggested in the recommendations written out in the section that follows in the paragraph below

Recommendations.

For the purpose of the future research, the project work can be improved upon. The following areas were highlighted for this purpose.

1. The whole circuitry can be reduced by making use of integrated circuit with higher scale of integration.
2. A higher scale integrated circuit can be used so that other means of authentication could be used to cut across to the less privileged in the society (e. g. visually impaired individual).
3. Moreover, it is recommended that students should be enlightened on new areas of technology that are yet to be addressed in order to bring solution to the various problems faced by man in his day to day activities.
4. A lot of work has been put into place to ensure that the voice recognition system is able to recognize the owner's voice and grant access to the owner. There is still a lot of work that needs to be done to ensure that the voice recognition system is able to perform well and optimally. Some key areas that should be considered when making improvements on the voice recognition system include:

i. Native language: Major Nigerian native languages such as Igbo, Yoruba, and Hausa currently do not have speech to text engine. The ability of the voice recognition to recognize the native language of the user can be worked and improved upon by training speech models of those language.

ii. Security: The exchange of data between the various components of the hardware and the software should be encrypted to ensure that the system cannot be hacked and data integrity is maintained. The recommendation is further that a custom encryption algorithm should be implemented.

7. REFERENCES

[1] Agbo David O, Madukwe Chinaza, Odinya Jotham O, (2017) designed and implemented a door locking system using android app

[2] Ahmed Moawad. (2012). Speech Recognition system.

[3] Ali Mansour Almadani. (2018). AI Speech Recognition System

[4] Arduino,(n.d.). (2011). Getting Started. https://www.arduino.cc/en/guide/introduction.

[5] Arnab, P., Sahidullah, M, & Goutam, S. (2018). Speaker verification with short utterances: a review of challenges, trends and opportunities. *IET Biometrics*, 7(2): 91-101.

[6] Avrmicrocontrollers. (2014). http://en.wikipedia.org/wiki/avr_microcontrollers.

[7] Bača, M., Petra, G. & Fotak, T. (2012). Basic principles and trends in hand geometry and hand shape biometrics. *New Trends and Developments in Biometrics*, pp 77-99

[8] Bengio, Y. (1991). Magnetic lock, artificial neural networks and their application to speech/sequence recognition. McGill University, Canada, Hu, Hongbing. https://www.indiamart.com

[9] Brunelli, R., & Poggio, T. (1993). Face Recognition: Features versus Templates. *IEEE Trans. on PAMI*, 10(15): 1042-1052.

[10] Cho, M., & Kang, J. (2017). Voice security on the rise: examining the path to secure voice automation. Alticast Inc. Colorado.

[11] De Vries, Cross, N. & Grant D. P. (1992). Design methodology and relationships with science: introduction. Eindhoven: Kluwer Academic Publishers. pp 32.

[12] Dharavath, K., Talukdar, F. A., & Laskar, R. H. (2013). Study on biometric authentication systems, challenges and future trends: A review. *IEEE International Conference on Computational Intelligence and Computing Research*, pp 1-7.

[13] Ehud Shapiro, Y. (1983). The fifth-generation project—a trip report. *Communications of the ACM*, 26(9): 637-641.

[14] George Loveday (1984). Essential electronics. Pitman, United Kingdom.

[15] Haşim Sak, Andrew Senior, Kanishka Rao, Françoise Beaufays, & Johan Schalkwyk. (2015). Google voice search: faster and more accurate. *Google Research Blog*, pp 35.

[16] Julio Sanchez (2007). Microcontroller programming. Second Edition, C.R.C Press, Taylor and Francis Group, Great Britian.

[17] Juang, B.H., & Rabiner, L.R. (2007). Automatic speech recognition – a brief history of the technology development. *Journal of Computer Science*, 3(5): 274-280.

[18] Kamelia, L., Noorhassan, A.S.R., Sanjaya, M., & Edi Mulyana, W.S. (2014). Door-automation system using bluetooth-based android for mobile phone. *Asian Research Publishing Network Journal of Engineering and Applied Sciences*, 9(10): 1759-1762.

[19] Kamoru Olarewaju Iyapo, Olukayode Michael Fasunla, Shadrack Alaba Egbuwalo,Akin James Akinbobola and Olatunji Temitope Oni, (2018) designed motion detector alarm and security system. Key Dimensions. (n.d.). Retrieved from http://www.dimensionsinfo.com/tabular-key-dimensions.

[20] Kim, Z. (2012). Reverse-engineered irises look so real, they fool eye-scanners. *ARPN Journal of Engineering and Applied Sciences*, 9(10): 17591762.

[21] Lin, & Kamis, Z. (2014). Biometric voice recognition in security system. *Indian Journal of Science and Technology*, **7(2)**: 104-112.

[22] Locks, C. (1958). History of locks. Encyclopaedia of Locks and Builders Hardware Macmillan, British English definition of voice recognition, Macmillan Publishers Limited, 2008.

[23] Maddock, R.J., & Calcutt, D.M. (1994). Electronics a course for engineers. Longman, United Kingdom.

[24] Maltoni, D., Maio, D., Jain, K., & Prabhakar, S. (2009). Handbook of fingerprint recognition. *Springer Science & Business Media*, pp 11.

