

프로세스 마이닝과 딥러닝을 활용한 구매 프로세스의 적기 입고 예측에 관한 연구

Exploring the Prediction of Timely Stocking in Purchasing Process Using Process Mining and Deep Learning

강 영 식 (Youngsik Kang) 명지대학교 경영정보학과
이 현 우 (Hyunwoo Lee) 명지대학교 경영정보학과
김 병 수 (Byoungsoo Kim) 영남대학교 경영학과, 교신저자

요 약

예측 분석을 전사 프로세스에 적용하는 것은 운영비용을 절감하고 생산성을 증대시킬 수 있는 효과적 방법이다. 이에 따라 비즈니스 프로세스의 행동과 성과지표를 예측하는 능력이 기업의 핵심역량으로 간주되고 있다. 최근에 순환신경망 형태의 딥러닝을 이용한 프로세스 예측 연구가 큰 관심을 받고 있다. 특히, 순환신경망을 이용하여 다음 단계의 액티비티를 예측하는 접근법이 우수한 결과를 내고 있다. 그러나 동적 순환신경망 형태의 딥러닝을 프로세스 성과지표의 예측에 적용한 연구는 부족한 상황이다. 이러한 지식의 공백을 메우기 위해 본 연구는 프로세스 마이닝과 동적 순환신경망 형태의 딥러닝을 이용하는 접근법을 개발했다. 국내 대기업의 실제 데이터를 활용하여 구매 프로세스의 중요한 성과지표인 적기 입고 예측에 개발된 접근법을 적용했다. 본 연구의 실험 방법과 결과, 연구의 시사점과 한계점이 제시되었다.

키워드 : 딥러닝, 프로세스 마이닝, 동적 순환신경망, 구매 프로세스, 물류

I. 서 론

보호무역주의 확대, 미국과 중국 간 무역분쟁 등 대외 시장 상황의 어려움으로 기업들은 힘든 상황에 직면해있다. 이런 상황을 타개하기 위해 기업들은 운영비용 절감과 생산성 향상을 통한 경쟁력 확보에 많은 노력을 기울이고 있다(Park and Chang, 2011; 박광오 등, 2018; 황채영, 서창교, 2015).

최근 정보시스템에 축적된 데이터를 활용한 예측 분석이 비용을 절감할 수 있는 효과적 방법으로 식되고 있다(Barbosa *et al.*, 2018; Evermann *et al.*, 2017; Van der Aalst, 2016). 특히 SCM(Supply Chain Management) 관련 구매나 물류관련 데이터 분석을 통한 정확한 납기일이나 재고 예측은 기업의 운영 비용을 절감하고, 자원을 효율적으로 활용할 수 있는 방법으로 주목을 받고 있다(Barbosa *et al.*, 2018). 하지만 구매와 물류 프로세스를 지원하는 정보시스템에 축적된 데이터를 활용하여 적기 입

† 본 논문은 정석물류학술재단 지원에 의하여 연구되었음.

고 시점을 예측하려는 기업들의 노력은 상대적으로 부족한 실정이다. 실제 국내 많은 기업들은 구매 프로세스의 적기 입고 관리를 여전히 기존 관행이나 경험에 의존하는 경우가 많다. 이런 경우 자재가 필요한 시점에 입고되지 않아 생산 결함이 발생하거나 조기 입고되어 장기 재고를 가지는 경우가 자주 발생하게 된다. 그래서, SCM 관련비용 절감을 위해 데이터 기반 예측 분석을 SCM에 적용하는 Smart SCM(Smart Supply Chain Management)이 제안되었다(Wu *et al.*, 2016). Smart SCM는 구매나 물류 프로세스 분야에서 올바른 의사 결정을 수행할 수 있도록 도와주는 것으로 알려져 있다(Waller and Fawcett, 2013; 박성민, 박찬권, 2018). Waller and Fawcett(2013) 연구에서 SCM 데이터를 활용하여 예측 분석을 수행한 기업은 고객대응을 긴밀하게 수행하고, 재고 및 인력 등의 낭비도 제거할 수 있음을 밝혔다. 여러 SCM 연구들에서 구매한 자재들이 창고에 머무르는 시간이 최소화되도록 적기 입고를 예측하는 것이 매우 중요하다고 주장하고 있다(강영식, 2014; 송인성, 조남욱, 2015). 송인성, 조남욱(2015) 연구에서는 물류관리시스템의 데이터를 분석하여 적시 입고 수준을 파악하는 것이 중요하다고 제안하였다. 또한 유영석 등(2016) 연구에서도 조선 산업에서의 데이터 분석을 통해 영업 리스크의 많은 부분이 적기 입고 문제로 발생함을 지적하였다. 입고고 데이터를 분석하여 필요한 자재를 필요한 시점에 제공해줄 수 있다면 운영 비용을 절감하고 생산성도 향상시킬 수 있다. 서영복 등(2012) 연구에서는 중심 기업이 구매 자재의 적기 입고 시기를 예측하고 협력업체와 관련 스케줄을 공유한다면 SCM이 효율적으로 운영될 수 있음을 밝혔다. 본 연구는 구매 프로세스의 중요한 성과 지표인 적기 입고를 구매 및 물류 관련 데이터를 바탕으로 예측해보고자 한다.

기업에 축적된 데이터들을 분석하고 예측하기 위해 프로세스 마이닝과 순환 신경망(RNN: Recurrent Neural Networks)이 자주 활용되고 있다. 프로세스 마이닝은 프로세스 진행 과정에서 발생 한 이

벤트 로그로부터 프로세스를 분석하여 작업자의 업무 실행을 개선할 수 있는 방법이다. 최근 Breuket *et al.*(2017) 연구에서는 프로세스 마이닝에 RNN과 같은 예측 기법을 도입하면 더 많은 시사점과 정보를 얻을 수 있다고 주장하였다. RNN은 시계열 데이터 가지고 있는 특징을 학습하여 향후 시점 값을 예측하는 기법으로, 입력된 데이터를 신경망 노드와 활성화 함수를 활용하여 최적의 값을 도출한다. Tax *et al.*(2017) 연구에서는 기업에 축적된 이벤트 로그 파일을 RNN을 기반으로 분석하면 업무 사이클 타임이나 최적 프로세스를 예측할 수 있음을 밝혔다. 이 연구에서 기업의 업무 프로세스는 순차적으로 이루어져 있기 때문에, RNN과 같은 순서를 고려한 분석이 올바른 예측을 할 수 있다고 제안하였다. 업무 프로세스는 액티비티 수가 다른 경우가 빈번하기 때문에, 시퀀스 길이에 상관없이 모든 케이스를 활용할 수 있는 동적 RNN을 적용하는 것이 적합하다. 그래서 본 연구에서는 프로세스 마이닝과 동적 RNN을 적용하여 구매 프로세스의 중요한 성과지표인 적기 입고 시기를 예측해보고자 한다.

본 연구의 목적은 다음과 같다. 첫째 국내 대기업의 ERP 시스템 내 구매 및 물류 데이터를 프로세스 마이닝과 동적 RNN을 결합하여 적기 입고 시기를 예측하고자 한다. 둘째 프로세스 마이닝과 동적 RNN을 결합한 방법의 예측 정확도가 다른 예측 방법에 비해 더 높음을 보이고자 한다. 본 연구 결과를 통해 실무자들은 기업에 축적된 구매 및 물류 데이터를 바탕으로 최적의 입고 시기를 예측하여 운영 비용을 절감하고 자원을 효율적으로 이용할 수 있을 것으로 기대된다.

