

Final Project - report - Semester A 2022

Faculty of Engineering Sciences Department of Industrial Engineering & Management

Multi-Agent System for Scheduling Operating Rooms using Reinforcement Learning

2022-01-222

Barak Lavi

Advisor: Dr. Roie Zivan

8.1.2022

Content

[1. Introduction 3](#_Toc92028646)

[2. Literature Review 5](#_Toc92028647)

[2.1 Distributed Artificial Intelligence 5](#_Toc92028648)

[2.2 Multi-Agent Systems 5](#_Toc92028649)

[2.3 Distributed Constraint Optimization 5](#_Toc92028650)

[2.4 Search Algorithms 6](#_Toc92028651)

[Simulated annealing (SA) 6](#_Toc92028652)

[Distributed Stochastic Algorithm (DSA) 6](#_Toc92028653)

[Query Response Daily Schedule Algorithm (QADSA) 7](#_Toc92028654)

[2.5 Reinforcement Learning 7](#_Toc92028655)

[3. Planning 8](#_Toc92028656)

[3.1 Main Components of Procedure 8](#_Toc92028657)

[3.2 Cost Function 8](#_Toc92028658)

[3.3 Reward Function 8](#_Toc92028659)

[3.3 RL Implementation 8](#_Toc92028660)

[4. Implementation 8](#_Toc92028661)

[4.1 Surgery Schedule Problem SSP Formalization 8](#_Toc92028662)

[4.2 Reinforcement Learning application 9](#_Toc92028663)

[References 9](#_Toc92028664)

# 1. Introduction

In recent years there has been a technological advancement and research of intelligent systems. Those systems allow us to present problems from the real world with more accurate allocations. Algorithms and models of those systems can be used to represent and solve distributed combinatorial problems, one of them is surgery operating room scheduling in large hospitals.

Hospitals operate a significant number of surgeries every day, operations over a single day are performed by multiple doctors, belonging to multiple wards performed surgeries on patients that need to be prepared for the specific operation they signed on. To improve cost and procedure efficiency, while maintaining a quality level of service. This scheduling problem is highly challenging due to numerous constraints and variables between all the different participants on hand.

This project will work with a model of a distributed multi-agent system and will use optimization and constraint satisfaction algorithms. The relevant agents for the problem are the head of wards, head nurses, anesthetists, and surgical equipment coordinators. These agents need to cooperate and communicate to achieve a valid and acceptable operations schedule. Agents interact and share their actions to achieve a global goal on the one hand and their own local goal on the other. The global goal relates to cost-effectiveness and patient care. The agent’s local goals are their personal constraints, each agent has its own constraints to optimize its local allocation for higher stability and better prioritization that can support frequent dynamic changes.

The Surgery Schedule Problem (SSP) for a single day of operations was modeled by Noam Gaon in her thesis (Gaon, 2021) as a multi-agent distributed model. The model takes the surgeries’ requests for the operation room, and the coordination agents (anesthetists, nurses, required equipment). All of these are shared resources among all the surgical wards. Agents representing each group of resources are called constraint element (CE), need to decide on an allocation of the operation rooms schedule under the effort of satisfying constraints of each ward agent (WR) while optimizing the global goal of the problem.

This project will use the two most efficient algorithms that were modeled, distributed stochastic algorithm (DSA) and query response daily schedule algorithm (QRDSA). In contrast to standard DSA, the graph's structure of the different agents in the SSP model is bipartite, therefore, agents send their assignment only to agents of the other type, i.e., WRs to CEs and vice versa.

The algorithms use a stability factor (SF) that penalizes a change in the assignment of an operation, i.e., whenever an operation that was fully scheduled got changed, there was a reduction in the agent’s utility.

Furthermore, both algorithms used Single variable Change with Exploration methods (sc\_e). A random variable is selected. The agent tries to replace the current variable's assignment with an alternative value to improve its utility. If the variable change does not improve, the variable's initial value is returned, and the following random variable is chosen until a stop condition is met.

To assign every surgery request in its best allocation, the WRs evaluate its constraints for each surgery request with a cost function. Using the surgery request and patient features, the agent can estimate which surgery to assign and when to increase its utility. Those cost functions weight each feature’s importance and effect. For example, a surgery request with a big number of previous cancellations is more likely to affect the agent's decision of early assignment.