[25] Massachusetts Institute of Technology. (n.d.). (2013). Inventor of the week archive. massachusetts institute of technology http://web.mit.edu/invent/iow/yale.html.

[26] Margolis Michael (2011), Arduino cookbook. First edition, O'Reilly Media, Inc., Usa. pp 301.

[27] Mehta, V.K. (2003). Principles of electronics. Second Edition, S.Chand and Company Ltd, India.

[28] Muthuselvi, G., & Saravanan, B. (2014). Real time speech recognition based building automation system. *ARPN Journal of Engineering and Applied Sciences*, 9: 2831-2839.

[29] Nitzan Lebovic. (2015). Biometrics or the power of the radical center. *Critical Inquiry*, 41(4): 841–868.

[30] Robert Boylestad, l., & Louis Nashelsky. (2002). Electronic devices and circuit theory. Eight Edition, Prince-Hall, United Kingdom. pp 875.

[31] Shah, H.N.M., Ab Rashid, M.Z., Abdollah MohdFairus, Kamarudin Muhammad, Nizam, C.K.

[32] Wildstrom, S. (2014). Nuance Exec on iPhone 4S, Siri, and the Future of Speech, http://tech.pinions.

[33] Schlage's History of Locks. (n.d.). Retrieved from http://www.locks.ru/germ/informat/schlagehistory.htm.

[34] Shirriff Ken (2016). The surprising story of the first microprocessors. *IEEE Spectrum*, 53(9): 48-54.

[35] Tokheim, l.R. (2005). Digital electronics: principles and application. Sixth Edition, tata McGraw-Hill, New York.

[36] Wahyudi Astuti W., & Mohamed, S. (2007). Intelligent voice-based door access control system using adaptive-network-based fuzzy inference systems (ANFIS) for building security. *Journal of Computer Science*, **3(5)**: 274-280.

[37] Wolverhampton City Council. (2005). Lock Making. Chubb & Son's Lock & Safe Co Ltd". http://www.wolverhamptonhistory.org.uk/work/industry/lock_making.

[38] Zaid A. Mundher , Khalida Basheer, Safaa Najeeb, Rami Zuhair, (2019) build a Real-Time Home Security Alarm System Using a Kinect Sensor.

International Journal of Science and Engineering Applications
Volume 10-Issue 09, 140 - 143, 2021, ISSN:- 2319 - 7560

Study the effects of Process Parameters on Overcut of Al6061 Alloy by (Fe_2O_3) Nano- Powder-Mixed Micro-EDM

Nagwa Mejid Elsiti
Benghazi University,
Industrial and Manufacturing
System Engineering
Benghazi, Libya

Abstract: The effect of process parameters on micro-EDM namely current (I), voltage (V) and pulse on time (TON) were studied based on Tool Overcut (OC). The effect of Fe_2O_3 nano-powder mixed dielectric on the overcut for AL6061 alloy was studied using a Die Sinking (EDM) Machine. Effect of the process parameters was examined by 2-Level Factorial Design using Design of Experiment (DOE) software, whereas the level of importance was statistically evaluated using ANOVA. The results indicate discharge current and pulse on time significantly influenced the EDM process compared to voltage. It was also concluded that the use of Fe_2O_3 nano-powder mixed micro-EDM decreased the overcut values.
Keywords: Fe_2O_3 Nano-powder; Micro-EDM; Overcut, Al6061 alloy

1. INTRODUCTION

Electrical discharge machining (EDM) has become the workhorse of the tool making industry due to the precise machining of the workpiece that conducts electricity. It plays a major role in machining of dies, tools and other products manufactured from difficult-to-machine materials [1]. In principle, the material erosion mechanism employed in EDM is largely based on the conversion of electrical energy to thermal energy. This is accomplished by the release of discrete electric charges between the electrode (tool) and the workpiece immersed in a dielectric fluid [2]. The repetitive spark discharge generates considerably high temperatures in the spark zone which melts and vaporizes the workpiece material during EDM [3]. However, in Micro-EDM (scaled-down version of EDM), the range of process parameters is lowered to ensure spark discharge in a micro-joule range. This reportedly improves the machining of hard-to-machine materials in micro domain (chip or debris size in microns)[4].The technique has a number of unique advantages such as exerting a small force between the work piece and tool electrode essential for fabricating hard-to-cut materials [5].

