

Prediction of Tier in Supply Chain Using LSTM and Conv1D-LSTM

KyoungJong Park[†]

Department of Business Administration, Gwangju University

LSTM 및 Conv1D-LSTM을 사용한 공급 사슬의 티어 예측

박 경 중[†]

광주대학교 경영학과

Supply chain managers seek to achieve global optimization by solving problems in the supply chain's business process. However, companies in the supply chain hide the adverse information and inform only the beneficial information, so the information is distorted and cannot be the information that describes the entire supply chain. In this case, supply chain managers can directly collect and analyze supply chain activity data to find and manage the companies described by the data. Therefore, this study proposes a method to collect the order-inventory information from each company in the supply chain and detect the companies whose data characteristics are explained through deep learning. The supply chain consists of Manufacturer, Distributor, Wholesaler, Retailer, and training and testing data uses 600 weeks of time series inventory information. The purpose of the experiment is to improve the detection accuracy by adjusting the parameter values of the deep learning network, and the parameters for comparison are set by learning rate (lr = 0.001, 0.01, 0.1) and batch size (bs = 1, 5). Experimental results show that the detection accuracy is improved by adjusting the values of the parameters, but the values of the parameters depend on data and model characteristics.

Keywords : Supply Chain, Deep Learning, Information, Learning Rate, Batch Size

1. 서 론

2019년 12월에 발생한 코로나19(COVID-19) 감염증은 2020년 1월부터 아시아뿐만 아니라 전 세계로 확산되면서 공포와 두려움을 안겨주고 있다. 코로나19의 영향으로 세계 곳곳에서의 인적 이동은 제한되거나 불가하게 되었다. 이로 인해 제조공장이 멈추면서 제품 생산을 위한 원자재가 필요한 시기에 공급되지 않아 공급 사슬 전체가 심각한 영향을 받으면서 생산 차질로 인한 손실이 증가하고 있다.

공급 사슬은 외부에서 발생하는 천재지변 또는 재난과 같은 환경 변화에 치명적인 타격을 입게 되고, 개별 공급 사슬만으로는 전 세계적으로 동시에 발생하는 문제를 해결하기가 쉽지 않기 때문에 국가 수준에서 대응해야 한다. 그러나 공급 사슬 내에서 발생하는 문제는 공급 사슬 관리자에 의해 해결되어야 공급 사슬의 전체 최적화가 가능하다. 공급 사슬의 전체 최적화를 위해서는 공급 사슬에 참여하는 개별 기업들의 적극적인 도움 없이는 불가능하다는 것을 모든 참여 기업들이 알고 있지만 쉽지 않다.

그러므로 본 연구는 공급 사슬의 관리자가 공급 사슬에 참여하는 개별 기업들이 정확한 정보를 주지 않아도 공급 사슬에서 발생하는 데이터를 관찰하여 그 데이터의 특성이 설명하는 개별 기업들을 찾아내어 관리하는 방안을 제시하고자 한다.

기업이 가진 데이터의 특성을 파악해서 예측하거나 패턴을 찾아내는 방법이 인공지능 기술 및 컴퓨터 성능 향상으로 빠르게 발전하고 있으며, 이를 해결하기 위해 딥러닝(Deep Learning) 기법이 다양한 분야에서 적용되고 있다[6].

공급 사슬은 시간의 변화에 민감한 동적 특성을 가지기 때문에 얻어지는 정보는 시계열 데이터의 특성을 가진다. 딥러닝 기법 중에서 시계열 데이터 분석에는 Recurrent Neural Network(RNN)의 Long Short-Term Memory(LSTM-RNN) 기법이 적극적으로 사용되고 있다[1, 5, 10]. 본 연구는 공급 사슬에 포함된 개별 기업에서 발생하는 시간에 따른 주문-재고 데이터를 수집하여 딥러닝을 통해 데이터의 특성이 설명하는 기업을 찾아내는 방안을 제안한다. 적용되는 딥러닝 기법은 기본형인 순차 네트워크로 구성되는 Seq-DNN 모델과 시계열 데이터 분석에 우수한 성능을 보여주는 LSTM-RNN 모델이다. 또한, 2차원 데이터의 이미지 분석에 효과적이라고 알려진 Convolution Neural Network(CNN)가 시계열 데이터의 특성을 추출하는 장점이 있으므로 LSTM과 결합한 Conv1D-LSTM 모델도 포함한다[3, 6]. 딥러닝 모델의 파라미터에 의한 성능 차이를 비교하기 위해 학습률(learning rate)과 배치 크기(batch size)의 변화를 분석한다.

