University of Pannonia
Faculty of Information Technology
Doctoral School of Information Science and Technology

Online monitoring and optimization of business processes using process mining and artificial intelligence algorithms

Theses of Ph.D. dissertation

Zsuzsanna Nagy

Supervisor: dr. Ágnes Werner-Stark

2025

1 Introduction

During my doctoral research, I worked on business process related problems, mainly in the area of manufacturing and transportation processes.

1.1 Manufacturing processes

Process mining, which aims to improve processes, provides a way to analyze information from business process event data [1]. Conformance checking is a main type of process mining where event data is compared with the process model so that deviations can be detected and explained [2]. Nowadays, the basic technique accepted in practice is to calculate alignments [3]. The differences accurately marked by the alignments can be interpreted by experts, allowing them to draw conclusions and take action to improve future process executions.

There are several kinds of alignment-based approaches to conformance checking. Multi-perspective Conformance Checking (MCC) solutions consider other perspectives (e.g. time, resource or data) in addition to the control-flow perspective (e.g. balanced MCC [4]). Online Conformance Checking (OCC) solutions are able to detect deviations from the expected behaviors even before the process instance terminates (e.g. prefix-alignment based OCC [5]). In the case of manufacturing processes, it is crucial that deviations are detected in all perspectives and as quickly as possible. However, such a Multi-perspective Online Conformance Checking (MOCC) solution is currently not available.

The visualization of process data is crucial to quickly identify deviations. However, existing event data visualization solutions do not handle temporal information, such as overlapping events, adequately. In addition, the graphical representation of deviations detected by conformance checking algorithms is still an open challenge [6].

1.2 Transportation processes

Transportation processes are often modeled as a Capacitated Arc Routing Problem (CARP), which is a combinatorial optimization NP-hard problem [7]. The objective of CARP is to serve a set of designated road segments (edges and/or arcs) of a road network (graph) with a given capacity and quantity of homogeneous vehicles executing the tasks. The objective is to find optimal route plans (service plans) in which all designated road sections are served while the total cost of the vehicles is minimized.

Since the CARP is an NP-hard problem, instead of using exact methods, heuristic and metaheuristic approaches are used in the literature to achieve optimal or near-optimal solutions. However, the existing CARP solvers are either too slow (e.g. Hybrid Metaheuristic Approach (HMA) [8]) or unable to find solutions of acceptable quality (e.g. Ant Colony Optimization algorithm with Path Relinking (ACOPR) [9]). The Artificial Bee Colony (ABC) algorithm is a swarm intelligence-based optimization method that was inspired by the foraging behavior of honey bees [10]. It has been successfully applied to many combinatorial optimization problems similar to CARP, such as the Travelling Salesman Problem (TSP), and has been shown to outperform many evolutionary algorithms [11]. However, the ABC algorithm has not yet been applied to solve the CARP. A further problem is that

the search move operators available for CARP either make too small or too large changes, limiting the ability to exhaustively explore the search space.

Normal CARP assumes a static problem, which is not realistic. During the execution of a service plan, dynamic changes (unexpected events) can occur, altering the problem, feasibility, and total cost of the current service plan [12]. Such an event could be, for example, the appearance of a new task or the breakdown of a vehicle in service. The Dynamic Capacitated Arc Routing Problem (DCARP) aims at handling such events in order to ensure a problem-free execution of the service. Recently, a framework for applying static CARP solving algorithms to solve DCARPs has been presented [13]. This was shown to be more effective than the earlier approaches, which could handle only a few dynamic events and were also slow. For the framework, a Hybrid Local Search (HyLS) algorithm was developed, which performs local search until no better solutions can be found [14]. This solution helps to adapt quickly in a dynamic environment, but only allows a complete rescheduling of the current service plan, which may affect the route plans of several vehicles. An additional challenge is that current DCARP approaches assume that vehicles follow the assigned service plan perfectly, which does not reflect real-world conditions. For this reason, realistic benchmarking of the performance of (D)CARP solver algorithms remains a challenge.

