Review of Materialized Views Selection Algorithm for Cyber Manufacturing

Nor Amalina Mohd Sabri[1], Nurul A. Emran[1], Nor'azah Md Khushairi[1,2]
[1]*Computational Intelligence Technologies (CIT), Center of Advanced Computing Technologies (CACT), Universiti Teknikal Malaysia Melaka (UTeM), 76100 Durian Tunggal, Melaka, Malaysia.*
[2]*SilTerra Malaysia Sdn Bhd, 09000 Kulim, Malaysia*
nurulakmar@utem.edu.my

***Abstract*—Technological advancement in data transfer and connection has driven massive data growth. Within the semiconductor cyber manufacturing environment, in order to cope with rapid data transfer enabled by the Internet of Things (IoT) technology, rapid query processing becomes a priority. Especially, in the era of Industry 4.0, semiconductor manufacturing that operates within cyber-physical systems (CPS) relies heavily on the reporting function to monitor delicate wafer processing. Thus, delay in reporting which is usually caused by slow query processing is intolerable. Materialized views (MVs) are usually used in order to improve query processing speed. Nevertheless, as MVs requires database space and maintenance, the decision to use MVs is not determined by time factor only. Thus, MVs selection is a problem that calls for an efficient selection algorithm that can deal with several constraints at a time. In this paper, we reveal the criteria of optimisation algorithms that were proposed to deal with MVs selection problem. In particular, this paper attempts to evaluate the coverage and limitations of the algorithm under study.**

***Index Terms*—Materialised View Selection; Bio-Inspired Algorithm; Optimisation Algorithm; Cyber Manufacturing; Industry 4.0.**

I. INTRODUCTION

The prediction of rapid growth in manufacturing has come to reality due to technological advancement in "Internet of things" (IoT) that utilises high volumes of interconnected sensors and automated hardware instruments [1]. In the era of Industry 4.0, semiconductor manufacturing especially has been strategically located within cyber-physical systems (CPS). Semiconductor cyber manufacturing is employed within CPS to utilise transformative technologies to enable data translation from multiple interconnected systems into predictable operations for competitive performance [2]. Because semiconductor processes are delicate and require close monitoring, sensors are utilised to overcome human operators' weaknesses [2]. These sensors are installed within interconnected machines and hardware to record real-time data from a great number of complex fabrication processes [3].

Within CPS, semiconductor manufacturers are capable of achieving rapid data transfer and storage with the Internet of Data (IoD) [4]. However, rapid data transfer makes query processing a challenge, as it now becomes a priority. For instance, the requirement of rapid wafer fabrication and production for wafer semiconductor manufacturing industry like SilTerra Malaysia Sdn Bhd has made a delay in query processing intolerable [5]. This is because production monitoring relies heavily on reporting function whose performance is determined by query processing speed. The architecture of SilTerra's Reporting System is as shown in Figure 1.

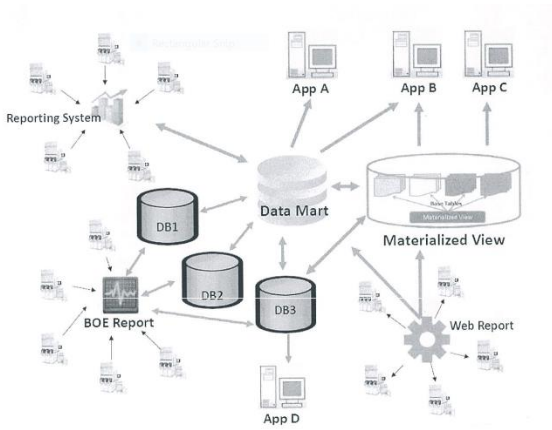

Figure 1: Architecture of Silterra's Interconnected Reporting System

The complexity of fabrication processes is caused by the diversity of product mix where any occurrence of disruptions requires fast handling [6]. Sensors record raw data produced by these processes and stored into a massive transaction history. These transaction history records are persistently stored in databases that grow over time. These records are queried to generate reports for monitoring. As every transaction that occurs in fabrication processes is recorded, the manufacturing industry has been reported as having the largest amount of data [7], [8], [9], [10]. With massive volumes of data, manual reporting is no longer feasible in cyber manufacturing. The manual reporting is not only labour intensive but also questionable in achieving time-sensitive production goal [11],[12]. Data that are extracted from databases through queries are used to generate reports. Using large databases to produce reports, the problem that hinders rapid data extraction (and thus rapid reporting) in this industry is slow query processing.