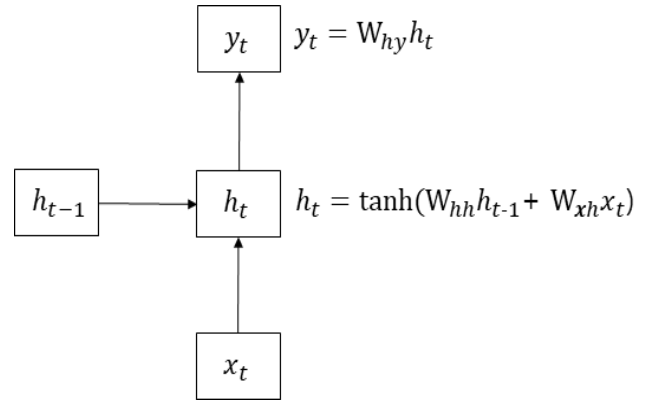
본 논문은 연구의 배경 및 필요성을 제 1장에서 설명했고, 제 2장에서는 분석을 위해 사용되는 LSTM-RNN과 CNN에 관해서 설명한다. 제 3장에서는 실험 모델 및 조건을 설명하며, 제 4장에서는 실험 및 결과 분석을 진행한다. 마지막으로 제 5장에서는 결론 및 추후 연구과제를 제시한다.

2. 딥러닝 기법

2.1 LSTM-RNN

RNN은 신경망의 꽃이라 불리고 시퀀스(sequence) 데이터를 다루는 신경망이다. 예를 들면, 언어(말)는 단순히 한 개의 단어를 통해서 이해될 수 없고 전체 문장으로 구성되는 시퀀스 데이터에 의해서 이해될 수 있다. RNN은 현재의 상태(state)는 바로 전 상태에 의해서 영향을 받는 시계열 데이터를 적용하면 우수한 결과를 얻는 방법이다. 가장 기본이 되는 RNN은 바닐라(Vanilla) RNN이며 다음의 <Figure 1>과 같은 네트워크 구조를 가진다.

<Figure 1>의 h_t 및 h_{t-1} 는 타임 스텝 t 및 t-1에서 셀의 은닉(hidden) 상태를 의미하며, x_t 및 y_t 는 타임 스텝 t에서 입력값과 출력값을 의미한다. W_{hh} 는 은닉층(hidden layer)



<Figure 1> Structure of RNN

h 에서 출력값 y 로 보내는 파라미터이고 W_{hh} 는 이전 은닉층 h 에서 다음 은닉층 h 로 보내는 파라미터이다. W_{xh} 는 입력값 x 를 은닉층 h 로 보내는 파라미터를 의미한다. RNN의 일반 구조는 현재 상태의 출력값 y_t 는 h_t 를 전달 받아 갱신되며 은닉 상태의 활성화 함수(activation function)는 비선형 함수인 하이퍼볼릭탄젠트(tanh)를 사용한다.

RNN의 적용 분야는 언어 모델링, 음성 인식, 기계 번역, 대화 모델링, 이미지 및 비디오 캡션 등이다. RNN의 장점은 시퀀스 데이터 처리이며 적용하는 문제 대부분은 시퀀스의 길이가 동일한 경우이지만 시퀀스의 길이가 같지 않을 수도 있다. 이런 경우에는 Dynamic RNN을 사용한다. RNN은 경사도를 중첩해 쉽게 고칠 수 있는 경사도 폭발(exploding gradients) 문제와 쉽게 고치기 힘든 기울기 소실(vanishing gradients) 문제를 가지고 있다. RNN에서 발생하는 기울기 소실과 같은 문제를 해결하여 성능을 향상시키기 위해서 사용되는 대표적인 모델은 LSTM과 Gated Recurrent Unit(GRU)[4]이다.