2 Goals and applied methods

In the first part of my dissertation, I focused on **multi-perspective online conformance checking** and process data visualization. The goal of my research was to create solutions that support the online monitoring and improvement of fixed processes (e.g. manufacturing processes) with short lead times and heterogeneous activities. The use of these solutions enables the detection of deviations from the expected behavior and their root causes as quickly as possible, thereby minimizing the number of faulty products and the frequency and duration of downtimes. For these problems, the process model and the event data (in the form of an event stream) are given.

In the second part of my dissertation, I focused on **static and dynamic capacitated arc routing problems**. The goal of my research was to create solutions that support online monitoring and improvement of flexible transportation processes consisting of homogeneous activities. These solutions generate a revised, modified service plan, if necessary, to minimize the total cost (service time and/or CO_2 emissions). For these problems, the mathematical model (road network, set of tasks, etc.) and the event data (in the form of an event stream) are given.

3 New scientific results

Thesis 1

In this thesis, I focused on developing solutions to support online monitoring and improvement of processes with short lead times (e.g. manufacturing processes). In this scope, I focused on one of the main types of process mining, conformance checking and process data visualization.

The corresponding publications are: [P1–P9]

Thesis 1.1

I proposed the first MOCC solution, which is a prefix-alignment based MOCC method. The proposed method aims to support online monitoring of processes from multiple perspectives. The solution calculates optimal multi-perspective (prefix-)alignments between a multi-perspective process model (given as a Data Petri Net) and the observed events of an event stream. For this purpose, I further developed the incremental A* algorithm available in the literature, which originally searches for optimal prefix-alignments considering only the control-flow perspective.

To reduce the computation time of the alignments, I added two features to the method. One of them is direct synchronizing, which speeds up the search process for perfectly fitting traces. The other one is the caching of solutions of optimal variable assignment problems, which reduces the computation time for not fully fitting traces. I also extended the MOCC method with a case caching method, which ensures that online alignments are computed for ongoing process executions and offline alignments for completed process executions.

I evaluated the effectiveness and robustness of the proposed MOCC method through two different kinds of experiments. In the first experiment, I compared the MOCC method with an OCC solution applied on a real process. The results showed that the MOCC method can detect more wrong behaviors than the OCC solution due to its multi-perspective nature, while performing at an acceptable speed. In the second experiment, I compared the alignments given by the MOCC method and an MCC solution applied on three real processes. The results showed that the MOCC method provided multi-perspective alignments of the same quality as the MCC method for the completed process executions. Furthermore, in contrast to the MCC solution, the MOCC method is able to perform conformance checking on process executions that are still ongoing.

Based on these results, I concluded that the MOCC method is suitable as both an online and a multi-perspective conformance checking solution. Consequently, it is ideal for monitoring real-world processes with short lead times and a prescriptive process model, but it can also be applied to processes with longer lead times.

Thesis 1.2

I developed two new process data visualization methods that are capable of graphically displaying the original event data and the MOCC outputs (alignment data) calculated from the event data. The purpose of the methods is to provide an intuitive and comprehensive view of the complex process data. This allows the process under investigation to be monitored, as well as quickly detecting deviations and

their root causes. The visualizations use a Gantt chart inspired mixed (fixed and duration) timeline-based visualization, which is able to handle overlapping objects appropriately. They also allow users to select the perspective from which they want to examine the data.

I implemented and evaluated the visualization methods in a test environment. I tested their utility and performance on a real manufacturing process. The results showed that by using the methods, otherwise difficult to detect anomalies can be easily found, as well as possible causes of problems (e.g. unexpected sudden short machine downtimes during the process execution cause the products become faulty).

Thesis 2

In this thesis, I focused on developing solutions to support online monitoring and improvement of transportation processes. In this scope, I focused on the Capacitated Arc Routing Problem (CARP) and its dynamic variant (DCARP).

The corresponding publications are: [P11–P14]

Thesis 2.1

I created a medium step-size move operator for solving the CARP, which I named the sub-route plan operator. This operator is based on TSP's Greedy Sub Tour Mutation (GSTM) operator and aims to make the local search process more efficient for the evolutionary algorithms developed for the CARP. The operator consists of two different greedy search methods (greedy reconnection and sub-route rotation) and a distortion method. The sub-route plan operator modifies the order and the service direction of one route plan task at a time, so it does not affect the feasibility of the service plan.

