The Importance of Information System for Optimization with Genetic Algorithms

IGOR BERNIK, ROBERT LESKOVAR, MOJCA BERNIK, MIROLJUB KLJAJIĆ
University of Maribor, Faculty of Organizational Sciences
Kidričeva cesta 55a, p.p. 165, 4000 Kranj
SLOVENIA

Abstract: Modern and competitive company collects information about company functioning and market demands in integrated information system (IS). At the decision making stage, management more and more use optimization and/or simulation methods, where the role of data from IS is very important. Reliable input data are needed for optimization and simulation, because they provide appropriate base for development and running optimization and/or simulation model. IS should also provide transfer of optimized data back from the optimization/simulation model to the IS. The purpose of integral IS and use of simulation and genetic algorithms optimization method is therefore for easier, quicker and better decision making. Modern approach to connecting corporate IS and method of optimization and simulation, necessary interfaces for data transfer between optimization simulation system and IS and time complexity of data optimization and simulation for management decision making will be described.

Key-words: Optimization, Genetic Algorithms, Simulation, Information System, Production Planning

1 Introduction

Modern day companies are encountering the optimal resource scheduling problem increasingly frequently [1]. Apart from the quantity and quality of products a very important category is the production costs, therefore companies seek ways to find a better net profit towards gross production or service costs ratio while struggling with the market's unstable conditions and a diversifying line of products. In such conditions the use of information systems (IS) coupled with modern business support methods (optimizing, simulation, expert systems, data mining systems, …) is a key to success and gaining strategic advantages over the competition. A common trend is to connect all such systems that aid quality business to build an integrated IS supporting the decision makers. When upgrading these systems with decision support systems (DSS) and artificial intelligence methods, the desire is to integrate them on a higher level – to establish an executive IS.

Optimizing and simulation are thus only a part of the entire system that needs to be coupled via the IS to form an overview of the business system. To use the optimizing and simulation methods successfully a company needs a quality IS and direct access to the corporate database – the basis of the integrated IS. The diagram in Fig. 1 shows the main data and information flows necessary for the successful implementation of a optimizing and simulation based scheduling system serving as a decision aid [2].

The market demands for certain products and services trigger a line of processes in the company that internally represent the market demands and are stored in the corporate database. According to the demands and requirements of the market and the amount of products in storage a set of possible schedules that satisfy the criteria needs to be prepared. The most suitable schedule is then accepted as the valid production or service schedule and becomes operational. After the schedule preparation the solutions, i.e. possible schedules, are tested in a simulation model of the system that is serving as a decision aid. An optimized schedule that is selected after the simulation test is entered into the corporate database via the IS and is later executed, unless unexpected changes in the market (and other) conditions appear. Thus in a case of an urgent order the process of scheduling and visual simulation is repeated. The newly selected schedule first meets the urgent order's requirements and moves the less urgent jobs to a later time.

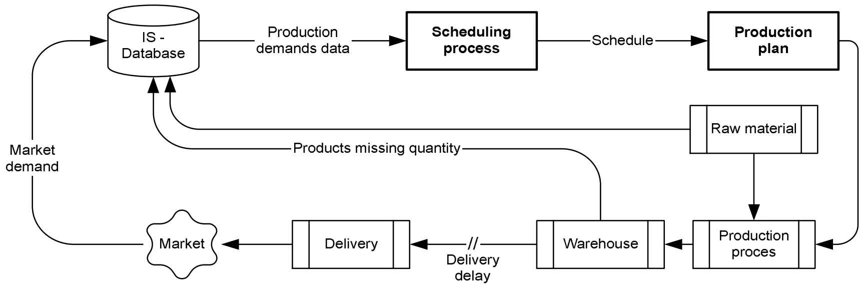


Fig. 1: Block diagram of the links between IS, optimizing and the real-world system

In the process of schedule optimization there are several interconnected sub processes that at the end of optimization process produce possible job schedules (the possible and high ranking scenarios) that most favorably meet the criteria derived from the IS data input.