Despite its unique advantages, Micro-EDM has some drawbacks one of which is overcut. The overcut is the clearance per side between the electrode and workpiece. It is also defined as the difference between the magnitude of the electrode and the cavity created during machining. As a result, it is essential to minimize overcut to ensure the dimensional accuracy of finished products [6]. In order to improve the performance output of EDM, like the low material removal rate (MRR) and relatively poor surface finish, a modified process using suspended powder particles in dielectrics has been developed. The addition of the powder improves MRR and reduces the tool wear rate (TWR), however, it may result in increased overcut size. The powder addition process in conventional EDM is called powder mixed EDM (PMEDM) [7, 8]. Numerous studies has been carried out to improve the performance of EDM. The parameters typically examined include; MRR, TWR, EWR, EW, SR, SF, taper and very few for overcut, dimensional precision, surface quality, RCL, HAZ, depth of the micro cracks and HV [9]. Pradhan [10]; investigated the effect of material removal rate (MRR), tool wear rate (TWR) and overcut(OC) with AISI D2 tool steel on EDM performance. The weights of the responses were analyzed by GRA, PCA and response surface methodology (RSM). Similarly, Belgassim et al.,[11]; examined and optimized EDM parameters using L9 orthogonal array based on the Taguchi method and Analysis of Variance (ANOVA). The EDM parameters investigated included; pulse current (Ip), pulse on time (T_{on}), pulse off time (T_{off}), and gap voltage (V_g). In addition, machining responses based on surface roughness, machined surface and overcut were examined. Consequently, the optimum surface finish of the EDM surface was deduced from the input parameters investigated in the study. Kumar et al.,[12] investigated overcut using reverse polarity and powder metallurgy tool electrodes based on the Taguchi method. The results demonstrated that overcut and pulse-on time are maximized by applying minimal current. Likewise, the findings demonstrated that overcut can be improved at average values of duty cycle and gap voltage.

The review of literature demonstrates that there is limited research on the size of overcut during PMEDM. Therefore, the objective of this work is to investigate the effect of process parameters namely; current (C), Voltage (V),Pulse ON-time and nano-powder mixed dielectric on the size of overcut (OC) during the process of PMEDM.

2. EXPERINMETAL PLAN AND PROCEDURES

The experiments in the current research were performed by an AG40L Sinker EDM (Fig.2), and the material of workpiece samples was Al6061 alloy. The samples were cut into slides with dimensions of 50mm x 90mm x 1mm, and the electrode was copper with 6mm length and 500µm diameter. The dielectric fluid was mixed with 4g/l percentage of Fe_2O_3 nano-powder. The

PMEDM experiments were carried out in the working tank with dimensions; 46cm × 35cm × 24cm fabricated from 1.5mm thick stainless steel sheets as illustrated in figure 2. During each test, a stirrer was employed to maintain the homogeneity of the Fe_2O_3 nano-powder suspension in the tank. The selected input variables of the study were; current, voltage, and pulse-on time.

Figure 1. AG40L Sodick electrical discharge machine

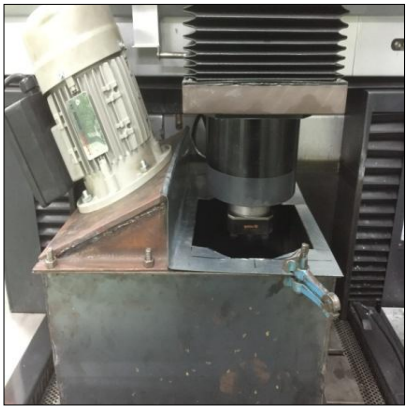


Figure 2. Working tank

The Design of Experiment (DoE) was carried out using 2 Level Factorial Design comprising 8 points (2^3), and 3 center points as presented in table 1.Consequently, a total of 11 experimental runs were conducted as outlined in table 2.

Table 1. Selected Machining Process Parameters

Symbol	Input Factors	Unit	level	
			Low	High
A	Current	A	4	6
B	Voltage	V	4	15
C	Pulse On	μs	10	120

The overcut (OC) was calculated for all experimental runs according to Eq. 1;

$$OC = DW/2 \quad (1)$$

Where the terms DW= D_1 – D; and D_1 represents the diameter of the entry hole and *D* – the diameter of the electrode (tool).

Table 2. Design of Experimental and Responses

no	Level parameters			Results	
	current	voltage	Pulse-ON time	(OC1) Free powder	(OC2) Fe_2O_3 4g/l
	Amp	v	s	mm	mm
1	4	4	10	17.78	1.04
2	6	4	10	119.61	9.32
3	4	15	10	23.74	2.98
4	6	15	10	157.87	15.26
5	4	4	120	40.29	4.56
6	6	4	120	119.45	7.24
7	4	15	120	46.63	9.75
8	6	15	120	125.61	14.27
9	5	4.5	65	38.39	6.71
10	5	4.5	65	94.18	7.24
11	5	4.5	65	99.22	9.97

3. RESULTS AND DISCUSSIONS

3.1 Effect of Process Parameters on OC

Figures 3- 4 presents the variation of overcut of micro-holes produced by various process parameters during machining. The main effect plots of OC1 and OC2 demonstrate the overcut of the micro-holes increase sharply with peak current. The minimum overcut was obtained at the corresponding discharge current of 4 A. In addition, it was observed that the peak current increased with increase in discharge energy density resulting in breakdown of debris into smaller particles. As a result, the smaller sized debris particles obtained easily seeped through the narrow passage between the micro-tool and wall of the micro-hole. This phenomena decreases the chances of a second spark and mechanical scratching of the internal walls of micro-hole resulting in lower overcut (OC) [13]. At higher peak current settings of 6A, the overcut was observably high. This is attributed to the ejection of large sized of debris from the micro-hole at high discharge energy. This may also be ascribed to the occurrence of the secondary sparking phenomenon and materials entering the side hole wall [13, 14]. The results also demonstrated that pulse-on time significantly influenced OC2. When the values of pulse on time increase the overcut values increase. However, the effect of voltage on OC1, was not significant, while it has a significant effect on OC2. The OC2 had a small value when voltage was 4v.

