INTEGRATED DELIVERY PLANNING AND SCHEDULING BUILT ON CLUSTER ANALYSIS AND SIMULATION OPTIMISATION

Galina Merkuryeva
Department of Modelling and Simulation
Riga Technical University
1 Kalku Street, Riga LV-1658, Latvia
E-mail: galina.merkurjeva@rtu.lv

KEYWORDS
Tactical planning, vehicle scheduling, cluster analysis, simulation, optimisation.

ABSTRACT

Integrated solutions for product delivery planning and scheduling in distribution centres are proposed and built on a cluster analysis and simulation optimisation methodology. A cluster analysis of product demand data of stores is used to identify typical dynamic demand patterns and product delivery tactical plans. Further, simulation optimisation techniques are applied to find optimal parameters of product transportation and vehicle delivery schedules. In the paper, a cluster analysis of the demand data by using the K-means clustering algorithm and silhouette plots mean values is performed, and an NBTree-based classification model is built. In order to define optimal parameters of vehicle schedules, a genetic algorithm is applied and interacts with a discrete-event simulation transportation model built in AnyLogic simulation environment. Integrated solutions are illustrated and adjusted to a specific business case.

INTRODUCTION

Product delivery planning and scheduling is a high commercial priority task in transport logistics. In real-life applications the problem has different stochastic performance criteria and conditions. Optimisation of transportation schedules itself is computationally time-consuming task which is based on the data from tactical planning of weekly deliveries. This research focuses on the methodology that will allow reducing the affect of the demand variation on the product delivery planning and scheduling, and avoid numerous time-consuming planning adjustments and high computational costs.

In the distribution centres (DC), this problem is related to deliveries of various types of goods to a net of stores, in predefined time windows, taking into account transportation costs and product demand variability. The problem has also a high number of decision variables, which complicates the problem solution process. Heuristic methods and commercial software that are usually applied could lead to non-effective solutions, high computational costs and high time consumption. In practice, product demand from stores is variable and not deterministic. As a result, the product delivery tactical plan that is further used for vehicle routing and scheduling has to be adjusted to real demand data, and product delivery re-planning supervised by a planner is often required. This task is very time-consuming and requires specific knowledge and experience of planning staff in this domain. Moreover, in practice a cluster analysis of the product demand data and potential tactical plans is not performed. But the most suitable delivery plan could be defined as a result of such an analysis that would ensure high quality solutions to schedule optimisation problem and reduce computational costs of the problem solution.

The paper presents an integrated approach to product delivery planning and scheduling built on a cluster analysis and simulation optimisation techniques. In the paper, a cluster analysis is performed by using the K-means clustering algorithm. To define an appropriate number of clusters, silhouettes plots are built and their mean values are estimated. As far as the demand is dynamic and variable, a classification model that assigns an appropriate demand cluster is presented by an NBTree, which induces a hybrid of decision-tree and Naive-Bayes classifiers. In order to find an on optimal grouping of stores into regions based on their geographical locations and aimed to leverage the total product demand over regions, a multiobjective optimisation algorithm NSGA-II is used in (Merkuryeva et al, 2011). In simulation optimisation of vehicle schedules, a genetic algorithm (GA) is designed to search for the best combination of schedule parameters. GA interacts with an AnyLogic-based simulation model which is used to simulate product transportation schedules and estimate their fitness values. All algorithms are applied to a specific business case.

INTEGRATED APPROACH

The methodology for integrated delivery planning and scheduling is aimed at selecting an effective product delivery tactical plan for the upcoming week and optimising product transportation routes and delivery schedules. The proposed methodology integrates a cluster analysis that defines typical product dynamic demand patterns, identifies an appropriate demand

Proceedings 26th European Conference on Modelling and Simulation ©ECMS Klaus G. Troitzsch, Michael Möhring, Ulf Lotzmann (Editors)
ISBN: 978-0-9564944-4-3 / ISBN: 978-0-9564944-5-0 (CD)

cluster and tactical weekly delivery plan; and simulation optimisation to optimise vehicle delivery schedules.