The input data represent the difference between the market demand and the products in storage. In this the IS plays a major part since only a fast and efficient access and updating of centralized database provides fresh information that can serve as a basis for the optimizing and simulation tasks. We need to realize that no optimization method can work when the data are not suitable (garbage in – garbage out).

The data for the schedule optimization and system simulation are based on the data stored in the IS. The optimization process shown in this paper is executed with the use of genetic algorithms (GA), and the schedule verification is done with simulation methods. The bigger the number of parameters in the simulation model, the bigger the possibility of finding a suitable solution with the use of GA. Among the general search strategies the GA are considered to have a good search quality to time consumption ratio. After the optimization the simulation enables us to observe the systems behavior for the selected schedule. The system enables us to analyze various situations that are a basis for a consistent business actions designation. The process of testing or verification of different schedules in the simulation models is repeated until a proper solution is formed. The simulation results are evaluated according to the given utility function. We shall describe the IS related demands of the optimization and simulation methods, the basics of GA based optimization and a short description of the simulation execution in the solution search. The results will be shown in relation to various optimization parameters.

2 Information system

In the beginning of the construction of simulation model and the scheduling application the system structure data are integrated in the model. Thus it is important that the model is built in a way that enables us to alter and update the data. Some organizations have precise the information about their processes, such as machine failure statistics or servicing time. If such data is unavailable during the model construction phase it is necessary either to:

- obtain the help of experts acquainted with the system being modeled,
- gather the necessary information internally, or
- carefully define assumptions about the system structure.

Regardless of the input data quality, one should always also perform the model sensitivity analysis. With that procedure we can determine the magnitude of various hypotheses' influence on the simulated systems behavior. The sensitivity analysis can sometimes prove that a certain hypothesis has a very strong influence on the results and it is thus wise to gather additional information on the system and its functioning.

The data necessary for the preparation of scheduling simulations and the operative production plan are usually collected at the company's information system. As many companies, especially smaller ones, don't possess an appropriately sophisticated IS, the IS needs to be upgraded to support effective execution of optimizations and simulations. One should also consider the transfer of data between the IS and the simulation models. Two examples of the modern methods of data transfer are the ODBC (Open DataBase Connectivity) and JDBC (Java DataBase Connectivity).

Open database connectivity and Java database connectivity are established as standard interfaces for the applications that require access to relational and non-relational data management systems. By using ODBC and JDBC one can access optimization and simulation data stored in various databases on different platforms running on any of the common operating systems [e.g. 3].

ODBC and JDBC enable applications to access a large set of data sources without the knowledge of the data format or structure in the source database. ODBC and JDBC are able to discern a certain amount of information on any given data source. In most cases the two offer the user application excellent access to the data sources. An application is also able to share its data with other data access applications. As the ODBC and JDBC functions handle the details of accessing various data sources, the user applications data handling can be largely independent of the used data source.

The data in the IS that are used for optimization and simulation should be valid if we wish to perform good quality operative scheduling. If the IS data is not updated according to the market demand dynamics (as shown in Fig. 1), the updates are delayed, necessitating repeated optimization. Due to time complexity this decreases the quality of schedules. In the case described later on, a procedure of optimization and simulation is shown, which is using data from an IS that is appropriately handling the data and transactions in the integration of the production system and the optimizing and simulation system.

3 Optimization and Simulation

In the example of GA based scheduling we are analyzing a job shop type production with limited resources. Several types of products can be handled by the production but each line or station can process only one type of a product, and that poses a scheduling problem. When the basic schedule is complete, it is accepted only as a working schedule that may be altered in case of changing demands. In case of urgent orders that arrive after scheduling has been completed, some jobs are assigned a higher priority, and in case of order cancellation or equipment failures the scheduling has to be redone and a new working schedule produced [2]. The scheduling problem is a complex task due to resource limitations, time limitations, machine setup times, priorities and system history.

The complete schedule is composed of a finite number of feasible non-overlapping jobs ordered in a changeable sequence. The sequence may change at any time by adding new jobs until the moment it is being executed in the production. The presented type of scheduling can be understood as a combinatorial, meta-heuristic problem as essentially the mission is to sort a set of jobs into a sequence that satisfies the given criteria. The task is a complex one since not all possible schedules (sequences) are acceptable as some violate system limitations.