II. 이론적 배경

2.1 구매 및 물류 프로세스 예측

최근에 기업들은 방대한 양의 데이터를 전자적으로 통합하고 분석함으로써 운영 비용감축,

제조 품질 개선 및 가치 사슬의 효율성을 제고하고 있다(예: Manyika *et al.*, 2011; Van der Aalst, 2016). 예를 들어, Manyika *et al.*(2011) 연구는 제조업의 가치사슬에서 빅데이터 분석을 통해 비용 절감과 수익 증대를 얻을 수 있다고 제안하였다. Sowar and Gromley(2011)는 철강산업의 SCM에서 발생한 데이터를 분석하여 비용을 절감하는 것이 철강 회사의 경쟁력을 증진하는 요인임을 주장하였다. Wu *et al.*(2016) 연구는 축적된 SCM 데이터 분석을 통해 더욱 정교한 수요 및 재고 예측을 할 수 있고, 이를 통해 기업의 비용 절감이나 기업 경쟁력 제고를 할 수 있음을 제안하였다

최근에는 기업 내의 정보시스템에 축적된 이벤트 로그 데이터를 활용하여 프로세스 행동을 예측하는 분야인 프로세스 지능화의 중요성이 커지고 있다(Evermann *et al.*, 2017). 특히 구매 및 물류 프로세스에 적용된 프로세스 지능화는 물류 비용의 절감에 크게 기여할 수 있을 것이다. 일반적으로 구매 및 물류 프로세스는 크게 3개의 하위 프로세스(즉, 입고 관리, 재고 관리, 출고관리)로 이루어지며, 각 하위 프로세스는 다시 여러 하위의 작업들로 구성된다. 구매 및 물류 프로세스를 구성하는 일련의 작업들이 병목없이 원활하게 진행되어 비용을 절감하기 위해서는 필요한 물품이 정시에 정량으로 제공될 필요가 있다.

구매 및 물류 프로세스를 개선하기 위해서는 현재의 업무수행 방식을 분석하여 문제점을 파악하고, 물류 프로세스의 행동을 올바르게 예측하는 것이 무엇보다도 중요하다. 하지만 최종 제품의 생산에 투입되는 다양한 자재가 동시 다발적으로 구매되어 관리되고, 다양한 협력업체가 자재를 공급하기 때문에 표준화된 프로세스를 발견하고, 정시 또는 정량 공급과 관련된 프로세스 성과지표를 정확하게 예측하는 것은 매우 어려운 과제이다.

이러한 과제를 해결하는 방안으로 담당자 인터뷰와 설문조사와 같은 정성적인 방법이 주로 활용되어 왔다. 그러나 이러한 방법은 경험에 기

반한 정보를 제공하고 표준화된 구매 및 물류 프로세스를 발견할 수 없다는 단점을 가진다. 기존 방법의 이러한 한계를 극복하기 위해 최근에는 구매 및 물류 프로세스 수행을 지원하는 정보시스템에 저장된 데이터를 활용하여 현행 프로세스를 분석하고, 각 자재 유형별 구매 및 물류 프로세스를 예측하여 미래에 발생할 리스크를 선제적으로 줄이는 방안을 다룬 다양한 연구들이 수행되고 있다. 예를 들어, 이동하 등(2014) 연구에서는 조선 산업의 블록조립 계획 시스템에 저장된 데이터를 바탕으로 다양한 조립 블록에 대한 필요 시기 및 리스크 관리 모형을 도출하였다. 송인성, 조남욱(2015) 연구에서는 물류관리시스템의 시계열 데이터를 ARIMA 분석을 통해 품목별 출고 횟수 및 수량을 예측하였다. 유영석 등(2016) 연구는 조선 산업에서의 빅데이터 분석을 통해 영업 리스크 발생 요인들을 상시 모니터링 할 수 있는 방안을 제안했고, 구매 프로세스를 관리할 때 자주 발생하는 체크포인트를 9가지로 구분하였다. 자재구매단가 변경, 구매오더 삭제, 구매수량 변경, 납기일 변경, 납기일 지연, 검수 후 합격품 사후 관리, 대금지불조건 변경을 주요 체크포인트로 제시하였다. Lee(2017) 연구에서는 선박 건조 단계에 따른 참조 모델을 제안하였고, 각 시기에 따른 자재량 예측을 제시하였다. 오민재 등(2018) 연구에서는 조선업 데이터 분석을 통해 배관 자재 프로세스를 분석하고, 배관 자재량의 구매 수요량을 예측하였다.

2.2 프로세스 마이닝

프로세스 마이닝이란 기업이 구축한 정보시스템에서 프로세스 진행 과정에서 발생한 이벤트 로그로부터 의미 있는 지식을 선별해 내는 과정이다(Van der Aalst, 2016). 즉, 특정 프로세스 액티비티 실행 과정에서 축적된 기록을 분석하여 프로세스의 개선이나 재설계에 필요한 의미있는 지식을 추출하는 것이다. 프로세스 진행 과정에서 발생한

이벤트 로그를 분석하는 방법에는 통계 기법, 인공지능 기법, 사회적 네트워크 기법 등 다양한 기법들이 사용된다. 프로세스 마이닝을 통해 발견된 프로세스 모델과 지식은 작업자의 업무 실행을 효율적으로 개선하거나 업무 재설계에 활용할 수 있다. Yoo *et al.*(2016) 연구에서는 고객들의 병원 대기 시간 및 체류 시간을 줄이기 위해 고객들의 병원 체류 시간 관련 데이터를 프로세스 마이닝을 수행하였다. Andrews *et al.*(2018) 연구에서도 병원 응급실에서의 데이터를 프로세스 마이닝을 통해 환자들의 이동 흐름을 증진할 수 방법을 제시하였다.

프로세스 마이닝은 다음과 같은 점에서 데이터 마이닝이나 BPM(business process management)과 차별화된다. 데이터마이닝은 대용량의 복잡한 데이터에서 의미있는 정보를 추출하고, 유용한 규칙이나 패턴을 찾는 과정이다(레퍼런스 필요). 활용되는 기술은 크게 군집(clustering), 분류(classification), 예측(prediction)으로 나눌 수 있다. 프로세스 마이닝은 구축된 정보시스템에 존재하는 데이터를 분석하여 지식을 얻는다는 점에서 데이터 마이닝과 유사하다고 할 수 있다. 그러나 프로세스 마이닝은 축적된 이벤트 로그의 시간 정보를 활용하여 각 부문의 프로세스를 분석하여, (중단간) 프로세스 모델을 도출한다는 점에서 데이터 마이닝과 차별화 된다. 한편, BPM은 기존 프로세스 모델을 바탕으로 분석하거나 사람에 의해 정의된 모델을 분석하지만, 프로세스 마이닝은 실제 업무 프로세스를 데이터를 바탕으로 도출한다는 점에서 다르다(Van der Aalst, 2016).