In order to handle reporting delay, materialised views (MVs) are usually used to speed up query processing. Using MVs, complex queries that require a longer time to complete can be pre-computed in advanced, and thus less time is taken for query processing.

Nevertheless, MVs requires database space because the results of the pre-computed queries are stored in the database. Concern regarding storage space is crucial especially for organisations that are moving towards green data management [13]. In addition, as MVs are created based on existing database tables, MVs must be refreshed for any updates made against their base tables. Without the refresh process, data extracted from MVs are inconsistent with the data in the base tables. Thus, the decision to use MVs is not determined by the time factor only. Other constraints are usually taken into consideration in selecting the optimal MVs.

MVs selection is a problem that calls for an efficient selection algorithm that can deal with several constraints at a time [14]. MV selection problem refers to the selection of MV which is suitable to materialise to reach stability between the factors of increased query performance and low computational cost [15]. Also, the main reason for view selection problem is to decrease cost function or either one of the constraints. In another study by Karde and Thakare (2010) the view selection problem is defined as a problem to select a set of views to materialise which can minimise the sum of the total response time of query and also the maintenance of selected views [16]. Hence, the optimal query performance for a given query workload is achieved.

Selecting MV problem becomes a complex selection problem due to the enormous number of MVs [17]. Therefore, artificial intelligence optimisation algorithms are increasingly used to solve the problem.

In this paper, we aim to reveal the criteria of optimisation algorithms (bio-inspired and non-bio inspired) that were proposed to deal with MVs selection problem. In particular, this paper attempts to study the coverage and limitations of the algorithms under study.

In the next section, several optimisation algorithms for MV selection will be presented. In Section III, the coverage of criteria (constraints) of the selection algorithms will be reported. Finally, Section IV, concludes this paper.

II. Optimization Algorithms

Several algorithms are used to optimise MVs in terms of speed or response time. Metaheuristic algorithms are widely used in the optimisation field for its better performance as compared to heuristic algorithms (see for example in software defect prediction [18]). Metaheuristics algorithms are often effective in solving difficult optimisation problems and often mimicking some successful characteristics in nature (nature-inspired) [19]. Metaheuristics efficiently explore search space to find near best or optimal solution [20]. Examples of metaheuristic optimisation algorithms are Ant Colony Optimization (ACO), Genetic Algorithm (GA) and Particle Swarm Optimization (PSO). These algorithms are also known as bio-inspired (or nature-inspired) algorithms. An example of heuristic algorithms is the greedy algorithm [21].

ACO and GA have been reported as among the powerful bio-inspired algorithms [22]. GA is said as suitable to solve MV selection problem as it works to find an optimal solution [23]. In another study, (PSO) have been reported for their potential to solve similar selection problem [24]. Even though (PSO) is least explored in MV selection, this algorithm has shown better performance as compared to the heuristic algorithm and non bio-inspired algorithm.

In the study by Sun and Wang (2009), the performance of PSO has been evaluated against a greedy-based, Heuristic algorithm (HA) and GA in MV selection. The results showed that PSO achieves better performance than the others [24]. Meanwhile, Zhang, Sun and Wang (2009), compared the Memetic algorithm (MA) with GA to deal with MV selection problem [25]. The researchers claimed that MA could find better optimal MV regarding storage space as compared to GA and heuristic algorithm.

Karde and Thakare (2010) tested the Tree-based MV selection algorithm and node selection algorithm to speed up MV selection in a distributed environment [16]. The results obtained suggest that the shortest processing time can be revealed by this algorithm, where the total cost of query processing has been considered as the selection constraint.

