I L S 2020
INTERNATIONAL CONFERENCE ON INFORMATION SYSTEMS, LOGISTICS & SUPPLY CHAIN
Austin, Texas April 22-24, 2020

Improvement of freight consolidation with a data mining technique

Zineb Aboutalib[1,2], Bruno Agard[1,2]

[1] Laboratoire en Intelligence des Données,
Département de Mathématiques et génie industriel,
École Polytechnique de Montréal, 2900 Boulevard Edouard-Montpetit, Montréal, Québec, Canada.
[2] CIRRELT – Interuniversity Research Centre on Entreprise Networks, Logistics and Transportation,
3520-2920 Ch de la Tour, H3T 1J4, Montréal, Québec, Canada.
zineb.aboutalib@polymtl.ca, bruno.agard@polymtl.ca

Abstract. At first glance, freight consolidation may appear simple, but it is supported by many variables and a large diversity of methods and strategies, which make it a complex notion to understand and to apply in many real situations. Freight consolidation consists of grouping different products into a single batch of goods or areas, for different purposes such as maximizing loading space, the quantity delivered in the same vehicle, the quantity delivered to the same destination, or reducing the distribution costs. Through a literature review, it was noticed that operational research is often used to explain the advantages of freight consolidation. The present study improves freight consolidation using the data mining approach. A case study, in a real situation, supports the conclusions. The results show that applying data mining techniques such as association rules provides some insights and helps a manager in the decision-making process.

Keywords: Freight consolidation, data mining, association rules, transportation costs, Logistics 4.0.

1. Introduction

In the area of big data and mass customization, customers are becoming more demanding and versatile. A race toward efficiency is taking place to acquire and retain them. Shorter planning cycles and accelerated deliveries at a lower cost are expected from companies in order to remain competitive. A tradeoff between stocking costs and delivery costs needs to be found. Freight consolidation comes as a plausible solution to the problem.
Even though it is a fairly old concept, freight consolidation is still drawing the interest of many researchers, [1-4] leading to better supply chain management. That's why companies should watch the evolution of technology and techniques that could enhance the efficiency of their processes and allow them to take advantage of the massive amounts of data stored. The choice of literature for solving this tradeoff has mainly used operational research. This paper aims to use data mining techniques as an alternative solution to improve freight consolidation.
The remainder of this paper is organized as follows: section 2 discusses a literature review of freight consolidation and data mining techniques. Section 3 presents a proposed methodology regarding the consolidation using association rules. Section 4 presents a case study using data from our industrial partner, and in section 5 the results will be presented and discussed, followed by the conclusions in section 6.

2. Literature review

This section is an overview of freight consolidation and data mining concepts; each subpart shows the pertinent definitions, objectives and tools that will facilitate an understanding of the case study.

2.1. Freight consolidation

According to [1], "*The consolidation concept has been known for hundreds of years and the practices are widely used in rail, ground, sea, and air transport*". Various definitions of freight consolidation have appeared over the years. This section will present some of them and discuss the objectives and classifications of freight consolidation.

2.1.1. Definition and objectives

Freight consolidation was defined by [1, 5] as the combination of many small shipments so that a larger and more economical load of goods could be grouped together on the same vehicle. Traditionally, the consolidation has been classified as temporal, spatial or product based [11].
The objectives of consolidation are to take advantage of the price reduction associated with larger loading sizes, common to the railways, road haulers and airlines [6, 7]. Consolidation permits managing distribution costs efficiently and effectively [2, 4, 6, 8, 9, 10]. Grouping and combining products at different locations and at different times in a single load or tour can increase earnings and reduce transportation and inventory losses.
To conclude, a compromise between the benefits and the losses of using consolidations should be examined [6]. The literature provides different classifications of freight consolidation that provide solutions to various contexts.