To prevent a conflicting schedule from appearing we can use an algorithm that employs restrictions. The scheduling algorithm selects the first job on the list and schedules it accordingly to its priority. Then the algorithm selects the next job and schedules it according to the limitations and the existing schedule. This process is then repeated for all jobs and leads to the creation of a feasible schedule that satisfies all the criteria for all jobs in the list. When the basic schedule is completed, it needs to be optimized, and that is done with the use of GA. When the optimization is finished, the schedule is verified on the simulation model and made operational.

3.1 Optimization with Genetic Algorithms

The scheduling algorithm builds the starting population of schedules using the order data in the IS. Optimization is done with the use of GA. The GA combine the survival of the fittest in a set of individuals with a structured, yet random information exchanges in the form of a search algorithm. The algorithm first randomly places the jobs to form a first generation schedule. The population is made up of a finite and predetermined quantity of individuals (schedules) that form a new generation with every iteration. With the help of the simulation model that provides the utility function, the fitness of every individual is evaluated. The selection process makes sure that only a part of the population “survives” the evaluation and serves as the basis for the evolution of new individuals. The evolution is implemented via the genetic operators that transform a population into a next generation population. The evolution is repeated until the generated schedules satisfy the criteria function. The last generation of schedules is simulated on the visual simulation model and thus the most suitable or an optimal schedule is selected for the current data in the examined production process.

To construct an optimization program that operates on the basis of GA we need to define four components that are necessary for the GA operation: chromosome syntax, chromosome interpretation and evaluation and the chromosome operators. The necessary data is transferred via the ODBC or JDBC interface from the IS database. We also need to define other parts of the GA application such as parent selection; however that part is usually less relevant and depends on the problem being solved. A more detailed description of the GA is available in [2].

We should also decide on the number of population evaluations. In the case of a large population we can execute all evaluations in a single run or divide the population into several subpopulations and execute a series of evaluations. Which method is better can be determined by experimentation as it is highly dependant on the problem. Also, the criteria function should be able to determine the quality of a schedule, but unfortunately that task is not an easy one, as there may be conflicting criteria and criteria with relative values.

The scheduling results for different genetic operators and selection types, and number of generations and population size, are shown here on a real-world example [4]. In the scheduling phase the goal is to find the shortest job execution time from production start to storage with the help of GA and obeying all limitations. The quality of optimization depends on the input data and the choice of GA parameters. The parameter selection method is described later. As in the start of the GA optimization a random population is generated, all the shown results are an average of thirty iterations.

The population size influences the operation of GA in two aspects. The first aspect is the number of individuals in each generation that provides material for the next generations. The second aspect is that with the same number of generations but larger populations we can achieve better results. Fig 2 shows the population size influence on solution convergence for twenty generations.

It can be easily seen that the result quality improves with population size. Of course that may increase search time which may be limited. We can conclude from the results of population size influence tests that a population size of sixty individuals is appropriate for the studied example. Should we want to decrease the time of search we may execute the evolution with a smaller number of generations, since the quality of the solution does not change significantly after the first ten generations, as can be seen in the Fig. 2. Also, good solutions appear as early as in the fifth and sixth generations.

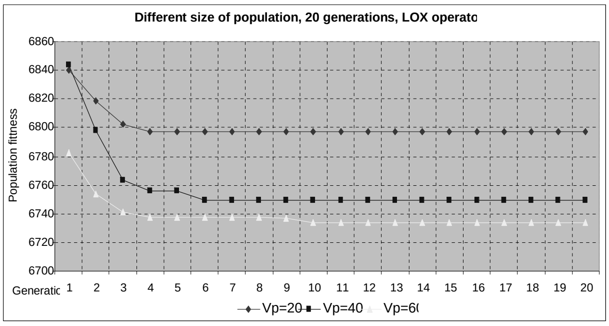


Fig. 2: Population size influence on convergence

The method of selecting individuals from the current population and placing them in the next generation largely depends on the available data; however most of the GA research uses the elitist selection type [e.g. 5, 6]. The elitist selection type provides an ample diversity of individuals and a relatively fast convergence towards the optimal solution. A more important factor is the selection of the genetic operator. Its choice depends largely on the number of possible solutions in the studied problem.