LSTM은 RNN의 은닉 상태에 셀-상태(cell-state)를 추가한 구조이며 셀-상태는 컨베이어 벨트 역할을 수행하기 때문에 상태가 오래 지속되어도 기울기가 상대적으로 잘 전파된다. LSTM은 가중치를 수정하는 방법으로 이전 정보를 잊거나 기억하도록 만들 수 있다. 이는 LSTM에 장단기 기억 능력을 제공하여 기울기 소실 문제를 해결하도록 하며, 비유(활성화) 함수로 tanh를 사용한다.

2.2 CNN

CNN은 동물의 시각피질(visual cortex) 구조에서 영감을 얻어 만들어진 딥러닝 신경망 모델로 동물의 계층적 특징 추출과 시각인식 체계를 참조하여 만들어졌다[11]. CNN은 컴퓨터 비전, 음성인식, 영상인식 등의 분야에서 많이 사용되며, 종류는 AlexNet, GoogleNet, VGGNet, ResNet 등이 있다.

CNN은 특징을 추출하기 위한 층(feature learning layers)의 집합과 이것을 분류하기 위한 층(classification layers)의 집합으로 구성된다[7]. 특징을 추출하기 위한 층은 Convolution Layer, ReLu Layer, Pooling Layer를 반복하면서 네트워크 구조를 쌓아간다. Convolutional Layer는 2D 필터를 통해 영상의 특징을 추출하고, ReLu Layer는 비유 함수에 의해 뉴런의 출력값을 결정하며, Pooling Layer는 인접한 공간의 값들에 대해 하나의 대표값을 사용하여 표본의 수를 줄여 준다.

분류하기 위한 층은 Fully connected Layer, Softmax Layer, Classification Layer로 구성되며, 전 단계에서 넘겨받은 특징을 입력값으로 받아 기대값과 출력값 간의 오차를 최소화하도록 오류역전파 방법을 통해 뉴런 간의 가중치를 갱신한다. Softmax Layer는 출력 노드의 값이 해당 부류에 속할 확률값인 분류기를 신경망으로 구현된 것으로 신경망의 출력 노드의 출력값이 0 이상이면서 전체 출력값의 합은 1이 되도록 구성한다. 이때 특징을 추출하기 위한 층에 속한 Convolutional Layer의 필터 계수도 함께 갱신된다.

CNN은 이미지와 같은 2차원 배열의 데이터를 일반적으로 사용하기 때문에 2D(Conv2D) CNN 모델을 적용한다. 그러나 본 연구는 1차원 시계열 데이터를 사용하기 때문에 컨볼루션 층이 1차원인 1D(Conv1D) CNN 모델을 적용한다.

3. 실험 모델

본 논문에서 사용하는 공급 사슬 모델은 실험 데이터에 대한 검증을 고려하여 Strozzi et al.[12]의 모델을 사용한다 (<Figure 2> 참조). <Figure 2>의 공급 사슬 모델은 Factory, Distributor, Wholesaler, Retailer로 구성되며, Retailer는 고객과 직접 접촉이 가능하며 고객으로부터 주문을 받아서 생산된 제품을 다시 고객에게 판매한다. 고객으로부터의

주문 정보는 Retailer → Wholesaler → Distributor → Factory로 보내지고 완성된 제품은 Factory → Distributor → Wholesaler → Retailer로 전달되어 최종적으로 고객에게 배송된다.

Retailer에게 주문하는 고객의 수요는 확률적(stochastic) 수요를 가정하며 AR(1) 모델을 사용한다(식 (1)).