구매 및 물류 분야에서도 프로세스 마이닝을 적용하여 비용을 절감할 수 있다(김현경, 신광섭, 2014). 특히, 구매 및 물류 프로세스는 입고 관리, 재고관리, 출고 관리 등의 일련의 서브 프로세스로 이루어지기 때문에, 프로세스 마이닝을 통해 병목을 살펴보거나 재작업 등의 비용 발생요인을 찾을 수 있다. 예를 들어, 강영식(2014)은 SAP ERP 프로세스 마이닝을 대기업의 구매 프로세스 개선에 적용하였다. 이 연구에서는 구매오더 생성에

서부터 자재 입고까지의 조달프로세스를 살펴보고, 적기 입고를 위한 프로세스 개선 방안을 제안하였다. 전대옥, 배혜림(2015) 연구에서는 컨테이너 터미널에서 컨테이너 취급 프로세스를 분석하기 위한 개념적 분석 프레임워크를 제안하였다. Wang *et al.*(2014)에서는 중국 벌크 항구에서 물류 프로세스 마이닝 기법을 적용하였다. 그러나 입고 및 출고데이터를 바탕으로 프로세스 마이닝 기법을 적용한 연구는 미흡한 실정이며(김현경, 신광섭, 2014), 프로세스 마이닝에서 발견한 통찰력과 딥러닝의 예측 역량을 결합하려는 시도도 부족한 상황이다.

2.3 순환신경망: RNN과 LSTM

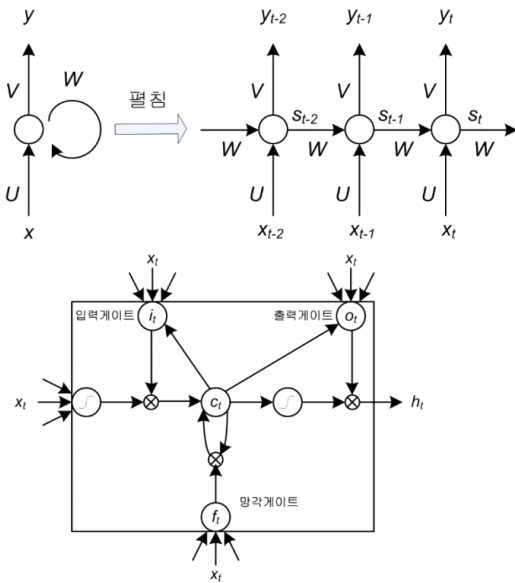
최근에 딥러닝 접근법을 통해 프로세스를 예측하려는 다양한 연구들이 수행되고 있다(예: Evermann *et al.*, 2017; Tax *et al.*, 2017). 이러한 연구들은 시간 흐름이 있는 데이터 분석에 활용되는 RNN 알고리즘을 활용하고 있다. RNN은 시간 흐름이 있는 데이터의 패턴을 분석하여 특정 분야에 대한 예측 모형을 만들기 위해 활용되는 알고리즘이다. RNN과 기존 신경망 모형과의 가장 큰 차이점은 시간 흐름에 따른 데이터 패턴을 분석할 수 있고 특정 시점의 결과 예측이 이전 시점의 데이터들에 영향을 받는다는 점이다(Evermann *et al.*, 2017). <그림 1>에서 볼 수 있듯이, RNN은 다른 인공 신경망 구조와 비슷하게 입력층, 은닉층, 출력층으로 이루어져 있다. 각 층에는 데이터 패턴과 이에 해당하는 분류 결과를 학습하기 위한 행렬이 있으며, 출력 값과 실제 값의 차이를 최소화하기 위해 행렬의 가중치들을 찾는 것을 목표로 한다. RNN 모델의 시간 t 에 대한 활성화 함수는 다음과 같다.

$$\text{은닉층: } s_t = \tanh(Ux_t + Ws_{t-1} + \text{bias})$$

$$\text{출력층: } y_t = \text{sigmoid}(Vs_{t-1} + \text{bias})$$

x 는 입력변수이고 y 는 출력변수이며 s 는 은닉

층의 상태를 의미한다. 그리고 U, W, V 는 네트워크의 가중치이다. 이 함수에서 가중치와 편차(bias)는 임의로 초기화되며, 경사하강법을 이용한 역전파(back-propagation) 알고리즘을 통해 학습된다. 각 반복시점에서 출력 값과 실제 값과의 차이인 오류 값은 출력층으로부터 은닉층과 이전 시점으로 역방향 전송된다(트란 광 카이, 송사광, 2017). 하지만 전통적 RNN은 데이터의 흐름이 길어질수록 역전파 알고리즘의 장기의존성으로 학습이 어렵다(Bengio et al., 1994). 즉, 특정 시점에서 먼 과거 정보를 예측 결과에 반영하지 못한다는 구조적 단점을 가지고 있다. 이런 장기 의존성 문제를 해결하기 위한 알고리즘으로 LSTM(Long Short-Term Memory)이 제안되었다(Hochreiter and Schmidhuber, 1997).



〈그림 1〉 RNN과 LSTM

LSTM의 핵심은 RNN의 은닉층이 메모리 셀 연산으로 대체되고, 과거 정보의 망각과 유지를 담당하는 메모리 구조를 갖는다는 것이다. 즉, LSTM은 먼 과거 시점을 포함한 과거 셀 상태 정보와 은닉층의 연산 결과를 다음 시점 은닉층으

로 전달한다. 셀 연산을 위한 세 개의 게이트가 존재하며 이들은 입력 데이터를 선택적으로 학습한다. LSTM 메모리 셀 연산과 활성화 함수는 다음과 같다:

$$\text{입력게이트: } i_t = \sigma(U_i x_t + W_i s_{t-1} + bias_i)$$

$$\text{망각게이트: } f_t = \sigma(U_f x_t + W_f s_{t-1} + bias_f)$$

$$\text{출력게이트: } o_t = \sigma(U_o x_t + W_o s_{t-1} + bias_o)$$

$$\text{블록층: } g_t = \sigma(U_g x_t + W_g s_{t-1} + bias_g)$$

$$\text{셀 상태(Cell State): } c_t = c_{t-1} \times f_t + g_t \times i_t$$

RNN에 기반을 둔 예측 분야의 기존 연구들은 정적(Static) RNN에 기반을 두고 프로세스 예측 알고리즘과 모델을 개발하였다. 서지혜, 용환승(2017) 연구에서는 소비자들의 추천시스템을 RNN, LSTM, GRU를 바탕으로 학습시켜 각 알고리즘의 성능을 비교하였다. 방성혁 등(2018) 연구에서는 RNN을 통해 고객들의 행위 의도 예측 모델을 개발하였다. Tax et al.(2017) 연구에서는 기업의 업무 프로세스를 LSTM을 기반으로 분석하여 최적의 업무 프로세스를 설계하였다. 하지만 정적 RNN을 활용한 프로세스 예측 알고리즘과 모델은 실제 프로세스의 행동을 제대로 반영하지 못하고 특정 프로세스에 적용할 때마다 많은 노력과 시간을 요구한다는 한계를 가진다. 또한 정적 RNN은 시퀀스 길이보다 작은 액티비티 수를 가진 데이터를 학습과 예측에 활용할 수 없을 뿐만 아니라 설정된 시퀀스 길이보다 많은 액티비티를 갖는 경우도 학습과 예측에 포함되기 힘들다(김정연 등, 2018). 이러한 한계를 극복하기 위해 동적(Dynamic) RNN을 활용하여 프로세스 예측 모델을 개발할 필요가 있다. 정적 RNN과 달리, 동적 RNN은 시퀀스 길이에 상관없이 모든 케이스를 활용할 수 있다는 장점을 가진다. 이와 함께, Du(2017) 연구에서는 정적 RNN은 예측 모델이 복잡해지면, 메모리에 저장해야 할 양이 늘어나 예측을 하는데 긴 시간이 걸림을 밝혀내었으며, 동적 RNN은 유동적인 시퀀스 길이를 다룰 수 있고, 복

잡한 모형에서는 정적 RNN보다 더 빠른 속도로 결과 값을 도출할 수 있음을 보였다.

