Article

Data-Driven Design and Optimization for Smart Logistics Parks: Towards the Sustainable Development of the Steel Industry

Yaqiong Lv ¹ , Shangjia Xiang ¹, Tianyi Zhu ¹ and Shuzhu Zhang ^{2,*}

¹ School of Logistics Engineering, Wuhan University of Technology, Wuhan, Hubei 430070, China; Y.Q.LV@whut.edu.cn (Y.L.); S.J.XIANG@outlook.com (S.X.); darkzty@whut.edu.cn (T.Z.)

² Department of Information Management and Artificial Intelligence, Zhejiang University of Finance & Economics, Hangzhou, Zhejiang 310018, China

* Correspondence: shuzhu.zhang@connect.polyu.hk

Received: 31 July 2020; Accepted: 25 August 2020; Published: 28 August 2020

Abstract: The design of steel logistics parks acts as fundamental infrastructure supporting the operations of storage, allocation, and distribution of steel products in the steel logistics industry, which actually lags behind the development of other logistics industries, such as e-commerce logistics, due to its large lot bulk storage, low turnover rate, and costly transportation and operations. This research proposes a data-driven approach for a specific steel logistics park, aiming to improve its operational efficiency in terms of product layout and allocation in multiple yards. The entry and delivery order data are analyzed comprehensively so as to determine the products with high operational frequency and the corresponding relevancy among them. Experimental results show that, among the 69 steel specifications, 14 high-frequency products are identified, and the correlation among the 14 identified high-frequency products possesses evident distribution characteristics concerning their brands and specifications. The identified frequency and correlation among various products can not only facilitate the product layout and allocation in steel logistics parks, but also advance the vehicle scheduling efficiency for product pick-up and delivery. Moreover, the research methodology and framework can provide managerial insights for other industries with mass data processing requirements.

Keywords: smart logistics; FP-growth; data-driven design; layout planning

1. Introduction

The market condition of the steel industry continues to deteriorate due to a variety of reasons. On the one hand, overcapacity in the steel industry is becoming increasingly serious, which results in excessive product storage and a high storage cost correspondingly. On the other hand, the steel price holds on at a high level, and the export volume of steel falls significantly, further exacerbating market deterioration. In order to maintain sustainable development, more and more steel enterprises are seeking new profit options, among which steel logistics services become a promising choice. A multitude of large steel enterprises intend to handle their logistics activities, such as material procurement, production, and product distribution, internally by establishing their own steel logistics parks, which can help avoid the high cost of activity outsourcing. However, in contrast to other logistics industries, such as e-commerce logistics, the development of steel logistics lags due to its inherent industrial features. The product life cycle of steel is relatively long, and the throughput of a steel logistics park is significantly affected by the market condition. The profit of a steel logistics enterprise is reduced as well because of the high steel storage cost and low operational efficiency.

The design of steel logistics parks acts as fundamental infrastructure supporting the operations of storage, allocation, and distribution of steel products in the steel logistics industry. With the expansion of steel warehousing varieties and spot trading varieties, and the increase of e-commerce transactions, the delivery volume of a steel logistics park increases substantially. However, most of the current layout design cannot meet the increasing requirements of inventory management and facility utilization. The operational efficiency of steel yards decreases due to the excess of vehicle turnaround time and unexpected waiting time, and unnecessary frequent facility operations. Queuing for product pick-up frequently appears because of insufficient storage capacity and inappropriate product allocation, which also increases the warehouse operating cost. All the above problems necessitate the improvement and optimization of the layout and operation of steel logistics parks.

Moreover, both social development and technological progress have stimulated and advocated the operational and managerial paradigm shift of the steel industry towards sustainability development. Environmental considerations have been promoted due to social development and become the topmost consideration for industries, especially heavy industries, such as the steel industry [1]. On the one hand, the continuous consumption of non-renewable resources, such as iron ore and carbon coke, undermines the development of the steel industry. On the other hand, environmental awareness rises as an increasing number of consumers are paying attention to product production and transportation. The operations of steel products are commonly environmentally hazardous in terms of noise pollution, carbon emission, air contamination, and so on. Concerning the technological progress, with the application of various Internet of things (IoT) devices in steel logistics, a data-rich environment is generated in terms of large data volume, high data velocity, and multiple data variety [2]. For instance, the location of in-transit steel products can be precisely acquired, providing a more accurate product arrival estimation. The operations inside the steel logistics park can be digitalized and further exploited for optimization.