2.1.2. Freight consolidation classification and strategies

Several classifications concerning freight consolidation have been presented in the literature, Brennan's classification [11] considers three principal subdivisions: spatial, temporal and product freight consolidation strategies. In the study that follows, only temporal and spatial strategies will be presented.
a) Temporal consolidation.
Temporal consolidation is the grouping of small orders over time, balancing customer service and inventory costs against delivery costs [6]. It is also called pure consolidation, and can be implemented without the need for coordination [13]. According to [12], temporal consolidation can be applied using the following policies:

- Time-based policies that consist of sending a consolidated shipment every "T" period. These release all pending orders. [13] gave an example of the computer industry, explaining that this policy applies for items with low volume and high value.
- Quantity-based policies that consist of sending a consolidated shipment when an economic shipping quantity is reached. Based on the same example, [13] states that this policy applies for items with high volume and low value.

Another type of temporal consolidation that consists of combining the two policies above to achieve fast delivery and optimized loading is hybrid consolidation [13]. It appears most often in day-to-day operations management associated with accelerated orders [12].
b) Spatial consolidation.
Spatial consolidation is the process of determining the starting and arriving point, the route and the grouping of small orders to be shipped in one large shipment [14]. Spatial consolidation studies focus on determining the minimum cost associated with combining small shipments using a consolidation terminal. [14] considers that spatial consolidation is similar to that of a multi-warehouse vehicle tour problem. [9] on the other hand, explains that a consolidation problem is different from a vehicle routing problem, first because the consolidation can be selective and does not need to combine all products, which can reduce costs and improve service. Second, because vehicles are not necessarily making round-trips, especially if they use for-hire carriers.
Although [14] and [9] have a different perspective, [14] joins [9] in explaining that the spatial consolidation problem includes vehicle touring using a consolidation point rather than a direct tour from deposits to the

customer. To conclude, spatial consolidation takes into account several parameters and its application allows for potentially important gain.

c) Product consolidation

According to [14], product consolidation is the combination of different types of products into a single shipment. When a customer buys different products, consolidation may increase the quantity delivered to each customer by delivery and decrease the delivery costs, resulting in big gains [15]. Product consolidation can also acknowledge that two or more distribution centers can share the same distribution channel and that each customer can request different products in small quantities [11].

Finally, freight consolidation strategies can be classified as upon the repartition given in Table 1 [18]. It should be noted that [6, 11, 16] propose less consolidation classifications than those of [10]. In the state of the art given by [17], redundancy was pointed out in the repartition of [10], who explains it as the likeness between the consolidation of units of the vehicle and by containers. This study will consider Brennan's [11] spatial consolidation and temporal consolidation strategies.

Table 1: Summary of the repartition of various consolidation strategies [18].

Brennan [11]	Cooper [16]	Sheffi [10]	Hall [6]
Spatial consolidation	Customer Consolidation	Vehicle consolidation	Inventory consolidation
Product consolidation	Product consolidation	Container consolidation	Vehicle consolidation
Temporal consolidation	Time unit consolidation	Chanel consolidation	Terminals consolidation
		Network consolidation	
		Time unit consolidation	
		Tour consolidation	

2.2. Data mining

Data mining (DM) is a discipline that emerged with the information revolution and the creation of large amounts of data. Furthermore, using the internet as a channel of distribution in an open market increases the competition between companies in various areas. DM has mainly been used for prediction and description [19] to extract knowledge that will enhance the efficiency and the productivity of companies. Depending on the companies' objectives and the data available, classification, regression, clustering, prediction, association, etc. can be used in decision making [19]. [20] gave an overview and a classification of the existing techniques and their use in an industrial context.

One of the techniques used for descriptions and prediction is the association rules (AR). AR aims to find patterns or co-occurrences in a dataset [21]. It is used in different areas such as: the market basket analysis for commercial planning, the filtering and predicting of anomalies in telecommunication, diagnosis in medical research, etc. To understand how AR works, let us take two item sets U and V, according to [22], an association rule is defined as the rule: if U then V or $\{U\} \rightarrow \{V\}$. Three criteria are used to evaluate the quality of an AR: Support, which can be expressed as the probability that a transaction will contain U and V; Confidence, which is the probability that a transaction will contain V knowing that U is already present, and Lift, which indicates whether the correlation between U and V is higher than the hazard.