Moreover, vehicle routing and scheduling optimisation is based on the data from tactical planning for a week delivery. At the same time, a weekly delivery plan itself is dependent on the data about the number of goods to be delivered to stores in a particular day of a specific week and stores' geographical allocation. In practice, historical data of store demands shows that often the real demand can be very different from expected or average one, which is determined in the predefined or base plan. Thus, significant changes should be made in the base delivery plan for each new week. It seems reasonable to specify typical patterns of dynamic daily demand for different planning weeks and introduce several base plans each representing an appropriate product delivery time table for a specific demand pattern. This will reduce the work of adjusting a typical or base delivery plan to the current situation. Since there are now more typical delivery plans that are based on typical demand patterns, the work will be reduced to making a decision, which delivery plan should be used for the next week and small adjustments of it still may be required. In addition, selecting the most suitable delivery plan may ensure better scheduling solutions and reduce their computational costs.

The proposed scheme for an integrated solution (Fig. 1) includes the following tasks (Merkuryeva et al, 2011):

- Definition of typical dynamic demand patterns by clustering historical daily demand data available for different planning weeks;
- Grouping of stores based on their geographical locations to leverage the total product demand over regions.
- Tactical weekly delivery planning performed for each group of stores and each pattern demand by using combinatorial meta-heuristic optimisation techniques.
- Identification of a specific demand pattern based on the classification model created for typical dynamic demand patterns, and selection an appropriate tactical delivery plan for the new week.
- Adjustment of a selected tactical weekly delivery plan to a new or forecasted demand.
- Vehicle routing and scheduling by using scheduling optimisation meta-heuristics (Merkuryeva and Bolshakov, 2011).

CLUSTER ANALYSIS OF DYNAMIC DEMAND DATA

Here, a cluster analysis (Seber, 1984) is aimed: (1) to find a number of typical dynamic demand patterns and corresponding clusters of planning weeks; (2) to construct a classification model that for any week allows determining an appropriate demand pattern, allocating a specific week to one of previously defined clusters and determining correspondent product delivery plan. In the business case, the historical data on daily number of delivered products for 52 weeks are used and specified by weekly demand time-series each representing a sequence of points - daily number of the product deliveries for a specific week (see Fig. 2).