This research aims to improve the operational efficiency of yard management in steel logistics parks by redesigning the location and allocation schemes of different steel varieties according to their relevance degree. Statistical methods and data mining approaches are applied to analyze the warehousing entry and delivery data so as to find the operational frequency and correlation among different steel varieties. Steel varieties with high operational frequency are suggested to be located near the entrance/exit. Steel varieties with high relevance degree are suggested to be allocated in nearby yards considering the available yard capacities. Meanwhile, steel products are suggested to be deployed in multiple yards evenly, which can balance the usage rate of gantry cranes and improve the overall operational efficiency.

The rest of this paper is organized as follows. Related literature is reviewed and examined in Section 2 in terms of both the logistics park design and methodology application in related areas. Section 3 describes the case scenario in this paper. Then, algorithm application and data analysis are conducted in Section 4. Finally, conclusions are drawn in Section 5.

2. Literature Review

As a significant role of supply chain management, the design of a logistics park is a strategic issue for large-scale enterprises, which is indeed a complex and integrated issue consisting of logistics management systems, advanced information systems, and cooperative freight transportation systems [3,4]. A well-designed logistics park can facilitate reduction in inventory costs and ease of freight pick-up so as to enhance the overall operational efficiency. As it is acknowledged that the establishment of a new logistics park is considerably costly, especially for the steel industry, the operational optimization of steel logistics parks becomes critically important, attracting interest from an increasing number of researchers and practitioners towards the development of smart logistics [5,6]. Moreover, in practice, enterprises in the steel industry are commonly large-scale conglomerates, such as POSCO, TATA steel, JFE, Nippon Steel Corporation, Baosteel Group, etc., which exemplify the conventional operations of the steel industry and supply chain. Wu et al. [7] studied the

supply chain and logistics patterns of these enterprises and concluded that the distinguishing features of iron and steel logistics indicate the importance of proper operational and tactical planning.

Both qualitative and quantitative approaches have been applied to facilitate the design and operation of steel logistics parks. For instance, Wang et al. [8] proposed a rule-based ontology reasoning method to support the decision-making of steel logistics managers. Yilmaz et al. [9] applied a fuzzy analytical network process to handle the material operations for in-plant logistics. Regarding the quantitative solutions, the systematic layout planning (SLP) method is frequently utilized for solving the logistics park planning problems [10,11]. Moreover, many researchers tend to combine SLP and other heuristic algorithms in the logistics parks study. For example, Zhang et al. [12] integrated a genetic algorithm (GA) and SLP for solving the functional area layout of a railway logistics park. Chen et al. [13] designed an improved adaptive GA with scatter search to handle the SLP problem in a non-rectangular logistics park with split lines. Palominos et al. [14] integrated SLP with quality function deployment (QFD) for the layout design of a service-oriented facility.

In addition to the conventional approaches for logistics park planning, data-driven methods become a promising choice with the support from data mining techniques due to the application of various data-enriched devices and facilities. For instance, He and Xue [15] applied fuzzy clustering for solving the layout planning problem of regional logistics parks. Hsieh and Huang [16] developed a k-means batching (KMB) and self-organization map batching (SOMB) mechanism to improve total travel distance and average vehicle utility in logistics parks. Kulak et al. [17] proposed a novel tabu search (TS) algorithm integrated with a novel clustering algorithm to solve the order batching and picker routing problems jointly for multiple-cross-aisle warehouse systems. Chiang et al. [18] presented a new association measure approach, named weighted support count (WSC), to analyze the intensity and nature of the relationships among different products in a distribution center in order to facilitate the order picking efficiency. Pang and Chan [19] applied a data mining-based algorithm for the storage location assignment of piece picking items in a randomized picker-to-parts warehouse by extracting and analyzing the association relationships between different products in customer orders.

In contrast to the aforementioned research work, frequent pattern growth (FP-growth) is applied in this research to explore and exploit the entry and departure orders. FP-growth is a classical association rule mining approach, which was initially introduced by Han et al. [20,21], as a successful and popular extension of the conventional Apriori method [22]. Ever since its first introduction, the FP-growth approach has gained much popularity due to its unique design of FP-tree and high processing capability. Kuo et al. [23] combined FP-growth and an artificial immune network (AIN) for supplier selection and order allocation in the logistics industry. Feng et al. [24] proposed an expert recommendation algorithm based on the Pearson correlation coefficient and FP-growth. Wu and Zhang [25] designed an analysis model for electronic evidence based on an FP-growth algorithm. An initial investigation suggests that although FP-growth has been applied to solve miscellaneous specific problems from multiple industries, little empirical research pertaining to the data analytics for the steel industry using data mining techniques has been conducted, which motivates this research.