A hybrid algorithm that integrates GA and ACO has been proposed by Zhou, Geng and Xu (2011) aims to obtain the least maintenance cost and rapid response of user queries [26]. The hybrid algorithm has been proved to become practical tools for data warehouse evolution, by gaining near-optimal solutions in limited time. Drias (2011) also utilised a hybrid method by combining ACO and Tabu search to improve queries performance [17]. The algorithms were implemented to take up scalability challenge for searching process. GA has also been used in MV selection within data warehouse context where the concept of the vector has been embedded in the algorithm [27].

In 2013, ACO’s contribution to solving MV problems can be seen for example in the study by Tiwari (2013) where hybrids of ACO is applied in a distributed database. In this study, the algorithm seeks an optimal solution in solving database volume issue in relation to MVs [28]. The study also supports that the hybrid ACO improves the performance of distributed query optimisation.

Datta and Dey (2015) used the Apriori algorithm to generate optimal MV candidates by considering the frequencies of the attributes queried [29]. The algorithm is used to design a method that identifies data sets that will be materialised based on their frequencies and dependencies on other data. The algorithm has been regarded as scalable and dynamic by fixing the frequencies of attributes occurrences. In another work by Arun and Kumar (2015), an improved algorithm to select near-optimal sets of views for materialisation has been proposed [30]. The improved algorithm is Bee Colony Optimization (BCO) algorithm, that is able to minimise the response time of analytical queries for efficient strategic decision-making. As shown in Table 1, the optimisation algorithms proposed between the year 2009-2016 fall under bio-inspired and non-bio-inspired algorithm. While some of these algorithms have been proposed individually to solve MV selection problem, we can also see the proposals of hybrid algorithms (where the functionality of different algorithms are integrated for better performance).

As shown in Table 1, it can be observed from the studies covered in this paper that, MVs selection problems have been addressed by algorithms that are mostly under the bio-inspired category. These algorithms are GA, PSO, MA, ACO, Novel Shuffled Frog Leaping (SFL). Another bio-inspired algorithm like Bat algorithm has been used in solving query processing problem [20] but has not been tested in MV selection. In fact, the algorithms in this category have been chosen to address MV selection problems due to several reasons.

Table 1
Previous Works on Materialized Views Selection

Year	Algorithm	Type of Algorithm
2009	GA [28]	Bio inspired
	GA and PSO [19]	
	MA [22]	
2010	Tree based mv selection algorithm [21]	Non bio inspired
	Greedy Algorithm [18]	
2010	ACO compared to GA [29]	Bio inspired
	(SFL) [30]	
2011	I-mine techniques [31]	Non bio inspired
	GA and ACO [23]	Bio inspired
	Vector Evaluated GA (VEGA) [24]	
	ACO and Tabu Search [15]	Bio inspired and Non-bio inspired
2012	Improved GA [32]	Bio inspired
2013	Hybrid ACO [25]	
2015	Improved BCO [27],[33]	
	Apriori [26]	Non bio inspired
	Polynomial Greedy Algorithm (PGA) [34]	
2016	GA [20]	Bio inspired

For example, GA's is used as it is good in solving NP-hard problems. However, although the algorithm is popular in solving NP-hard problems, the algorithm is reportedly demonstrated lower performance than ACO in selecting MVs under varying space limitation constraint [31].

MA has a simple concept and easy to implement, and it offers computational efficiency and better intensification power of local search as compared to evolutionary algorithms. This algorithm has been successfully implemented for several NP-hard combinatorial optimisation problems with confirmation of efficient results. Moreover, this algorithm is similar to GA with different names of chromosome elements, which are called memes, not genes. However, unlike GA, all offspring and chromosomes in MA are allowed to gain experience.

PSO can control robustness with easy parameters even though it has a simple concept. It also provides computational efficiency.

ACO is useful for query optimisation in a distributed database, and its performance can be improved if combined with heuristic [17]. Furthermore, the algorithm has characteristics such as intelligent search techniques, robustness, distributed computing, global optimisation and the ability to integrate with other heuristics algorithm.

BCO uses the concept of artificial bees that collaborate to solve difficult combinatorial optimisation problem and require very low computation time. BCO potential can be expanded if combined with other algorithms [32].