I evaluated the efficiency of the sub-route plan operator in comparison to the small step-size move operators known from the literature (inversion, insertion, swap, and two-option). The experiment was performed on three CARP instances of different complexity levels, from the EGL and EGL-Large benchmark test sets (egl-e1-A, egl-s1-A, and egl-g1-A), which are instances that are based on real road networks and tasks. The results showed that the sub-route plan operator has a higher probability of achieving a better solution than the other move operators. This is especially true in cases where the current solution requires major modifications to achieve the optimal solution.

Thesis 2.2

I proposed the first ABC algorithm for the CARP (CARP-ABC algorithm). In total, I created two versions of the algorithm: one focused on exploration and one focused on discovery. The discovery-focused version aims to provide solutions as diverse as possible, so it is recommended to be used as a static CARP solver. In contrast, the exploration-focused version aims to find the best solution as quickly as possible, so it is recommended to be used as a DCARP solver.

I measured the performance of the discovery-focused CARP-ABC algorithm compared to other CARP solving algorithms available in the literature (HMA and ACOPR) on five CARP instances of different complexity levels (kshs1, egl-e1-A, egl-s1-A, egl-g1-A, and egl-g2-A), one of which is synthetic and the other ones are

real-life based instances. The results showed that the algorithm can be considered competitive with the currently most accurate CARP solver, the HMA, when the maximum allowed runtime of the algorithms is limited to about one minute. The efficiency of the exploration-focused CARP-ABC algorithm was tested using the data-driven DCARP framework (see Thesis 2.3).

Thesis 2.3

I proposed a minimal rerouting (RR1) algorithm to solve DCARP instances. The goal of the RR1 algorithm is to make an infeasible service plan feasible, with a minimum number of changes to the route plan, as quickly as possible.

To improve the practical applicability of the RR1 algorithm and other DCARP solvers, and to evaluate their performance in a realistic simulation environment, I developed a data-driven DCARP framework. The framework can handle all possible dynamic events. It updates the problem instance and the service plan after each event observation and invokes its rescheduling module if it is necessary, which applies the predefined DCARP solver algorithm(s). To perform the simulations, two versions of the framework were created. The first version uses only an event generator module (DCARP framework with an event generator), the second one uses a traffic simulation software for a more realistic simulation (DCARP framework with traffic simulation). I measured the performance of the exploration-focused CARP-ABC algorithm, the RR1 algorithm and the (D)CARP solver algorithms (HMA and HyLS algorithms) available in the literature using both frameworks, compared to each other, with a one-minute runtime limit.

The performance of the algorithms (RR1, exploration-focused CARP-ABC and HMA) was examined using the DCARP framework with an event generator to investigate the performance of the algorithms (RR1, exploration-focused CARP-ABC and HMA) on the occurrence of critical events ("task appearance", "demand increase" and "vehicle breakdown") generated from a medium-sized real-world CARP example (egl-e1-A). The results showed that the exploration-focused CARP-ABC algorithm was the most efficient in terms of solution quality, while the RR1 algorithm performed best in terms of runtime.

The performance tests performed with the DCARP framework with traffic simulation analyzed 12 different scenarios based on real road networks and tasks. Three types of experiments were performed with combinations of traffic events generated by the traffic simulation and the different number of task appearances. The results showed that the exploration-focused CARP-ABC algorithm was more efficient on average, especially when combined with the RR1 algorithm, while the HyLS and HMA algorithms performed weaker.

4 Publications

The described results have been presented in journals, international conferences, and conference proceedings. I list them below.

Multi-perspective Online Conformance Checking

[P1] **Zs. Nagy**, Á. Werner-Stark, and T. Dulai, "An industrial application using process mining to reduce the number of faulty products", in *New Trends in Databases and Information Systems: ADBIS 2018 Short Papers and Workshops, AI* QA, BIGPMED, CSACDB, M2U, BigDataMAPS, ISTREND, DC, Budapest, Hungary, September 2–5, 2018, Proceedings*, Springer, Cham, 2018, pp. 352–363.