3. Case Description

A case study is conducted in this research aiming to provide a straightforward and visualized understanding of the operations of steel logistics parks. The steel logistics service provider involved in this research, with the alias as GC logistics, is a subsidiary enterprise of a large Iron and Steel Corporate, which is among the top five Iron and Steel Corporates in China. The GC logistics park plays a vital role in the operations of the Iron and Steel Corporate, accounting for around 20% of sales volume of the entire corporate. The existence of the GC logistics park not only serves as a regional logistics center to provide storage and distribution services of steel products, but also serves as a favorable sales channel, which generates a huge amount of sales profits to the Iron and Steel Corporate every year. However, along with market development of the steel trade, the steel warehouse operations in the GC logistics park are facing a growing number of challenges. For example, queuing for product

pick-up occurs frequently, which decreases the product turnover rate substantially. The workload imbalance among multiple yard cranes is frequently observed as well. Moreover, the allocation for arrival products are mainly conducted based on personal experience lacking optimization operations. Therefore, the management team of GC logistics parks are seeking various approaches to improve the operational efficiency and reduce the operational cost with the purpose of sustainable development.

As shown in Figure 1, the GC logistics park possesses six yards, among which the first five yards are designed for rebar storage purposes, and Yard 6 is for backup and storage of other spare and scarce steel products. Each yard has one gantry crane for product stacking. When steel products arrive at this steel logistics park, the warehouse manager needs to assign storage units for different steel products considering their brands and specifications. When a delivery order arrives, the logistics manager needs to schedule proper vehicles and facilities to pick up steel products from different yards. An initial investigation finds that most of the entry and delivery operations are conducted based on the knowledge and experience of warehouse and logistics managers, which lacks quantitative support and optimization. The increase of entry and delivery orders and steel varieties necessitate the redesign of product layout and optimization of logistics activities.

Figure 1. The layout of the steel logistics park.

4. Data Investigation and Analysis

In order to handle the location and allocation problem of various steel products in steel logistics parks, this research is conducted in two dimensions, i.e., steel order analysis and steel correlation analysis. The steel order data are firstly investigated and explored in view of the operational frequency of different steel types. Then, correlation among the steel types with high operational frequency is further examined using the FP-growth approach. The frequency identification and correlation analysis can facilitate the layout and deployment of different steel varieties, and further improve the overall efficiency of steel logistics parks.

4.1. Data Investigation

The collected order data comprises steel entry records and steel delivery records from 2017.10 to 2018.8. The steel entry data are relatively simple, which contains 5725 records. The format of the steel entry record is fixed, as shown in Table 1. Each entry number contains only one type of rebar. By comparison, the complexity of steel delivery data is much more complicated, which contains 77,018 records, as illustrated in Table 2. One of the major differences between entry operations and delivery operations is that the entry operations are mainly conducted in batch scale, while the delivery operations are more scattered and diversified. One delivery number commonly contains multiple steel types. Averagely, there are around 15 entry operations and more than 200 delivery operations on a daily basis. Moreover, the delivery operations are more time-consuming and laborious. Therefore, the steel delivery data is further analyzed in detail to explore and exploit the correlation among high-frequency steel types.

Table 1. Examples of entry records.

Entry Time	Entry Number	Brand	Steel Type	Specification	Unit	Weight (Ton)
2017/10/1	RK1710010002	Rebar	HB400	12*9	34	59.772
2017/10/1	RK1710010003	Rebar	HB400	12*9	32	56.256
2017/10/1	RK1710010004	Rebar	HB400	12*9	32	56.256
2017/10/1	RK1710010004	Rebar	HB400	12*9	106	186.348
2017/10/1	RK1710010005	Rebar	HB400	22*9	18	48.276
2017/10/1	RK1710010006	Rebar	HB400	16*9	5	8.53
2017/10/1	RK1710010007	Rebar	HB400	22*9	18	48.276
2017/10/2	RK1710020001	Rebar	HB400	12*9	110	193.38
2017/10/2	RK1710020001	Rebar	HB400	12*9	183	321.714
2017/10/2	RK1710020001	Rebar	HB400	12*9	150	263.7

Table 2. Examples of delivery records.