III. 접근법

3.1 프로세스 마이닝과 프로세스 성과지표 예측의 연계 방안

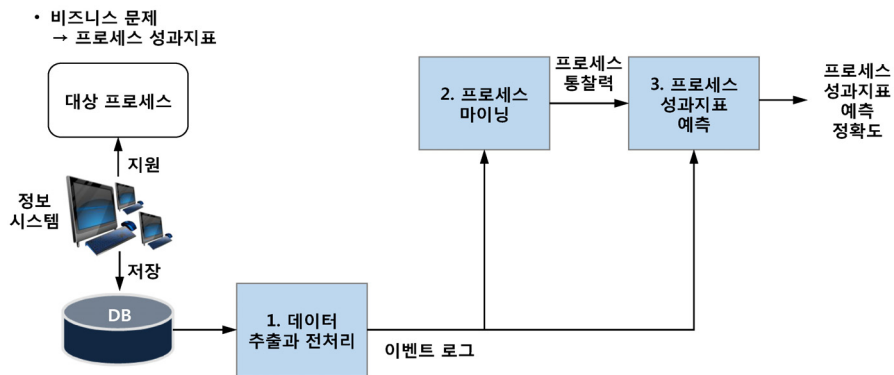
프로세스 마이닝과 프로세스 성과지표를 연계하기 위한 개념적 모델이 <그림 2>에 제시되어 있다. 일반적으로 조직의 정보시스템은 분석과 예측 대상인 프로세스의 업무 수행을 지원하며 업무 수행과 관련된 다양한 데이터가 데이터베이스에 저장된다. 예를 들어, 사례 기업인 A 대기업의 SAP ERP 시스템은 구매 프로세스의 특정 업무를 수행할 때마다 발생하는 이벤트 데이터를 기록하고 있다(Breuker *et al.*, 2016). 각 이벤트는 발생시점을 나타내는 타임스탬프와 특정 프로세스 인스턴트(케이스)에 속함을 표시하는 케이스 아이디 속성을 가진다(Van der Aalst, 2016).

3.2 프로세스 성과지표 예측을 위한 구조

RNN 형태의 딥러닝을 활용하여 프로세스의 다음 이벤트(즉, 액티비티)를 예측하는 접근법은 우수한 성과를 달성하였다(예: Evermann *et al.*,

2017; 김정연 등, 2018). 일반적으로 특정 프로세스를 구성하는 수십 개 액티비티들의 수행 순서는 학습을 통해 수행 패턴을 파악할 수 있기 때문이다. 이러한 접근법을 확장하여 본 연구는 동적 RNN 형태의 딥러닝을 활용하는 접근법을 구현했다. 구현된 접근법은 <그림 2>의 ‘3. 프로세스 성과지표 예측’에서 활용되었다. 그러나 구현된 접근법을 활용하여 다음 액티비티의 수행 시점(예, 구매 프로세스에서 ‘최초입고’ 액티비티의 수행 시점)과 같은 프로세스 성과지표를 예측하면 학습이 제대로 수행되지 않거나 예측 정확도가 현저히 떨어졌다. 특정 액티비티의 수행 시점은 너무나 많은 경우의 수를 가지기 때문에 프로세스의 다음 액티비티를 예측하는 접근법을 활용하는 것은 한계를 지닌다.

이러한 한계를 극복하기 위해 본 연구는 특정 케이스에서 발생한 이벤트 간의 시간 차이를 범주화하여 예측에 활용하는 방식을 적용했다. 구체적으로 설명하면 다음과 같다. 예측에 활용되는 데이터에 포함된 각 케이스에 소속된 이벤트 간의 시간 차이의 모든 최댓값(예, ‘900’)과 최솟값(예, ‘0’)을 구해서 최대 시간 차이를 구한다. 만약 최대 시간 차이가 900(= 900-0)이고, 시간 범주의 수를 30이라고 설정하면(학습할 때 전체 시간 범주를 일종의 하이퍼 파라미터로 간주해서 다양한 값들을 시험함) 첫 번째 범주(즉, ‘범주 1’)는



<그림 2> 프로세스 마이닝과 프로세스 성과지표 예측의 연계를 위한 개념적 모델

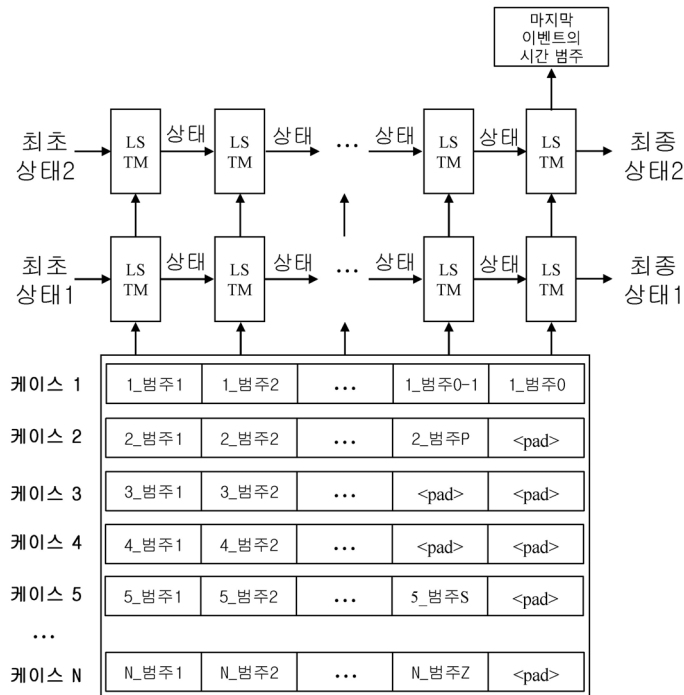
<표 1> 특정 케이스에서 발생한 이벤트 간의 시간 차이를 범주화하는 방법

케이스 아이디	액티비티	타임스탬프	변환된 타임스탬프	시간 차이	시간 범주
40721927-1	구매오더 생성	2012/06/29 21:14:03	2012/06/29	0	범주 1
40721927-1	변경(계약납기일)	2012/06/29 21:12:05	2012/06/29	0	범주 1
40721927-1	공급업체변경확인	2012/08/20 11:42:27	2012/08/20	52	범주 2
40721927-1	자재입고	2014/05/29 20:04:05	2014/05/29	647	범주 22

0~30일의 시간 차이에 할당되고, 마지막 범주(즉, ‘범주30’)는 871~900일의 시간 차이에 할당된다. 예를 들어, 구매 프로세스에서 ‘최초입고’ 액티비티의 수행 시점을 예측하는 경우에 케이스 아이디 ‘40721927-1’에 소속된 4개 이벤트의 타임스탬프는 <표 1>과 같다. 구매 자재의 최초 입고일을 예측하는 것이 목표이므로 최초 타임스탬프에서 ‘시:분:초’ 정보를 삭제하고 ‘년/월/일’ 정보만을 이용한다. 그런 다음에 ‘변환된 타임스탬프’를 기준으로 직전 이벤트와의 시간 차이를 구한다. 각

케이스의 첫 번째 이벤트는 직전 이벤트가 없기 때문에 시간 차이를 ‘0’으로 할당한다. <표 1>의 두 번째 이벤트는 첫 번째 이벤트와 같은 날에 수행되었기 때문에 시간 차이가 ‘0’으로 할당되었다. ‘시간 범주’는 이벤트 간의 ‘시간 차이’ 값에 따라 할당된다.