In the current SSP model, the weights of the feature in the cost functions are determined in advance as an assumption of validation, without an expert’s help or any learning algorithms. Therefore, the actions each agent takes for the best assignment are not with the best features weights allocation. Furthermore, the SF penalty value is randomly assigned as well and can affect the agent’s actions.

For the agents to learn what is their best cost function allocation to apply, this project will implement a Reinforcement Learning (RL) algorithm, a machine learning method of rewarding and punishing actions of the agents to achieve an optimal solution.

In this project, using RL, the agent actions will explore different scaling for each feature’s weight to achieve optimal cost functions (policy) of surgery assignment in the daily schedule. RL require a Reward function to evaluate each schedule given from the SSP DCOP solution.

The project will define reward function and redefine its cost function with the use of Composite Desirability (CD), an optimization tool used to evaluate environments and performance of material production, similar complex problems with parameter combination who affect each other, with the goal of determine the optimal value of parameters that can provide superior results in terms of performance. The advantage of CD, it is easy to understand and incorporate individual weights for different features.

The project develops over time, therefore it is probable for more methods and applications to be applied going forward.

# 2. Literature Review

## 2.1 Distributed Artificial Intelligence

Interaction between multiple entities can be noticed in many realistic problems. These problems are usually distributed by nature. An interaction is defined by how an entity is influenced from the decisions and actions of other entities (Bond & Gasser, 2014). In Distributed Artificial Intelligence (DAI) an entity will be referred to as an agent. An agent can be a physical or virtual entity that can act, perceive its environment, and communicate with other agents (Ferber & Weiss, 1999; Russell & Norvig, 2016). A major focus in DAI is the coordination between multiple intelligent agents. This coordination is described by the interaction between behaviors, knowledge, goals, skills and programs of the agents. (Bond & Gasser, 2014)

## 2.2 Multi-Agent Systems

Multi-agent system (MAS) contains an environment, objects, and agents (the agents being the only ones to act), relations between all the entities, a set of operations that can be performed by the entities and the changes of the universe in time and due to these actions (Ferber & Weiss, 1999). From the point of view of the agent in the system, the environment is dynamic and changes according to the activity of other elements in the system. Systems in which several agents make use of interaction to maximize utility and jointly solve tasks are called Cooperative MAS (Bond & Gasser, 2014; Stone, 2007)

## 2.3 Distributed Constraint Optimization

Distributed Constraint Optimization Problems (DCOPs) is a standard structure used to characterize Distributed Combinatorial Optimization Problems. These problems usually involve many interdependent agents and have a broad range of application in Multi – Agent Systems (Netzer, Meisels, & Grubshtein, 2010)

The following description of a DCOP is consistent with the definitions in many DCOP studies e.g., (Modi, Shen, Tambe, & Yokoo, 2005). A DCOP is a tuple . A is a finite set of agents {A1, A2, …, An}. X is a finite set of variables {X1, X2, …, Xm}. Each variable is held by a single agent. D is a set of domains {D1, D2, …, Dm}. Each domain Di contains the finite set of values that can be assigned to variable Xi. An assignment of value d∈ Di to Xi is denoted by an ordered pair 〈Xi, d 〉. R is a set of relations (constraints). Each constraint C ∈ R defines a nonnegative cost for every possible value combination of a set of variables and is of the form C ∶ Di1× ­Di2 × … × Dik → ℝ+ ∪ {0}. A binary constraint refers to exactly two variables and is of the form Cij ∶ Di× Dj → ℝ+ ∪ {0}. A binary DCOP is a DCOP in which all constraints are binary. A partial assignment (PA) is a set of value assignments to variables, in which each variable appears at most once. Vars (PA) is the set of all variables that appear in PA. A constraint C ∈ ℛ of the form C∶ Di1× Di2­ × … × Dik→ ℝ+ ∪ {0} is applicable to PA if Xi1, Xi2 ­, …, Xik ∈vars (PA). The cost of a PA is the sum of all applicable constraints to PA over the assignments in PA. A complete assignment (or a solution) is a partial assignment that includes all the DCOP’s variables (vars (PA) = X). An optimal solution is a complete assignment with minimal cost.