Delivery Time	Delivery Number	Brand	Steel Type	Specification	Unit	Weight (Ton)
2017-10-01 08:37	LK1709300261	Rebar	HB400	12*9	3	5.274
2017-10-01 08:37	LK1709300261	Rebar	HB400	16*9	1	1.706
2017-10-01 08:37	LK1709300261	Rebar	HB400	16*9	5	8.53
2017-10-01 08:37	LK1709300261	Rebar	HB400	20*9	9	20.007
2017-10-01 08:49	LK1709300278	Rebar	HB400	14*9	18	31.356
2017-10-01 09:23	LK1709300268	Rebar	HB400E	14*9	14	24.388
2017-10-01 09:23	LK1709300268	Rebar	HB400E	16*9	6	10.236
2017-10-01 09:39	LK1709300139	Rebar	HB400	14*9	7	12.194
2017-10-01 09:41	LK1710010005	Rebar	HB400	28*12	12	31.296
2017-10-01 09:58	LK1710010012	Rebar	HB400	14*9	17	29.614

An initial data investigation shows that six steel types are operated in this steel logistics park, i.e., HB400, HB400E, HB500, HB500E, HTB600, and HTB600E. Each steel type contains different specifications in terms of different diameters (centimeter) and lengths (meter). Twenty-four specifications are involved, as 12*12, 12*9, 14*12, 14*9, 16*12, 16*9, 16*7, 18*12, 18*9, 18*7, 20*12, 20*9, 20*7, 22*12, 22*12, 22*9, 22*7, 25*14, 25*12, 25*9, 25*7, 28*12, 28*9, 32*12, and 32*9. Totally, there are 69 combinations of steel types and steel specifications. Herein, the term steel specification is used to represent the combination of steel type and steel specification in the following context for simplification.

The delivery frequency and weight of various steel specifications are demonstrated in Figures 2 and 3. It is noticed that among the 69 steel specifications, the delivery operational frequency of the first 14 steel specifications are far beyond the rest, which is 88.72% of the total delivery operations. The delivery weight distribution is in a similar pattern, in which the first 14 steel

specifications outweigh the rest substantially, which is 82.77% of the total delivery weight. Further analysis shows that the two 14 steel specifications are actually the same combination, although the permutation is slightly different. The first 14 steel specification in terms of delivery frequency are HB400-12*9, HB400-16*9, HB400-14*9, HB400-20*9, HB400-18*9, HB400-25*9, HB400E-12*9, HB400E-16*9, HB400E-14*9, HB400-22*9, HB400E-18*9, HB400E-25*9, HB400E-20*9, and HB400E-22*9.

Figure 2. Delivery frequency analysis.

Figure 3. Delivery weight analysis.

4.2. FP-Growth Analysis

Correlation analysis among the 14 frequent steel specifications is conducted in this section using the FP-growth approach. FP-growth (frequent pattern) is a classical and popular data mining approach, which was initially introduced by Han, Pei and Yin [20]. One of the inherent features of the FP-growth approach is the design of an FP-tree structure, which facilitates the data processing significantly in contrast to the previous Apriori algorithm. The frequent itemset can be discovered after two-time data traversal due to the application of the FP-tree. Moreover, the FP-growth approach comprises a header table and a node list enabling the construction and analysis of the FP-tree. The major operations and

steps of the FP-growth approach are described in Figure 4 comprising two main steps, i.e., FP-tree generation and FP-tree mining.

Figure 4. Frequent pattern (FP)-growth flowchart.

As described in the previous section, there exist 77,018 records in the delivery data. However, only 37,458 delivery numbers exist, which means one delivery number contains multiple records. Therefore, data cleaning is firstly implemented to consolidate the delivery records with the same delivery number. After that, the FP-growth approach is applied to process the consolidated delivery data. Figure 5 illustrates an example of an FP-tree when considering the first 7 high-frequency steel specifications with support degree as 400. The notations in Figure 5 are refined for display purposes. The first and second line in each node represent the steel brand and specification, respectively. The third line represents the support degree related to this specification. The FP-tree with 14 high-frequency specifications is not provided due to the page limit and its complicated structure.