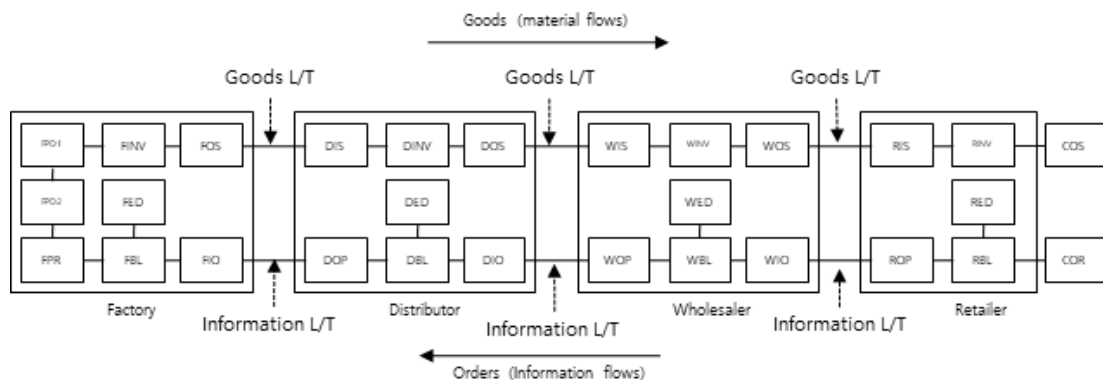
$$D_t = \mu + \rho D_{t-1} + \epsilon_t, |\rho| < 1 \tag{1}$$

식 (1)에서 D_t 및 D_{t-1} 는 시점 t 및 t-1의 고객 수요, μ 는 음수가 아닌 상수, ϵ_t 는 평균이 0, 분산이 σ^2 인 독립적이고 동일한 정규분포를 따르는 오차항, ρ 는 1차 자기상관계수(autocorrelation coefficient)이다. 식 (1)의 σ 와 ρ 는 고객 수요의 프로세스에 영향을 미치게 된다. ρ 는 수요 추세를 의미하며 전체 시간 동안의 수요의 상관관계를 설명한다. σ 는 수요의 변동을 의미하며 수요의 분산을 설명한다[8, 9, 13].

<Figure 2>의 공급 사슬에서 각 티어 간의 주문량은 Strozzi et al.[12] 모델에서 제시한 수식을 이용하여 계산한다. 본 연구의 실험 모델에서 얻어진 각 티어의 재고량 데이터와 Strozzi et al.[12] 모델에서 얻어진 각 티어의 재고량 데이터에 대한 동일성이 검증되어 딥러닝 실험에 사용한다. <Figure 2>의 공급 사슬 모델의 재고량 계산을 위해 사용된 수식에 대한 추가적인 설명은 Strozzi et al. [12]을 참조한다.

4. 실험 및 결과 분석

<Figure 2>의 Factory, Distributor, Wholesaler, Retailer로 구성된 각 티어에서 수집되는 재고량은 식 (1)로 표현되는 고객 수요에서 수요의 추세를 의미하는 ρ 값을 -0.6, -0.3, 0.0, 0.3, 0.6으로 변화시키면서 얻어진 600주 동안의 시계열 데이터이다. 수집된 학습 데이터의 크기는 12,000(4티어×5개 ρ 값×600주) 개다.



<Figure 2> Supply Chain Model

검증 데이터는 딥러닝을 통해 학습 데이터가 정확하게 학습되었는지를 평가하면서 기존에 학습되지 않은 새로운 데이터에 대한 정확도를 동시에 평가되도록 구성한다. 검증 데이터의 크기는 학습에 사용된 학습 데이터 전체와 학습 데이터 수집에 사용했던 동일한 조건에서 새롭게 얻어진 데이터로 구성하는 것을 가정한다. 그러므로 검증 데이터의 크기는 학습 데이터의 2배인 24,000개이다.

