Intelligent service capacity allocation for cross-border-E-commerce related third-party-forwarding logistics operations: A deep learning approach

Shuyun Ren\textsuperscript{a}, Tsan-Ming Choi\textsuperscript{b}, Ka-Man Lee\textsuperscript{c,*,} Lei Lin\textsuperscript{d}

\textsuperscript{a} Guangdong University of Technology, Guangzhou, China
\textsuperscript{b} Business Division, Institute of Textiles and Clothing, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong
\textsuperscript{c} Faculty of Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong
\textsuperscript{d} Goergen Institute for Data Science, University of Rochester, Rochester, NY 14623, USA

\textbf{ARTICLE INFO}

\textbf{Keywords:}
Cross-border e-commerce
Multi-product newsvendor
Logistics service capacity (LSC) allocation
Third-party forwarding logistics (3PFL)
Deep learning

\textbf{ABSTRACT}

With the rise of “cross-border-e-commerce”, the third-party-forwarding-logistics (3PFL) service becomes increasingly popular. Different from the traditional third-party-logistics (3PL) service, the 3PFL company provides forwarding services cost-effectively by consolidating orders from different e-tailers/platforms. The random arrivals of orders create a big challenge. Different from most of the existing studies, a deep learning based one-step integration optimal decision making approach S2SCL (Seq2Seq based CNN-LSTM) is proposed in this paper which intelligently integrates inventory optimization and demand-forecasting process. The Seq2Seq based forecasting architecture, which integrates CNN and LSTM network, is able to model the system dynamics and dependency-relations in varying demand for logistics services. Besides generating the point forecasting results, the proposed approach can quantify demand uncertainty via a dynamic distribution and make optimal decision on logistics service capacity allocation. Through a case-study analysis with real data obtained from a 3PFL company in China’s Great Bay Area, we compare the proposed S2SCL with two benchmark models, including a one-step statistics based integration approach ARIMA and a two-step optimization based approach PSO-ELM, for two tasks: (1) point forecasting and (2) optimal logistic service capacity (LSC) allocation. Experimental results show that S2SCL outperforms the two benchmark models in both tasks significantly.

1. Introduction

1.1. Background

With the cross-border e-commerce (CBEC)’s increasing popularity and thriving globally (Wang et al., 2019), the demand of international logistics has also experienced rapid growth (Lin et al., 2014b, 2014a; W. Liu et al., 2018; X. Liu et al., 2018). However,
due to the characteristics of CBEC, the traditional third-party logistics (3PL) cannot meet up with the rapid development and requirements of CBEC. Given the overwhelming trend in global trade, which can be induced by the advances in logistics, globalization, multinational corporations, e-commerce and m-commerce, the international third-party-forwarding logistics (3PFL) services are becoming increasingly popular in recent years. The 3PFL service providers like various freight forwarders, shipping forwarders and airfreight forwarders are also emerging. One category of logistics service providers of this kind is serving as a logistics carrier for arranging related forwarding logistics services (such as FedEx Corporation, an American multinational courier with a market cap of over $40 billion\(^1\); and SF Express (Group) Co., Ltd., which is the second largest courier in China). The other category is the 3PFL service provider which is specialized in forwarding business for cross-border e-commerce (such as tracking and tracing the delivery cross border, custom clearance and international payment). 3PFL service providers may consolidate various international orders received from different consignors into one full load for road transport or air transport. It thus can help substantially enhance the flexibility and efficiency of the whole system (Arabzad et al., 2015). Different from the conventional 3PL, the orders requesting the 3PFL service arrive randomly. To make the situation even more challenging, both the arrival time and the amount of last order are stochastic. It is hence extremely difficult for the 3PFL service providers to obtain the useful and reliable information in a timely manner that can be used to estimate the market demand. For instance, 3PFL service providers cannot get access to the demand information e-tailers shared since 3PFL service providers receive orders from e-commerce customers who have forwarding needs rather than e-tailers. The information being available for demand forecasting is only from the platform of the 3PFL service providers themselves. That leads to the consequence that demand of logistics is hard to be predicted efficiently and accurately. As a result, the capacity allocation of logistic facilities (including the pick-up store, different kinds of lockers and vehicles) in different logistics regions is a very challenging task for 3PFL service providers. On one hand, the capacities of logistics facilities should be sufficient to satisfy the customers’ demand in each logistics region. On the other hand, the amount of facilities should be minimized so as to reduce the expected cost of facilities maintenance and operations. Therefore, the optimal logistics service capacity allocation, as an operational planning function, needs to determine and allocate sufficient capacity to meet the logistics demand requirements in the next planning cycle. The 3PFL company hence needs to identify the optimal flexible logistics service capacity allocation in the distribution network economically and efficiently.

Prior studies related to the distribution network for e-commerce mostly focus on the “last mile problem”, including optimization problems on network design, transfer processes, location-inventory, or routing problem (Mofidi et al., 2018). To be specific, the location-inventory problem is to find the best locations for setting up distribution centers (DCs) in order to deliver products to customers according to the provided plan with the minimal transportation as well as stocking costs (Arabzad et al., 2015). The classic location-inventory model explored in the literature aims to find the optimal quantity and warehouse location under given locations of suppliers. A more detailed review can be found in Farahani et al. (2015). The problem explored in this paper, which focuses on service capacity allocation, aims to identify the optimal quantities of logistics resource allocation (i.e. with respect to the number of fulfillment order) in different distribution regions with uncertain demand. That can be regarded as an inventory problem with the considerations of logistics services. However, there is a limited amount of studies, especially in analytical (i.e., mathematical) modelling domain, which derives optimal decision making in service capacity allocation problems although it is critically important for cost saving in operations management. The most similar work is by Mofidi et al. (2018) who investigate an order-fulfillment resource allocation model to derive analytically the optimal sets of SKUs of logistics resources and the respective quantities to satisfy demand. A two-stage newsvendor framework is analytically proposed to determine which SKUs and in what quantities should be prepared with the goal of minimizing the additional handling cost from the operations on inbound logistics. In addition, Crainic et al. (2016) propose an analytical modeling framework introducing the logistics capacity optimization problem under volatile demand.

The traditional solution schemes for inventory control models, such as the classical newsvendor or EOQ problem (Liu et al., 2015), involve two steps: (i) demand is predicted at the first step, and (ii) the optimal quantity is then determined in the second step. However, this may lead to a serious problem that errors in the first step will create errors in the second step on inventory quantity optimization (Ban and Rudin, 2018). In fact, recently, Ban and Rudin (2018) investigate a single step solution (called data-driven newsvendor model) for the newsvendor problem. With the use of various “machine learning algorithms”, the authors obtain the optimal solution by considering the exogenous variables (such as seasonality weather, location and economic indicators) when the inventory decision is made in a single step. Following this idea, in this paper, we propose a one-step solution for the logistics service capacity allocation problem by integrating both demand uncertainty prediction and inventory decision together. In our proposed framework, the logistics capacity in different distributing region is modeled as a single period multi-product newsvendor problem, and demand distribution is obtained from real observation rather than assuming demand distribution.

Over the last few decades, various theoretical models and methods have been studied and investigated for demand forecasting in different domains. Despite achieving satisfactory results in areas such as accuracy, stability, running speed, user-friendly, and so on (Ren et al., 2017), these methods which base on single-value prediction cannot be integrated with other operational decision making process (such as inventory) easily. Sometimes, a better forecasting result of which the effectiveness is measured by error may not lead to a better operations decision (Lin et al., 2018b). Recently, a prediction interval (PI) method with an “upper bound and lower bound” consideration is suggested to be an accurate and reliable way to quantify demand uncertainty (Lin et al., 2018b). In fact, PI helps to enhance decision making by identifying the approximate optimal value within the range between the upper and lower bounds. But, this proposal has some drawbacks: (1) it cannot guarantee to find the globally optimal solution for the problem in all

---

possible scenarios; (2) it is difficult to determine parameters to guarantee the performance of the optimal process; (3) it cannot provide the absolute optimal solution via solving the newsvendor problem because no assumption of the distribution can be made in advance. To overcome these problems, in this paper, we propose a “deep-learning based” demand distribution quantifying method. In our proposal, the demand distribution can be directly obtained from the historical demand observations or transaction record. We first propose a novel deep learning forecasting system to accommodate for the operational need. Following that, uncertainty information is approximated by an effective method developed recently, based on the proposed Seq2Seq based CNN-LSTM (S2SCL) model without changing the network’s architecture.