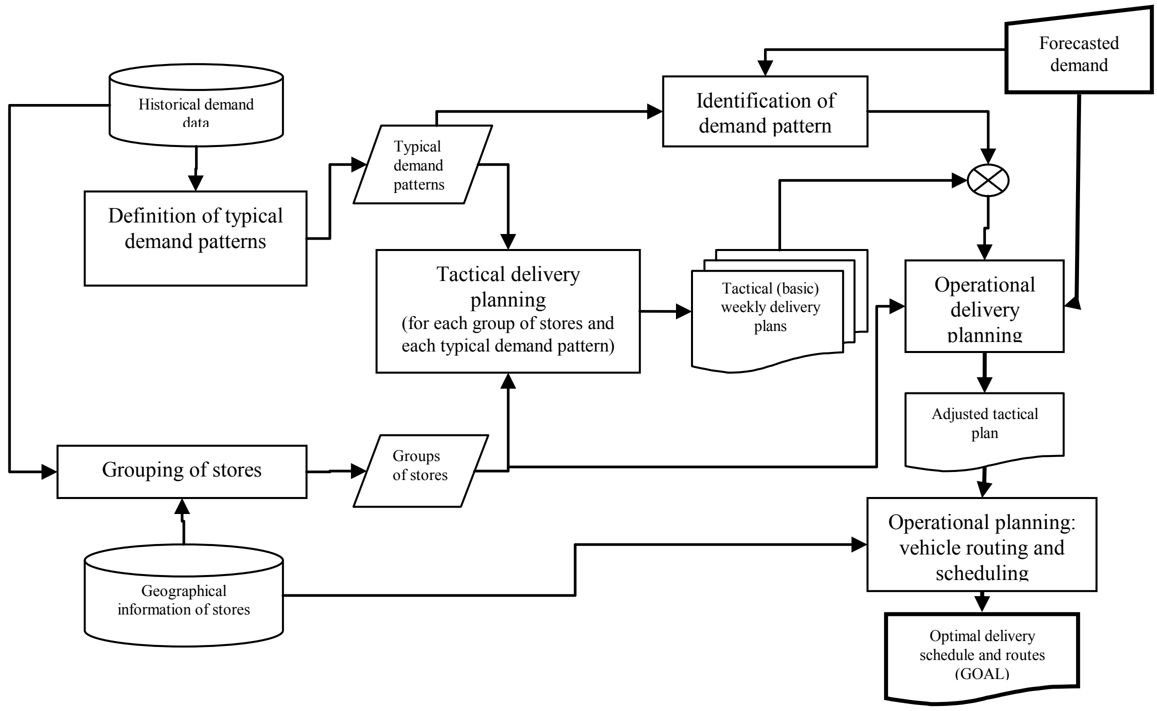

Figure 1: Scheme of Integrated Solution

Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
2885	3390	3891	4115	4612	4687	3371
2831	3553	3859	3785	4432	4899	3527
2763	3548	4067	4631	4838	5057	3511
2951	3820	3987	4360	5075	[illegible]	3345
[illegible]	2731	3101	2988	3385	3524	2643
3150	3459	4339	4377	5187	4956	3545
2934	3229	3643	3693	4018	4411	3583

Figure 2: Sample Demand Data

Here, a cluster analysis of input data provides an opportunity to divide a variety of planning weeks into clusters and to find the number of clusters that represent weeks with specific demand patterns. It also gives information for a construction of the classification model in order to decide which weekly delivery plan would be preferable for next week.

The K-means clustering algorithm (MacQueen, 1967) is used in the paper. It is aims to divide *n* observations into a user-specified number *k* of clusters, in which each observation belongs to a cluster with the nearest mean representing a cluster centroid. The result is a set of clusters that are as compact and well-separated as possible. Here, an appropriate number of *k* clusters, or typical demand patterns is defined by using silhouette plots (Kaufman and Rousseeuw, 1990). In this method, a numerical measure of how close each point is to other points in its own cluster compared to points in the neighbouring cluster is defined as follows:

$$s_i = \frac{b_i - a_i}{\min(a_i, b_i)}, \qquad (1)$$

where s_i is a silhouette value for point i, a_i is an average dissimilarity of point i with the other points in its cluster, and b_i is the lowest average dissimilarity between point i and other points in another cluster. Higher mean values of silhouettes show better clustering results that determine better clusters giving the best choice for a number of clusters.

In the research, k-means clustering experiments have been performed for the number of clusters from 2 to 8. Then for each clustering experiment, silhouette plots have been built, and mean values of silhouettes per cluster have been calculated (Fig. 3). Analysis of silhouettes mean values leads to the conclusion that the best cluster separation could be done at k=4 with a silhouette mean value equal to 0.558. Clusters 1 to 3 seem to be appropriately clustered. However, silhouettes values for a cluster 4 are negative. Theoretically, weekly demands assigned to this cluster could be better allocated to another cluster. These weeks are unlike in the demand dynamics and in specific days, where demand peaks observed.