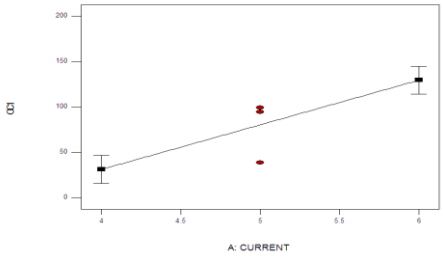
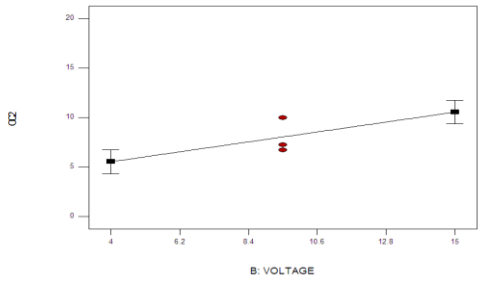
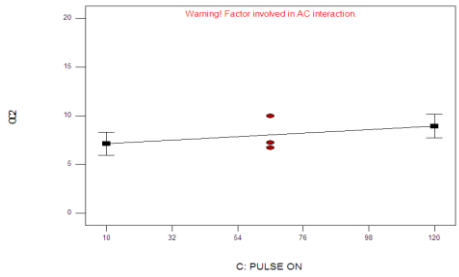


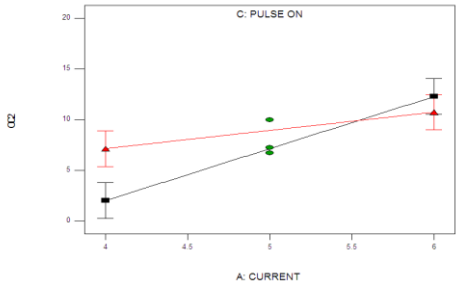
Figure. 3 Variation of OC1 with process parameters



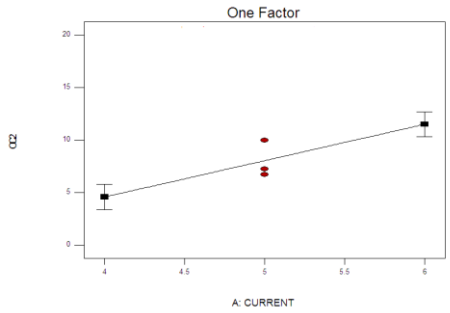
(a)



(b)



(c)



(d)

Figure. 4 From (a-d) variation of OC2 with process parameters

3.2 ANOVA Analysis and Optimal Conditions

The accuracy and relative significance of the process responses OC1 and OC2 were examined by Analysis of Variance (ANOVA). Tables 3 – 4 present ANOVA test results for the defined responses. From the ANOVA tables, it can be concluded that peak current had a significant contribution to OC1 and OC2. However, the pulse on time and voltage had a significant effect on only OC2.

Table 3. Analysis of Variance (ANOVA) test for (OC1)

Source	SS	DF	Mean Sq	F-value	P-value
A(current)	19414.35	1	19414.35	45	0.0002
Curvature	36.84	1	36.84	0.077	0.7889
Residual	3846.2	8	480.78	--	--
Lack of fit	1566.8	6	261.13	0.23	0.9324
Total	23297.4	10			

R^2 = 83.33% R^2 adjusted = 81.48% R^2-predicted = 77.01%

Table 4. Analysis of Variance (ANOVA) test for (OC2)

Source	SS	DF	Mean Sq	F-value	P-value
A(current)	96.33	1	96.33	36.16	0.0018
B(voltage)	50.50	1	50.50	18.96	0.0073
C(pulse	6.52	1	6.52	2.45	0.1786
AC	22.31	1	22.31	8.37	0.034
Curvature	0.014	1	0.014	0.000513	0.9457
Residual	13.32	5	2.66		
Lack of fit	7.2	3	2.4	0.78	0.6026
Total	188.99	10			

R^2 = 92.94% R^2adjusted = 88.24% R^2-predicted = 73.30%

3.3 Effect of Fe_2O_3 nano-powder on Overcut

Figure 5 shows the effect of Fe_2O_3 powder concentration on the magnitude of overcut. The results clearly indicate the size of the overcut decrease with increasing powder concentration.

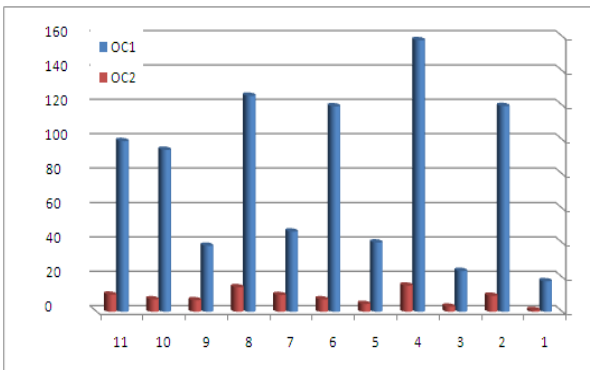


Figure. 5 Comparison between Overcut values

4. CONCLUSION

In the present study, attempts were made to machine micro-holes on Al6061alloy with lower overcut (OC) using EDM and Fe_2O_3 nano-powder mixed dielectric at different concentrations. The following conclusions were made;

- The ANOVA analysis for OC1 and OC2 responses indicated current, voltage and pulse on time were

significant factors. The lowest OC was obtained at 4A (current), 4V (voltage) and 10μs (pulse-on time).