3. Methodology

This section presents a methodology that aims to improve freight consolidation with the use of AR. The proposed methodology consists of extracting valuable information and insights to help managers make data-driven, day-to-day decisions regarding freight consolidation. Figure 1 presents the freight consolidation methodology improved with AR. The methodology can be approached in two ways. Either by making a freight consolidation methodology (highlighted in blue) using well established concepts. Or, by including association rules in the process of steps 4 and 5 for enhanced results.

The freight consolidation methodology is depicted in 6 steps: *step 1- Data preparation*, *step 2 - Grouping and filtering of geocoded data*, *step 3- Consolidation cycle algorithm,* which is performed in parallel with step 2. Afterward, *step 4 - Spatial consolidation* and *step 5- Temporal consolidation* are also performed in parallel, respectively, following steps 2 and 3. The method ends with *step 6- Spatio-temporal consolidation*.

The AR freight consolidation methodology is depicted by step 4 and 5, which can be conducted using association rules following the steps highlighted in green. Association rules follow 3 steps: *step A-Contingency matrix preparation,* which recaps "*the conditional frequencies of two attributes and shows how these attributes are dependent on each other*" [23]. *Step B- Frequent transaction study,* which consists of exploring frequent transactions; transactions are defined as the orders made frequently at the same time by customers who share geographical units by applying the Apriori algorithm on transaction data. Thus, the rules generated by the algorithm are extracted. S*tep C- Significant rules selection* involves analysing the rules and selecting significant ones according to the AR evaluation criteria.

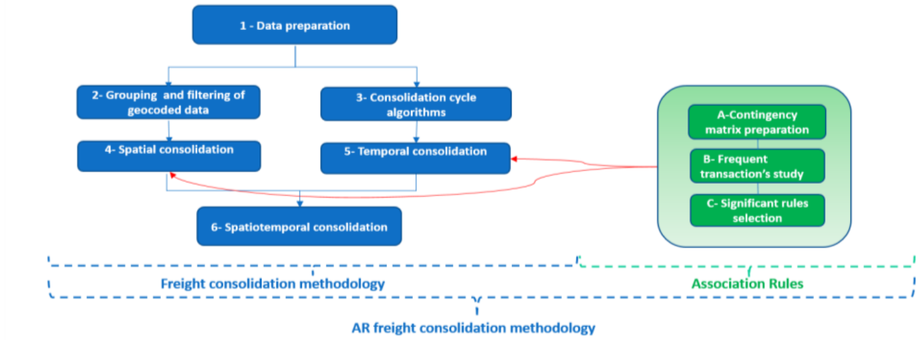

Figure 1: AR freight consolidation methodology.

4. Case Study

This case study will evaluate improvements in freight consolidation strategies using association rules. The data considered are real reception orders from an industrial partner, **Logistik Unicorp**, who supports the uniform program of various Canadian companies both organizational and governmental. Activities of **Logistik Unicorp** are the design, manufacturing, and quality monitoring of those uniforms. The study will focus on governmental clients, using a "clothing online" service, which is similar to e-commerce.

A database that includes several millions of transactions from 2006 to 2018 was provided by **Logistik Unicorp**. Those transactions have been processed and clients' orders were delivered throughout Canada from one unique warehouse. The objective is to evaluate improvements in freight consolidation strategies using real data from our industrial partner.

4.1. Descriptive analysis and data visualization

The reception orders data includes information about the location, entry dates, client code, quantity, and more. Delivery dates are not available in the data; therefore, it is considered to be the day after the entry date.

This subsection starts with the evaluation of tendencies related to the number of orders received from 2012 to 2017. 2018 was excluded from the analysis due to incomplete records.