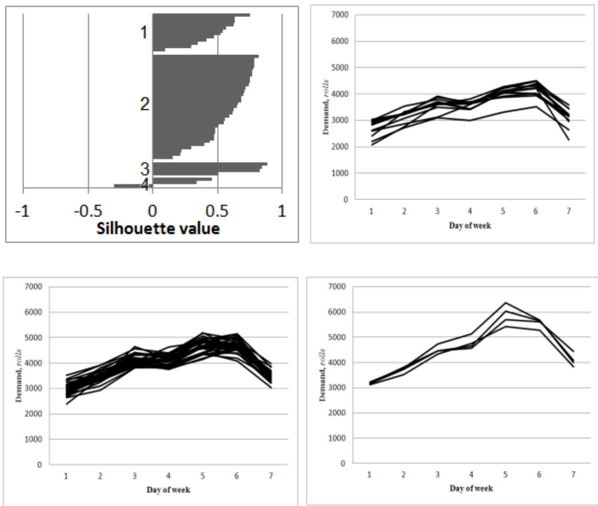

Figure 3: Silhouette plots for the number of clusters k=4 and demands patterns with a mean value greater than 0.5

Reallocation of ‘unlike’ weeks avoids receiving negative silhouette values (see Fig. 4). However, this does not provide an increase of the silhouette mean value as might be expected. In this case, ‘unlike’ weekly demands behave as a ‘noise’ in their ‘native’ clusters, decreasing silhouette values. Then, clustering experiments have been performed with 49 weeks, where three ‘unlike’ weeks have been excluded from a cluster analysis. This has allowed us to increase the silhouette mean value up to 0.5822 while getting the same groups of data clusters 1-3.

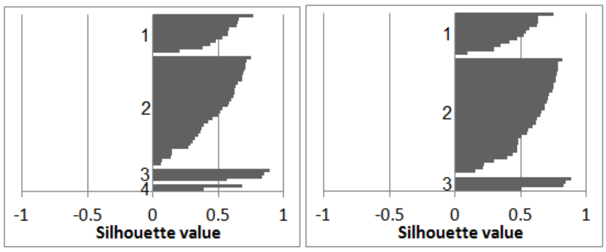

Figure 4: Silhouette plots for a number of clusters k=4 with reallocation of ‘unlike’ weeks and for the number of clusters k=3 and 49 sample weeks

As a result, the number of clusters is fixed and set to k=4. It is worth noting that a tactical weekly delivery base plan could be defined for a cluster with a silhouette mean value greater than 0.5. In this case, a tactical product delivery base plan is selected, adjusted or build for the first three clusters, and not analysed for the last one. Dynamic patterns received for clusters from 1 to 3 are presented in Fig. 3.

A classification model that assigns an appropriate demand cluster is presented by a NBTree, which induces a hybrid of decision-tree and Naive-Bayes classifiers. This algorithm is similar to classical recursive partitioning schemes, except that leaf nodes created are Naive-Bayes categorizers instead of nodes predicting a single class (Seber, 1984).

For a specific week and demand time-series, a cluster is identified by determining a proper leaf number C according to the decision tree. When the leaf number is known, a cluster is estimated by a formula:

$$C = \arg\max_{c_j = C} P(c_j) \prod_{i=1}^{m} P(a_i \mid c_j), \qquad (2)$$

where $P(c_j)$ defines the probability that weekly demand belongs to cluster c_j, and $P(a_i / c_j)$ defines a conditional probability that demand in day a_i belong to cluster c_j. Probabilities $P(c_j)$ are calculated from clustering results, while $P(a_i / c_j)$ defined from the classifier according to the above determined leaf number.

To improve performance of the classification model, the number of weeks has been increased up to 156. Two demand time-series were generated for each planning week by uniformly changing its daily number of delivered products by ±5%. In a similar way, input data for another 52 weeks have been generated and used to validate a classification model itself. Built on this data the NBTree-based classification model with an example of the leaf Naive Bayes classifier is given in Fig. 5. In this case, 10-fold cross-validation showed that only eight weeks have not been classified correctly, which produced an error value of about 5%.

```
FRI <= 4294.5
|   TUE <= 3380.5: 1 cluster
|   TUE > 3380.5: Leaf Number 1
FRI > 4294.5
|   TUE <= 3700.5
|   |   MON <= 3062: 2 cluster
|   |   MON > 3062
|   |   |   THU <= 4368.5: Leaf Number 2
|   |   |   THU > 4368.5: Leaf Number 3
|   TUE > 3700.5: Leaf Number 4
```