Therefore, the research issues which we plan to address are listed below:

1. The optimal quantity decision making problem for logistics capacity allocation in the cross-border e-commerce 3PFL business operations is explored, by considering demand fulfillment and cost saving;
2. One-step based integration optimal decision making approach is proposed, which intelligently integrates inventory optimization and the demand forecasting process together;
3. The quantitative model for demand uncertainty of logistics service is developed in a real-time scenario, with the help of a novel deep learning forecasting system.

1.2. Contribution

This paper aims to address the above research issues and contributes in the following ways:

First, as far as we know, this paper is the pioneering study on the optimal logistics capacity allocation problem for cross-border e-commerce related 3PFL business operations. The capacity allocation problem derived in this work is considered as an inventory problem and the use of the single period multi-product newsvendor model is innovative. We can also find the optimal capacity allocation in different regions which helps to achieve the minimum expected cost.

Second, instead of adopting a two-step process of having demand forecasting first and then conducting inventory quantity, a deep learning based one-step intelligent optimal decision making approach is proposed by integrating inventory optimization and demand forecasting process together.

Third, we demonstrate how demand uncertainty can be quantified in a real-time scenario for logistics service demand management with the help of a novel deep learning forecasting system. This is an important contribution because most of the quantitative measurements in the literature such as interval forecasting are extracting features from a point estimation rather than training distribution from real observations.

Finally, a real case analysis is conducted to illustrate the effectiveness and robustness of our proposed deep learning based integration approach for capacity allocation problem in 3PFL operations and some important findings are also obtained.

2. Literature review

This paper relates to several fields of studies. In the following, we concisely review some related prior research and also show the literature positioning of this paper.

2.1. Forwarding Logistics Services (FLSs)

Benefiting from the rapid development of information technology (IT), cross-border e-commerce is becoming more and more popular and thriving globally (Wang et al., 2019a). With consumers’ demand escalating, competition between retailers has gradually transformed from “production related demand” to “service related demand” (e.g., logistics) (Niu et al. 2019). As an important criterion to satisfy consumers, logistics service is crucial and serves as a competitive lever for e-commerce operations (Tang and Veelenturf, 2019; Barenji et al., 2019). The international third-party-forwarding logistics (3PFL) services, which include various freight forwards, shipping forwards and airfreight forwards, have emerged as a major player in international logistics in recent years. For FLSs, the consolidation problem has been investigated in airfreight FLSs (Huang and Chi, 2007). Given the high cost charged, the authors explore the problem facing the highly complex air cargo rate. They develop a recursive algorithm, and prove that it is highly useful. Tongzon (2009) investigates the critical factors affecting the port choice in the optimal design of the freight FLSs. The author conducts a specific case study on a Southeast Asian freight forwarder and uncovers that efficiency plays a very crucial role in the decision-making process of freight FLS companies. Li and Zhang (2015) examine the price competition problem between shipping FLS firms. To be specific, the authors focus on a duopoly scenario in which two competing forwards can purchase shipping capacity from each other. They find that this arrangement can yield a win–win situation and show the respective market conditions. Our study also relates to this stream of logistics research but the specific problem is different from all prior studies.

2.2. Logistics Service Capacity (LSC) allocation problems

A 3PL service provider can help fulfill the requested demand for logistics services. This is a critical problem in the field of transportation logistics (Ren et al., 2019), warehousing, distribution, stocking and cross docking, reverse logistics, etc. The major competitive edge established by a 3PL service provider is known to mainly come from its ability to effectively integrate the related logistics services to help its clients properly manage their supply chains systems. As a special form of 3PL, the “3PFL service providers” play a central role
in managing the entire channel and thus have tremendous potential to enhance the channel's performance.

In the literature, several prior studies have reported scientific research on the “third-party warehousing operations”. For example, Chen et al. (2001) analyze the third-party warehousing contract with the considerations of commitments and revenue sharing. The authors determine the optimal capacity allocation decision. Gong and de Koster (2011) conduct a sophisticated review on various stochastic analytical models related to warehousing operations. For capacity allocation studies, in the context of maritime logistics, Mofidi et al. (2018) investigate the optimal capacity resource allocation so as to optimize the performance of an order fulfillment system. The authors reveal that a “proactive strategy” may not necessarily be a wise one as it may not outperform the commonly adopted reactive strategy. Most recently, Li and Jia (2019) analytically explore the inventory-fulfillment-allocation problem with the consideration of transshipment for e-tailers. The authors consider the situation when a nearby facility may help to fulfill orders via transshipment. From the perspective of product allocation, Holzapfel et al. (2018) establish an analytical product allocation problem. The authors develop a mixed-integer programming model to help allocate the “stock keeping units” (SKUs) to distribution centers. Considering fairness behaviors, Liu et al. (2018a,b) investigate the optimal “order allocation problem” in a logistics service supply chain system. Following that, Janjevic et al. (2019) propose an integration method for establishing a “collection-and-delivery points” system in the distribution network, by taking the changes in demand patterns into account. However, there is no doubt that the literature on the LSC allocation is very limited even though it is a problem of a very high real world relevance (Holzapfel et al., 2018). This paper hence fills this important research gap in the literature.

2.3. Demand forecasting in business operations

Supply chains face uncertainty in demand. It is crystal clear that demand forecasting is a fundamental yet challenging problem for virtually all real world supply chain systems, including supply chain with cross-border e-commerce companies (van der Laan et al., 2016). Prior research suggests that demand forecasting can yield a significant cost improvement for operations (Choi et al., 2011). Over the past few decades, a large number of forecasting methods have been developed for a wide variety of real world applications in different domains. One mainstream method for demand forecasting is based on “time series statistical prediction models”, such as ARIMA and SARIMA (Lin et al., 2013; Liu et al., 2013). These methods are easy to implement, intuitive, quick in conducting forecasting, and can be expressed in closed-form which means they can be easily integrated into other decision making models in operations (e.g. inventory problems). However, time series forecasting methods usually present unsatisfactory performance when demand is non-linear, unstable or largely dependent on exogenous factors that cannot be captured by simple statistics. Another category of forecasting methods is by machine learning, which can take the exogenous impact factors into account. For example, the Artificial Neural Networks (ANN) together with the improved neural networks (NN) approaches (Luo et al. 2019) are widely used for demand prediction (Yu et al., 2011). Although machine learning methods commonly can obtain good performance, they are time consuming for model training and the need of having sufficient training data is non-trivial (Ren et al., 2014).

To obtain a more accurate and reliable forecasting result, there is a proposal of using the prediction interval (PI) forecasting method in which both the upper bound and lower bound are developed. The PI forecasting model has been applied in various industrial domains, such as traffic, medical, power system, and so on (Lin et al., 2018b) with good results. Different from point forecasting, PI forecasting aims to capture demand uncertainty within a given interval with a given probability. This kind of forecasting is hence very much useful in decision making. One of the most crucial applications of demand forecasting in operations management is to assist decision making in inventory management (Choi et al., 2019). Following a recent influential study by Ban and Rudin (2018), we know that there are several main approaches to help model the decision making under demand uncertainty, such as the Bayesian approach, and the data-driven approach (Choi and Luo 2019). This paper belongs to the data driven approach, and the specific problem we attempt to address is new.

Finally, this paper explores the cross-border e-commerce problem in logistics. This is an under-explored area and interested readers can refer to Liu et al. (2018a,b) for more details. We also examine some related studies in Section 3, when we introduce the problem under exploration in this paper.

3. Cross-border E-commerce 3PFL order fulfillment system

In this section, we present the fulfillment process of cross-border 3PFL and describe the logistics service capacity (LSC) allocation problem with respect to different regions in a distribution network. Firstly, we describe the basic order fulfillment process of cross-border 3PFL and identify some challenges in the respective fulfillment system. Then, our proposed framework aiming to improve the operational efficiency from cost view will be presented.