The support degree in FP-growth is tuned tentatively so as to find meaningful combinations among these high-frequency steel specifications. Meanwhile, the complexity of the FP-tree structure diminishes along with the increase of support degree setting. Table 3 shows that 18 frequent 5-itemsets are found with the setting of support degree as 100. When the support degree increases from 100 to 200, no frequent 5-itemsets exist. Instead, 10 4-itemsets are found as presented in Table 4. Table 5 actually is the first two lines of Table 4, as the frequent 4-itemsets still exist. However, when the support degree increases from 300 to 400, the optimal frequent itemset contains only three items as shown in Table 6. Furthermore, Table 7 presents the analytical results when the support degree increases to 900, in which only frequent 2-itemsets can be discovered.

With the increase of support degree, the correlation among high-frequency items strengthens accordingly. It is noticed that among the 14 high-frequency steel specifications, the performance gap between HB400 and HB400E is straightforward and unambiguous. Concerning the steel brand HB400, HB400-12*9 is of the highest delivery frequency. When there are multiple items involved, [HB400-12*9, HB400-14*9], [HB400-12*9, HB400-16*9, HB400-14*9], [HB400-12*9, HB400-16*9, HB400-14*9, HB400-18*9], and [HB400-12*9, HB400-16*9, HB400-14*9, HB400-20*9, HB400-18*9] are gradually formed as the k increases from 2 to 5. Concerning the steel brand HB400E, HB400E-12*9 is the highest frequent one. The combination of [HB400E-12*9, HB400E-16*9], [HB400E-16*9, HB400E-14*9],

[HB400E-12*9, HB400E-14*9], and [HB400E-16*9, HB400E-18*9] are in a similar scale in the case of the frequent 2-itemsets. Moreover, the combination of [HB400E-12*9, HB400E-16*9, HB400E-14*9] and [HB400E-16*9, HB400E-18*9, HB400E-20*9] behave approximately in the case of the frequent 3-itemsets. The above results are actually consistent with practical situations, as HB400 and HB400E are commonly applied for different construction purposes.

Figure 5. FP-tree illustration considering the first 7 high-frequency items with support degree as 400.

Table 3. Frequent itemset with support degree as 100.

Index	Itemset	k	Support
1	[HB400-12*9, HB400-16*9, HB400-14*9, HB400-20*9, HB400-18*9]	5	191
2	[HB400-12*9, HB400-14*9, HB400-20*9, HB400-18*9, HB400-22*9]	5	137
3	[HB400-16*9, HB400-14*9, HB400-20*9, HB400-18*9, HB400-22*9]	5	130
4	[HB400-12*9, HB400-16*9, HB400-20*9, HB400-18*9, HB400-22*9]	5	129
5	[HB400-12*9, HB400-16*9, HB400-14*9, HB400-18*9, HB400-22*9]	5	124
6	[HB400-12*9, HB400-16*9, HB400-14*9, HB400-20*9, HB400-22*9]	5	123
7	[HB400-12*9, HB400-20*9, HB400-18*9, HB400-25*9, HB400-22*9]	5	116
8	[HB400-12*9, HB400-16*9, HB400-20*9, HB400-18*9, HB400-25*9]	5	112
9	[HB400-16*9, HB400-20*9, HB400-18*9, HB400-25*9, HB400-22*9]	5	112
10	[HB400-12*9, HB400-14*9, HB400-20*9, HB400-25*9, HB400-22*9]	5	108
11	[HB400-12*9, HB400-14*9, HB400-20*9, HB400-18*9, HB400-25*9]	5	107
12	[HB400-12*9, HB400-16*9, HB400-14*9, HB400-20*9, HB400-25*9]	5	107
13	[HB400-12*9, HB400-16*9, HB400-20*9, HB400-25*9, HB400-22*9]	5	107
14	[HB400-16*9, HB400-14*9, HB400-20*9, HB400-18*9, HB400-25*9]	5	107
15	[HB400-12*9, HB400-16*9, HB400-14*9, HB400-18*9, HB400-25*9]	5	104
16	[HB400-14*9, HB400-20*9, HB400-18*9, HB400-25*9, HB400-22*9]	5	103
17	[HB400-16*9, HB400-14*9, HB400-20*9, HB400-25*9, HB400-22*9]	5	101
18	[HB400-12*9, HB400-16*9, HB400-14*9, HB400-25*9, HB400-22*9]	5	100

Table 4. Frequent itemset with support degree as 200.