학습 및 검증을 위해 사용되는 딥러닝 모델에서 은닉층의 활성화 함수는 `lelu`를 사용하고, 주어진 실험이 `Factory, Distributor, Wholesaler, Retailer`를 판별하는 문제이기 때문에 출력층의 활성화 함수는 다중 분류(`multi classification`) 문제에 적용되는 `softmax`를 사용한다. 오차 (`loss`) 함수는 다중 분류 문제에 적용되는 범주형 교차 엔트로피(`categorical cross entropy`)를 사용하며, 최적화를 위한 경사하강법(`gradient method`)은 `adam`을 사용한다. 종료 조건(`stopping condition`)은 검증 데이터의 손실값(`loss value`)이 20세대(`patience = 20`) 동안 연속적으로 개선되지 않으면 멈추도록 설정하고 `epoch`는 400으로 가정한다. 고정되는 파라미터 값은 일반적으로 추천되는 `beta1 = 0.9, beta2 = 0.999, eps = 1e-08`로 설정한다[2]. `beta1` 및 `beta2`는 첫 번째 및 두 번째 모멘트 추정치에 대한 지수 붕괴율이다. `eps`는 0으로 나누어지는 것을 방지하기 위한 매우 작은 수이다.

Seq-DNN 모델의 활성화 함수는 `relu` → `softmax`로 구성하였고, LSTM-RNN 모델의 활성화 함수는 `relu` → `LSTM (tanh)` → `softmax`로 구성하였고, Conv1D-LSTM의 활성화 함수는 `relu` → `Dropout(0.5)` → `Conv1D(padding, relu)` → `MaxPooling1D` → `LSTM(tanh)` → `softmax`로 구성하였다.

앞에서 설명한 것처럼 실험에 사용되는 딥러닝 모델은 Seq-DNN, LSTM-RNN, Conv1D-LSTM이며, 다음의 <Table 1>은 3개의 딥러닝 모델에 대한 네트워크 구성 및 파라미터값에 따른 실험 결과를 보여준다.

<Table 1>은 배치 크기가 1과 5, 학습률이 0.001, 0.01, 0.1일 때 Seq-DNN, LSTM-RNN, Conv1D-LSTM 모델에 대한 종료 시점과 정확도를 보여주고 있다. Seq-DNN은 배치 크기가 5일 때의 정확도가 0.975로 1일 때의 정확도보다 높고, 전체 정확도 평균은 0.933이다. LSTM-RNN은 배치 크기가 1일 때의 정확도가 0.492이고 5일 때의 정확도가 0.508로 5일 때의 정확도가 더 높고, 전체 정확도 평균은 0.500이다. Conv1D-LSTM은 배치 크기가 1일 때의 정확도가 0.683이고 5일 때의 정확도가 0.525로 1일 때의 정확도가 더 높고, 전체 정확도 평균은 0.604이다.

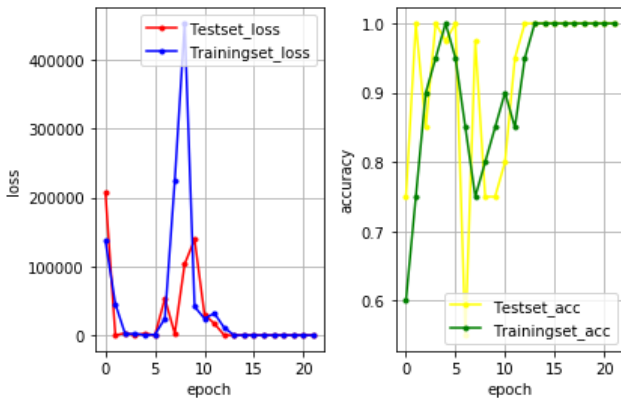
<Table 1>의 정확도 비교 분석에서는 기본형 딥러닝 모

<Table 1> Experimental Conditions and Results of Deep Learning Model

network	batch size	learning rate	result			
			stop (Epoch)	accuracy		
Seq-DNN	1	0.001	24	0.975	0.892	0.933
		0.01	22	1.000		
		0.1	23	0.700		
	5	0.001	28	0.950	0.975	
		0.01	40	1.000		
		0.1	25	0.975		
LSTM-RNN	1	0.001	92	0.900	0.492	0.558
		0.01	34	0.325		
		0.001	51	0.250		
	5	0.001	119	0.825	0.625	
		0.01	64	0.775		
		0.1	36	0.275		
Conv1D-LSTM	1	0.001	65	0.900	0.683	0.613
		0.01	57	0.250		
		0.1	65	0.900		
	5	0.001	98	0.900	0.542	
		0.01	44	0.300		
		0.1	44	0.425		