3.1. Order fulfillment process

Compared with the traditional “domestic e-commerce operations”, the order fulfillment process of cross-border e-commerce is much more complicated (Choi et al., 2013). It typically involves international transportation, customs declaration, and inspection that altogether would lead to great challenge for logistics operations. As a professional logistics service provider, 3PL companies are an ideal choice to help undertake transportation, facility arrangement, distribution logistics and “customs declaration and inspection” in an efficient way (Liu et al., 2015a). Recently, cross-border e-commerce experienced an unprecedented and rapid development, which gives the 3PL companies a huge potential market. As a special kind of 3PL, 3PFL has taken up a growing share of the cross-border e-commerce logistics market, especial for the China market, including both “inbound and outbound” e-commerce. Different
from the traditional 3PL service providers, the 3PFL service providers first receive orders from e-commerce customers who buy products from different online stores or even different platforms. Then, they pack various products as one package and deliver to customers’ hands. On customer side, it’s a clearly convenient and cost saving way to receive orders by 3PFL service providers. On supply chain side, consolidating various international orders belonging to the same customer from different e-tailers as “one parcel” can help substantially enhance the operation efficiency. The order fulfillment process of 3PFL is illustrated in Fig. 1. Comparing with the conventional 3PL, the main feature of 3PFL fulfillment is that the orders arrive randomly and both the arrival time of last order and the amount of orders waiting to be packaged are unknown. As a result, the operational decision making process is challenging for 3PFL systems as what we elaborate further in the following two sub-sections.

3.1.1. Warehouse management

Nowadays, warehouses are designed to achieve various functions, which include fast unit-load operations and efficient, responsive, and flexible order fulfillment operations (Tappia et al. 2019). The order fulfillment process for 3PFL starts from receiving orders online and ends on distributing orders to customers’ hands either by “pick up point” or “delivery to door”. Different from the traditional 3PL service that is able to arrange the order fulfillment process according to the information shared from the e-tailers, the 3PFL service provider faces great challenges in warehouse operations and delivery planning because the online orders are highly volatile and they arrive randomly. Warehousing operations management for 3PFL is scaled and customized with respect to customer needs including the packing requirements and delivery requirements for their orders. There are two main tasks for warehousing operations, namely layout optimization and operational efficiency improvement (Wang et al., 2015). Layout optimization is to optimize space utilization and equipment utilization in the orders receiving, storing, packing, picking/retrieving and shipping process. While the operational efficiency improvement aims to enhance customer services during the order proceeding process. However, due to the great uncertainty arising from the time and amount of customers’ last order, 3PFL companies commonly face enormous challenges in warehouse operations management. For layout planning, it makes the estimation of space requirement and the layout arrangement difficult that the packing orders arrive randomly. According to Little’s Law, the average space required is difficult to determine. Random arrival of the last packing order leadsto the unpredictable inventory time of the previous arriving orders that wait to be packaged. As a result, it becomes very difficult to properly arrange the inventory receiving, storage and even the warehouse operations management.

3.1.2. LSC allocation problem in a distribution network

In the e-commerce “order fulfillment process”, it is the last and the least efficient stage to deliver the parcels to the hands of consumers. How to improve the efficiency of this stage in an economically sound way is commonly called the “Logistics Last Mile Problem” (LLMP) (Deutsch and Golany, 2018). As the most expensive part of the whole logistics process, LLMP has aroused a great number of studies from different view-points. One key factor in this process is the procurement of sufficient LSC at different geographical locations in the network so as to satisfy the volatile demand at a lower level of expected cost. For 3PFL companies, it is in fact very challenging to decide the optimal capacity by balancing demand satisfaction and cost saving.

A. Highly volatile demand

Demand is highly unstable and volatile, mainly because of the unknown time and amount for the last order arrival. The uncertain demand leads to the fact that the number of orders proceeded are imbalanced among different stores in the vast distribution network.
The stores in busy areas may face a large amount of proceeding orders, while the stores in leisure areas may face a very small order volume.

B. The costly rental of pick up stores

“Pick up (in) store” is the traditional and most commonly adopted way for distributing parcels to customers. However, the rental fees of stores are very costly in places such as the city centers. “Parcel locker” is another commonly used solution for the last mile logistics in reducing the high cost of home delivery and transport operators. Parcel locker refers to a group of lockers, usually placed in blocks of different apartments, or even near some traffic stations like metro. In some cases, the lockers have electronic device supported locks. So, they can be used by different consumers, at any time it is convenient to them (Deutsch and Golany, 2018).

From the perspective of the 3PFL company, parcel locker is a promising facility to address the “last-mile logistics problem” by reducing the number of vehicles, the cost of operating pick-up store rental operations, and hired deliverers needed to cover a geographic service zone. However, the acceptance of collecting parcels from parcel lockers is closely related to the distance from home to the site of parcel locker. In the literature, Deutsch and Golany (2018) suggest that the percentage of customers who would agree to travel and collect their parcels from parcel lockers other than waiting at home decreases with the distance. As a result, for the 3PFL company, the efficiency improvement of logistics network by using “parcel lockers” to replace “pick up store” will be at the expense of losing potential customers who do not agree to travel for picking up the merchandise.

Therefore, it is challenging for the 3PFL company to decide how many stores to operate (and close), how many parcel locker sites to open, and the optimal number of logistics service capacity allocated in each logistics region by considering the uncertainty demand and operational cost.

3.2. Deep learning based integration solution

The problem investigated in this study is to decide the optimal LSC in the given logistics region (in other words, “which” and “how many” different warehouses should be established from many warehouse candidates). The LSC is composed of multiple warehouses such as the pick up stores, reception boxes, parcel lockers, and so on, which are assumed to be able to satisfy the logistics demand in each region. Therefore, the aim of LSC allocation in a 3PFL distribution network is to pursue a better balance between the proceeding capacity allocation and uncertain market demand for each region in order to minimize the expected cost. It can be considered as an inventory problem and modelled as the newsvendor problem.

In the traditional newsvendor model, it is assumed that demand follows a certain distribution such as the normal distribution with the mean and variance being updated by the latest information (if we follow the Bayesian approach). The newsvendor problem is commonly solved in a “two-step process”: First, estimate a demand distribution (e.g., based on statistics or advanced deep learning based model (Lin et al., 2018a, 2018b). The optimal order quantity is then optimized in the second step by finding the “critical fractile solution”. The important drawback of this kind of two-step solution approach is that errors in the first step will lead to problems in the optimization step (Ban and Rudin, 2018). In order to overcome this drawback and for the demand model to be “standardized” easily, Ban and Rudin (2018) innovatively propose a single step solution for the newsvendor problem using the Empirical Risk Minimization (ERM) and Kernel-weights Optimization (KO) methods. They also argue and find that their proposed method performs better than the best practice benchmark approach. Following the solution approach in Ban and Rudin (2018), we propose a one-step decision making process for the LSC allocation problem stated in this work, in which the optimal quantity is determined by a deep learning algorithm without finding the demand distribution first.

4. Analytical models

4.1. LSC allocation model

The notation and mathematical expressions used for the development of optimal LSC models are given in Table 1. It is assumed that the 3PFL company wants to make decisions on the logistics service procurement with uncertain logistics service demand (LSD) by minimizing the expected cost in a single ordering period (Chiu and Choi, 2016) according to Eq. (1):

$$K(q^o_i) = c_j + c_i(q^o_i - q'_i) + p E[\max(y'_i - q^o_i, 0)] + h E[\max(q^o_i - y'_i, 0)].$$  

(1)

In $K(q^o_i)$, the “first order loss function” $E[\max(y'_i - q^o_i, 0)]$ captures the expected shortage LSC; its complement, $E[\max(q^o_i - y'_i, 0)]$, denotes the expected product quantity in stock at the end of the period (Chiu and Choi, 2016). On the basis of this expected cost function, we note that the determination of the optimal LSC level is simply an expected cost minimization problem. As the expected cost function is strictly convex, the optimal LSC which is equivalent to the “optimal newsvendor quantity” can thus be solved by the standard first order condition following the classical newsvendor problem:

$$q^* = F^{-1}\left(\frac{p' c_i}{p' + h}ight),$$  

(2)

where $F^{-1}$ denotes the inverse cumulative distribution function of $y'_i$. 