Leaf Number: 4 Naive Bayes Classifier

Class	Monday	Tuesday	Wednesday (-∞-4406]	Wednesday (4406-5152.5]	Wednesday (5152.5-∞)	Thursday (-∞-5467.5]	Thursday (5467
2 0.348	1.000	1.000	0.727	0.182	0.091	0.900	
3 0.522	1.000	1.000	0.067	0.867	0.067	0.929	
4 0.130	1.000	1.000	0.167	0.167	0.667	0.200	

Figure 5: Detailed of NBTree Classification Model

For a specific week, an NBTree allows identifying an appropriate cluster and then choosing weekly tactical delivery base plan corresponding to this cluster. Then, selected weekly delivery plan is used for optimisation of parameters of vehicle schedules.

SIMULATION OPTIMISATION OF VEHICLE SCHEDULES

A vehicle schedule defines a schedule of deliveries of various types of goods from DC to a net of stores, in predefined time windows. All delivered goods are divided into three groups. Input data includes three data sets about stores, vehicles and trips. Each store is described with its daily demand and time windows for delivery of each group of goods. Vehicle capacities are limited and known. Each trip is determined by a sequence of stores (trip points) and average time intervals for vehicle moving between these points, loading and unloading processes and, and types of goods to be carried.

Decision variables are introduced to assign vehicles to trips and define a start time for each trip. The problem operational constraints include vehicle capacity constraints and delivery time constraints defined by time windows. The objective function is defined by the total idle time for all vehicles and is minimised. Idle time is the time between two sequential trips performed by a vehicle.

Express analysis shows that the problem could have many solutions not feasible within defined constraints. To increase optimisation efficiency all constraints are converted in soft constraints and fitness function is modified with penalties (Merkuryeva et al, 2010). In the paper, dimensions of the problem are 37 trips, 17 vehicles and 36 stores.

The vehicle schedule model (Fig. 6) is built as a discrete-event simulation model (Merkuryeva and Bolshakov, 2010). Each vehicle is modelled as an active object. Its behaviour is described by a state chart that defines vehicle states (e.g., parking, loading, moving and unloading) and transitions between them.

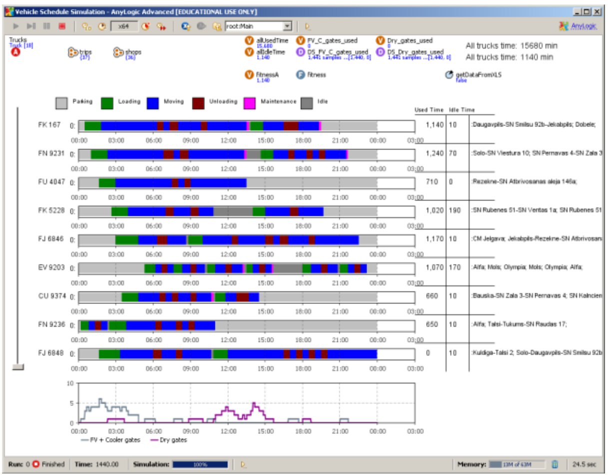

Figure 6: Vehicle Scheduling Model Gantt Chart

Input data are defined in MS Excel spreadsheets and transferred into the simulation model within its initialisation. The vehicle schedule parameters are introduced as control variables in the model. During simulation, constraint violations are fixed.