## 2.4 Search Algorithms

Local Search algorithms run using a single node and generally progress only through the neighbors of this node. These algorithms are not systematic. The two major advantages of LSA are that they use very small amount of memory, and they typically find sufficient solutions in large or infinite state spaces. Additionally, LSA are valuable for solving optimization problems, in which the intention is to find the best state corresponding to an objective function. An optimal algorithm always finds a global minimum/maximum; a complete LSA always finds a goal if one exists (Russell & Norvig, 2016).

### Simulated annealing (SA)

Simulated annealing solves combinatorial optimization problems by an analogy to statistical mechanics. States in thermodynamic correspond to solutions in the combinatorial optimization problem. The cost function in simulated annealing is the energy in thermodynamics. The ground state, change of state, and temperature in thermodynamics transform to the optimal solution, a neighboring solution, and the control parameter in SA, correspondingly. (Kirkpatrick, Gelatt, & Vecchi, 1983)

In each step in the algorithm, SA selects a random move. If the move improves the current state, it is accepted and changed. If not, the algorithm takes the move with a probability calculated as the energy in thermodynamics. The probability decreases exponentially with the number of iterations, it is more likely to perform a bad change for exploration in the begging of the algorithm then the end.

### Distributed Stochastic Algorithm (DSA)

The DSA algorithm is uniform, all procedures are the same and have no identities to differentiate one another. All procedures advance in synchronized steps, in each step it sends and receives messages from its neighbor agents and then operates local computations, e.g., changing local state if necessary (Tel, 2000).

For a given number of iterations. In each iteration, the agent chooses a random allocation value and sends it to its neighbors. Then, if not done, receive information from neighbors, calculate best alternative allocation, if new allocation improves the agent solution, change allocation with some probability (usually 0.7) and send to neighbors. Only if the agent finds a new value that improves or maintains its current state quality, it may change to it based on a stochastic scheme (Fitzpatrick & Meertens, 2001).

### Query Response Daily Schedule Algorithm (QADSA)

QRDSA main difference from standard DSA is the query response structure. WR agents start with selecting assignment for their local problem using SA. Then, a WR agent sends its selected schedule to its CE neighbors and waits for their response. Once these responses are received, it updates its local information and revises its local assignment before sending it again. On the other hand, the CE agent waits for the assignments of its WR neighbors to arrive before it performs its computation. It updates its local operation schedule and proposes its corresponding assignment of constrained elements to this schedule. Each of its neighboring WR agents sends the projection of its assignment on the schedule relevant to the neighbor (Gaon, 2021)

## 2.5 Reinforcement Learning

Reinforcement learning is a machine learning (RL) method concerning how an agent decides which action to make in a current state based on its environment to maximize rewards and minimize penalty. The goal of RL is for the agent to find balance between exploration of its environment and exploitation of its knowledge. RL discover which actions yield the most reward by trial and error, with the purpose of effect not only the immediate reward but also learn what will be the best actions in the future (Sutton & Barto, 2018). In RL problems, agent learns a policy (i.e., a complete mapping between situations and actions) by trying actions out without any domain expert has told it, as in many other forms of machine learning (Naeem, Rizvi, & Coronato, 2020).

## 2.6 Composite Desirability (CD)

CD mostly use in material production. The statistical desirability tool can be used to compare the environmental performance of different material based on the features considered by various proposed indicators. This statistical function allows converting multiple response variables into a single value to be considered by the end-user in decision-making processes. Another advantage is that methods based on desirability are easy to understand and flexible to incorporate individual weights for different features (de Souza et al., 2021).

Initially, all the experimental values of individual output response are normalized by characteristics using Eq. (1). The normalized value of each output characteristic is termed as the desirability index (di). Then, Composite Desirability (CD) is calculated using Eq. (2) (Sharma, Rana, Singh, Singh, & Chattopadhyay, 2021). W is the weight assigned to each feature, eq2 represent equal weights for all features while eq3 takes different weights.

תמונה שמכילה טקסט

התיאור נוצר באופן אוטומטי

## תמונה שמכילה טקסט התיאור נוצר באופן אוטומטי

Eq3

The desirability function allows the users to choose the most appropriate w values, based on the relative importance or priority assigned to each response variable. Based on the application of weights between (0.1 ,1 or 10) (de Souza et al., 2021).

# 3. Planning

The main components concerning this project are the cost function (policy), reward function and the implementation of the RL algorithm. With the goal of the procedure is to provide a quality schedule while maximize the utility function.