각 케이스의 이벤트에 대응되는 ‘시간 범주’ 데이터를 활용해서 프로세스 성과지표를 예측하는 동적 RNN 기반의 구조는 <그림 3>과 같다. 전체 케이스 중에서 이벤트 수가 가장 많은 케이스(예:



<그림 3> 프로세스 성과지표 예측을 위한 동적 순환신경망 기반의 구조

‘케이스 1’)에 맞추어 시퀀스 길이를 설정하고, 나머지 케이스들에 대해서는 <그림 3>과 같은 패딩(padding) 작업을 해 준다. 각 이벤트의 ‘시간 범주’와 함께 데이터에 포함된 속성과 프로세스 마이닝을 통해 도출한 프로세스 행동과 관련된 속성 정보가 동적 RNN의 LSTM 셀에 입력될 수 있다.

3.3 프로세스 성과지표 예측을 위한 구조의 구현

<그림 3>에 제시된 구조를 구현하기 위해 오픈소스 딥러닝 프레임워크인 텐서플로우와 파이썬 프로그래밍 언어를 이용했다. 개발된 프로그램은 CSV 형식의 데이터를 파싱하여 각 이벤트의 ‘시간 범주’를 구하고, 패딩 처리를 담당하는 파싱 모듈을 포함하고 있다. 앞에서 언급한 것처럼, 배치 크기, 에포크 수, 레이어 수, 활성화 함수와 최적화 함수와 같은 하이퍼 파라미터의 설정과 함께 시간 범주의 수도 일종의 하이퍼 파라미터로 설정할 수 있도록 구현했다. 개발된 프로그램은 예측된 시간 범주의 중간값을 직전 이벤트의 변환된 타임스탬프에 더하여 최종 예측일(예: 구매 프로세스의 ‘최초 입고일’)을 구한다.

IV. 접근법 검증

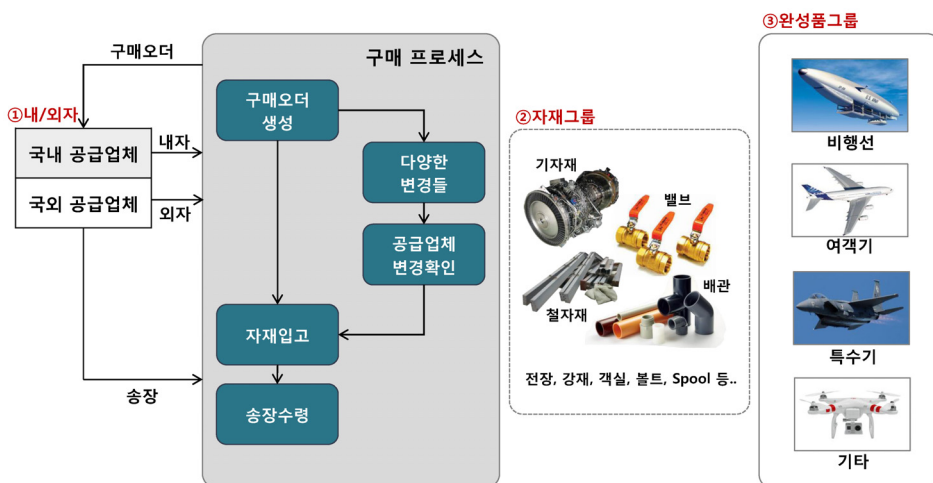
4.1 대상 프로세스와 예측 성과지표

연구대상 프로세스인 A 대기업의 구매 프로세스는 <그림 4>와 같다. A 대기업의 표준 구매 프로세스는 매우 단순하나, <그림 5>에 제시한 것과 같이 프로세스 마이닝을 통해 발견한 프로세스 맵은 이보다 훨씬 더 복잡함을 알 수 있다.

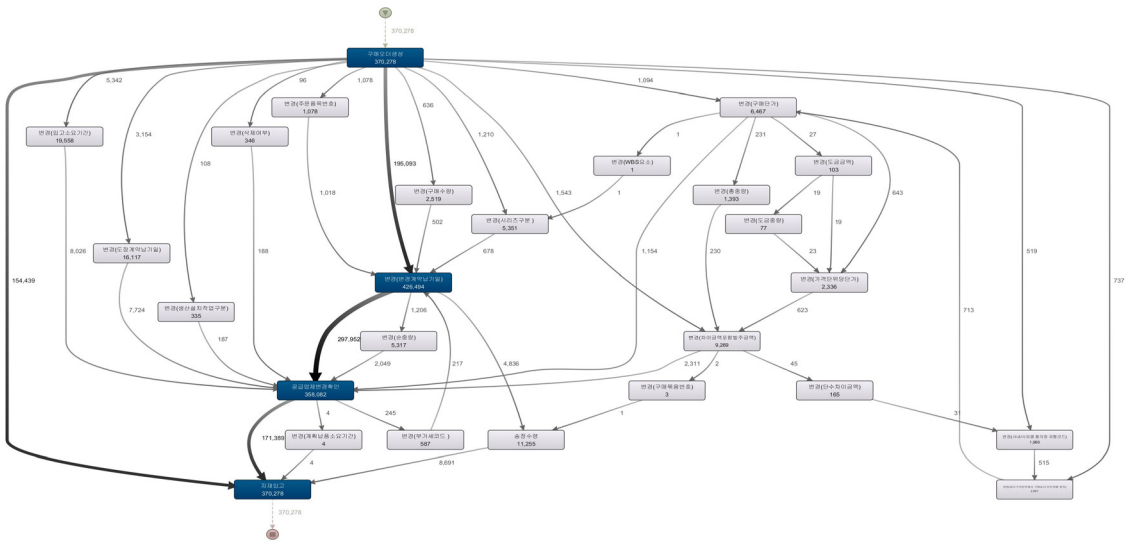
A 대기업과의 인터뷰 결과, 구매 프로세스와 관련하여 두 가지 핵심 이슈의 해결을 원하고 있었다.

- 정상 리드타임으로 발행된 구매주문서에 대해 지연 입고되는 자재가 발생하여 생산 결품이 발생함.
- 생산계획일 대비 조기 입고되는 자재가 발생하여 장기 재고가 발생함.

이러한 두 가지 핵심 이슈를 해결하기 위해 본 연구는 구매 자재의 지연 또는 조기 입고와 관련된 프로세스 성과지표인 ‘최초 입고일’을 예측하고자 한다. A 대기업과의 협의를 통해 예측된 ‘최초 입고일’과 SAP ERP 시스템에 기록된 ‘최종 계약 납기



<그림 4> A대기업의 표준 구매 프로세스



<그림 5> A 대기업의 발견된 프로세스 맵(액티비티: 100%, 경로: 0%, 절대빈도 표시)



<그림 6> 최초 입고일과 최종 계약납기일 간 차이에 기반을 둔 7개의 적기입고 수준

일' 사이의 차이를 기반으로 7개의 적기 입고 수준을 정의하였고, <그림 6>에 제시하였다.