- The overcut (OC) decreased when Fe_2O_3 nano-powder was added to dielectric liquid.
- Lower overcut (OC) was observed at Fe_2O_3 nano-powder concentration of 4g/l.

5. RFERENCES

1. Sharma, R. and J. Singh, Effect of Powder Mixed Electrical Discharge Machining (PMEDM) on Difficult-to-machine Materials – a Systematic Literature Review, in Journal for Manufacturing Science and Production 2014. p. 233.
2. Kolli, M. and A. Kumar, Effect of Boron Carbide Powder Mixed into Dielectric Fluid on Electrical Discharge Machining of Titanium Alloy. Procedia Materials Science, 2014. **5**: p. 1957-1965.
3. Dwivedi, A.P. Comparative Assessment of MRR, TWR and Surface Integrity in Rotary and Stationary Tool EDM for Machining AISI D3 Tool Steel. 2015. (01.07.2015).
4. Jahan, M.P., M. Rahman, and Y.S. Wong, A review on the conventional and micro-electrodischarge machining of tungsten carbide. International Journal of Machine Tools and Manufacture, 2011. **51**(12): p. 837-858.
5. Moghanizadeh, A., Reducing side overcut in EDM process by changing electrical field between tool and work piece. The International Journal of Advanced Manufacturing Technology, 2016: p. 1-8.
6. Vikas, et al., Effect and Optimization of Machine Process Parameters on MRR for EN19 & EN41 Materials Using Taguchi. Procedia Technology, 2014. **14**: p. 204-210.
7. Talla, G., S. Gangopadhyay, and C.K. Biswas, Effect of Powder-Suspended Dielectric on the EDM Characteristics of Inconel 625. Journal of Materials Engineering and Performance, 2016. **25**(2): p. 704-717.
8. Batish, A., A. Bhattacharya, and N. Kumar, Powder Mixed Dielectric: An Approach for Improved Process Performance in EDM. Particulate Science and Technology, 2014. **33**(2): p. 150-158.
9. Choudri, P.B., Experimental Investigations into the Effect of Process Parameters on Performance Measures of Sink EDM Process- A Review till the year 2010 and Future Work. International Journal of Latest Trends in Engineering and Technology (IJLTET), 2016. **6**: p. 504-511.
10. Pradhan, M.K., Estimating the effect of process parameters on MRR, TWR and radial overcut of EDMed AISI D2 tool steel by RSM and GRA coupled with PCA. The International Journal of Advanced Manufacturing Technology, 2013. **68**(1-4): p. 591-605.
11. Belgassim, O. and A. Abu-Saada, Optimization of EDM Parameters in Machining AISI D3 Tool Steel by Grey Relational Analysis. Applied Mechanics and Materials, 2013. **330**: p. 747-753.
12. Dinesh Kumar, N.B., Anil kumar, Study Of overcut during electric discharge machining of hastelloy steel with different electrodes using the taguchi method. International Journal of advanced engineering technology, 2011(E-ISSN 0976-3945).
13. Shivakotia, I., et al., Multi-objective optimization and analysis of electrical discharge machining process during micro-hole machining of D3 die steel employing salt mixed de-ionized water dielectric. journal of computational and applied research, 2013. **3**: p. 27-39.
14. Zhao, F.L., H. Wang, and Z.Z. Lu, Calculating the Overcut in Electro-Discharge Machining. Key Engineering Materials, 2005. **291-292**: p. 561-566.

International Journal of Science and Engineering Applications
Volume 10-Issue 09, 144 - 147, 2021, ISSN:- 2319 - 7560

A Sampling Approach based on Set Coverage Algorithm

Huiling LI
School of Computer Science and Technology, Shandong University of Technology, Zibo, 255000, China

Xuan SU
School of Computer Science and Technology, Shandong University of Technology, Zibo, 255000, China

Shuaipeng ZHANG
School of Computer Science and Technology, Shandong University of Technology, Zibo, 255000, China

Abstract: Massive amounts of business process event logs are collected and stored by modern information systems. Model discovery aims to discover a process model from such event logs, however, most of the existing approaches still suffer from low efficiency when facing large-scale event logs. Event log sampling techniques provide an effective scheme to improve the efficiency of process discovery, but the existing techniques still cannot guarantee the quality of model mining. Therefore, a sampling approach based on set coverage algorithm named set coverage sampling approach is proposed. The proposed sampling approach has been implemented in the open-source process mining toolkit *ProM*. Furthermore, experiments using a real event log data set from conformance checking and time performance analysis show that the proposed event log sampling approach can greatly improve the efficiency of log sampling on the premise of ensuring the quality of model mining.