---

*Note: The above text is a continuation of the previous discussion, focusing on the analysis and optimization of logistics service capacity allocation in a 3PFL distribution network.*
Table 1  
Definitions of parameters.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>( c_j )</td>
<td>The fixed cost in region ( i ). This cost always exists for the 3PFL company such as the cost related to pick up stores rented in different regions.</td>
</tr>
<tr>
<td>( c'_j )</td>
<td>The variable cost in region ( i ). This cost expresses the logistics service cost related to such as the delivery vehicle and self-pick up locker.</td>
</tr>
<tr>
<td>( q^*_i )</td>
<td>The quantity of LSC for the 3PFL company in region ( i ) at time step ( t ). This parameter includes the initial LSC such as the pick up stores rented in different regions as well. If no more logistics service is procured, then this quantity is equal to the initial quantity of LSC.</td>
</tr>
<tr>
<td>( q^*_i )</td>
<td>The initial quantity of LSC for the 3PFL company in region ( i ). We assume that the 3PFL company has a fixed level of LSC at the beginning of each LSD period.</td>
</tr>
<tr>
<td>( p' )</td>
<td>The penalty cost in region ( i ). If there is less procured LSC than needed to satisfy the demand, this is the penalty cost of the unsatisfied demand.</td>
</tr>
<tr>
<td>( y[i] )</td>
<td>A random variable with the cdf ( F ) representing the uncertain LSD in region ( i ) at time step ( t ).</td>
</tr>
<tr>
<td>( \hat{y}^i )</td>
<td>The expected value of random variable ( y[i] ).</td>
</tr>
<tr>
<td>( h^i )</td>
<td>The over procured LSC cost in region ( i ).</td>
</tr>
<tr>
<td>( x_{ip} )</td>
<td>The LSD input vector.</td>
</tr>
<tr>
<td>( c_1, \ldots, c_q )</td>
<td>Vector output from the 1st convolutional sub-layers to the ( q )th sub-layers in CNN layer.</td>
</tr>
<tr>
<td>( b )</td>
<td>The bias for each feature map in CNN layer.</td>
</tr>
<tr>
<td>( w )</td>
<td>The weight of the kernel in CNN and LSTM layers.</td>
</tr>
<tr>
<td>( m )</td>
<td>The index value of the filter in CNN layer.</td>
</tr>
<tr>
<td>( \phi )</td>
<td>The activation function in CNN layer.</td>
</tr>
<tr>
<td>( p )</td>
<td>The pooling vector output from each neuron cluster in the previous layer in CNN layer.</td>
</tr>
<tr>
<td>( T )</td>
<td>The pooling stride that decides how far to move the area of input data in CNN layer.</td>
</tr>
<tr>
<td>( R )</td>
<td>The pooling size of the last input in CNN layer.</td>
</tr>
<tr>
<td>( t^i )</td>
<td>The input gate in LSTM layer.</td>
</tr>
<tr>
<td>( F_t^i )</td>
<td>The forget gate in LSTM layer.</td>
</tr>
<tr>
<td>( F_t^i )</td>
<td>The output gate in LSTM layer.</td>
</tr>
<tr>
<td>( h^i )</td>
<td>The hidden state in LSTM layer.</td>
</tr>
<tr>
<td>( c_t )</td>
<td>The memory cell state in LSTM layer.</td>
</tr>
<tr>
<td>( u_{ij}^k )</td>
<td>Similarity measuring vector in Seq2Seq decoder.</td>
</tr>
<tr>
<td>( W_t^j )</td>
<td>Trainable weight matrix in Seq2Seq decoder.</td>
</tr>
<tr>
<td>( q_t^j )</td>
<td>Dimension adjusting vector in Seq2Seq decoder.</td>
</tr>
<tr>
<td>( h_{l-1}^i )</td>
<td>The encoder hidden state in Seq2Seq decoder.</td>
</tr>
<tr>
<td>( \tilde{h}_{l+1}^i )</td>
<td>The attention hidden state in Seq2Seq decoder.</td>
</tr>
<tr>
<td>( s_{k+1}^j )</td>
<td>The attention weighted state of the keys in Seq2Seq decoder.</td>
</tr>
<tr>
<td>( W_r )</td>
<td>The weighted parameter matrix in attention in Seq2Seq decoder.</td>
</tr>
<tr>
<td>( a_{l-1}^j )</td>
<td>The normalization weight coefficient of ( u_{ij}^k ) in Seq2Seq decoder.</td>
</tr>
<tr>
<td>( b_i^j )</td>
<td>The intercept parameter in Seq2Seq decoder.</td>
</tr>
<tr>
<td>( \hat{y}_{l+1}^i )</td>
<td>The point forecasting value of the S2SCL model.</td>
</tr>
</tbody>
</table>

4.2. Deep learning based LSD forecasting and uncertain quantifying system

The key of solving the newsvendor problem defined in Section 4.1 is to know the dynamic Gaussian distribution of the LSD with mean \( E[y^*_i] \) and variance \( \text{Var}[y^*_i] \) at each time step. We therefore utilize a deep learning-based LSD forecasting and uncertainty quantifying system, which can forecast the LSD as well as its dynamic mean \( E[y^*_i] \) and variance \( \text{Var}[y^*_i] \) so as to make an optimal decision for the LSC \( q^*_i \).

We call the respective system the S2SCL system. Our proposed S2SCL system includes the following two tasks: Firstly, considering the high uncertainty for the LSD of 3PFL companies, the S2SCL system is proposed to make the point forecasting according to the operational need. Then, in order to quantify the uncertainty of the LSD in real-time, the S2SCL is integrated with a recently proposed “dropout method” (Gal and Ghahramani, 2016) to help approximate “uncertainty information” from the proposed S2SCL model without changing the deep learning network’s architecture.

A. A novel deep learning forecasting system

Considering that the supply chain data represents a multivariate time series of LSD variables which in turn could be used to model and forecast the LSD, the “sequence to sequence architecture” (Seq2Seq) is good at mapping a fixed length input data set with a different length output data set (Hao et al., 2019). For example, it can utilize the 3PFL company’s LSD data from different regions to forecast LSD in a specified region. Besides, the supply chain data usually shows not only temporal but also spatial patterns. Similar to Ren’s PDPF system (Ren et al., 2014), we propose a “hybrid deep learning method” consists of “Convolutional Neural Networks (CNN)” & “Long Short-Term Memory (LSTM)” network to simultaneously capture the complex spatial dependencies (similar to a deep learning version of panel data) and non-stationary temporal dynamics (similar to a deep learning version of particle filter) in 3PFL.
demand forecasting. The overall architecture for the proposed “A deep learning based one-step integration optimal decision making approach S2SCL” is shown in Fig. 2.

From the right hand side part of Fig. 2, we can clearly observe the Seq2Seq architecture, which encodes the source sequence (weekly) into a “context vector” (CV) and decodes the CV into the target sequence (daily). The input datasets in different regions are encoded as a weekly panel containing not only spatial but also temporal patterns. The left hand side part of Fig. 2 is the CNN and the LSTM network, which are developed for taking-in the input spatial-temporal data and outputting the forecasting information to the Seq2Seq structure for further processing. To sum it up, the encoder of the proposed S2SCL system adopts a CNN-LSTM network, which receives data and produces a CV and the context vector acts as the “last hidden state” of the encoder. The decoder of the S2SCL takes the forecasting value of the previous time step as well as the CV as the input, and output the future values in different regions (according to the operational need) in parallel.

a. CNN-LSTM module

In order to incorporate the “weekly effect” into the model, the encoder of S2SCL first applies a simple sliding window algorithm (Shao et al., 2019) to reorganize the multivariate input data into the weekly window. Therefore, the input is a three-dimensional matrix with both the temporal and spatial features. Then the CNN layer mainly applies the “convolution mathematical operation” to help extract the spatial characteristics, and pool the results to the “LSTM layer” with noises removed and parameters reduced. The final step in the CNN-LSTM module is to capture the complex time-series dynamics and model the irregular patterns in the LSTM layer.

The CNN layer consists of convolution sub-layers and pooling sub-layers (Wang et al., 2019). Eqs. (3) and (4) are the results of the vector $c_{ij}^l$ output from the 1st convolutional sub-layers to the $l$th sub-layers (Kim et al., 2019).

$$c_{ij}^l = \varphi \left( b^l_j + \sum_{m=1}^{M} w_{m,j} x_{ij}^{l+m-1} \right),$$

(3)

$$c_{ij}^l = \varphi \left( b^l_j + \sum_{m=1}^{M} w_{m,j} x_{ij}^{l+m-1} \right),$$

(4)

where $x_{ij}^{l} = [x_i, x_j, \ldots, x_n]$ is called the LSD input vector. $i$ denotes the region, and $n$ represents the number of weeks, $c_{ij}^l$ is determined by the “output vector” $x_{ij}^l$ of the former layer, $b^l_j$ denotes the “bias” for the $j$th feature map, $w$ denotes the “weight of the kernel”, $m$ is the “index value of the filter”, and $\varphi$ is the “activation function” like the Linear Rectifier Unit (ReLU).