[P2] **Zs. Nagy**, Á. Werner-Stark, and T. Dulai, "Analysis of industrial logs to reduce the number of faulty products of manufacturing", in *OGIK 2018: 15. Országos Gazdaságinformatikai Konferencia - Az előadások összefoglalói, Sopron, Hungary, November 9–10, 2018*, L. Bacsárdi, G. Bencsik and Z. Pödör, Eds., Sopron, Hungary: Alexander Alapítvány a Jövő Értelmiségéért, 2018, pp. 26–27.

[P3] **Zs. Nagy**, Á. Werner-Stark, and T. Dulai, "Analysis of industrial logs to reduce the number of faulty products of manufacturing", in *OGIK'2018 Országos Gazdaságinformatikai Konferencia - Válogatott közlemények, Sopron, Hungary, November 9–10, 2018*, L. Bacsárdi, G. Bencsik and Z. Pödör, Eds., Sopron, Hungary: Alexander Alapítvány a Jövő Értelmiségéért, 2019, pp. 53–57.

[P4] **Zs. Nagy**, Á. Werner-Stark, and T. Dulai, "Using process mining in real-time to reduce the number of faulty products", in *Advances in Databases and Information Systems: 23rd European Conference, ADBIS 2019, Bled, Slovenia, September 8–11, 2019, Proceedings*, T. Welzer, J. Eder, V. Podgorelec, and A. K. Latifić, Eds., ser. Lecture Notes in Computer Science, vol. 11695, Springer, Cham, 2019, pp. 89–104.

[P5] **Zs. Nagy** and Á. Werner-Stark, "A multi-perspective online conformance checking technique", in *2020 6th International Conference on Information Management (ICIM), London, UK, March 27–29, 2020, Proceedings*, IEEE, New York, NY, USA, 2020, pp. 172–176.

[P6] **Zs. Nagy** and Á. Werner-Stark, "An alignment-based multi-perspective online conformance checking technique", in *Abstract book for the 16th MIKLÓS IVÁNYI INTERNATIONAL PHD & DLA SYMPOSIUM, Pécs, Hungary, October 26–27, 2020*, P. Iványi, Ed., Pécs, Hungary: Pollack Press, 2020, p. 126.

[P7] **Zs. Nagy** and Á. Werner-Stark, "An alignment-based multi-perspective online conformance checking technique", *Acta Polytechnica Hungarica*, vol. 19, no. 4, pp. 105–127, 2022. [**IF: 1.7**]

[P8] **Zs. Nagy** and Á. Werner-Stark, "Real-time visualization of the outputs of multi-perspective online conformance checking analysis", in *Abstract book for the 18th MIKLÓS IVÁNYI INTERNATIONAL PHD & DLA SYMPOSIUM, Pécs, Hungary, November 3–4, 2022*, P. Iványi, Ed., Pécs, Hungary: Pollack Press, 2022, p. 80.

[P9] **Zs. Nagy** and Á. Werner-Stark, "Visualization methods to support real-time process monitoring", in *Proceedings of the 3rd International Workshop on Information Technologies: Theoretical and Applied Problems 2023 (ITTAP 2023), Ternopil, Ukraine, Opole, Poland, November 22–24, 2023*, ser. CEUR Workshop Proceedings, vol. 3628, CEUR-WS.org, Aachen, Germany, 2023, pp. 1–14.

Capacitated Arc Routing Problems

[P10] **Zs. Nagy**, Á. Werner-Stark, and T. Dulai, "Solving the capacitated arc routing problem by a special evolutionary optimization algorithm", in *OGIK'2021 Országos Gazdaságinformatikai Konferencia, Veszprém, Hungary, November 12–13, 2021*, Veszprém, Hungary: Platina Nyomda and Kiadó Kft., 2021, pp. 41–42.

[P11] **Zs. Nagy**, Á. Werner-Stark, and T. Dulai, "A data-driven solution for the dynamic capacitated arc routing problem", in *Proceedings of IAC in Budapest 2021, Budapest, Hungary, November 26–27, 2021*, H. Kratochvílová and R. Kratochvíl, Eds., Prague, Czech Republic: Czech Institute of Academic Education z.s., 2021, pp. 64–83.