Keywords: event logs; log sampling; quality measure; set coverage; conformance checking

1. INTRODUCTION

Process mining[1-3] is a novel discipline that connects data science and business process management. It aims to extract effective information about business processes from event logs and discover, monitor and improve real business processes[4]. Process mining also includes sub-areas such as process prediction [5]-[6] and business process automation [7]. Process discovery is one of the most challenging process mining tasks, which allows the discovery of process models from event logs without any prior information. In recent years, it has received extensive attention. Over the past two decades, domestic and foreign researchers have proposed various process discovery methods, for example, *Alpha Miner*[8], *Heuristics Miner*[9], *Heuristics Miner*[10], *Tsinghua Alpha* [11], *Split Miner*[12], etc. But most discovery methods are no longer suitable for using a single machine to process an entire large data set. With distributed platforms such as the well-known *MapReduce framework* [13]-[14], the process can be very time consuming, so a new approach is urgently needed to address these issues.

The event log sampling approach provides a feasible solution to the above problem. It takes the original event log as input and returns a sample log. At present, many event log sampling approaches have been proposed, such as an event log sampling approach based on graph sorting algorithm named *LogRank*[15]-[16] and an event log sampling approach based on trajectory similarity named *LogRank+*[17]. However, their performance still cannot meet the needs of practical application, for example, the quality of the model is still not ideal, meanwhile, with the increase of the original log size, the difference between the sum of the original log sampling time and the sample log mining time and the original log mining time becomes more and more obvious.

Inspired by the traditional set coverage and other related ideas, we propose set coverage sampling approach. Compared with the existing sampling methods, the set coverage sampling approach proposed in this paper can obtain simpler and higher quality process models. In addition, in order to verify the feasibility and efficiency of the four sampling approaches of event logs, related experiments are done from the aspect of conformance checking and time performance analysis. The quality of sample logs compared with the original logs can be obtained by the experimental results.

The remainder of this paper is organized as follows. Section 2 discusses the related work. Section 3 introduces set coverage sampling approach. Section 4 describes the tool implementation. Section 5 describes the data set used in the experiments, introduces the experiments and shows the results of the evaluation. Finally, Section 6 draws conclusions and points our future research scope.

2. PRELIMINARIES

Let S be a set. We use $|S|$ to denote the number of elements in set S. $B(S)$ is the set of all multisets over set S. $f \in X \rightarrow Y$ is a function, i.e., $dom(f)$ is the domain and $rng(f) = \{f(x) \mid x \in dom(f)\}$ is the range.

Definition 1 (Event, Trace, Event Log). Let A be a set of activities. A trace $\sigma \in A^*$ is a sequence of activities (also referred to as events). For $1 \leq i \leq |\sigma|$, $\sigma(i)$ represents the ith event of σ. $L \in B(A^*)$ is an event log.

An event log records the execution of a potential business process whose business process model is the task target of process mining, so it does not appear explicitly in the definition of the event log. The execution of a business process instance is represented by the corresponding traces. The events in the trace are recorded in the event log.

Definition 2 (Process Discovery). Let UM be the set of all process models, a process discovery method is a function γ mapped from an event log $L \in B(A^*)$ to a process model $pm \in UM$, i.e., $\gamma(L) = PM$. In general, the process discovery method can transform the event log into a process model represented by marked *Petri nets*, *BPMN*, *EPC*, etc. Regardless of the representation used by the process model, each trace in the input event log corresponds to a possible execution sequence in the discovered process model.

Definition 3 (Directly Follows Relation). Let a and $b \in A$ be two activities and $\sigma = \langle \sigma_1, .., \sigma_n \rangle$ is a trace in the event log. A directly follows relation from a to b exists in trace σ, if there is $i \in \{1, ..,$

n−1} such that $\sigma_i = a$ and $\sigma_{i+1} = b$ and we denote it by $a >_\sigma b$. For example, in $\sigma = \langle a, b, c, e, g \rangle$, we have $c >_\sigma e$, but $d \not>_\sigma a$.

Definition 4 (Start point set). The start event of each trace in the event log constitutes the start point set.

Definition 5 (End point set). The end event of each trace in the event log constitutes the end point set.

3. SAMPLING APPROACH

In theory, we can choose an arbitrary subset of the trace from the event log as its sample log, while the real challenge is to find sample logs that are representative enough so that a reliable process model can be found compared to the original event log. In response to this challenge, this paper propose a sampling approach based on set coverage algorithm named set coverage sampling approach. This sampling approach can get the sample log that is a representative subset of the original log. It can reduce the computational cost. At the same time, compared with the existing event log sampling approaches, set coverage sampling approach can not only ensure the quality of the process model mined from the sample logs, but also greatly shorten the sampling time and mining time and improve the efficiency of process discovery.

The set coverage sampling approach is mainly based on the greedy algorithm to solve the set coverage problem, so the idea of the Set coverage sampling approach is as follows: Input the original event log in the platform, firstly, the directly follows relation of all traces in event log are traversed. If the trace's directly follows relation has the biggest intersection with the log's directly follows relation, meanwhile this intersection is not empty, or the trace's start point has an intersection with the log's start point set, or the trace's end point has an intersection with the log's end point set, then put this trace into the sample log. Finally, delete the following three parts: (1) The intersection of the log's directly follows relation set and the trace's directly follows relation set in the trace's directly follows relation set; (2)The intersection of the start point set and the trace's start point; (3) The intersection of the end point set and the trace's end point. Trace traversal is stopped until the log's directly follows relation set, start point set and end point set are all empty. In the end, the platform outputs a sample log.