Note that the convolution sub-layer is the pooling sub-layer, which can enhance the efficiency and achieve simplicity of the CNN-
LSTM network. The idea of our proposed model is to combine the output of a “neuron cluster” in one layer into a neuron for the next layer so as to reduce the space size and network computation costs. Eq. (5) shows the operation of the max-pooling layer. denotes the stride that is used to decide how far to move the area of input data, and is the pooling size which is less than the size of the last input .

Define:

\[
p^1_j = \max_{r \in \mathbb{R}} c^1_{i-1,j+r}^1
\]

where is the pooling size, is the stride that decides how far to move the area of input data. The max-pooling selects the maximum value from each neuron cluster in the previous layer so as to address the over-fitting problem. The result obtained in the CNN layer is then passed to the LSTM layer, where it is mainly responsible for storing the temporal details of the LSD extracted from the CNN (see Kim et al., 2019 for more details).

The idea of adding the LSTM layer is to bridge long-term memory units to update the previous hidden state (Lin et al., 2018) and to realize the temporal features learning through three gate units: “input, output, and forget gate”. Eqs. (6)–(8) show the required operations of these three gates which help establish the “LSTM”:

\[
\begin{align*}
I^1_i &= \text{ReLU}(w_{ip}p^1_i + w_{ih}h^1_{i-1} + w_{ic}c^1_{i-1} + b), \\
F^1_i &= \text{ReLU}(w_{fp}p^1_i + w_{fh}h^1_{i-1} + w_{fc}c^1_{i-1} + b), \\
O^1_i &= \text{ReLU}(w_{op}p^1_i + w_{oh}h^1_{i-1} + w_{oc}c^1_i + b),
\end{align*}
\]

where is the “weight matrix of each gate unit” and includes the features of LSD. The hidden state and the memory cell state are updated every step with activation of every gate controlled by a continuous variable, bound between 0 and 1; denotes the bias vector (Kim et al., 2019).

\[
\begin{align*}
c^1_i &= f^1_i c^1_{i-1} + I^1_i \text{ReLU}(w_{pc}p^1_i + w_{ph}h^1_{i-1} + b), \\
h^1_i &= O^1_i \text{ReLU}(c^1_i).
\end{align*}
\]

Eqs. (9) and (10) elaborate on how the “cell states” and the “hidden states” update through the above mentioned three gates. It should be noticed that the LSTM layer adopts the ReLU as activation function to minimize the risk of gradient vanishing. In general, the CNN-LSTM network predicts LSD in different regions.

b. Seq2seq encoder & the context vector

The encoder of Seq2Seq architecture takes the multiple-dimensional structure of the input vector from the CNN-LSTM network. How the encoder part operates is demonstrated in Eq. (11). At the time step for, the previous hidden status is sent to the current “time stamp” and uses the input to determine (Zhang et al., 2019).

\[
h^1_{i,j} = \begin{cases} 
\text{Cell}_{encoder}(h^1_{i-1,j}, O^1_{j-1}), & j = m \\
\text{Cell}_{encoder}(h^1_{i-1,j}, O^1_{j}), & j \in [0, \ldots, m - 1] 
\end{cases}
\]

where is called the “initial hidden status”; denotes the encoder (see Zhang et al., 2019 for more information).

Another key component of the Seq2Seq architecture is the context vector, which is shown in Eq. (12). It serves as a bridge between the encoder and the decoder, which stores all the encoder’s details and information.

\[
c = h^1_i.
\]

c. Seq2seq decoder & the attention state

The decoder takes the forecasting value generated from the CNN-LSTM of previous “time step” and the CV as the input. Furthermore, the attention hidden state was attached to the hidden state in the model. As given by Eqs. (13)–(15) below, the attention function, is expressed in the following,

\[
u^k_{i,j} = q^T \tanh(h^1_{i,j} W_f h^1_{i,k}), \quad k = 0, 1, \ldots, m,
\]

\[
a^k_{i,j} = \text{soft} \max(u^k_{i,j}), \quad k = 0, 1, \ldots, m,
\]

\[
S^k_{i,j} = \sum_{k=1}^{m} a^k_{i,j} h^1_{i,k}.
\]

where measures the “similarity” between and determined by Eq. (15), and in this study, we employ the “Luong Attention” form (Luong et al., 2015) as the compatibility function with the trainable weight matrix and vector to help adjust the dimension of the associated result; is the normalization of and is further used as the weight coefficient with the
corresponding encoder hidden state $h_{t+j}^i$ to calculate $S_{t+j}$.

$$
\hat{r}_{t+j}^i = \tanh(W^e_r [S_{t+j}^i; h_{t+j}^i]),
$$

(16)

$$
\hat{y}_{t+j}^i = W^f \hat{r}_{t+j}^i + b^f_i.
$$

(17)

As shown in Fig. 2, the attention hidden state $\hat{h}_{t+j}^i$ includes the “attention vector” $S_{t+j}$ and the “original hidden state” $h_{t+j}^i$, as shown in Eqs. (16) and (17). An immediate result of the forecasting value $\hat{y}_{t+j}^i$ becomes the forecasting result of the next time step. Therefore, we clearly know that seq2seq can enjoy the “sequence relationship among predictions”.

B. Quantitative measurement of uncertainty

The most common approach to quantitatively measure the uncertainty is to apply the softmax function to get probabilities (Vlahogianni et al., 2007). Gal and Ghahramani (2016) point out that this approach is far from perfect because it tends to yield a big training bias. In other words, most of the quantitative measurements such as interval forecasting are extracting features from a point estimation rather than from the training data (Fei et al., 2011). In this paper, our target is not only on forecasting the point value or measuring the uncertainty of LSD, but also deciding the optimal value of LSC. It should be noticed that our proposed S2SCL is an integrated solution approach, which is different from other quantitative measurement such as interval forecasting that requires another step to translate the LSD forecast into LSC quantities.

A recent approach proposed by Gal and Ghahramani (2016) helps estimate the model uncertainty by a dropout model. Dropout is a regularisation technique which was first proposed to resolve the co-adaptation problem; its idea is to turn-off some of the hidden units with some probability $p$, namely the dropout probability (Hinton et al., 2012). Eq. (18) shows how forecasting could be conducted with a “Gaussian process”, $f$ is the function space. To be specific, we have the following details:

$X$ is the input dataset, which includes all historical $x_t$, and $Y$ is the training output dataset including all $y_t$. The expectation of $y^*$ is called the forecasting demand mean (see Gal and Ghahramani, 2016 for more information).

$$
p(y^*_t|x_t^*, X, Y) = \int p(y^*_t|f^*)p(f^*|x_t^*, X, Y)df^*.
$$

(18)

Analytically, we know that (19) and (20) are employed to obtain the approximate “mean” and “variance of ” $y_t^*$:

$$
E(y_t^*) \approx \frac{1}{T} \sum_{i=1}^{T} \hat{y}^i_t(x_t^*).
$$

(19)

$$
\text{Var}(y_t^*) \approx \tau^{-1}I_D + \frac{1}{T} \sum_{i=1}^{T} \hat{y}^i_t(x_t^*)\hat{y}^i_t(x_t^*) - E(y_t^*)^TE(y_t^*).
$$

(20)

Following Gal and Ghahramani (2016), we define $\tau$ as follows which is used to capture the model precision:

$$
\tau = \frac{\mu^2p}{2NA}.
$$

(21)

The motivation behind modelling the uncertainty of LSD is to make the optimal decision of the LSC. A benefit for applying S2SCL into supply chain analysis is to avoid the propagation of errors (see McAllister et al., 2017) and relates to the presence of the bullwhip effect (see, e.g., Choi et al., 2013). As a consequence, at time stamp $t$, the S2SCL is able to output a Gaussian distribution with a dynamic mean $E(y_t^*)$ and variance $\text{Var}(y_t^*)$ of the LSD besides its point forecasting results.

5. Case study

To show the performance of the proposed S2SCL model, we implement our models on a real-world logistics dataset. This section will introduce the dataset and the corresponding preliminary exploration first, following that, our proposed models will be built and compared computationally with other models from aspects of both point prediction and uncertainty quantification. The LSC optimization results based on various demand prediction results are then evaluated as well.

