Causality-driven dynamic scheduling of multipurpose batch plants

Taicheng Zhenga, Dan Lia, Jie Lia\*

a Centre for Process Integration, Department of Chemical Engineering, School of Engineering, The University of Manchester, Oxford Road, Manchester, M13 9PL, United Kingdom

\*Corresponding author email: jie.li-2.manchester.ac.uk;

Abstract

Modern multipurpose batch plants often require non-periodic rescheduling in response to random disturbances. The rescheduling strategy, that is, when to reschedule and how to reschedule, is crucial not only for reducing long-term cost, but also for the alleviation of nervousness within production environment. However, such two objectives often conflict in practice. In this study, we propose a causality-driven rescheduling strategy to address this challenge in multipurpose batch plants that widely exist in chemical, pharmaceutical, and food industries. The main idea is to use a directed Probabilistic Graphical Model (PGM) to formalize the temporal and spatial causal relationships between tasks. As the dynamic system evolves, given the evidence of observed disturbances, the PGM serves as a standby knowledge base to help scheduler infer the posterior distribution of impacted level of unobserved tasks. When the combined count of observed highly affected tasks and unobserved but highly likely to be impacted tasks exceeds a threshold, rescheduling is triggered. Also, the strategy alleviates system nervousness by fixing tasks that are less likely to be highly impacted. We compare this strategy with the classic exhaustively-minimize-cost strategy on a standard benchmark problem. Numerical results show that, comparatively, our strategy reduces 91.4% computational time and 91.2% count of task changes, at a sacrifice of 18.7% cumulative cost.

**Keywords**: dynamic scheduling, causality, directed Probabilistic Graphical Model, multipurpose batch plant.

* 1. Introduction

Practically, due to the presence of irregularly arriving disturbances, multipurpose batch plants often operate in a dynamic pattern (Ouelhadj and Petrovic, 2009). This pattern often involves a repeating process of scheduling, execution, observation, feedback, and next-round rescheduling. For example, in chemical, pharmaceutical, and food industries, during schedule execution, schedulers may need to frequently reschedule in response to various disruptions such as machine breakdown, task delay, raw material shortage, and demand variation. As indicated by the literature, in such dynamic environments, the rescheduling strategy is crucial for enhancing long-term performance. The rescheduling strategy includes two principal aspects: *when* to reschedule and *how* to reschedule. In the case of multipurpose batch plants, arguably because of the complexity of operational logic, previous studies mostly follow a periodically, completely rescheduling strategy (Gupta and Maravelias, 2019). However, as we will provide evidence later in the case study, such strategy often causes a high-level nervousness within the production environment. In plain terms, schedules before and after rescheduling are often quite different. Such nervousness often causes the plant personnel being strained to adjust to a constantly changing schedule, which has negative impacts on the production process (Pinedo, 2016).

When experienced human schedulers implement dynamic scheduling, they do not mechanically reschedule in a periodic manner. Instead, they often at first evaluate the impact of newly observed disturbances on the schedule by causal reasoning, and then adjust only the targeted part of the schedule. In multipurpose batch plants, spatially, if an upstream task was affected by a disruption, downstream tasks in the production process will be affected as well; temporally, if an earlier task is delayed, subsequent tasks on the shared machine will also be affected. Such causality-based expertise, which is mainly achieved by click-and-drag human-machine interaction in scheduling software, has been proved to be effective in many cases (Harjunkoski et al., 2014). However, to the best of our awareness, no literature has formalized this idea before. In this study, we propose a causality-driven rescheduling strategy for dynamic scheduling of multipurpose batch plants. The main idea is to use this logical causality to factorize the joint probability distribution of impacted level of to-be-executed tasks by a directed Probabilistic Graphical Model (PGM). When a disturbance is observed, the PGM performs inference algorithms to derive a posterior distribution that presents the overall impacted level of remaining unobserved part of the schedule. This posterior distribution is then used to further design the rescheduling strategy. We test the effectiveness of this idea in one of the standard benchmark problems in process scheduling community (Kallrath, 2002). The results show that the strategy can significantly reduce production nervousness in the measure of task change counts and computational time at a cost of 18.7% cumulative objective value.

