A decomposition based on constrained clustering algorithms for job shop scheduling problems

Student:
Mohammed El-Kholany, MSc.
Alpen-Adria-Universität Klagenfurt, Austria

Supervisor:
Assoc. Prof. Konstantin Schekotihin

April 2021

1 Introduction

Nowadays, the massive progress and development of the internet and online technologies, data generated by machines and devices, product development, quality and inventory management systems, or production planning systems has become huge and is expected to increase in the coming years. Hence, to capture long-term revenues and sustainable competitive advantages, companies must manage the knowledge and have the valuable information to make the right decision at the right time in [1].

It can be assumed that knowledge management is a crucial issue in the industry. To extract implicit, unknown, and potentially useful information from data, we use data mining techniques that have been responsible for many of artificial intelligence's recent successes. Clustering is one of these techniques; clustering has always been an exploratory but critical task in the knowledge discovery process, with applications ranging from image processing, production systems, and information retrieval in [1]. Clustering is a technique that aims to group datasets (objects) into clusters by defining similar objects and aggregate them in the same cluster while the dissimilar objects should be belonging to different clusters.

Furthermore, from an optimization perspective, the main objective is to minimize the distance between objects falling in the same cluster and maximize the distance between the others that belong to different clusters. Since clustering does not use a subset of the dataset as labels to learn a classification model, clustering differs from classification in [2]. In other words, with the terminology of machine learning, clustering is a form of an unsupervised task, which calculates the similarity between data objects without knowing the proper distribution in [3]. Due to these unsupervised characteristics, clustering is known as one of the most challenging tasks in [1].

Clustering has been widely applied in several disciplines; one is production-planning systems, more specifically scheduling. Scheduling is one of the most complex problems in the industry. Scheduling of operations involves resource allocation over a period of time to perform a series of tasks, where one of the most critical problems is the Job-shop Scheduling Problem (JSP). The JSP has been known as a complex and combinatorial optimization problem, and it has been proven to be NP-hard in [4, 5]. Therefore, it is very difficult to reach

optimal solution in a reasonable time, even for small instances. However, from a practical perspective, the JSP scale is large, where the number of operations can be up to 10000 in some workshops in [6].

2 Project Objective

The main objective is to find the most appropriate sequence of operations to optimize the typical performance indicator (makespan). Several approaches have been studied to solve JSP, and one of the most effective approaches used is decomposition in [6, 7]. The main idea of the decomposition is to split the problem into a series of sub-problems (windows) and solve each of them separately in a sequence, and then obtain the solution of the original problem.

This project intends to apply the clustering technique to decompose the problem into sub-problems, where the number of clusters corresponds to the number of sub-problems (windows), and the operations correspond to data objects. For the JSP, An instance consists of a set of jobs to be processed on machines. The number of jobs is n, and the number of machines is m. Each job contains m operations with the operation precedence constraints. Each operation should be executed on machine m for a certain amount of time. Since the problem at hand has precedence constraints, it's required to apply a constrained clustering approach for the decomposition in [2].

In the context of partitioning algorithms, instance-level constraints are helpful to express a priori knowledge about which operations should or should not be grouped. Consequently, two types of pairwise constraints will be considered:

1. Must link constraints specify that two operations should be in the same window.

2. Cannot link constraints specify that two operations must not be on the same window.

It can be said that the attributes of the classical JSP are quite a few and not sufficient to define the similarity and dissimilarity between the operations. The lack of these attributes guides me to extract more features that could specify which operations are similar and should be in the same window and which not that must not be in the same window. The main contribution of this project

is to extract some features that have a significant impact on the decomposition process and accordingly on the makespan. More specifically, I need to extract some features to define which operations are similar to be scheduled on the same window and get a near-optimal solution in a reasonable time.

2.1 Feature extraction

Some features could be extracted:

1. The earliest starting time (EST) of an operation, the smaller the EST difference between the two operations, the more similarity.

2. Work Remaining (WR) of a job, an operation that belongs to a higher WR job, will have a higher probability of being in an earlier window.

3. Rank of operations sharing the same machine; the smaller difference the rank, the more similarity.

4. Distance between operations sharing the same machine (using EST); the smaller distance, the more similarity.

3 Dataset Selection

I select a set of benchmark instances generated in [8] as a dataset to work on. Many researchers have studied these instances and presented different approaches to solving them. I also developed different decomposition strategies to solve those. Therefore, it will be possible to measure the success/failure of my proposed model by comparing the results of the strategies developed in the past with those obtained from the proposed one.

References

[1] A. C. Benabdellah, A. Benghabrit, and I. Bouhaddou, "A survey of clustering algorithms for an industrial context," *Procedia computer science*, vol. 148, pp. 291–302, 2019.

[2] K. Wagstaff, C. Cardie, S. Rogers, S. Schroedl, *et al.*, "Constrained k-means clustering with background knowledge," in *Icml*, vol. 1, pp. 577–584, 2001.

[3] P. Li, O. Niggemann, and B. Hammer, "A geometric approach to clustering based anomaly detection for industrial applications," in *IECON 2018-44th Annual Conference of the IEEE Industrial Electronics Society*, pp. 5345–5352, IEEE, 2018.

[4] K. R. Baker, *Introduction to sequencing and scheduling*. John Wiley & Sons, 1974.

[5] J. K. Lenstra and A. R. Kan, "Computational complexity of discrete optimization problems," in *Annals of discrete mathematics*, vol. 4, pp. 121–140, Elsevier, 1979.

[6] R. Zhang and C. Wu, "A hybrid approach to large-scale job shop scheduling," *Applied intelligence*, vol. 32, no. 1, pp. 47–59, 2010.

[7] Y. Zhai, C. Liu, W. Chu, R. Guo, and C. Liu, "A decomposition heuristics based on multi-bottleneck machines for large-scale job shop scheduling problems," *Journal of Industrial Engineering and Management (JIEM)*, vol. 7, no. 5, pp. 1397–1414, 2014.

[8] E. Taillard, "Benchmarks for basic scheduling problems," *european journal of operational research*, vol. 64, no. 2, pp. 278–285, 1993.