Index	Itemset	k	Support
1	[HB400-12*9, HB400-16*9, HB400-14*9, HB400-18*9]	4	346
2	[HB400-12*9, HB400-14*9, HB400-20*9, HB400-18*9]	4	310
3	[HB400-12*9, HB400-16*9, HB400-14*9, HB400-20*9]	4	295
4	[HB400-12*9, HB400-16*9, HB400-20*9, HB400-18*9]	4	290
5	[HB400-16*9, HB400-14*9, HB400-20*9, HB400-18*9]	4	286
6	[HB400-12*9, HB400-20*9, HB400-18*9, HB400-22*9]	4	224
7	[HB400-16*9, HB400-20*9, HB400-18*9, HB400-22*9]	4	210
8	[HB400-12*9, HB400-14*9, HB400-20*9, HB400-22*9]	4	205
9	[HB400-14*9, HB400-20*9, HB400-18*9, HB400-22*9]	4	205
10	[HB400-20*9, HB400-18*9, HB400-25*9, HB400-22*9]	4	201

Table 5. Frequent itemset with support degree as 300.

Index	Itemset	k	Support
1	[HB400-12*9, HB400-16*9, HB400-14*9, HB400-18*9]	4	346
2	[HB400-12*9, HB400-14*9, HB400-20*9, HB400-18*9]	4	310

Table 6. Frequent itemset with support degree as 400.

Index	Itemset	k	Support
1	[HB400-12*9, HB400-16*9, HB400-14*9]	3	861
2	[HB400-12*9, HB400-16*9, HB400-18*9]	3	584
3	[HB400-12*9, HB400-14*9, HB400-18*9]	3	580
4	[HB400-16*9, HB400-14*9, HB400-18*9]	3	556
5	[HB400-12*9, HB400-20*9, HB400-18*9]	3	553
6	[HB400-12*9, HB400-14*9, HB400-20*9]	3	536
7	[HB400-16*9, HB400-20*9, HB400-18*9]	3	531
8	[HB400-12*9, HB400-16*9, HB400-20*9]	3	510
9	[HB400-16*9, HB400-14*9, HB400-20*9]	3	502
10	[HB400-14*9, HB400-20*9, HB400-18*9]	3	494
11	[HB400E-12*9, HB400E-16*9, HB400E-14*9]	3	439
12	[HB400-20*9, HB400-18*9, HB400-22*9]	3	438
13	[HB400-20*9, HB400-25*9, HB400-22*9]	3	404
14	[HB400E-16*9, HB400E-18*9, HB400E-20*9]	3	402

Table 7. Frequent itemset with support degree as 900.

Index	Itemset	k	Support
1	[HB400-12*9, HB400-14*9]	2	1903
2	[HB400-12*9, HB400-16*9]	2	1801
3	[HB400-16*9, HB400-14*9]	2	1767
4	[HB400-12*9, HB400-18*9]	2	1260
5	[HB400-20*9, HB400-18*9]	2	1260
6	[HB400-16*9, HB400-18*9]	2	1257
7	[HB400-12*9, HB400-20*9]	2	1208
8	[HB400-16*9, HB400-20*9]	2	1173
9	[HB400-14*9, HB400-18*9]	2	1110
10	[HB400-14*9, HB400-20*9]	2	1056
11	[HB400E-12*9, HB400E-16*9]	2	998
12	[HB400E-16*9, HB400E-14*9]	2	976
13	[HB400-20*9, HB400-25*9]	2	974
14	[HB400-20*9, HB400-22*9]	2	967
15	[HB400E-12*9, HB400E-14*9]	2	952
16	[HB400E-16*9, HB400E-18*9]	2	941
17	[HB400-12*9, HB400-25*9]	2	922

4.3. Discussion

The target of this study was to improve the overall operational efficiency and performance of steel logistics parks by exploiting the association rules among different steel products using data mining approaches. Based on the procedures carried on in the mentioned case study, a generic prototype design for data-driven optimization is proposed, as shown in Figure 6. Multiple data sources, i.e., product entrance records, product delivery records, and interior operations records, are taken into account to support the implementation of a data analytics module. The data analytics module comprises three steps of data preprocessing, model building, and parameter tuning, and identifies the operational frequency and association patterns among multiple steel products. After that, the output of the data analytics module underpins the operations management in view of the features of incoming orders and the real-time utilization of yards so as to achieve the optimal allocation and deployment of steel products in this steel logistics park.