델인 Seq-DNN이 가장 우수하고, 다음은 Conv1D-LSTM 이고, LSTM-RNN이 가장 낮다. 배치 크기 비교 분석에서는 Seq-DNN 및 LSTM-RNN은 배치 크기가 5일 때가 1일 때보다 우수하고 Conv1D-LSTM은 배치 크기가 1일 때가 5일 때보다 우수하다. 학습률 비교 분석에서는 학습률이 0.001일 때가 Seq-DNN, LSTM-RNN, Conv1D-LSTM의 경우에 전반적으로 우수한 결과를 보여주며, Seq-DNN 모델에서는 배치 크기에 상관없이 학습률이 0.01일 때 정확도가 100%로 가장 우수하다. 종료 시점 비교 분석에서는 기본형 딥러닝 모델인 Seq-RNN이 LSTM-RNN 및 Conv1D-LSTM 모델보다 전반적으로 이른 시간에 종료됨을 알 수 있다.

딥러닝 모델이 종료 조건에 의해 정확하게 종료되는지를 확인하는 것도 필요하다. 다음의 <Figure 3>은 Seq-DNN 모델에서 배치 크기가 1이고 학습률이 0.001일 때 종료 조건 `patience=20`에 따른 에포크(`epoch`) 변화를 보여준다. 검증 데이터의 손실값(`Testset_loss`)이 3번째 에포크에서 0으로 떨어져서 20 에포크 동안 계속해서 개선되지 않아 22 에포크에서 종료된다. 검증 데이터의 정확도(`Testset_acc`)는 3번째 에포크에서 1.0이 되어 변동되다가 13번째 에포크부터는 지속해서 1.0을 유지하다가 22 에포크에서 종료됨을 알 수 있다.



<Figure 3> Loss Value and Accuracy of Seq-DNN(batch size = 1, learning rate = 0.001)

5. 결론 및 추후 연구과제

본 연구는 공급 사슬의 각 티어에서 발생하는 데이터이지만 어떤 티어에서 발생하는지를 공개하지 않은 재고량 데이터를 딥러닝 기법을 통해 데이터가 발생한 티어를 찾아내고 관리하는 방안을 제시하였다.

공급 사슬의 각 티어에서 발생하는 데이터는 기본적으로 시계열 데이터의 특성을 가지기 때문에 시계열 데이터 분석에 가장 효과적이라고 알려진 딥러닝 모델 LSTM-RNN을 적용하였다. 2차원 데이터의 이미지 분석에 효과적이라고 알려진 CNN이 시계열 데이터의 특성을 추출하는 데 장점이 있어서 LSTM과 결합한 Conv1D-LSTM 모델도 적용하였다. 그리고 기본형 딥러닝 모델 Seq-DNN도 같이 적용하여 어떤 딥러닝 모델이 공급 사슬의 티어 판별에 가장 효과적인지를 분석하였다.

분석을 통한 결과는 기본형 딥러닝 모델인 Seq-DNN이 시계열 데이터의 분석에 효과적이라고 알려진 LSTM-RNN 및 Conv1D-LSTM보다 정확도는 더 우수하고 종료 시점이 더 빠르다는 것을 알 수 있었다. 또한, Seq-DNN이 배치 크기 및 학습률과 관계없이 다른 모델들보다 더 우수하다는 것을 보여주었다.