4. TOOL IMPLEMENTATION

In this experiment, we use a laptop with a 2.70 *GHz CPU*, *Windows* 10 Professional, *Java SE* 1.8.0_281 (64-bit) , Python 3.7.6 (64-bit) and allocate 12 *GB* of *RAM*. In addition, the drawing software *Origin* 2021 *Pro* version is used to show the experimental results.

The open source process mining tool platform *ProM* provides a fully pluggable experimental environment for process mining. It can be extended by adding plugins and currently contains more than 1600 plugins. The tool and all plugins are open source. Set coverage sampling approach proposed in this paper has been implemented in *ProM* platform as plugin, which called *Business Process Event Log Sampling Plugin*. The snapshot of this tool is shown in Figures 1. It takes an original event log as input and outputs a sample log when the sampling approach is selected.

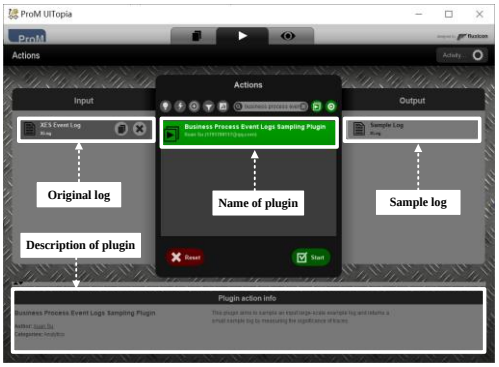


Figure. 1 The instance of *ProM* plugin

In the conformance checking experiment to verify the effectiveness of the set coverage sampling approach, the plugin called *Replay a Log on Petri Net for Conformance Analysis* implemented in *ProM* as shown in Figure 2 is used. It takes original event log and the process model mining from the sample log as input.

Figure. 2 Plugin for fitness index

5. EXPERIMENTAL EVALUATION

5.1 Experimental data sets

In this experiment, a real event log data set is used to evaluate the proposed set coverage sampling approach. Table 1 details some major statistics of this event log, including the trace number, event number and activity number and so on.

Table 1. Major statistics of event logs

Event log	Trace number	Variant number	Event number	Activity number	Trace length		
					Minimum value	Average value	Maximum value
BPIC_2012_A	13087	32	146044	20	6	11	20

·**BPIC_2012_A data set:** This data set is a real-life log, taken from a Dutch Financial Institute. Apart from some anonymization, the log contains all data as it came from the financial institute. The process represented in the event log is an application process for a personal loan or overdraft within a global financing organization. The amount requested by the customer is indicated in the case attribute AMOUNT_REQ, which is global, i.e. every case contains this attribute. The event log is a merger of three intertwined sub processes. The first letter of each task name identifies from which sub process (source) it originated from. Feel free to run analyses on the process as a whole, on selections of the whole process and/or the individual sub processes.

5.2 Conformance checking

To verify the availability of the set coverage sampling approach, we measure it in terms of conformance checking. The conformance checking experiment associates events in the event log with activities in the process model and compares them. The goal of it is to find commonalities and differences between the modeled behavior and the inspected behavior. In this experiment, we use fitness degree as quality standard, which is the measure related to conformance. Firstly, the process model of sample log is mined by using the version of *IM* algorithm with noise threshold of 0.9, then the process model and the original event log are measured for fitness degree. The experimental results are shown in Figure 3.

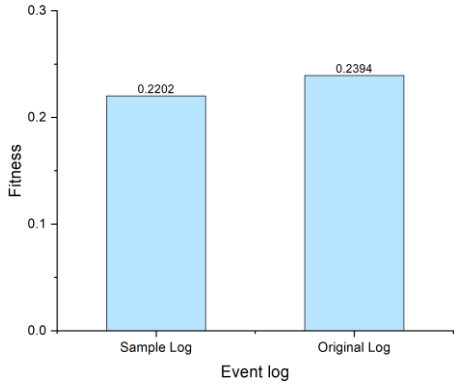


Figure. 3 The result of conformance checking

The experimental results show that the fitness of sample log obtained by set coverage sampling approaches is very closely to the fitness of original event log. It is proved that this set coverage sampling approach can extract sufficiently representative sample logs to a large extent and further prove its availability.

5.3 Time performance analysis

Time performance analysis experiment measures and records three types of time: (1) the original event logs' mining time; (2) the sampling time by using four sampling methods; (3) sample logs' mining time. Due to the computer internal environment every time may be different, so in order to guarantee the accuracy of experimental results, we measure each data for 5 times to get the average. Finally, the sampling time of each of the set coverage sampling approach is summed up with the sample log mining time, then compare with the original logs' mining time. The experimental result is shown in Figure 4.

The experimental results show that compared with the original event log, the sample log obtained by set coverage sampling approach can use less time to mining process models when the log's scale is large. Meanwhile, with the increase of the event log's scale, the difference between them becomes more and more obvious. It can be proved that the operation efficiency can be greatly improved by using sample log instead of original log, meanwhile the set coverage sampling approach of event log have high efficiency.