- Best('BT'): 최초 입고일과 최종 계약납기일 간 차이가 -7일~+7일
- 지연_Good('L_GD'): 최초 입고일과 최종 계약납기일 간 차이가 +8일~+14일
- 지연_Bad('L_BD'): 최초 입고일과 최종 계약납기일 간 차이가 +15일~+28일
- 지연_Worst('L_WT'): 최초 입고일과 최종 계약납기일 간 차이가 +29일 이상
- 조기_Good('E_GD'): 최초 입고일과 최종 계약납기일 간 차이가 -8일~-14일

약납기일 간 차이가 -8일~-14일

- 조기_Bad('E_BD'): 최초 입고일과 최종 계약납기일 간 차이가 -15일~-28일
- 조기_Worst('E_WT'): 최초 입고일과 최종 계약납기일 간 차이가 -29일 이상

구매 자재의 적기 입고 수준을 예측하는 것은 구매 및 물류 프로세스에서 자재공급량 부족이나 공급계획 차질을 줄이는 것에 기여할 수 있다. 또한 정확한 입고일을 예측할 수 있다면 기업들은 재고 낭비를 최소화할 수 있고, 인력 및 공간의 효율성을 제고할 수 있다(송인성, 조남욱, 2015;

유영석 등, 2016). 송인성, 조남욱(2015) 연구에서도 물류관리시스템에서 적기 입고 수준을 파악하는 것이 중요하다고 제안하였다. 또한 유영석 등(2016) 연구에서도 조선 산업에서의 빅 데이터 분석을 통해 영업 리스크 발생 요인들로 적기 입고의 중요성을 밝혀내었다.

그런데 이러한 구매 프로세스를 분석하고 예측할 때의 중요한 도전과제가 존재한다. 강영식 등(2015) 연구에서 지적한 것과 같이, 실제 대기업의 구매 프로세스는 ① 자재그룹, ② 완성품그룹, ③ 내·외자 등의 구분에 따라서 별도로 분석되고 개선될 필요가 있다. 이에 따라, 본 연구도 A 대기업과의 협의를 통해 ‘지연_Worst’가 가장 빈번하게 발생하는 ‘벨브’, ‘기자재’ 자재그룹의 외자 중 ‘비행선’과 ‘여객기’ 완성품그룹을 위한 구매 자재의 ‘최초 입고일’을 예측하여 ‘적기입고수준’을 최종적으로 예측하고자 한다.

4.2 활용 데이터와 전처리

제안한 접근법을 검증하기 위해 본 연구는 국내 A 대기업의 구매 프로세스 데이터를 활용했다. SAP ERP 시스템에서 추출한 데이터 중에서 ‘벨브’, ‘기자재’ 자재그룹의 ‘비행선’과 ‘여객기’ 완성품그룹을 위한 외자 구매와 관련된 데이터가 프로세스 마이닝 분석과 예측에 활용되었다. <표 2>에 제시한 것과 같이 본 연구는 최초 입고일을 예측하는 것이 목표이므로 각 케이스의 최초 ‘자재입고’ 이후에 발생한 이벤트 데이터를 제거했다. 이러한 제거 과정을 거친 데이터는 48,744개의 이벤트와 7,373

개의 케이스를 포함하고 있다. 제4.3절에 설명할 프로세스 마이닝 결과를 반영하여 동시에 수행된 액티비티들의 그룹화(grouping)를 통해 1개의 액티비티(즉, ‘변경(WBS요소)’)를 제외하였다.

반복 실험을 통해 가장 높은 예측 정확도를 가지는 시간 범주의 수가 50개임을 파악하였다. 41번째부터 50번째의 시간 범주를 가지는 케이스의 수는 32개(0.46%)에 불과하기 때문에 이러한 케이스들을 극단치(outlier)로 간주하여 제거했다. 최종적으로, 분석과 예측에 활용된 7,341개의 케이스는 최소 2개부터 최대 179개의 이벤트를 포함하고 있다. 이들을 임의로 나누어서 학습 데이터(70%, 5,138개의 케이스)와 테스트 데이터(30%, 2,203개의 케이스)를 구성했다. 7,341개 케이스의 평균 수행기간(duration)은 40.3주이고, 중간 수행기간(duration)은 39.6주로 파악되었다. <표 2>에 활용 데이터들의 정보를 요약하여 제시하였다. 마지막으로, 7,341개 케이스의 적기 입고 수준 비율은 <표 3>에 제시하였다. ‘지연_Worst’의 비율이 57.7%로 가장 높음을 알 수 있다.

<표 3> 적기 입고 수준 분포

적기 입고 수준	빈도	비율(%)
Best(‘BT’)	406	5.5
지연_Good(‘L_GD’)	175	2.4
지연_Bad(‘L_BD’)	262	3.6
지연_Worst(‘L_WT’)	4,238	57.7
조기_Good(‘E_GD’)	99	1.4
조기_Bad(‘E_BD’)	339	4.6
조기_Worst(‘E_WT’)	1,822	24.8
전체	7,341	100

<표 2> 활용 데이터 요약

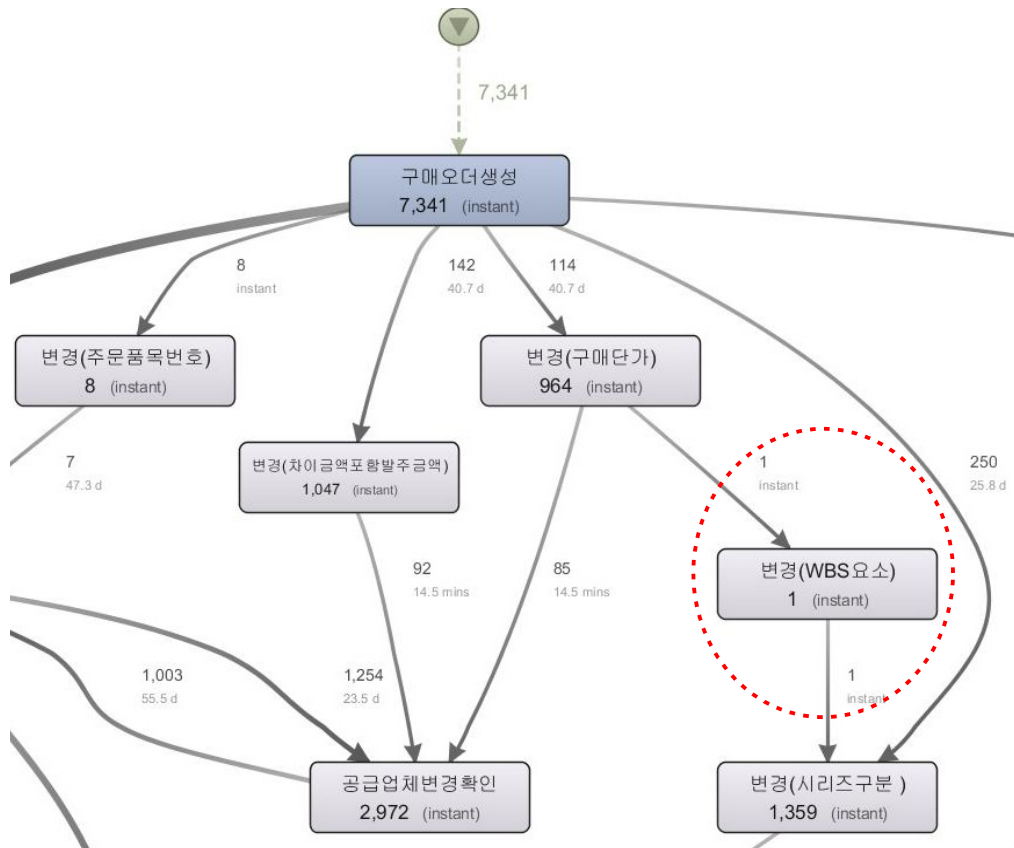
구분	값	구분	값
최초 이벤트 수	48,744개	최초 케이스 수	7,373개
학습 데이터의 케이스 수	5,138개	테스트 데이터의 케이스 수	2,203개
제외 전 액티비티 종류의 수	17개	제외 후 액티비티 종류의 수	15개
케이스의 최대 이벤트 수	179개	케이스의 최소 이벤트 수	2개
시작 일시	2009/07/25 00:00:00	완료 일시	2014/07/25 08:44:35
평균 수행기간	40.3주	중간 수행기간	39.6주

4.3 프로세스 마이닝 분석 결과의 활용

프로세스 성과지표(즉, ‘최초 입고일’) 예측의 정확도를 높이기 위해 프로세스 마이닝에서 도출한 과거의 업무 수행에 관한 통찰력을 2가지 측면에서 활용하였다. 첫째, <그림 7>과 같이 프로세스 마이닝을 통해 발견한 프로세스 맵을 살펴보면 동시에 수행되는 액티비티들이 포함되어 있음을 알 수 있었다. 이러한 액티비티들의 포함은 액티비티 간 대기시간의 하한값을 왜곡할 수 있기 때문에 ‘최초 입고일’ 예측에 영향을 줄 수 있다. 그러므로 동시에 수행된 액티비티들의 그룹화

(grouping)를 통해 1개의 액티비티(즉, ‘변경(WBS 요소)’를 제외하였다.