각 모델에 대해 실험을 하기 전에는 LSTM-RNN 또는 Conv1D-LSTM이 가장 우수한 결과를 줄 것으로 예측하였으나 기본형 딥러닝 모델인 Seq-DNN이 동일한 조건에서 가장 우수한 결과를 보여주었다. 이러한 결과는 적용하는 딥러닝 모델뿐만 아니라 데이터의 특성 및 구조 또는 다른 요인이 결과에 영향을 줄 수 있음을 설명하고 있다.

그러므로 추후 연구과제는 데이터의 특성 및 구조가 모델에 어떤 영향을 주는지를 분석하는 것이 필요하지만 그 전 단계의 연구로서 LSTM보다 더 우수한 성능을 보

여주는 양방향(bidirectional) LSTM 모델을 적용하여 단방향 LSTM 모델, 기본형 RNN 모델(SimpleRNN) 및 기본형 딥러닝 모델의 성능과 비교하는 것이 필요하다고 판단된다.

Acknowledgement

This study was conducted by research funds from Gwangju University in 2020.

References

- [1] Althelaya, K.A., Alfay, E.M., and Mohammed, S., Evaluation of bidirectional LSTM for short-and long-term stock market prediction, *Proceedings of the 2018 9th International Conference on Information and Communication Systems(ICICS)*, Irbid, Jordan, 2018, pp. 151-156.
- [2] Brownlee, J., Gentle Introduction to the Adam Optimization Algorithm for Deep Learning, <https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/>.
- [3] Chen, Y.H., Krishna, T., Emer, J.S., and Sze, V., Eyeriss : An energy-efficient reconfigurable accelerator for deep convolutional neural networks, *IEEE Journal of Solid-State Circuits*, 2017, Vol. 52, No. 1, pp. 127-138.
- [4] Cho, K.H., Merriënboer, B.V., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y., Learning phrase representations using RNN encoder-decoder for statistical machine translation, *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, Doha, Qatar, 2014, pp. 1-15.
- [5] FukuoKa, R., Suzuki, H., Kitajima, T., Kuwahara, A., and Yasuno, T., Wind speed prediction model using LSTM and 1D-CNN, *J-STAGE*, 2018, Vol. 22, No. 4, pp. 207-210.
- [6] Jung, H.C., Sun, Y.G., Lee, D.G., Kim, S.H., Hwang, Y.M., Sim, I., Oh, S.K., Song, S.H., and Kim, J.Y., Prediction for energy demand using 1D-CNN and bidirectional LSTM in internet of energy, *Journal of Institute of Korean Electrical and Electronics Engineers*, 2019, Vol. 23, No. 1, pp. 134-142.
- [7] MathWorks Korea, <https://m.blog.naver.com/matlablove/220928458225>.
- [8] Park, K.J., Effect of customer demand on total inventory cost and order fill rate in a supply chain, *Journal of*

Society of Korea Industrial and Systems Engineering, 2009, Vol. 32, No. 3, pp. 93-98.

- [9] Park, K.J., Effect of information quality level and customer demand on performance measures in a supply chain, *Journal of Society of Korea Industrial and Systems Engineering*, 2012, Vol. 35, No. 2, pp. 136- 144.
- [10] Park, K.J., Performance evaluation of LSTM network through supply chain tier prediction, *Asia-pacific Journal of Multimedia Services Convergent with Art, Humanities, and Sociology*, 2019, Vol. 9, No. 8, pp. 799-808.
- [11] Sallehuddin, R., Shamsuddin, S.M., Hashim, S.Z.M., and Abraham, A., Forecasting time series data using hybrid gray relational artificial neural network and autoregressive integrated moving average model, *Neural Network World*, 2007, Vol. 17, No. 6, pp. 573-605.
- [12] Strozzi, F., Bosch, J., and Zaldivar, J.M., Beer game order policy optimization under changing customer demand, *Decision Support Systems*, 2007, Vol. 42, pp. 2153-2163.
- [13] Zhang, C. and Zhang, C., Design and simulation of demand information sharing in a supply chain, *Simulation Modelling Practice and Theory*, 2007, Vol. 15, No. 1, pp. 32-46.

ORCID

KyoungJong Park | <http://orcid.org/0000-0002-8079-6461>