While each proposal for MV selection algorithm has demonstrated the algorithm efficiency regarding computational performance, the coverage of constraints under consideration might be of interest for the query optimisation practitioners, especially in a cyber manufacturing context. This is because the decision regarding the adoption of MVs in speeding up reporting should consider the costs associated with it. In the next section, we examine the coverage of the algorithms that deal with MV selection problems. The focus is given to bio-inspired algorithms.

III. MATERIALIZED VIEWS SELECTION CRITERIA

There are three constraints commonly used in solving MV selection problem. As shown in Table 2, these three constraints are storage space, the time taken for query processing, and maintenance cost. ACO, as proposed by Gao and Song (2010) [31], has considered all three criteria in MV selection. Hybrid ACO (2011) by Drias (2011) and Tiwari (2013), however, did not cover maintenance cost in their algorithm [17], [28].

GA proposed by Chaves in 2009 and by Talebian and Kareem in 2011 covers space and time [33], [27]; by Zhou, Geng, and Xu in 2011 and by Zhou, He, Li in 2012, both covers time and cost [26], [34].

Furthermore, PSO by Sun and Wang (2009), MA by Zhang (2009) and BCO by Arun and Kumar (2015) only focused on time constraint [24], [25], [30]. BCO proposed in 2015 that covers only time constraint [30] has been improved by the same authors by adding space constraint [35]. Finally, SFL proposed in 2010 covered time and cost constraints.

Table 2
Previous Work with Fulfilled Criteria

Year	Space	Time	Maintenance Cost	Algorithm
2009 [33]	✓	✓		GA
2009 [24]		✓		PSO
2009 [25]		✓		MA
2010 [31]	✓	✓	✓	ACO
2010 [36]		✓	✓	SFL
2011 [17]	✓	✓		ACO + Tabu Search
2011 [27]	✓	✓		Vector Evaluated GA (VEGA)
2011 [26]		✓	✓	Improved GA
2012 [34]		✓	✓	GA
2013 [28]	✓	✓		Hybrid ACO
2015 [30]		✓		BCO
2015 [35]	✓	✓		Improved BCO

In comparison, ACO has shown the most coverage in terms of the types of constraints considered in MV selection. In fact, this algorithm is used to solve several hard problems such as in improving response query time, maintaining the least cost, determining best views, managing storage and dealing with increasing database size. The algorithm has its potential in solving MV selection problems either individually, or by integrating it with other algorithms.

Nevertheless, Tiwari (2013) highlighted the limitation of ACO where, given unsystematic information in distributed database queries, the algorithm has shown slow performance in convergence speed [28]. This limitation, however, can be offset by other ACO's strong characteristics such as intelligent search algorithms, distributed computing, global optimisation, robustness and ability to combine with other heuristics. Furthermore, this algorithm uses a quick genetic operator and selects the next state to accelerate actions [22].

In ACO, the ants find better solutions by updating pheromones. The pheromone is additional information in the algorithm that is used to decrease exploration ability in the algorithm. Moreover, an ant colony is regarded as an intelligent entity due to the great level of self-organisation and the ability to perform complex tasks. It also inspired many researchers to develop new clarification for problem optimisation in computer science [37]. Nevertheless, to empower the ability of ACO, combination with other algorithms might be necessary. El-Sawy and Zaki (2013) suggest that the combination of metaheuristics algorithms with other optimisation algorithms can offer more efficient behaviour and higher flexibility when dealing with large-scale problems in the real world [38].

IV. CONCLUSION

In this paper, the problem of MV selection which is driven by the requirement of rapid query processing in cyber manufacturing domain has been presented. The algorithms that were proposed to deal with MV selection (from 2009-2016) have revealed the popularity of bio-inspired algorithm in solving the problem. Nevertheless, most of these algorithms have a limitation regarding the coverage of selection constraints. While ACO seems promising in MV selection under the space, time and cost constraints, the ability of this algorithm to cover all of these constraints in its hybrid form has not been tested. Furthermore, the question of the practicality of ACO in supporting cyber manufacturing rapid reporting function (by speeding up MV selection) requires further investigation.

ACKNOWLEDGEMENT

We acknowledge financial assistance received from the Ministry of Higher Education Malaysia under Fundamental Research Grant (FRGS/1/2015/ICT04/FTMK/02/F00289) and the support from Universiti Teknikal Malaysia Melaka.