5.1. Dataset and data analysis

In this paper, to show the effectiveness of our proposed approach, we investigate the real daily LSD data of a 3PFL company A in 3 regions, namely New Territories (NT), Kowloon (KLN), and Hong Kong Island (HKI) in Hong Kong SAR, the data of which under examination covers the years 2017 and 2018 (730 days in total). Company A is one of the most popular e-commerce logistics platforms in Hong Kong, which has two main services: (1) consolidation forwarding; and (2) purchasing agent. For consolidation forwarding, the company consolidates and forwards customers’ order by providing customers a shipping address to receive the goods delivered from different e-platforms/countries. Data of the first 660 days from the three regions are used as the training data, and the last 70 days are used as the testing data.

From Fig. 3, it could be observed clearly that the sequential demand data exhibits temporal dependencies. Thus, deep learning
based modules (e.g., LSTM) that are good at processing sequential data and able to capture temporal features are much more powerful than convention models.

Except the temporal dependencies, the correlation test results show that those demand data in the three regions are spatially correlated. The test results can be observed in Table 2:

To investigate the correlation among those regions further, a Granger Causality Test (see Ren et al., 2014 for more details) is adopted, and results are shown in Table 3. From the computational results, it can quickly be concluded that demands in different regions exhibit spatial dependencies substantially. Thus, deep learning based modules (e.g., CNN) that are good at processing sequential data and able to capture spatial features are more suitable than convention models.

5.2. Point forecasting comparison

The preliminary exploration of the data shows the existing of temporal and spatial dependencies in the dataset and helps justify the design of our S2SCL deep learning model. The following discussions will further investigate and study the performance of the proposed model with other state-of-the-art machine learning models.

The datasets of the first 730 days operations from the three regions are taken as the training data, and the last 70 days are the testing data. We have conducted computational experiments to show that the proposed method outperforms other commonly adopted techniques in both point forecasting and optimal LSC allocation decision. The statistics based integration approach ARIMA (Choi et al., 2013), optimization based approach PSO-ELM (Lin et al., 2018b) and deep learning based integration approach S2SCL are adopted for time series based forecasting. These choices are not taken arbitrarily. According to Lin et al. (2018b), ARIMA has been

Table 2

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Region</th>
<th>HKI</th>
<th>KLN</th>
<th>NT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariance</td>
<td>HKI</td>
<td>1078.699</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correlation</td>
<td></td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>t-Statistic</td>
<td></td>
<td>–</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Probability</td>
<td></td>
<td>–</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Covariance</td>
<td>KLN</td>
<td>4295.30</td>
<td>21376.17</td>
<td></td>
</tr>
<tr>
<td>Correlation</td>
<td></td>
<td>0.89</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>t-Statistic</td>
<td></td>
<td>53.95</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>Probability</td>
<td></td>
<td>0</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>Covariance</td>
<td>NT</td>
<td>1827.37</td>
<td>8285.24</td>
<td>4374.39</td>
</tr>
<tr>
<td>Correlation</td>
<td></td>
<td>0.84</td>
<td>0.86</td>
<td>1.00</td>
</tr>
<tr>
<td>t-Statistic</td>
<td></td>
<td>41.95</td>
<td>44.80</td>
<td>4374.39</td>
</tr>
<tr>
<td>Probability</td>
<td></td>
<td>0</td>
<td>0</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Fig. 3. LSD differenced data in different regions.
applied in both short term point forecasting and interval prediction. In our problem, the predictive “mean” and “variance” of the demand are used to help decide the optimal LSC via solving the newsvendor problem at each time step. The PSO-ELM requires neither “statistical inference” nor assumption on demand distribution. It focuses on deriving the prediction intervals (PIs) to minimize a bi-objective function, in which the objectives are the “reliability” and “interval sharpness”. The optimal LSC can be searched within the PIs at each time step.

Table 4 compares the LSD forecasting performance of the statistics based integration approach ARIMA, optimization based approach PSO-ELM, and deep learning based integration approach S2SCL. Experimental results using real data demonstrate that the proposed S2SCL model achieves a much better performance than ARIMA and PSO-ELM in all 3 regions under study. In fact, the proposed S2SCL system can improve the ARIMA and PSO-ELM’s forecasting RMSE by at least 19.32% and 14.77%, respectively. This is a significant improvement.

Fig. 4 plots the predicted results of the three methods examined above. These results show that the S2SCL method can capture the variation of LSDs better than other methods. Observe that ARIMA fails to provide a good forecast since the non-linearity and the presence of correlated items are ignored. The result shows that modeling performance for the variation characteristics of LSDs is poor. Further observe that the hybrid PSO-ELM approach, which has utilized (i) the “heuristic and population based optimization capability of PSO” and (ii) ELM, also does not perform better than our proposed S2SCL. Although the PSO-ELM can well model the regular trends of LSDs in all three regions, it fails to predict most of the irregular spikes. The proposed S2SCL well models the complex time-series pattern. In addition, it helps to predict the most “frequently appeared” peak which occurs frequently. The reason that the proposed S2SCL system predicts the variation characteristics well might be due to the fact that the deep learning based approach can well predict the systems dynamics clearly.

---

### Table 3
Results of “Granger Causality Tests”.

<table>
<thead>
<tr>
<th>Pairwise “Granger Causality Tests”</th>
<th>Sample: 1/01/2017 10/22/2018</th>
<th>Lags: 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>The “Null Hypothesis”</td>
<td>No. of Observations</td>
<td>F-Statistic</td>
</tr>
<tr>
<td>HKI → ALL</td>
<td>660</td>
<td>0.24636</td>
</tr>
<tr>
<td>ALL → HKI</td>
<td>660</td>
<td>6.64655</td>
</tr>
<tr>
<td>KLN → ALL</td>
<td>660</td>
<td>1.38643</td>
</tr>
<tr>
<td>ALL → KLN</td>
<td>660</td>
<td>3.49777</td>
</tr>
<tr>
<td>NT → ALL</td>
<td>660</td>
<td>1.60448</td>
</tr>
<tr>
<td>ALL → NT</td>
<td>660</td>
<td>15.3201</td>
</tr>
<tr>
<td>KLN → HKI</td>
<td>660</td>
<td>4.96882</td>
</tr>
<tr>
<td>HKI → KLN</td>
<td>660</td>
<td>0.48203</td>
</tr>
<tr>
<td>NT → HKI</td>
<td>670</td>
<td>7.14500</td>
</tr>
<tr>
<td>HKI → NT</td>
<td>670</td>
<td>7.36436</td>
</tr>
<tr>
<td>NT → KLN</td>
<td>670</td>
<td>4.34395</td>
</tr>
<tr>
<td>KLN → NT</td>
<td>670</td>
<td>15.8279</td>
</tr>
</tbody>
</table>

*Note that “X → Y denotes X does not Granger Cause Y”.

### Table 4
Point forecasting results by different methods.

<table>
<thead>
<tr>
<th>NT</th>
<th>MSE</th>
<th>MAE</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIMA</td>
<td>9514.09</td>
<td>76.20</td>
<td>97.54</td>
</tr>
<tr>
<td>PSO-ELM</td>
<td>8890.70</td>
<td>71.38</td>
<td>94.29</td>
</tr>
<tr>
<td>S2SCL</td>
<td>6193.34</td>
<td>53.98</td>
<td>78.70</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>KLN</th>
<th>MSE</th>
<th>MAE</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIMA</td>
<td>17425.15</td>
<td>102.16</td>
<td>132.00</td>
</tr>
<tr>
<td>PSO-ELM</td>
<td>19131.94</td>
<td>116.08</td>
<td>138.32</td>
</tr>
<tr>
<td>S2SCL</td>
<td>8888.69</td>
<td>65.44</td>
<td>94.28</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>HKI</th>
<th>MSE</th>
<th>MAE</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIMA</td>
<td>1091.59</td>
<td>25.16</td>
<td>33.04</td>
</tr>
<tr>
<td>PSO-ELM</td>
<td>974.98</td>
<td>25.82</td>
<td>31.22</td>
</tr>
<tr>
<td>S2SCL</td>
<td>708.27</td>
<td>19.62</td>
<td>26.61</td>
</tr>
</tbody>
</table>
Fig. 4. Point predicted results of different methods.
5.3. LSC allocation

Another main contribution of our study is the proposed of an integration solution approach to quantitatively measure the uncertainty of LSDs so as to decide its optimal LSC allocation in three regions simultaneously at each time step. At the time stamp $t$, the S2SCL is able to output the cost-optimal quantity of LSC $q^*$ by solving a classical newsvendor problem as defined in Equation (2) based on the Gaussian distribution with a dynamic mean $E(y^*)$ and variance $\text{Var}(y^*)$ of the LSDs in the three regions. The results of the optimal quantity of LSC using the proposed S2SCL are presented and compared against two different approaches: statistics based integration approach, and optimization based approach.