Figure 2 shows the number of orders as a weekly time series for each of the 6 years. The curves follow a similar trend and seasonality. Two peaks in 2014 near week 20 were identified and linked to a special project that took place in the company during that period.

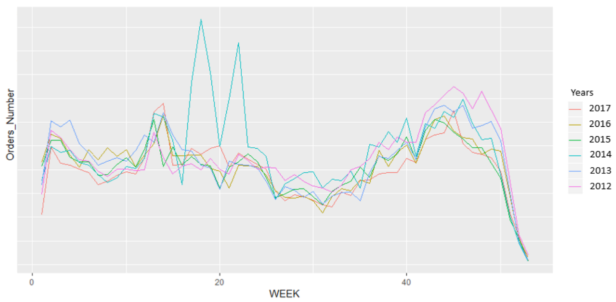

Figure 2: weekly variation in the number of orders (years 2012: 2017).

Curves show a clear trend over the years. Thus, using the last complete year in the development of a consolidation technique should be representative of the current state of orders. For scaling purposes, we also consider using a medium-density city, Winnipeg.
For the city of Winnipeg and the year 2017, the data contains more than 25,000 ordered items, in 2,000 orders, for 1,000 customers.
Figure 3 shows a distribution of customers in the city of Winnipeg. Orange dots represent single customers. Red dots represent set of customers that share the same postal code. The customer distribution map has been modified for confidentiality, but still represents typical behavior.

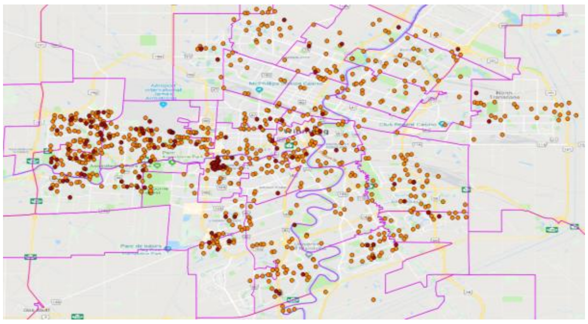

Figure 3: Typical distribution of customers in the city of Winnipeg (an FSA repartition).

4.2. Application of the methodology for consolidation

4.2.1. Freight consolidation methodology

Step 1- Data preparation
The first step in data preparation is formatting and cleaning the data. Location, entry dates, and quantities were formatted into factor, date and numeric structures, respectively. No missing values nor duplicates were recorded during the analysis. The second step of data preparation is data engineering. New variables were created, including forward Sortation Areas (FSA) based on location, year, month, week and day based on entry date and city province, latitude, longitude based on postal code.
The location information initially available for each order were the postal codes, the variable was extracted from it by selecting its first three characters resulting in a new geographical unit. For the consolidation strategy, FSA is the selected aggregation level. It was chosen instead of postal codes because it groups customers sharing the first three characters, which facilitate the logistics by delivering orders to a relay point rather than delivering to each customer.
Step 2- Grouping and filtering of geocoded data: This step groups and filters geocoded data. It leads to setting up the reception order data in step 4.
Step 3- Consolidation cycle algorithm: This step supports four different consolidation cycles (Cc). The "Cc1" scenario groups and delivers orders every 2 days. In the "Cc2" scenario, every 3 days. For the "Cc3" scenario, orders received Monday-Tuesday-Wednesday will be delivered on Thursdays, and those received Thursday-Friday-Saturday-Sunday will be delivered on Mondays. Finally, in the "Cc4" scenario, customers have a 7 day consolidation period. The choice of delivery days in the Cc3 scenario can be arbitrary. In this specific case, it is based on the tendencies noticed in **Logistik Unicorp**'s data. It has been noticed that Cc3 allowed the company to have a balanced number of deliveries for each of the proposed days for delivery**.**
Based on those scenarios, five attributes were added to the database. The first one transforms the entry date into the order day of the week. The remaining four attributes are based on the "consolidated delivery date" for each scenario. Thereby, four consolidated dates are presented for a given order.
Step 4- Spatial Consolidation: This step aggregates small geographical units into a larger one. In this case, postal codes are aggregated into FSA. Concretely, it translates into adding a relay point where customers receive deliveries, instead of delivering to each customer separately at their own address. This results in a reduction in the number of trips performed per week. The optimal relay point will not be discussed in this study.
Step 5- Temporal Consolidation: For each scenario, this step regroups orders that share the same consolidated delivery date. Thus, it enhances the quantity of orders for each delivery.