REFERENCES

[1] Economic Planning Unit, “Complexity analysis study Of Malaysia’s manufacturing industries,” 2014.

[2] J. Lee, E. Lapira, B. Bagheri, and H. Kao, “Recent advances and trends in predictive manufacturing systems in big data environment,” *Manuf. Lett.*, 2013.

[3] S. Munirathinam and B. Ramadoss, “Big data predictive analytics for proactive semiconductor equipment maintenance,” in *2014 IEEE International Conference on Big Data (Big Data)*, 2014, pp. 893–902.

[4] L. Wang and G. Wang, “Big Data in Cyber-Physical Systems, digital manufacturing and Industry 4.0,” *Int. J. Eng. Manuf.*, vol. 4, pp. 1–8, Jul. 2016.

[5] T. Ponsignon and L. Monch, “Architecture for simulation-based performance assessment of planning approaches in semiconductor manufacturing,” in *Simulation Conference (WSC), Proceedings of the 2010 Winter*, 2010, pp. 3341–3349.

[6] S. R. A. Rahim, I. Ahmad, and M. A. Chik, “Technique to improve visibility for cycle time improvement in semiconductor manufacturing,” in *2012 10th IEEE International Conference on Semiconductor Electronics (ICSE)*, 2012, pp. 627–630.

[7] Mckinsey Global Institute, “Big data: The next frontier for innovation, competition, and productivity,” 2011.

[8] D. Boyd and K. Crawford, “Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon,” *Information, Commun. Soc.*, vol. 15, no. 5, pp. 662–679, May 1986.

[9] R. Dubey, A. Gunasekaran, and S. Childe, “The impact of big data on world-class sustainable manufacturing,” *Int. J. Adv. Manuf. Technol.*, vol. 84, no. 1–4, pp. 631–645, Aug. 2016.

[10] J. Lee, H. Kao, and S. Yang, “Service innovation and smart analytics for industry 4.0 and big data environment,” *Procedia Cirp*, vol. 16, Jun. 2014.

[11] N. M. Khushairi, N. A. Emran, and M. M. Mohd Yusof, “Database performance tuning methods for manufacturing execution system,” *World Appl. Sci. J.*, vol. 30, no. 30 A, pp. 91–99, 2014.

[12] T. Wilschut, I. J. B. F. Adan, and J. Stokkermans, “Big data in daily manufacturing operations,” in *Proceedings of the Winter Simulation Conference 2014*, 2014, pp. 2364–2375.

[13] N. A. Emran, N. Abdullah, and M. N. M. Isa, “Storage Space Optimisation for Green Data Center,” in *Procedia Engineering*, 2013, vol. 53, pp. 483–490.

[14] T. Nalini, A. Kumaravel, and K. Rangarajan, “A comparative study analysis of materialized view for selection cost,” *World Appl. Sci. J.*, vol. 20, no. 4, pp. 496–501, 2012.

[15] Alka and A. Gosain, “A Comparative Study of Materialised View Selection in Data Warehouse Environment,” in *2013 5th International Conference and Computational Intelligence and Communication Networks}*, 2013, pp. 455–459.

[16] R. P. P. Karde and R. V. M. Thakare, “Materialized View Selection Approach Using Tree Based Methodology,” *Int. J. Eng. Sci. Technol.*, vol. 2, no. 10, pp. 5473–5483, 2010.

[17] H. Drias, “Generating Materialized Views Using Ant Based Approaches And Information Retrieval Technologies,” in *2011 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)*, 2011, pp. 276–282.

[18] R. S. Wahono, N. Suryana, and S. Ahmad, “Metaheuristic Optimization based Feature Selection for Software Defect Prediction,” *J. Softw.*, vol. 9, no. 5, 2014.

[19] X.-S. Yang, *Nature Inspired Metaheuristics Algorithms*. Luniver Press, 2010.

[20] I. Khennak, H. Drias, and S. Kechid, “New Modeling of Query Expansion Using an Effective Bat-Inspired Optimization Algorithm,” *IFAC-PapersOnLine*, vol. 49, no. 12, pp. 1791–1796, 2016.