* 1. Methodology



Figure 1. Flowchart of causality-driven dynamic scheduling.

In the flowchart, while squared boxes denote mathematical models or algorithms, cylindrical containers denote datasets.

We first describe the dynamic scheduling problem of multipurpose batch plants and then illustrate our methodology by going through the flowchart in Fig. 1. We consider a discrete-time dynamic system $s\_{t+1}=T\left(s\_{t},a\_{t},i\_{t}\right)$ of a multipurpose plant that can be characterized by an encapsulating parameter $λ$ (include plant layout, recipes, and supply/demand profile). In the system, starts from $t\_{0}$ and ends at $t\_{f}$, $s\_{t}$ is vectorized notation for system state; $a\_{t}$ is action set that the system is scheduled to take within horizon $[t,t+H)$; $i\_{t}$ is informational state that has been revealed to the scheduler; and $T$ denotes transition function. For informational state $i\_{t}$, we assume that (1) we have known information about the distribution of $i\_{t}$; (2) we have a certain length of *vision* $v$over $i\_{t}$. That is, at $v$-time-periods ahead of the instantiation of disturbance $i\_{t}$, the information is revealed to the scheduler; (3) after $i\_{t}$ is instantiated, it will not change. The target of the problem is to find a policy $π $such that $a\_{t+1}=π\left(s\_{t},a\_{t},i\_{t}\right)$ to minimize the cumulative cost $c=\sum\_{t\_{0}}^{t\_{f}}c\_{t}$ and the cumulative action difference $d=\sum\_{t\_{0}}^{t\_{f}-1}∆(a\_{t},a\_{t+1})$, where $∆$ is a difference-measuring function that maps the tuple $(a\_{t},a\_{t+1})$ to a non-negative integer.

At the system-level, the system dynamics starts with $t\_{0}$, $s\_{t\_{0} }$, $i\_{t\_{0} }$, and $a\_{t\_{0}}=∅$. If $a\_{t}=∅$ or binary rescheduling decision (viii) is triggered, the system state (ii) is fed to a Mixed-Integer Linear Programming (MILP) (ix) to generate a to-be-executed schedule (x); otherwise (viii) is not triggered, the system evolves to a new system state $s\_{t+1}$ as time advances. The essential part of the causality-driven policy is the active interaction between the rest modules. Specifically, after a schedule (x) is output by (ix), rather than directly advance the system by taking actions from (x), we first perform multiple episodes of Monte Carlo simulation (xi) to sample different realizations of unobserved disturbances in the horizon of (x). Then, for the set of tasks that are in (x) but fall beyond vision $v$, which is denoted by $\tilde{X}=\{X\_{1},X\_{2},…,X\_{n}\}$, we construct a PGM $G=(\tilde{X},A)$, such that each node $X\_{i}$ is a random variable that describes the impacted level of unobserved task $X\_{i}$ and the set of directed arcs $A$ is the temporal and spatial causal relationship that described in Sec. 1. The PGM factorizes the joint probability distribution of $\tilde{X}$ in the semantics that $P\left(\tilde{X}\right)=\prod\_{i=1}^{n}P(X\_{i}|u(X\_{i}))$, where $u(X\_{i})$ denotes the parent nodes of $X\_{i}$. Then, by feeding the training data (xiv) that was generated from (xi) and processed by feature engineering (xiii), we can estimate a full-fledged PGM as a buffered knowledge base to help the scheduler to infer the posterior distribution $P\left(\overbar{X}\right),∀i\in \tilde{X}\\overbar{X}$ from evidence (iv), where $\overbar{X}$ is the set of tasks that has been observed being affected by disturbance. In other words, the distribution of impacted level of remain unobserved tasks can be updated by realizations of $\overbar{X}$. Finally, the framework iterates between the three different abstraction levels and advance until $t\_{f}$.