**Compared with statistics based integration approach.** As the structure is extracted from the “raw training dataset” (Guo et al. 2014), the classical statistical forecasting method ARIMA is chosen as a benchmark method to highlight the proposed deep learning based integration approach S2SCL’s performance. Similar to our proposed S2SCL model, ARIMA is able to generate a demand distribution as well, which can be used to solve for the optimal solution in newsvendor model.

**Compared with the optimization based approach.** Different from ARIMA and S2SCL, the machine learning based approach PSO-ELM firstly predicts the point forecasting and the interval of LSD, then searches the optimal LSCs within the interval. It’s difficult for PSO-ELM to assume and extract the demand distribution of LSD, that means the PSO-ELM based approach has to optimize and determine the solution in a two-step process.

It’s interesting to investigate whether a better PIs prediction result yields a better LSC decision. Before reviewing the results of the optimal LSC solved by different approaches, we conduct a comparison for PIs with 95% Prediction Interval Nominal Confidence (PINC) level. Assuming that the LSDs follow Gaussian distributions, the PIs of S2SCL and ARIMA are simply calculated as $\text{PI}_i^t = [y_i^t - 1.96 \times \text{Var}(y_i^*), y_i^t - 1.96 \times \text{Var}(y_i^*)]$ (Lin et al., 2018b).

From Fig. 5, it can be noticed that the PI’s ranges (from upper PI to lower PI) of S2SCL are the narrowest and the ones of ARIMA are the widest. Not surprisingly, in our analysis, the “PI Coverage Probability” (PICP), which is determined by the “70 observations” in the dataset within the PIs, has a negative correlation with the PI’s ranges (we call the PI’s range the PI length (PIL)). While ARIMA always yields the highest PICP, it has the largest PICP and mean PI length (MPIL) among the three models. Although S2SCL is used to generate a narrower interval, it can cover some “0” value without reducing the lower PI to a negative value, as ARIMA does, which makes no sense in real practice. Table 5 summarizes PICP and MPIL of these three methods. The PICP and MPIL mentioned here refer to the average distances between the “upper bound” and “lower bound” of the PIs which help us clearly show the effectiveness of PSO-ELM in predicting the PIs. It can be concluded that PSO-ELM performs best with generating a reasonable PICP, and at the same time it can also keep the MPIL as narrow as possible.

The results of the optimal quantity of LSC based S2SCL are presented and compared against different approaches including ARIMA and PSO-ELM. It should be noticed that the optimal LSCs of ARIMA and S2SCL are generated according to Eq. (4) with the point forecasting result simultaneously. However, the optimal LSCs of PSO-ELM are searched and identified from the predictive PIs. The optimal LSCs of ARIMA, PSO-ELM and S2SCL and the 70 daily LSDs (indicated as black bars) in the three regions are shown in Fig. 6. The results uncover several insights (i) ARIMA tends to provide a stable LSC allocation gradually even the LSDs fluctuate significantly. This may imply that the assumptions of ARIMA are not valid; (ii) Since the S2SCL performs very well in point forecasting, it accurately makes “0” decisions for LSC when LSD is “0” from time to time, at which the other two models cannot achieve. Last but not least, the LSCs of PSO-ELM is closest to the ones of S2SCL in NT and KLN most of the time, but PSO-ELM’s LSC values in HKI seem quite aggressive in HKI when comparing with S2SCL and ARIMA. The inconsistency of the strategies of PSO-ELM might be explained by the randomness of searching optimal value within PIs which cannot guarantee to find the globally optimal solution to the defined problem in all cases.

Table 6 shows the total and daily average cost and the percentage of cost saving with the use of PSO-ELM and S2SCL over ARIMA. The results obviously show that the S2SCL approach performs the best in all regions. Specifically, the S2SCL approach can reduce the costs of ARIMA by at least 25%. Even though the PSO-ELM predicts the PIs better than S2SCL, its operations costs are higher than the ones of S2SCL. This interesting finding can be explained by the fact that the purpose of the S2SCL is to forecast the step-by-step value of LSDs as well as the dynamic distribution, which may be applicable to solve the optimal LSC allocation problem via solving the newsvendor problem, rather than predicting the PIs. In this case, even the PI forecasting results are not the best, its dynamic distribution can yield the optimal LSCs because it extracts features from the training data rather than from point estimations (such as using the PSO-ELM approach).

6. Conclusion and remarks

6.1. Conclusion

Business operations have entered the e-commerce era with O2O (Li et al. 2017) and CBEC all being popular. Motivated by the importance of CBEC, we conduct an analysis with a focus on the related logistics service capacity (LSC) allocation problem. To be specific, the LSC problem derived in this paper can be considered as a single period multi-product newsvendor model. A deep learning based one-step optimal decision making approach is proposed by integrating the inventory optimization and demand forecasting process together. Demand uncertainty will be quantified in real-time from end-to-end logistics service demand management with the help of a novel deep learning forecasting system. Our goal is to provide a versatile deep learning-based integration approach to support decision making in LSC allocation.

To tackle the challenges of logistics service demand management under a highly uncertain environment, we have established a
Fig. 5. Pls of different methods with the 95% PINV level.

Table 5
PICP and mean PI length (MPIL) of different models.

<table>
<thead>
<tr>
<th>PINC = 95%</th>
<th>ARIMA</th>
<th>PSO-ELM</th>
<th>S2SCL</th>
<th>ARIMA</th>
<th>PSO-ELM</th>
<th>S2SCL</th>
</tr>
</thead>
<tbody>
<tr>
<td>NT</td>
<td>78.57%</td>
<td>54.29%</td>
<td>24.29%</td>
<td>224.46</td>
<td>123.12</td>
<td>32.69</td>
</tr>
<tr>
<td>KLN</td>
<td>100.00%</td>
<td>71.43%</td>
<td>37.14%</td>
<td>1108.78</td>
<td>269.16</td>
<td>60.40</td>
</tr>
<tr>
<td>HKI</td>
<td>97.14%</td>
<td>62.86%</td>
<td>30.00%</td>
<td>123.67</td>
<td>60.74</td>
<td>14.16</td>
</tr>
</tbody>
</table>
deep learning based one-step integration optimal decision making approach S2SCL. The Seq2Seq based forecasting architecture, which integrates CNN and LSTM network, is able to model the system dynamics and dependency-relations in varying demand for logistics services. Besides generating the point forecasting results, the proposed S2SCL can quantify the demand uncertainty via a dynamic distribution and make optimal decision on logistics service capacity.
To evaluate the proposed method, we computationally compare it with two models, namely ARIMA and PSO-ELM, in the two tasks: (1) point forecasting and (2) optimal LSC allocation. The analysis is based on a cross-border e-commerce 3PFL company’s dataset. The classical statistical forecasting method ARIMA is chosen as a benchmark method since it can generate a demand distribution, which can be used to solve the optimal solution from newsvendor model by one-step. Different from ARIMA and S2SCL, the recent popularly developed machine learning based approach PSO-ELM firstly predicts the point forecasting and the interval of LSD, then searches the optimal LSCs within the interval.

Computational results clearly show that our proposed S2SCL method outperforms ARIMA and PSO-ELM methods in the LSD point forecasting in all 3 regions. The proposed S2SCL system improves the ARIMA and PSO-ELM’s point forecasting RMSE by at least 19.32% and 14.77%, respectively. Based on the proposed model, we further explore the LSC allocation problem for cross-border e-commerce 3PFL business operations. The computational results reveal that the optimal quantity of LSC using the proposed S2SCL can achieve the lowest costs in all regions compared with ARIMA and PSO-ELM. Besides, it’s interesting to investigate whether a better PIs prediction result yields a better LSC decision. From the PICP and MPIP results, PSO-ELM performs better than S2SCL in all regions. However, PSO-ELM’s LSC costs are always higher than the ones obtained under S2SCL.