## 3.1 Cost Function

In the current SSP model, the cost function is defined as follows:

לכתוב את הפונקציה מהתזה של נועם, כולל כולם

The functions examine the current parameters concerning only the max value of the various features in the current ward surgeries requests. For that the motivation to apply CD function to consider the minimum values as well.

Furthermore, order\_surgery is influence by the duration of the surgery, there is not sufficient reason to take this feature under consideration, thus, the feature will be removed.

אחרי שנסביר את המשתנים שאני רוצה לשנות, נציג את הפונקצייה החדשה

## 3.2 Reward Function

To estimate each action of the RL agent, reward function was created with CD to evaluate the schedule desirability. The function examines the current schedule state with its schedule surgeries concerning the unscheduled surgeries left to assign for another day, with the reduction of no good (ng) surgeries requests who unable to allocate based on the SSP DCOP approach.

For example, operation with high urgency who didn’t get schedule while there is a lower urgency surgery that got assign is not the best allocation for this feature. Maximize CD of the relevant features representing a good schedule and quality of service will be the RL agent reward function.

The features concerning a good schedule are choose with knowledge-base and therefor might change.

* Urgency
* Cancellations
* Deadline duo

Reward function here

לכתוב שבפונקציה הזו אין משקלים ועדיפויות

## 3.3 RL Agent

הסוכן משתמש בפונקציית הפרס ומחליט על אקשן...

The RL agent action will change the weights of the different features. In each iteration the agent will change one variable of weight and will learn check the reward function change in the next state.

מרגיש שיש עוד דברים להוסיף, לעשות תנאים על הדלתא של הפונקצייה, אם גדול מאפס זה טוב ולהמשיך להגדיל בעוד הפוך אם לא, אם אין שינוי או שינוי קטן אפסילון אז לעצור

יש להגדיר את העתק הקפיצות של האלגוריתם ואת מצב ההתחלה

# 4. Implementation

## 4.1 Surgery Schedule Problem SSP Formalization

להעתיק מהדוח הסופי של נועם של תואר ראשון

## 4.2 Reinforcement Learning application

לכתוב פסאדו קוד של האלגוריתם

References

Bond, A. H., & Gasser, L. (2014). *Readings in distributed artificial intelligence* Morgan Kaufmann.

de Souza, A. M., de Lima, Gustavo Emilio Soares, Nalon, G. H., Lopes, M. M. S., de Oliveira Júnior, André Luís, Lopes, G. J. R., . . . de Carvalho, José Maria Franco. (2021). Application of the desirability function for the development of new composite eco-efficiency indicators for concrete. *Journal of Building Engineering, 40*, 102374.

Ferber, J., & Weiss, G. (1999). *Multi-agent systems: An introduction to distributed artificial intelligence* Addison-Wesley Reading.

Fitzpatrick, S., & Meertens, L. (2001). An experimental assessment of a stochastic, anytime, decentralized, soft colourer for sparse graphs. Paper presented at the *International Symposium on Stochastic Algorithms,* 49-64.

Gaon, N. (2021). Multi-agent system for scheduling   
operating rooms .

Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science, 220*(4598), 671-680.

Modi, P. J., Shen, W., Tambe, M., & Yokoo, M. (2005). ADOPT: Asynchronous distributed constraint optimization with quality guarantees. *Artificial Intelligence, 161*(1-2), 149-180.

Naeem, M., Rizvi, S. T. H., & Coronato, A. (2020). A gentle introduction to reinforcement learning and its application in different fields. *IEEE Access,*

Netzer, A., Meisels, A., & Grubshtein, A. (2010). Concurrent forward bounding for DCOPs. Paper presented at the *Proceedings of the Twelfth International Workshop on Distributed Constraint Reasoning (DCR’10),* 65-79.

Russell, S. J., & Norvig, P. (2016). Artificial intelligence: A modern approach. malaysia.

Sharma, V. K., Rana, M., Singh, T., Singh, A. K., & Chattopadhyay, K. (2021). Multi-response optimization of process parameters using desirability function analysis during machining of EN31 steel under different machining environments. *Materials Today: Proceedings, 44*, 3121-3126.

Stone, P. (2007). Learning and multiagent reasoning for autonomous agents. Paper presented at the *Ijcai, , 7* 13-30.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* MIT press.

Tel, G. (2000). *Introduction to distributed algorithms* Cambridge university press.

stylefix