* 1. Case study

We test the effectiveness of our causality-driven policy on one of the standard benchmark problems in process scheduling community (Kallrath, 2002). The State-Task Network (STN) representation of the problem is shown in Fig. 2. The processing time, minimum and maximum batch size parameter are identical to the supplementary material of (Li et al., 2022). In addition, we assume that there is a $1 setup cost for each task and proportional cost of batch size is 0. The capacity for raw material k0 is infinity, for k11 and k21 is 25 ton, and for the rest is 15 ton. The price for each material is listed in Table 1. The inventory cost per time-period is 0.15 multiple of price and backlog cost for products is 1.5 multiple of price.

Table 1. Price for each material (unit: $)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Material | k0 | k11 | k21 | k22 | k31 | k41 | k42 | k43 | k44 | k51 |
| Price | 1 | 5 | 10 | 10 | 10 | 10 | 10 | 15 | 10 | 15 |
| Material | k52 | k61 | k62 | k63 | p1 | p2 | p3 | p4 | p5 |  |
| Price | 15 | 15 | 20 | 20 | 50 | 50 | 50 | 50 | 50 |  |



Figure 2. State-Task Network (STN) representation of the benchmark problem (Kallrath, 2002).

For system dynamics, we assume that the horizon of each schedule is 72 h (3 days), and we perform a rolling-horizon dynamic scheduling experiment over 336 h (2 weeks) timespan. For raw material k0, the regular supply is 20 ton every hour. For each product, there is a regular demand of 12 tons per 12h that starts from $t\_{0}$. We assume five types of possible disturbances (Table 2), and the scheduler has a vision of 36h over each type of disturbance.

Table 2. Assumed disturbances in the dynamic scheduling problem

|  |  |  |  |
| --- | --- | --- | --- |
| Disturbance | Raw material supply change | Demand change | Processing time change |
| Distribution | $N\_{clipped}(1, 0.5^{2})$ a | $Poisson\_{clipped}\left(0.02\right)\*U(0.5, 1.5)$ b | $N\_{clipped}(1, 0.5^{2})$ c |
| Disturbance | Machine breakdown (occurrence) | Machine breakdown (duration) | Yield change |
| Distribution | $Bern(0.02)$ d | $U(2, 5)$ e | $U(0.5, 1)$ f |

a: The raw material supply will be affected by a clipped Gaussian random variable such that the maximum purchasable amount of raw material will be limited to $N\_{clipped}(1, 0.5^{2})$-multiple of the regular supply. The random variable is clipped (that is, when the random variable is sampled beyond the boundary, the value will set to its closest boundary value) between [0.5, 1] for the consideration of nonnegativity. b: The unexpected arrival of demand will be count of orders (clipped Poisson random variable) multiplies amount for each order (uniform random variable). The Poisson random variable is clipped between [0, 2] to avoid infinity. c: The processing time for tasks will be affected by a clipped Gaussian random variable such that the varied processing time will be $\left⌈τ∙N\_{clipped}(1, 0.5^{2})\right⌉$, where $τ$ is the nominal processing time and $N\_{clipped}(1, 0.5^{2})$ is Gaussian random variable that is clipped between $\left[\frac{1}{τ}, 2\right]$. d: Whether there is a breakdown disruption on a machine is a Bernoulli random variable with $p=0.02$. In other words, for each time-period and each machine, there is 0.02 chance that there will be a machine breakdown disruption. e: When a machine breakdown is occurred, the duration of the breakdown is a uniform random variable $U(2, 5)$. f: The yield of each task is $U(0.5, 1)$-multiple of the original batch size.