We wish to discover the genetic operator, that provides a reliable convergence towards a final and satisfactory solution while assuring the evaluation of a wide set of solutions. We demonstrate the influence of the genetic operator measured in a population of sixty individuals, as the experiments with different sizes of the populations have showed that an evolution of twenty generations can be executed in an acceptable amount of time, and the solutions found are satisfactory and suitable for practical use.

Three crossover operators and one mutation operator have been tested. Permutation Crossover (PermutX), executes permutations after the crossing of two individuals, Local Order Crossover LOX) executes local gene ordering, while the Partially Matched Crossover (PMX) does partial matching. The mutation can be done in different ways, and for the scheduling problem the insertion mutation (MutVrine) is a very suitable choice.

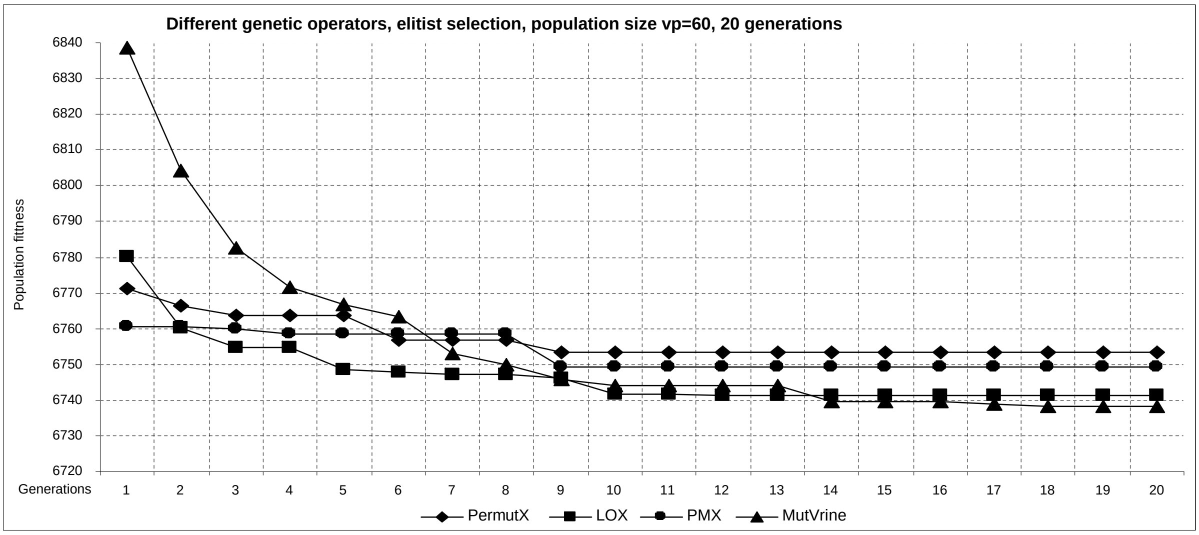
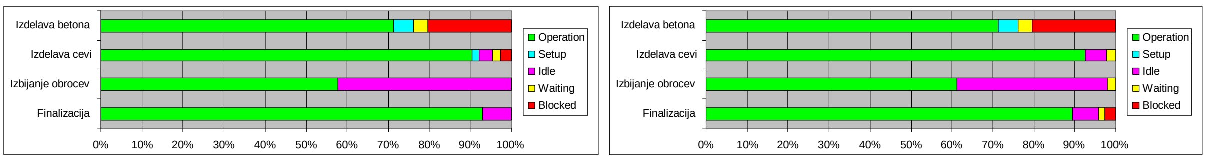


Fig. 3: Influence of genetic operator type on solution convergence in elitist selection

The results of all tested crossover operators approach a satisfactory solution within the first few generations. In the fifth or sixth generation the mutation operator reaches a similar quality solution. Therefore the most suitable operators are the LOX operator and the mutation operator. For the implementation of a scheduling system we should decide which type of operator to use in the application.