둘째, 적절한 예측 변수의 포함은 예측 정확도 향상에 크게 기여할 수 있다(Evermann *et al.*, 2017). 예를 들어, 데이터에 포함된 속성(예, ‘자재그룹’, ‘완성품그룹’, ‘내·외자’, ‘공급업체코드’, ‘최종 계약납기일’, ‘구매 주문서등록자 아이디’)만을 예측 변수로 포함하는 것보다 프로세스 마이닝에서 도출한 프로세스 행동과 관련된 속성(예: 케이스별 ‘공급업체변경확인 횟수’, ‘총 변경 횟수’, ‘구매단가 변경횟수’)을 포함하는 것이 예측 정확도 향상에 기여할 수 있다.



<그림 7> ‘비행선’과 ‘여객기’ 완성품그룹을 위한 ‘밸브’와 ‘기자재’ 자재그룹의 외자 구매 프로세스 맵의 일부 (액티비티: 100%, 경로: 0%, 절대빈도와 평균 수행시간 표시)

4.4 학습 방법

제4.2절에서 설명한 5,138개의 케이스를 포함한 데이터를 학습하기 위해 우리가 구현한 동적 RNN은 2개의 레이어(layer)를 갖도록 설정했다. 각 레이어의 시퀀스 길이(sequence length)는 179(케이스의 최대 이벤트 수)가 될 것이다. 각 LSTM셀의 유닛(unit) 수를 40으로 설정하였다. <표 4>에 하이퍼 파라미터 설정값을 제시하였다. 배치 크기는 예측 정확도에 큰 영향을 주지 않았으나 200으로 설정했을 때 가장 좋은 예측 정확도를 얻을 수 있었다. 5,138개의 케이스를 포함한 학습 데이터가 50번 반복하여 학습되도록 에포크(epoch) 수를 50으로 설정했다. 학습률은 0.01로 설정되었다. 마지막으로, 과적합(overfitting)을 방지하기 위해 3겹 교차검증 기법을 적용하였다.

<표 4> 하이퍼 파라미터 설정 값

하이퍼 파라미터	값
시간 범주의 수	50
겹(fold)	3
에포크	50
배치	200
학습률	0.01
레이어	2
셀의 유닛 수	40

4.5 예측 결과

테스트 데이터를 이용하여 적기 입고 수준을 예측한 정확도를 <표 5>에 제시하였다. 앞에서 언급한 것처럼, 적기 입고 수준 중에서 ‘자연_Worst’의 비율(57.7%)이 가장 높게 나타났다. 이것은 구매한 자재의 입고 수준을 ‘자연_Worst’라고 임의로 예측하면 57.7%의 정확도를 가질 수 있음을 의미한다. 본 연구에서 제안한 동적 RNN을 활용한 프로세스 성과지표 예측 모형은 이보다 훨씬 높은 예측 정확도(<표 5>의 ② 71.0%, 설

명력 13.3% 증가)를 가짐을 알 수 있다. 이와 함께, 프로세스 마이닝에서 발견한 2가지 통찰력을 반영하는 것은 예측 정확도의 향상(<표 5>의 ⑤ 74.0%)에 기여함을 알 수 있다. 특히, 프로세스 행동과 관련된 속성을 예측 변수로 이용하는 것이 더욱 효과가 있었다.

<표 5> 테스트 데이터를 이용한 예측 정확도

구분	예측 정확도(%)
① ‘자연_Worst’의 비율	57.7
② 데이터에 포함된 속성만을 예측 변수로 이용	71.0
③ ②에서 ‘변경(WBS 요소)’ 액티비티를 제거함	71.2
④ 데이터에 포함된 속성 및 프로세스 행동과 관련된 속성을 모두 이용	73.2
⑤ ④에서 ‘변경(WBS 요소)’ 액티비티를 제거함	74.0

V. 결론 및 시사점

최근 경영정보와 SCM 분야 연구자들은 프로세스 마이닝과 RNN같은 딥러닝 기법을 결합하여 주요 성과 지표를 예측할 수 있다면 기업들은 더 많은 시사점과 정보를 얻을 수 있다고 주장하고 있다. 하지만 구매 및 물류 프로세스를 지원하는 정보시스템에 축적된 데이터를 활용하여 SCM의 주요 성과 지표인 적기 입고 시점을 예측하려는 연구는 거의 진행되지 않았다. 구매 및 물류 데이터를 분석하고 구매 프로세스의 적기 입고 수준을 예측하는 것은 기업의 비용 절감이나 경쟁력 확보에 필수적 요소가 되고 있다. 데이터 분석이 활용되지 않으면, 필요한 자재가 부족하거나 계획보다 빨리 입고되어 재고로 쌓이는 경우들이 발생하게 된다. 그래서 본 연구에서는 이러한 지식 공백을 메우기 위해 본 연구는 프로세스 마이닝과 동적 RNN의 예측 역량을 결합하여 적기 입고 시점을 예측해보고자 한다. 구매 프로세스는 순차적으로

이루어져 있기 때문에, 순차적 프로세스 분석을 활용할 수 있는 RNN을 활용하였다. 더욱이 구매 프로세스는 프로세스의 액티비티 수가 다른 경우가 많기 때문에, 동적 RNN을 적용하는 것이 적합하다. 제안된 접근법을 국내 대기업의 실제 데이터에 적용하였으며, 제안한 접근법을 이용한 적기 입고 수준에 대한 예측 정확도(74.0%)가 다수 클래스(majority class)인 '지연_Worst'의 비율(57.7%)보다 160% 더 높게 나타났다. 이와 함께, 본 연구는 프로세스 마이닝에서 도출한 과거의 업무 수행에 관한 통찰력이 예측 정확도의 향상에 기여할 수 있음을 증명하였다. 본 예측 결과를 통해 구매 관리 부서들은 프로세스 마이닝과 동적 RNN을 통해 적기 입고 수준을 예측하여 재고를 줄이고 결과적으로 운영 비용을 감소시킬 수 있을 것으로 기대된다.