* 1. Results

There are various measures to quantify how effective a rescheduling strategy performs in a dynamic environment. To set up a standard benchmark for the numerical results, we consider the following metrics:

* $c^{\*}$. The optimal objective cost of the static version of scheduling problem under *nominal* conditions. That is, we regard the dynamic scheduling problem as static by assuming that there is no disturbance through the entire dynamic scheduling timespan. Since the occurrence of a disturbance event will usually deteriorate objective, $c^{\*}$ is served as a reference lower bound for the cumulative objective cost in our case study.
* $c^{\infty }$. The optimal value of the dynamic scheduling problem with the assumption that the scheduler has infinite vision over uncertain events. In other words, we assume that the scheduler can foresee all realizations of uncertain events at $t\_{0}$. $c^{\infty }$ is obtained by solving a MILP that involves all realized uncertainties through the timespan. $c^{\infty }$ is served for two purposes: (1) when comparing to $c^{\*}$, $c^{\infty }$ reflects how unstable a system is due to the existence of disturbance. Specifically, the larger gap between $c^{\*}$ and $c^{\infty }$ is, the more unstable the system is. (2) $c^{\infty }$ is the theoretical lower bound for any dynamic scheduling policies because a longer vision will at least not worsen the cumulative objective value.
* $c^{v}$. The cumulative objective value of a dynamic scheduling policy with finite vision $v$. $c^{v}$ is the measure for the effectiveness of dynamic scheduling policies in practice. In this study, we compare $c^{v}$ of the *causality-driven* policy with two other classic policies: *exhaustively-minimize-cost* and *periodically-minimize-cost*. To explain:
	1. $c\_{causal}^{v}$. The cumulative cost of the proposed causality-driven policy. In this policy, we assume that a task is categorized as highly impacted either there is a breakdown on the execution machine, or the task is delayed over 50% of the nominal processing time, or the yield loss exceeds 40% of the original batch size. Rescheduling is triggered when the count of observed highly impacted tasks and unobserved tasks but with over 50% probability being highly impacted exceeds half of the task count in the schedule. When rescheduling is triggered, those not highly impacted tasks are fixed.
	2. $c\_{exhaust}^{v}$. The cumulative cost of exhaustively-minimize-cost policy. That is, the objective cost is minimized through MILP every hour. For the consideration of computational time, the MILP solver is terminated either when the relative MIP gap is below 10% or runtime exceeds 300s. Also, the value of integer variables are hinted from previous solutions.
	3. $c\_{periodic}^{v}$. The cumulative cost of periodically-minimize-cost policy. That is, the objective cost is minimized every 24h and the count of task changes is minimized every hour. When two types of minimization problems overlap, the former takes priority. The parameter for MILP solver is set identical to that in $c\_{exhaust}^{v}$.

In addition to cumulative cost, another metric that is of our interest is the cumulative value of difference-measuring function $d=\sum\_{t\_{0}}^{t\_{f}-1}∆(a\_{t},a\_{t+1})$. $d$ is the measure for plant nervousness of a dynamic scheduling process in practice. That is, the larger value $d$ takes, more frequent schedule adjustment is required in production. However, it is not straightforward to define $∆$ since there are some subtleties in the notion of task change in multipurpose batch plant. First, unlike discrete manufacturing where the number of tasks is often predetermined, multipurpose batch plants may have schedules with variable number of tasks. Second, it can be challenging to determine whether a task is shifted from previous time slots or is a completely new addition. To setup a unified metric for comparison, we define $∆(a\_{t},a\_{t+1})$ as follows: for the overlapping horizon of $a\_{t}$ and $a\_{t+1}$, $∆(a\_{t},a\_{t+1})$ is the sum of (1) number of tasks appear in $a\_{t}$ but not in $a\_{t+1}$, and (2) number of tasks appear in $a\_{t+1}$ but not in $a\_{t}$.