Based on previous results we may conclude that it is sensible to select an operator that proves suitable for our studied system. In our case that is LOX operator, which achieves very good results in less than ten generations. After the genetic operator and evolution parameter selection we can build the scheduling system that simplifies the man-machine selection and enables the operator to use simple interaction and experimentation. After the scheduling is complete the operator can use their experience and knowledge of the production system to make the obtained optimal schedules operational.

After the GA optimization the application finds the most suitable schedules with shortest production time, based on the input data. In the last, twentieth population obtained with the LOX genetic operator and starting with a population of sixty individuals there are five schedules that are suitable for use. The time necessary for optimization until the final solution is found is 56 seconds. The selected schedules were verified on a visual simulation model [6] and the system behaviour was observed. There are no noticeable results in the animation of each schedule, therefore we compared the statistical results on each schedule. Fig. 4 shows the statistical results for the best 2 schedules.



1 2

Fia. 4: Work station utilization for best 2 schedule

From the results shown we can quickly discern that there are no important differences in system operation for the two schedules. Therefore the system should function well regardless of the schedule selected. Let's assume the operator selects schedule 2. That schedule is now the operative schedule. Its execution is started immediately and continues until it's complete. However if major changes in the system appear during execution, such as failures, hold-ups or urgent orders, a certain part of the schedule is completed while we are preparing a new schedule based on the most current information - that is the new operational schedule, and subject to change.

4 Conclusion

In this paper we describe the importance of an integrated IS for high quality optimization and simulation in a company that wishes to achieve or preserve a strategic advantage over its competitors. An integrated IS represents a basis for the connection between the data in the corporate database and the methods of optimization and simulation. The access to the IS data should be open and enable the connection between data and applications with modern interfaces such as ODBC and JDBC. Optimization is shown on a GA based scheduling example. Simulation is used to evaluate the quality of each schedule after the optimization, and the operator selects the most suitable schedule to be made operational.

The connection between GA methods for scheduling and the simulation has proven to be efficient and successful. The experience obtained will accelerate further development of the integrated management's information system and provide strategic advantages to companies. In this way new possibilities are open in the application of GA in intelligent decision system and control systems where the system complexity and related evolution time is of essential importance. The time complexity of scheduling in the presented system is negligible compared to the time a skilled and educated planner may need to prepare a schedule without such support. The experts in the company where the studied process is executed have confirmed that the presented scheduling system has improved the flow of products in the system, and reduced management and production costs.

The system enables scheduling and provides aid for decision making in real systems. It is an efficient tool aiding quick business decisions in strategic and operative planning and in process control, observing various criteria and business scenarios of environment's future influence on the system. The system of schedule optimization using genetic algorithms and simulation enables the studying of the systems behaviour as a whole and the interaction with participants in the decision making process, improvements in the task or resource assignment process, improvements in the planning and decision making processes and the training of professional for the planning, scheduling and management tasks.

References:

[1] Silver, E. A., Pyke, D. F.: Inventory management and production planning and scheduling, John Wiley& Sons, 1998

[2] Kljajić, M., Bernik, I., Breskvar, U.: Production planning using simulation and genetic algorithms in multi-criteria scheduling optimization, editors: Florjančič, J., Paape, B.; Organisation and management: selected topics. Frankfurt am Main P. Lang, 2003, str. 193-208.

[3] Papler, A., Leskovar, R., Nemec, A., Rejec, V.: Interactive Multicriteria Production Scheduling, Proceedings of the 6th International Symposium on Operational Research, str. 153-158, 2001

[4] Kljajić, M., Bernik, I., Škraba, A.: Simulation Approach to Decision Assessment in Enterprises, Simulation 75:4, pp. 199-210, 2000

[5] Syswerda G.: Uniform Crossover in Genetic Algorithms, Proceedings of the 3rd International Conference on Genetic Algorithms, 1989

[6] Syswerda G.: Schedule Optimization Using Genetic Algorithms, editor: Davis L.: Handbook of Genetic Algorithms, Van Nostrand Reinhold,1991

Acknowledgment:

This research was supported by the Republic of Slovenia, Ministry of Education, Science and Sport, project no.: L5-5197.