[21] T. V. V. Kumar and M. Haider, “A Query Answering Greedy Algorithm for Selecting Materialized Views,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2010, pp. 153–162.

[22] H. Ismkhan, “Accelerating The Ant Colony Optimization by Smart Ants, using Genetic Operator,” 2014.

[23] Y. Xia, T. T. Luo, X. Zhang, and H. Y. Bae, “A Parallel Adaptive Partial Materialization Method of Data Cube Based on Genetic Algorithm,” *Adv. Sci. Technol. Lett.*, vol. 123, no. Cst, pp. 150–156, 2016.

[24] X. Sun and Z. Wang, “An Efficient Materialized View Selection Algorithm Based on PSO,” *Isa’09*, pp. 11–14, 2009.

[25] Q. Zhang, X. Sun, and Z. Wang, “An efficient MA-based materialized views selection algorithm,” in *Proceedings - 2009 IITA International Conference on Control, Automation and Systems Engineering, CASE 2009*, 2009, pp. 315–318.

[26] L. Zhou, H. Geng, and M. Xu, “An improved algorithm for materialized view selection,” *J. Comput.*, vol. 6, no. 1, pp. 130–138, 2011.

[27] S. H. Talebian and S. A. Kareem, “Materialized View Selection Using Vector Evaluated Genetic Algorithm,” in *International Conference on Computer Engineering and Technology, 3rd (ICCET 2011)*, 2011, p. 10.

[28] P. Tiwari and S. V. Chande, “Optimization of Distributed Database Queries Using Hybrids of Ant Colony Optimization Algorithm,” *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 3, no. 6, pp. 609–614, 2013.

[29] E. Datta and K. N. Dey, “Materialized View Generation Using Apriori Algorithm,” *Int. J. Database Manag. Syst.*, vol. 7, no. 6, pp. 17–27, 2015.

[30] B. Arun and T. V. V. Kumar, “Materialized View Selection Using Improvement based Bee Colony Optimization,” *Int. J. Softw. Sci. Comput. Intell.*, vol. 7, no. 4, pp. 35–61, 2015.

[31] L. Gao and X. Song, “An Ant Colony based algorithm for optimal selection of Materialized view,” *Int. Conf. Intell. Comput. Integr. Syst.*, pp. 534–536, 2010.

[32] D. Teodorović, P. Lucic, G. Markovic, and M. Dell’Orco, “Bee Colony Optimization: Principles and Applications,” *8th Semin. Neural Netw. Appl. Electr. Eng.*, vol. 0, pp. 151–156, 2006.

[33] L. Chaves, E. Buchmann, F. Hueske, and K. Böhm, “Towards materialized view selection for distributed databases,” *EDBT '09 Proc. 12th Int. Conf. Extending Database Technol. Adv. Database Technol.*, pp. 1088–1099, 2009.

[34] L. Zhou, X. He, and K. Li, "An improved approach for materialized view selection based on genetic algorithm," *J. Comput.*, vol. 7, no. 7, pp. 1591–1598, 2012.

[35] B. Arun and T. V. V. Kumar, "Materialized View Selection using Marriage in Honey Bees Optimization," *Int. J. Nat. Comput. Res.*, vol. 5, no. 3, pp. 1–25, 2015.

[36] X. Li, X. Qian, J. Jiang, and Z. Wang, "Shuffled Frog Leaping Algorithm for Materialized Views Selection," *2nd Int. Work. Educ. Technol. Comput. Sci. ETCS 2010*, vol. 3, pp. 7–10, 2010.

[37] N. Abd-Alsabour, "A review on evolutionary feature selection," in *Proceedings - UKSim-AMSS 8th European Modelling Symposium on Computer Modelling and Simulation, EMS 2014*, 2015, pp. 20–26.

[38] R. M. R.-A. Ahmed Ahmed El-Sawy , Elsayed M. Zaki, "A Novel Hybrid Ant Colony Optimization and Firefly Algorithm for Solving Constrained Engineering Design Problems," *J. Nat. Sci. Math.*, vol. 6, no. 1, pp. 1–22, 2013.