Step 6- Spatio-temporal consolidation: This step is a combination of steps 4 and 5. Orders are grouped according to the spatial preparation and the consolidated delivery date. The benefit that can be earned from this strategy will be discussed in section 5. In addition, a comparison between this spatio-temporal consolidation strategy and the spatio-temporal strategy using association rules will also be made.

4.2.2. AR freight consolidation methodology

In this section, association rules are explained in the context of freight consolidation. AR are used to discover relationships between orders made by customers in a spatio-temporal setting. Then, patterns that may enhance the performance of freight consolidation strategies are identified. The rules linking FSA to order entry date are evaluated as follows:
Step A- Contingency matrix preparation: This step is required for the pretreatment of data in order to make it into the appropriate form for the association rule package. The package requires a transaction matrix that was built by transforming the orders' reception database into a boolean contingency matrix and converts it to a transaction matrix (See figure 4).
Step B- Frequent transactions study: This step consists of exploring a set of frequent FSAs using the transaction matrix. An apriori algorithm is applied to generate rules. Different supports and confidences are set to explore significant implications.
Step C- Significant rules: In this step, significant rules are selected, based on the aggregation level expected. The operational manager of the company helps to select those rules, depending on the quality of the rules and the degree of tolerance for uncertainties. Thus, it provides the operational manager with good insight for setting when and which geographical units' orders will be consolidated and delivered.
The AR freight consolidation methodology consists of incrementing selected rules of step C into steps 4 and 5. The resulting consolidation rules will provide a starting point to the manager for an initial geographical consolidation strategy.

ENTRY_DATE	R2C	R2G	R2H	R2J	R2K	R2L	R2M	R2N	R2P	R2R	R2V	R2W	R2X	R2Y	R3A	R3B
2017-01-01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2017-01-02	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
2017-01-03	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
2017-01-04	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2017-01-05	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
2017-01-06	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
2017-01-07	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2017-01-09	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0
2017-01-10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
2017-01-11	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 2: Sample of contingency matrix.

5. Results

In this study, an analysis of FSAs according to "order number" and "client density" was used in addition to AR selection. The analysis gave an overview of how the rules chosen by the operation manager could impact freight consolidation.
The FSA rules generated by Apriori with a minimum support of 0.1 and a minimum confidence of 0.7, resulted in sets composed by 3 FSA in the form of {U, V} → {W} as presented in figure 5.

```
                  rules   support confidence      lift
{R3J,R3Y} => {R3N} 0.1810345  0.8181818 1.382171
{R3K,R3N} => {R3J} 0.1293103  0.9574468 1.348953
{R3J,R3K} => {R3N} 0.1293103  0.7894737 1.333674
{R2Y,R3N} => {R3J} 0.1982759  0.9324324 1.313710
{R2Y,R3Y} => {R3J} 0.1034483  0.9230769 1.300529
```

Figure 5: Top 5 rules

An example of one of the top five rules is [R3N, R3K] => R3J.
This can be read as follows: if **Logistik Unicorp** receives customers' orders that live in postal codes grouped into FSA "R3N" and "R3K", it will have 95% certainty that it will receive orders from customers who live in region "R3J" at the same time. In other words, customers from "R3J" always order when customers from "R3N" and "R3K" order as well.