The numerical results are presented in Table 3. In such experimental setup, we observe that $c^{\infty }$ is 8.94 multiple of $c^{\*}$. This large gap between $c^{\*}$ and $c^{\infty }$ indicates that this system is highly unstable due to the existence of disturbance that described in Table 2. Also, we notice that $c\_{exhaust}^{v}$ is even slightly lower than the theoretical lower bound $c^{\infty }$. This is because the computation of $c^{\infty }$ is terminated due to 3600s time limit and there is still 1.27% MIP gap for $c^{\infty }$. Therefore, $c\_{exhaust}^{v}$ is very close to the theoretical lower bound. However, the exhaustive strategy suffers from a high value of $d$ and a very long computational time. As a compromise, if we reduce the rescheduling frequency from every hour to every 24 h, as shown by $c\_{periodic}^{v}$, we can see an 89.2% reduction of computational time and a 72.6% decrease of $d$. However, $c\_{periodic}^{v}$ is 76.2% higher than $c\_{exhaust}^{v}$, in which is often unacceptable practice. While the trade-off between $c$ and $d$ seems hard to balance, the causality-driven strategy reduces $d$to 830, which is even significantly lower than that in periodic strategy. Also, $c\_{causal}^{v}$ is only 18.7% higher than exhaustive policy. Finally, notably, the computational time for the causality-driven policy is only 8.57% of the exhaustive one.

Table 3. Numerical results for the case study

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Metrics | $c^{\*}$  | $c^{\infty }$  | $c\_{causal}^{v}$  | $c\_{exhaust}^{v}$  | $c\_{periodic}^{v}$  |
| $c$ a (unit: M$) | 0.235c | 2.102d | 2.482 | 2.091 | 3.684 |
| $d$ b | -  | -  | 830 | 9412 | 2581 |
| Computational time (unit: CPUs) | 3600 | 3600 | 657 | 7666 | 923 |

a: Cost. For $c^{\*}$ and $c^{\infty }$, $c$ denotes for the objective value of the static MILP solution. For $c\_{causal}^{v}$, $c\_{causal}^{v}$, and $c\_{periodic}^{v}$, $c$ is the cumulative cost. The lower $c$ is, the lower long-term cost of a rescheduling strategy can achieve. b: Cumulative value of difference-measuring function $∆(a\_{t},a\_{t+1})$. The lower $d$ is, less frequent schedule adjustment is needed in dynamic scheduling. c: MILP solver terminates because 3600s time limit is reached with relative MIP gap of 20.68%. d: MILP solver terminates because 3600s time limit is reached with relative MIP gap of 1.27%.

* 1. Conclusions

In this study, we propose a causality-driven rescheduling strategy for dynamic scheduling problems in multipurpose batch plants. The main idea originates from the human expertise that schedule modification in practice is often driven by causal relationships between tasks. To formalize such idea, we use directed Probabilistic Graphical Model (PGM) to serve as a standby knowledge base to help scheduler quickly infer the posterior distribution of impacted level of unobserved tasks when disruption events occur. The causality-driven strategy is tested on the benchmark problem (Kallrath, 2002). The numerical results show that, compared to the exhaustively-minimize-cost strategy, our strategy reduces 91.4% computational time and 91.2% count of task changes, at a sacrifice of 18.7% cumulative cost.

* 1. Acknowledgements

Taicheng Zheng and Dan Li appreciate financial support from UOM-CSC (202106440020, 201908130170). Jie Li appreciates financial support from Engineering and Physical Sciences Research Council (EP/T03145X/1, EP/V051008/1).

References

Gupta, D., Maravelias, C.T., 2019. On the design of online production scheduling algorithms. Computers & Chemical Engineering 129, 106517.

Harjunkoski, I., Maravelias, C.T., Bongers, P., Castro, P.M., Engell, S., Grossmann, I.E., Hooker, J., Méndez, C., Sand, G., Wassick, J., 2014. Scope for industrial applications of production scheduling models and solution methods. Computers & Chemical Engineering 62, 161–193.

Kallrath, J., 2002. Planning and scheduling in the process industry. OR Spectrum 24, 219–250.

Li, D., Rakovitis, N., Zheng, T., Pan, Y., Li, J., Kopanos, G., 2022. Novel Multiple Time-grid Continuous-time Mathematical Formulation for Short-term Scheduling of Multipurpose Batch Plants. Ind. Eng. Chem. Res. 61, 16093–16111.

Ouelhadj, D., Petrovic, S., 2009. A survey of dynamic scheduling in manufacturing systems. J Sched 12, 417–431.

Pinedo, M.L., 2016. Scheduling. Springer International Publishing, Cham.