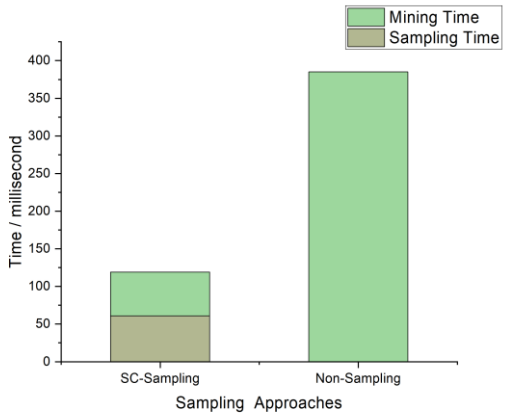


Figure. 4 The result of time performance analysis

6. CONCLUSIONS

In this paper, we propose set coverage sampling approach to effectively obtain sample logs with sufficient representability in large-scale event logs. Meanwhile we implemented set coverage sampling approach in the *ProM* platform. In addition, we assess the quality of the sample logs relative to the original logs in terms of conformance checking and time performance analysis. At the end, the experimental results on a real event log data set shows that compared with the existing sampling approaches, the proposed set coverage approach can not only greatly improve the efficiency of log sampling, but also ensure the integrity of the model.

As future work, we aim to apply the set coverage sampling approaches proposed in this paper to the event log for specific fields, such as education, medical care, finance, manufacturing, etc. It is also valuable to study the deployment of set coverage sampling approach in the distributed system because it is convenient to process the super-large event logs collected by other information systems in real life.

7. ACKNOWLEDGMENTS

This work was supported in part by National Natural Science Foundation of China under Grant 61902222, the Taishan Scholars Program of Shandong Province under Grant ts20190936 and Grant tsqn201909109, and Engineering and Technology R&D Center of IIOT in Colleges of Shandong Province (QingDao Technical College, Grant KF2019002).

REFERENCES

[1] VAN DER AALST W. Data science in action[M]//Process mining. Berlin, Germany: Springer-Verlag, 2016: 3-23.

[2] ZENG Q, SUN S X, DUAN H, et al. Cross-organizational collaborative workflow mining from a multi-source log[J]. Decision support systems, 2013, 54(3): 1280-1301.

[3] LIU C, DUAN H, ZENG Q, et al. Towards comprehensive support for privacy preservation cross-organization business process mining[J]. IEEE Transactions on Services Computing, 2019,12(4):639-653.

[4] VAN DER AALST W. Process Mining: Discovery, Conformance and Enhancement of Business Processes. Springer-Verlag, Berlin, 2011.

[5] POURBAFRANI M, VAN ZELST S J, VAN DER AALST W M P. Scenario-based prediction of business

processes using system dynamics[C]//OTM Confederated International Conferences" On the Move to Meaningful Internet Systems". Berlin, Germany: Springer-Verlag, 2019: 422-439.

[6] QAFARI M, VAN DER AALST W. Fairness-aware process mining[C]//OTM Confederated International Conferences" On the Move to Meaningful Internet Systems". Berlin: Springer, 2019: 182-192

[7] GAO J, VAN ZELST S J, LU X, et al. Automated robotic process automation: A self-learning approach[C]//OTM Confederated International Conferences" On the Move to Meaningful Internet Systems". Berlin, Germany: Springer-Verlag, 2019: 95-112.

[8] VAN DER AALST W, WEIJTERS T, MARUSTER L. Workflow mining: Discovering process models from event logs[J]. IEEE transactions on knowledge and data engineering, 2004, 16(9): 1128-1142.

[9] WEIJTERS A, RIBEIRO J T S. Flexible heuristics miner (FHM)[C]//2011 IEEE symposium on computational intelligence and data mining (CIDM). Washington, D. C., USA: IEEE, 2011: 310-317.

[10] LEEMANS S J J, FAHLAND D, VAN DER AALST W M P. Discovering block-structured process models from event logs-a constructive approach[C]//International conference on applications and theory of Petri nets and concurrency. Berlin, Germany: Springer-Verlag, 2013: 311-329.

[11] WEN L, WANG J, W.M.P. VAN DER AALST W, et al. A novel approach for process mining based on event types. Journal of Intelligent Information Systems, 32(2): 163-190, 2009.

[12] AUGUSTO A, CONFORTI R, DUMAS M, and ROSA M. Split miner: automated discovery of accurate and simple business process models from event logs. Knowledge and Information Systems, 1-34, 2018.

[13] CHENG L, LI T. Efficient data redistribution to speed up big data analytics in large systems[C]//2016 IEEE 23rd International Conference on High Performance Computing (HiPC). Washington, D. C., USA: IEEE, 2016: 91-100.

[14] Evermann J. Scalable process discovery using map-reduce[J]. IEEE Transactions on Services Computing, 2014, 9(3): 469-481.

[15] LIU C, PEI Y, ZENG Q, et al. LogRank: An approach to sample business process event log for efficient discovery[C]//International Conference on Knowledge Science, Engineering and Management. Berlin, Germany: Springer-Verlag, 2018: 415-425.

[16] LIU C, PEI Y, CHENG L , et al. Sampling business process event logs using graph-based ranking model[J]. Concurrency and Computation: Practice and Experience, 33(5):1-14, 2021.

[17] LIU C, PEI Y, ZENG Q, et al. LogRank+: A Novel Approach to Support Business Process Event Log Sampling[C]//International Conference on Web Information Systems Engineering. Berlin: Springer, 2020: 417-430.