To sum it up, we can conclude the following:

(1) For point forecasting, experimental results show that the proposed S2SCL model is an effective approach, which achieves superior performance than ARIMA and PSO-ELM in the LSD forecasting in all 3 regions.

(2) For PIs prediction, the optimization based approach PSO-ELM performs better than both the statistics based integration approach ARIMA and the deep learning based integration approach S2SCL. The PSO-ELM method can yield a reasonable PICP, and achieve a tight range for the PIs.

(3) For optimal decision making, the deep learning based integration approach S2SCL performs better than the statistics based integration approach ARIMA. This is mainly because the S2SCL is able to utilize the advantage of deep learning technologies to capture the variations of LSDs.

(4) For optimal decision making, the deep learning based integration approach S2SCL performs better than the optimization based approach PSO-ELM. This is mainly because the S2SCL can quantify the uncertainties based on the whole demand dataset rather than the point forecasting results in the PSO-ELM approach.

(5) A better PIs prediction result cannot guarantee a better LSC decision. Compared with S2SCL, the optimization based approach PSO-ELM achieves a better PIs prediction result but its LSC allocation decision result is worse. The reason explaining this phenomenon is that, the two-step optimization based approach does not guarantee the globally optimal solution will be achieved in all cases, and the one-step based S2SCL can extract features from the training data rather than from point estimations.

### 6.2. Managerial implications

In this paper, we propose an intelligent decision making process for the optimal LSC allocation problem with respect to demand fulfillment and cost saving. In this proposed process, demand uncertainty can be quantified by real-time information.

(1) Based on the location information of customers (e.g., GIS data), the 3PFL company is able to dispatch the logistic resources optimally according to the optimal LSC allocation mechanism provided by the proposed S2SCL.

(2) Considering the logistics resources, consisting of fixed components (e.g., pick up stores and couriers) and flexible components (e.g., vehicle and self-pick lockers), the 3PFL company can update the flexible component of the logistics resources planning to minimize the operational cost.

(3) In the long run planning, the 3PFL company can further utilize the advantage of Seq2seq architecture of the proposed S2SCL to adjust the forecasting step length (e.g. weekly and monthly) according to the planning need. In this case, some fixed component

<table>
<thead>
<tr>
<th>NT</th>
<th>ARIMA</th>
<th>PSO-ELM</th>
<th>S2SCL</th>
<th>KLN</th>
<th>ARIMA</th>
<th>PSO-ELM</th>
<th>S2SCL</th>
<th>HKI</th>
<th>ARIMA</th>
<th>PSO-ELM</th>
<th>S2SCL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sum($HKD)</td>
<td>609,596</td>
<td>523,771</td>
<td>457,271</td>
<td>1,251,756</td>
<td>524,667</td>
<td>441,617</td>
<td>212,121</td>
<td>210,531</td>
<td>156,830</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daily Average ($HKD)</td>
<td>8709</td>
<td>7482</td>
<td>6532</td>
<td>17,882</td>
<td>7495</td>
<td>6309</td>
<td>3030</td>
<td>3008</td>
<td>2240</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cost Saving (%)</td>
<td>–</td>
<td>14.08%</td>
<td>25.00%</td>
<td>–</td>
<td>58.09%</td>
<td>64.72%</td>
<td>–</td>
<td>0.75%</td>
<td>26.07%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

To evaluate the proposed method, we computationally compare it with two models, namely ARIMA and PSO-ELM, in the two tasks: (1) point forecasting and (2) optimal LSC allocation. The analysis is based on a cross-border e-commerce 3PFL company’s dataset. The classical statistical forecasting method ARIMA is chosen as a benchmark method since it can generate a demand distribution, which can be used to solve the optimal solution from newsvendor model by one-step. Different from ARIMA and S2SCL, the recent popularly developed machine learning based approach PSO-ELM firstly predicts the point forecasting and the interval of LSD, then searches the optimal LSCs within the interval.

Computational results clearly show that our proposed S2SCL method outperforms ARIMA and PSO-ELM methods in the LSD point forecasting in all 3 regions. The proposed S2SCL system improves the ARIMA and PSO-ELM’s point forecasting RMSE by at least 19.32% and 14.77%, respectively. Based on the proposed model, we further explore the LSC allocation problem for cross-border e-commerce 3PFL business operations. The computational results reveal that the optimal quantity of LSC using the proposed S2SCL can achieve the lowest costs in all regions compared with ARIMA and PSO-ELM. Besides, it’s interesting to investigate whether a better PIs prediction result yields a better LSC decision. From the PICP and MPIP results, PSO-ELM performs better than S2SCL in all regions. However, PSO-ELM’s LSC costs are always higher than the ones obtained under S2SCL.

To sum it up, we can conclude the following:

(1) For point forecasting, experimental results show that the proposed S2SCL model is an effective approach, which achieves superior performance than ARIMA and PSO-ELM in the LSD forecasting in all 3 regions.

(2) For PIs prediction, the optimization based approach PSO-ELM performs better than both the statistics based integration approach ARIMA and the deep learning based integration approach S2SCL. The PSO-ELM method can yield a reasonable PICP, and achieve a tight range for the PIs.

(3) For optimal decision making, the deep learning based integration approach S2SCL performs better than the statistics based integration approach ARIMA. This is mainly because the S2SCL is able to utilize the advantage of deep learning technologies to capture the variations of LSDs.

(4) For optimal decision making, the deep learning based integration approach S2SCL performs better than the optimization based approach PSO-ELM. This is mainly because the S2SCL can quantify the uncertainties based on the whole demand dataset rather than the point forecasting results in the PSO-ELM approach.

(5) A better PIs prediction result cannot guarantee a better LSC decision. Compared with S2SCL, the optimization based approach PSO-ELM achieves a better PIs prediction result but its LSC allocation decision result is worse. The reason explaining this phenomenon is that, the two-step optimization based approach does not guarantee the globally optimal solution will be achieved in all cases, and the one-step based S2SCL can extract features from the training data rather than from point estimations.

### 6.2. Managerial implications

In this paper, we propose an intelligent decision making process for the optimal LSC allocation problem with respect to demand fulfillment and cost saving. In this proposed process, demand uncertainty can be quantified by real-time information.

(1) Based on the location information of customers (e.g., GIS data), the 3PFL company is able to dispatch the logistic resources optimally according to the optimal LSC allocation mechanism provided by the proposed S2SCL.

(2) Considering the logistics resources, consisting of fixed components (e.g., pick up stores and couriers) and flexible components (e.g., vehicle and self-pick lockers), the 3PFL company can update the flexible component of the logistics resources planning to minimize the operational cost.

(3) In the long run planning, the 3PFL company can further utilize the advantage of Seq2seq architecture of the proposed S2SCL to adjust the forecasting step length (e.g. weekly and monthly) according to the planning need. In this case, some fixed component

<table>
<thead>
<tr>
<th>NT</th>
<th>ARIMA</th>
<th>PSO-ELM</th>
<th>S2SCL</th>
<th>KLN</th>
<th>ARIMA</th>
<th>PSO-ELM</th>
<th>S2SCL</th>
<th>HKI</th>
<th>ARIMA</th>
<th>PSO-ELM</th>
<th>S2SCL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sum($HKD)</td>
<td>609,596</td>
<td>523,771</td>
<td>457,271</td>
<td>1,251,756</td>
<td>524,667</td>
<td>441,617</td>
<td>212,121</td>
<td>210,531</td>
<td>156,830</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daily Average ($HKD)</td>
<td>8709</td>
<td>7482</td>
<td>6532</td>
<td>17,882</td>
<td>7495</td>
<td>6309</td>
<td>3030</td>
<td>3008</td>
<td>2240</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cost Saving (%)</td>
<td>–</td>
<td>14.08%</td>
<td>25.00%</td>
<td>–</td>
<td>58.09%</td>
<td>64.72%</td>
<td>–</td>
<td>0.75%</td>
<td>26.07%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
logistics resources planning can be decided in advance to minimize the overall operational cost.

(4) The rise of artificial intelligence in recent years is in fact a result supported by the success of deep learning. Considering that the proposed approach is deep learning based, both the academia and companies can apply this approach easily and conveniently via the help of python or other proper tools.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.tre.2019.101834.

References