Rules generated by Apriori with parameters set to a minimum support of 0.4 and a minimum confidence of 0.8, resulted in sets composed by 2 FSA in the form of {U} → {V}. It contains 2 rules, implying that 80%

of the customers who order from “R3J” and “R3N” are associated. Therefore, shipping orders for these two regions may be consolidated.

The gain evaluation over implementing AR on different freight consolidation strategies is presented in Table 2. This table represents two results for each scenario: the first one on the left shows the percentage of reduction in the number of deliveries according to a delivery date and/or the area of delivery compared to the current situation. The second result on the right (in bold) highlights the same information after the implementation of the previously described AR. Following the application of the different Cc scenarios on the Clothing Online orders, freight consolidation strategies were compared to a next day delivery policy.

Table 2: Freight consolidation strategies after applying a different consolidation cycle and implementing AR.

Freight consolidation strategies	Consolidation cycle			
	Cc1	Cc2	Cc3	Cc4
Temporal consolidation	41.6 | **41.6**	56.3 | **56.3**	69.8 | **69.8**	84.7 | **84.7**
Spatial consolidation	28.4 | **47.4**	28.4 | **47.4**	28.4 | **47.4**	28.4 | **47.4**
Spatio-temporal consolidation	39.4 | **55.6**	45.9 | **59.7**	48,7 | **62.1**	61.7 | **70.1**

Taking, for example, Cc2, which consists of delivering to customers every 3 days, according to Table 2, temporal consolidation decreases the number of deliveries by 56.3%, for spatial consolidation by 28.4% and for spatio-temporal consolidation by 45.9%. With the consideration of selected AR on freight consolidation, the number of deliveries does not change for temporal consolidation, but it decreases to 47.4% for spatial consolidation and to 59.7% for spatio-temporal consolidation.
We can observe that:

- Temporal consolidation, which does not consider geospatial information, gets no improvement with geospatial association rules.
- Spatial consolidation, which does not consider temporal cycles, remains constant across all scenarios, and improvements from AR stay constant. The reduction in the number of deliveries goes from 28.4% to 47.4%.
- AR improves both spatial consolidation and spatio-temporal consolidation strategies for all scenarios.
- In this case study, temporal consolidations provide better improvement than spatial consolidation because delivery points are not considered.
- Spatio-temporal consolidation is better than spatial consolidation in both cases (whether or not AR is used).
- For Cc1 and Cc2, spatio-temporal consolidation using AR provides better improvement than temporal consolidation.

According to those results, the proposed consolidation technique (using AR) is better than the classical technique for spatial and spatio-temporal strategies. However, the distribution manager can mix both techniques according to the activity and the context of the situation presented to maximize the profit.

6. Conclusion

The objective of this study was to show that data mining techniques, in particular association rules, could improve freight consolidation. The implementation of association rules into the development of consolidation strategies permits the number of deliveries to decrease. A comparison between the freight consolidation methodology and AR freight consolidation methodology proved to reduce the number of deliveries for both spatial and spatio-temporal strategies, but not temporal strategy. This method in turn could also increase the quantity delivered in the same vehicle or delivered to the same destination, thus reducing transportation costs and carbon footprint.
Limits of this method are the level of granularity of the spatial aggregation and the FSA assignation choice. A customer may geographically be located in one FSA, but better improve consolidation in another one. In future studies, those limitations would be taken into account by choosing the optimal dispatch to customer clusters to define a new granularity level. The capacity constraint of the number of consolidated orders that

can be loaded in a truck will be considered. Further improvements could be made to the association rules, such as predicting the Apriori derived rules and employing spatio-temporal AR that take into account longitude and latitude points.

Acknowledgements

The authors express their gratitude to their industrial partner (Logistik Unicorp) for their collaboration in this project and for the supply of data. The authors also thank MITACS (IT 12058) for the financial support. We would also like to express our gratitude to Camelia Dadouchi for sharing her expertise in the preparation of the document.