To be more specific, the frequency and correlation analysis among various steel specifications can facilitate the layout and deployment of the steel logistics park in several aspects. First of all, the steel specifications of high operational frequency are suggested to be deployed in Yard 1 and 3, which are near the entrance and exit area of the logistics park. Secondly, when multiple steel specifications arrive at this steel logistics park simultaneously, the items with high correlation are suggested to be allocated in the same yard considering the remaining capacity of each yard. Thirdly, the arrival of steel specifications with less correlation are advised to be assigned to yards with the same product categories. In addition, the products with low operational frequency or rare types are suggested to be stored in a separate yard.

Figure 6. Prototype design for data-driven optimization framework.

5. Conclusions

This research aimed to provide tactical management insights towards the layout and deployment of steel logistics parks. The entry and delivery order data of a steel logistics park was investigated and analyzed using a statistical and data mining approach. The operational frequency analysis can facilitate the classification of different steel specifications and improve the utilization of different yards. The correlation analysis among different steel specifications can help to identify the products with high relevance, which can affect the product allocation decisions. The analysis results act as a tactical reference, facilitating the performance improvement of this steel logistics park. Moreover, the proposed data analytics procedures and approaches can be applied to the other steel logistics enterprises in view of their similar operations and management strategies, and the prototype design for a data-driven optimization framework can shed light on the role of data analytics and facilitate the digital transformation of the steel industry.

Future research can be conducted in terms of the operational decision-making requirements. For instance, the specific product allocation for each arrival can be optimized considering the daily operational requirements and real-time capacity remaining for each yard. The order pick-up sequence and vehicle scheduling for delivery tasks can be further analyzed. The performance evaluation criteria of logistics parks can be one promising research area considering different industrial features. Moreover, the application of IoT devices in steel logistics can be another fruitful research direction towards industry digitalization.