For search of the best combination of vehicle schedule parameters, a genetic algorithm (GA) is applied. It is implemented as Java class and interacts with the simulation model via ‘Parameter variation’ experiment in AnyLogic (Merkuryeva and Bolshakov, 2011). GA chromosomes are implemented as strings of integer numbers that encode parameters of vehicle schedule, i.e. a vehicle number and start time for each trip. All genetic operators are customized for operating with the proposed structure of a chromosome. One-point crossover with rate of 75%; a mutation operator that changes on each iteration one random trip in the solution, with rate of 1%; and tournament selection with tournament size of two individuals are involved in the algorithm. The solution found allowed decreasing the total idle time for all vehicles comparing with the original delivery schedule in the business case. Better results have been achieved compared with ones received with a general-purpose optimiser OptQuest, which had stuck in local optima and could not find any solutions satisfied all constraints in the problem.

Further GA improvements through fitness landscape analysis are suggested in (Bolshakov et al, 2011). Here, a plug-in of the HeuristicLab optimisation framework (Wagner, 2009) has been implemented by maintaining the logic of the simulation model. Fitness landscape experiments has been performed to compare optimisation results for different GA operators with integer encoding of solutions, as well as to analyse the problem fitness landscapes for different types of solution representations and define most effective for the problem optimisation.

CONCLUSIONS

The proposed integrated approach to product delivery tactical planning and scheduling allows identifying typical dynamic demand patterns and corresponding product delivery tactical plans as well as finding the optimal parameters of product delivery schedules. This allows reducing the effect of product demand variation on the delivery planning process and avoids numerous time-consuming adjustments of the delivery tactical plans. Also, identifying demand pattern and an appropriate delivery plan ensure more qualitative solutions of the schedule optimisation task and cut down its computational costs.

REFERENCES

Bolshakov, V., Erik Pitzer, Michael Affenzeller, 2011. “Fitness Landscape Analysis of Simulation Optimisation Problems with HeuristicLab“. *Proc. 2011 UKSim 5th European Symposium on Computer Modeling and Simulation*, p. 107-112.

Kaufman, L., Rousseeuw, P. J. 1990. *Finding Groups in Data: An Introduction to Cluster Analysis*. Hoboken, NJ: John Wiley & Sons, Inc.

MacQueen, J. B. 1967.“Some Methods for Classification and Analysis of MultiVariate Observations” *Proc. of the 5th Berkeley Symposium on Math. Statistics and Probability,* Vol. 1, p. 281-297, 1967.

Merkuryeva, G., Bolshakovs, V. 2010. “Vehicle schedule simulation with AnyLogic,” *Proc. of 12th Intl. Conf. on Computer Modelling and Simulation,* 2010, p. 169–174.

Merkuryeva, G., Merkuryev, Y., Bolshakov, V. 2010. Simulation-based fitness landscape analysis for vehicle scheduling problem. *Proc. of the 7th EUROSIM Congress on Modelling and Simulation. September 6-10, 2010, Prague, Czech Republic,*7 p.

Merkuryeva, G., Bolshakov, V. 2011 “Simulation-based Fitness Landscape Analysis and Optimisation for Vehicle Scheduling Problem,” *Lecture Notes in Computer Science, EUROCAST 2011,* Part I, LNCS 6927, pp. 280–286, 2011.

Merkuryeva, G., Bolshakov, V., Kornevs, M. 2011. “An Integrated Approach to Product Delivery Planning and Scheduling“. *Scientific Journal of Riga Technical University, Computer Science, Information Technology and Management Science,* p. 97-103.

Seber G. A. F., 1984. *Multivariate Observations*. Hoboken, NJ: John Wiley & Sons, Inc.

Wagner, S., 2009.“Heuristic Optimization Software Systems - Modeling of Heuristic Optimization Algorithms in the HeuristicLab Software Environment”, PhD Thesis, Institute for Formal Models and Verification, Johannes Kepler University Linz, Austria, 2009.

AUTHOR BIOGRAPHY

GALINA MERKURYEVA is a full professor at Riga Technical University, Department of Modelling and Simulation, Latvia. She has research interests and experiences in discrete-event simulation, simulation metamodelling and optimisation, decision support systems, supply chain simulation and management, and simulation-based training.