7. References

1. J. C. Tyan, F.-K. Wang, and T. C. Du: An evaluation of freight consolidation policies in global third party logistics, Omega, 31(1), 55--62 (2003).
2. J. H. Bookbinder and J. K. Higginson: Probabilistic modeling of freight consolidation by private carriage, Transportation Research Part E: Logistics and Transportation Review, 38(5), 305--318 (2002).
3. W. Hu, A. Toriello, and M. Dessouky: Integrated inventory routing and freight consolidation for perishable goods, European Journal of Operational Research, 271(2), 548--560 (2008)
4. A. S. Hanbazazah, L. Abril, M. Erkoc, and N. Shaikh: Freight consolidation with divisible shipments, delivery time windows, and piecewise transportation costs, European Journal of Operational Research, 276(1), 187--201(2019)
5. S. Çetinkaya and C. Y. Lee: Optimal outbound dispatch policies: Modeling inventory and cargo capacity, Naval Research Logistics, 49(6), 531--556 (2002)
6. R. W. Hall: Consolidation strategy: inventory, vehicles and terminals, Journal of business logistics, 8(2), 57 (1987)
7. G. C. Jackson: Evaluating order consolidation strategies using simulation, Journal of Business Logistics, 2(2), 110--138 (1981)
8. Y. P. Gupta and P. K. Bagchi: Inbound Freight Consolidation Under Just-In-Time Procurement, Journal of Business Logistics, 8(2), 74-94 (1987)
9. J. Pooley and A. J. Stenger: Modeling and evaluating shipment consolidation in a logistics system, Journal of Business Logistics, 13(2), 153-174 (1992)
10. Y. Sheffi: Carrier Shipper Interactions in the Transportation Market: An Analytical Framework, Journal of Business Logistics, 7(1), 1-27 (1986)
11. J. J. Brennan: Models and Analysis of Temporal Consolidation, Ph.D.Thesis, University of California at Los Angeles (1981)
12. J. Higginson and J. H. Bookbinder: Policy recommendations for a shipment-consolidation program, Journal of Business Logistics, 15(1), 87--112 (1994)
13. S. Çetinkaya: Coordination of inventory and shipment consolidation decisions: A review of premises, models, and justification, in Applications of supply chain management and e-commerce research: Springer, 92(1),3--51(2005)
14. H. Min: The vehicle routing problem with product/spatial consolidation and backhauling, Ph.D.Thesis .The Ohio State University.
15. M. A. Ulku: Analysis of shipment consolidation in the logistics supply chain, Ph.D.Thesis, University of Waterloo (2009)
16. M. Cooper: Cost and delivery time implications of freight consolidation and warehousing strategies, International Journal of Physical Distribution & Materials Management, 14(6), 47-67 (1984)
17. H. Min and M. Cooper: A comparative review of analytical studies on freight consolidation and backhauling, Logistics and Transportation Review, 26(2), 149--169 (1990)
18. Z. Aboutalib and B. Agard: État de l'art sur la consolidation de marchandise, 13ème Congrès International de Génie Industriel,Montréal (Québec), Canada, 25-28 juin (2019)
19. A. K. Choudhary, J. A. Harding, and M. K. Tiwari: Data mining in manufacturing: a review based on the kind of knowledge, Journal of Intelligent Manufacturing, 20(5), 501--521 (2009)
20. B. Agard and A. Kusiak: Exploration des bases de données industrielles à l'aide du datamining–Perspectives, 9ème colloque national AIP PRIMECA, La Plagne, France, 5-8 avril (2005)
21. R. J. Bayardo Jr: Efficiently mining long patterns from databases, in ACM Sigmod Record, 27(2), 85-93 (1998)
22. R. Agrawal, H. Mannila, R. Srikant, H. Toivonen, and A. I. Verkamo: Fast discovery of association rules, Advances in knowledge discovery and data mining, 12(1), 307--328 (1996)
23. S. Tsumoto.: Statistical independence as linear independence, Electronic Notes in Theoretical Computer Science, 82(4), 274--285 (2003)