[P12] **Zs. Nagy**, Á. Werner-Stark, and T. Dulai, "An artificial bee colony algorithm for static and dynamic capacitated arc routing problems", *Mathematics*, vol. 10, no. 13, paper 2205, 2022. [**IF: 2.4**]

[P13] **Zs. Nagy**, Á. Werner-Stark, and T. Dulai, "Solving data-driven dynamic capacitated arc routing problems", in *Middle-European Conference on Applied Theoretical Computer Science (MATCOS-22), Koper, Slovenia, October 13–14, 2022, Booklet*, 2022, p. 16.

[P14] **Zs. Nagy**, Á. Werner-Stark, and T. Dulai, "Comparison of optimization algorithms for the dynamic capacitated arc routing problem", in *Abstracts of the International Conference on Optimization and Algorithms (OPAL 2023) Semi Online 25, Veszprém, Hungary, June 5–9, 2023*, 2023, p. 19.

MTMT profile

`https://m2.mtmt.hu/api/author/10068634`

References

[1] W. Van Der Aalst, *Process mining: data science in action*. Springer Berlin, Heidelberg, 2016, vol. 2.

[2] J. Carmona, B. van Dongen, A. Solti, and M. Weidlich, *Conformance checking*. Springer, Cham, 2018, vol. 1.

[3] S. J. van Zelst, A. Bolt, and B. F. van Dongen, "Tuning alignment computation: An experimental evaluation", in *Proceedings of the International Workshop on Algorithms & Theories for the Analysis of Event Data 2017 (ATAED 2017), Zaragoza, Spain, June 26–27, 2017*, vol. 1847, CEUR-WS.org, Aachen, Germany, 2017, pp. 6–20.

[4] F. Mannhardt, M. De Leoni, H. A. Reijers, and W. M. Van Der Aalst, "Balanced multi-perspective checking of process conformance", *Computing*, vol. 98, pp. 407–437, 2016.

[5] D. Schuster and S. J. van Zelst, "Online process monitoring using incremental state-space expansion: An exact algorithm", in *Business Process Management: 18th International Conference, BPM 2020, Seville, Spain, September 13–18, 2020*, Springer, Cham, 2020, pp. 147–164.

[6] J. Carmona, B. van Dongen, and M. Weidlich, "Conformance checking: Foundations, milestones and challenges", in *Process Mining Handbook*, Springer, Cham, 2022, pp. 155–190.

[7] B. L. Golden and R. T. Wong, "Capacitated arc routing problems", *Networks*, vol. 11, no. 3, pp. 305–315, 1981.

[8] Y. Chen, J.-K. Hao, and F. Glover, "A hybrid metaheuristic approach for the capacitated arc routing problem", *European Journal of Operational Research*, vol. 253, no. 1, pp. 25–39, 2016.

[9] C.-J. TING and H.-S. TSAI, "Ant colony optimization with path relinking for the capacitated arc routing problem", *Asian Transport Studies*, vol. 5, no. 2, pp. 362–377, 2018.

[10] D. Karaboga, "An idea based on honey bee swarm for numerical optimization", Computer Engineering Department, Engineering Faculty, Erciyes University, Kayseri, Turkey, Tech. Rep. tr06, Oct. 2005.

[11] D. Karaboga and B. Gorkemli, "Solving traveling salesman problem by using combinatorial artificial bee colony algorithms", *International Journal on Artificial Intelligence Tools*, vol. 28, no. 01, 2019.

[12] Á. Corberán, R. Eglese, G. Hasle, I. Plana, and J. M. Sanchis, "Arc routing problems: A review of the past, present, and future", *Networks*, vol. 77, no. 1, pp. 88–115, 2021.

[13] H. Tong, L. L. Minku, S. Menzel, B. Sendhoff, and X. Yao, "A novel generalized metaheuristic framework for dynamic capacitated arc routing problems", *IEEE Transactions on Evolutionary Computation*, vol. 26, no. 6, pp. 1486–1500, 2022.

[14] E. Babaee Tirkolaee, I. Mahdavi, M. M. Seyyed Esfahani, and G.-W. Weber, "A hybrid augmented ant colony optimization for the multi-trip capacitated arc routing problem under fuzzy demands for urban solid waste management", *Waste management & research*, vol. 38, no. 2, pp. 156–172, 2020.