Author Contributions: Conceptualization, Y.L.; methodology, S.X. and T.Z.; software, S.X. and T.Z.; validation, Y.L. and S.Z.; writing—Original draft preparation, Y.L. and S.Z.; writing—Review and editing, Y.L. and S.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China, grant number 71902171; in part by the Humanities and Social Science Foundation of Ministry of Education of China, grant number 19YJC630216 and 20YJC630096; in part by the Natural Science Foundation of Zhejiang Province, China, grant number LY19G010003.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Pinto, J.T.; Morales, M.E.; Fedoruk, M.; Kovaleva, M.; Diemer, A. Servitization in support of sustainable cities: What are steel's contributions and challenges? *Sustainability* **2019**, *11*, 855. [\[CrossRef\]](#)
2. Branca, T.A.; Fornai, B.; Colla, V.; Murri, M.M.; Streppa, E.; Schröder, A. The challenge of digitalization in the steel sector. *Metals* **2020**, *10*, 288. [\[CrossRef\]](#)
3. Tang, J.; Tang, L.; Wang, X. Solution method for the location planning problem of logistics park with variable capacity. *Comput. Oper. Res.* **2013**, *40*, 406–417. [\[CrossRef\]](#)
4. Yavas, V.; Ozkan-Ozen, Y.D. Logistics centers in the new industrial era: A proposed framework for logistics center 4.0. *Transp. Res. Part. E Logist. Transp. Rev.* **2020**, *135*, 101864. [\[CrossRef\]](#)
5. Escudero, M.I.; Balsera, J.V.; Fernández, F.O.; Montequín, V.R. Planning and scheduling with uncertainty in the steel Sector: A review. *Appl. Sci.* **2019**, *9*, 2692. [\[CrossRef\]](#)
6. Antuchevičienė, J.; Jafarnejad, A.; Mahdiraji, H.A.; Hajiagha, S.H.R.; Kargar, A. Robust multi-objective sustainable reverse supply chain planning: An application in the steel industry. *Symmetry* **2020**, *12*, 594. [\[CrossRef\]](#)
7. Wu, X.; Shao, Z.; Zhang, X.; Gao, J.; Jin, G. The operation of supply-chain logistics in iron and steel enterprises. In Proceedings of the 2010 2nd International Conference on Education Technology and Computer, Shanghai, China, 22 June 2010; IEEE: Piscataway, NJ, USA, 2020; Volume 3, pp. 103–106.
8. Wang, X.; Wong, T.; Fan, Z. Ontology-based supply chain decision support for steel manufacturers in China. *Expert Syst. Appl.* **2013**, *40*, 7519–7533. [\[CrossRef\]](#)
9. Yilmaz, O.F.; Oztaysi, B.; Durmusoglu, M.B.; Oner, S.C. Determination of material handling equipment for lean in-plant logistics using fuzzy analytical network process considering risk attitudes of the experts. *Int. J. Ind. Eng.* **2017**, *24*, 81–122.
10. Naqvi, S.A.A.; Fahad, M.; Atir, M.; Zubair, M.; Shehzad, M.M. Productivity improvement of a manufacturing facility using systematic layout planning. *Cogent Eng.* **2016**, *3*, 12. [\[CrossRef\]](#)
11. Suhardi, B.; Juwita, E.; Astuti, R.D. Facility layout improvement in sewing department with Systematic Layout planning and ergonomics approach. *Cogent Eng.* **2019**, *6*, 31. [\[CrossRef\]](#)
12. Zhang, Q.; Jiang, C.; Zhang, J.; Wei, Y. Application of genetic algorithm in functional area layout of railway logistics park. *Procedia Soc. Behav. Sci.* **2014**, *138*, 269–278. [\[CrossRef\]](#)
13. Chen, Y.; Jiang, Y.; Wahab, M.; Long, X. The facility layout problem in non-rectangular logistics parks with split lines. *Expert Syst. Appl.* **2015**, *42*, 7768–7780. [\[CrossRef\]](#)
14. Palominos, P.; Pertuze, D.; Quezada, L.; Sanchez, L. An extension of the systematic layout planning system using QDF: Its application to service oriented physical distribution. *EMJ-Eng. Manag. J.* **2019**, *31*, 284–302.
15. He, Q.; Xue, X. A Research on Layout Planning of Regional Logistics Park Based on Fuzzy Clustering. In Proceedings of the 2009 Second International Conference on Intelligent Computation Technology and Automation, Chansha, Hunan, China, 10 October 2009; IEEE: Piscataway, NJ, USA, 2009; pp. 907–909.
16. Hsieh, L.F.; Huang, Y.C. New batch construction heuristics to optimise the performance of order picking systems. *Int. J. Prod. Econ.* **2011**, *131*, 618–630. [\[CrossRef\]](#)
17. Kulak, O.; Şahin, Y.; Taner, M.E. Joint order batching and picker routing in single and multiple-cross-aisle warehouses using cluster-based tabu search algorithms. *Flex. Serv. Manuf. J.* **2011**, *24*, 52–80. [\[CrossRef\]](#)
18. Chiang, D.M.H.; Lin, C.P.; Chen, M.C. Data mining based storage assignment heuristics for travel distance reduction. *Expert Syst.* **2012**, *31*, 81–90. [\[CrossRef\]](#)
19. Pang, K.W.; Chan, H.L. Data mining-based algorithm for storage location assignment in a randomised warehouse. *Int. J. Prod. Res.* **2016**, *55*, 1–18. [\[CrossRef\]](#)
20. Han, J.; Pei, J.; Yin, Y. In Mining frequent patterns without candidate generation. In Proceedings of the International Conference on Management of Data, Houston, TX, USA, May 2000; pp. 1–12.
21. Han, J.-W.; Pei, J.; Yin, Y.; Mao, R. Mining frequent patterns without candidate generation: A frequent-pattern tree approach. *Data Min. Knowl. Discov.* **2004**, *8*, 53–87. [\[CrossRef\]](#)
22. Agrawal, R.; Imielinski, T.; Swami, A. Mining association rules between sets of items in large databases. In Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, Washington, DC, USA, 26–28 May 1993; pp. 207–216.

23. Kuo, R.; Pai, C.; Lin, R.; Chu, H. The integration of association rule mining and artificial immune network for supplier selection and order quantity allocation. *Appl. Math. Comput.* **2015**, *250*, 958–972. [[CrossRef](#)]
24. Feng, W.; Zhu, Q.; Zhuang, J.; Yu, S. An expert recommendation algorithm based on Pearson correlation coefficient and FP-growth. *Clust. Comput.* **2018**, *22*, 7401–7412. [[CrossRef](#)]
25. Wu, Y.; Zhang, J. Building the electronic evidence analysis model based on association rule mining and FP-growth algorithm. *Soft Comput.* **2019**, *24*, 7925–7936. [[CrossRef](#)]

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).