본 연구는 다음과 같은 연구 한계점을 가지고 있다. 공급사슬 파트너 기업 간 정보가 공유되고 이 데이터를 활용할 수 있다면 보다 더 정확한 자체들의 적기 입고 수준을 파악할 수 있을 것으로 예상된다. 황채영, 서창교(2015) 연구에서는 공급사슬 참여자들의 기업들이 데이터를 공유하면 수요예측의 정확성을 높을 수 있고, 외부 환경에 대한 불확실성에 대한 대응성을 강화할 수 있음을 제안하였다. 그래서 적기 입고 수준에 대한 예측 정확도를 높이기 위해서는 공급업체의 내부 데이터가 학습에 활용될 필요가 있다. 왜냐하면 공급업체의 생산계획과 일정에 관한 데이터를 활용할 수 없다면 '최초 입고일'과 이를 활용한 적기 입고 수준을 정확하게 예측할 수 없기 때문이다. 향후 연구는 공급업체의 내부 데이터를 포함함으로써 예측 정확도를 향상할 수 있을 것이다. 이와 함께, 향후 연구는 본 연구에서 제안하고 검증한 접근법을 다양한 종류의 프로세스 성과지표를 예측하는 것에 적용할 수 있을 것이다.

참고 문헌

[1] 강영식, *프로세스마이닝을 이용한 자재유형*

별 구매이력분석을 통한 조달프로세스 개선 방안 도출, 명지대학교 산학협력단, 2014.

- [2] 강영식, 박승범, 이보경, 강성구, “빅데이터 시대에 SAP ERP프로세스 마이닝과 프로세스 모니터링의 연계 방안에 관한 연구”, *e-비즈니스연구*, 제16권, 제4호, 2015, pp. 265-288.
- [3] 김정연, 윤석준, 이보경, 강영식, “동적 순환신경망을 이용한 프로세스 행동 예측에 관한 연구”, *2018 제20회 경영관련학회통합학술대회 한국경영공학회발표집*, 2018, pp. 43-54.
- [4] 김현경, 신광섭, “Process Mining 기법을 이용한 물류센터입출고 프로세스 분석 및 개선 방안 수립”, *한국경영과학회지*, 제39권, 제4호, 2014, pp. 1-17.
- [5] 박광오, 장활식, 정대현, “공급사슬 내의 파워 유형이 정보역량에 미치는 영향: CSV 가치창출 관점”, *정보시스템연구*, 제27권, 제3호, 2018, pp. 1-24.
- [6] 박성민, 박찬권, “환경 불확실성의 다차원성이 Smart SCM 요인 및 기업 경영성과에 미치는 영향에 대한 연구”, *로지스틱스 연구*, 제26권, 제2호, 2018, pp. 105-126.
- [7] 방성혁, 배석현, 박현규, 전명중, 김제민, 박영택, “순환신경망 기반의 사용자 의도 예측 모델”, *정보과학회논문지*, 제45권, 제4호, 2018, pp. 360-369
- [8] 서영복, 김봉기, 박찬권, 윤재희, “대기업과 중소기업의 상생협력을 통한 SCM 혁신 사례 연구: S사와의 협업을 통한 I사의 혁신사례”, *경영컨설팅연구*, 제12권, 제4호, 2012, pp. 495-517.
- [9] 서지혜, 용환승, “전자상거래 추천시스템을 위한 순환신경망 알고리즘들의 성능평가”, *정보과학회 컴퓨팅의 실제 논문지*, 제23권, 제7호, 2017, pp. 440-445.
- [10] 송인성, 조남욱, “예측 기반의 물류관리시스템 최적 저장위치결정모형”, *한국 SCM 학회지*, 제15권, 제1호, 2015, pp. 1-7.

- [11] 오민재, 노명일, 박성우, 김성훈, “빅데이터 분석을 이용한 해양 구조물 배관 자재의 소요량 예측”, *대한조선학회논문집*, 제55권, 제3호, 2018, pp. 243-251.
- [12] 유영석, 박한규, 백승훈, 홍성찬, “빅데이터 환경에서 프로세스 마이닝을 이용한 영업감사 상시 모니터링 강화에 대한 연구”, *인터넷정보학회논문지*, 제17권, 제6호, 2016, pp. 123-131.
- [13] 이동하, 박재훈, 배혜림, “조선 산업에서 프로세스 마이닝을 이용한 블록 조립 프로세스의 계획 및 실적 비교 분석”, *한국전자거래학회지*, 제18권, 제4호, 2014, pp. 145-167.
- [14] 전대욱, 배혜림, “인공신경망을 이용한 항만터미널에서 컨테이너의 비정상 이송 프로세스 예측”, *한국SCM학회지*, 제15권, 제2호, 2015, pp. 117-126.
- [15] 정진우, 이영신, 이보경, 김정연, 강영식, “프로세스 마이닝을 활용한 국내 중소기업 ERP 프로세스 분석에 관한 연구: 국내 화장품 제조기업의 사례를 중심으로”, *Information Systems Review*, 제20권, 제1호, 2018, pp. 81-98.
- [16] 트란 광 카이, 송사광, “딥러닝기반침수수위예측: 미국텍사스트리니티강사례연구”, *정보과학회논문지*, 제44권, 제6호, 2017, pp. 607-612.
- [17] 황채영, 서창교, “SCM통합과 기업성과의 인과관계 분석”, *정보시스템연구*, 제24권, 제1호, 2015, pp. 27-44.
- [18] Andrews, R., S. Suriadi, M. Wynn, A. H. M. ter Hofstede, and S. Rothwell, “Improving patient flows at St. Andrew’s war memorial hospital’s emergency department through process mining”, In: vomBrocke J. and Mendling J. (eds), *Business Process Management Cases. Management for Professionals*, Springer, Cham, 2018.
- [19] Barbosa, M. W., A. Vicente, M. B. Ladeira, and M. P. V. Oliveira, “Managing supply chain resources with big data analytics: A systematic review”, *International Journal of Logistics Research and Applications*, Vol.21, No.3, 2018, pp. 177-200.
- [20] Bengio, Y., P. Simard, and P. Frasconi, “Learning long-term dependencies with gradient descent is difficult”, *Journal of IEEE Transactions on Neural Networks*, Vol.5, No.2, 1994, pp. 157-166.
- [21] Breuker, D., M. Matzner, P. Delfmann, and J. Becker, “Comprehensible predictive models for business processes”, *MIS Quarterly*, Vol.40, No.4, 2016, pp. 1009-1034.
- [22] Du, M., *Empirical Study of Deep Neural Network Architectures for Protein Secondary Structure Prediction* (Master of Science), University of Missouri, 2017.
- [23] Evermann, J., J. R. Rehse, and P. Fettke, “Predicting process behaviour using deep learning”, *Decision Support Systems*, Vol.100, 2017, pp. 129-140.
- [24] Hochreiter, S. and J. Schmidhuber, “Long short term memory”, *Neural Computation*, Vol.9, No.8, 1997, pp. 1735-1780.
- [25] Lee, Y., *A Reference Model for Big Data Analysis in Shipbuilding Industry* (Master’s Thesis), Department of Management Engineering, UNIST, 2017.
- [26] Manyika, J., M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. H. Byers, *BigData: The Next Frontier for Innovation, Competition, and Productivity*, McKinsey GlobalInstitute, 2011.
- [27] Park, K. O. and H. S. Chang, “Studies on determinant factors of SCM performance: From the supplier perspective”, *Asia Pacific Journal of Information Systems*, Vol.21, No.1, 2011, pp. 1-27.
- [28] Sowar, N. and K. Gromley, *Sharper View: Analytics in the Global Steel Industry*, Deloitte, 2011.
- [29] Tax, N., I. Verenich, M. La Rosa, and M. Dumas, “Predictive business process monitoring with LSTM neural networks”, *International Conference on Advanced Information Systems Engineering*, 2017, pp. 477-492.
- [30] Van der Aalst, W. M. P., *Process Mining: Data*

Science in Action, Springer, 2016.

- [31] Waller, M. A. and S. E. Fawcett, "Data science, predictive analytics, and big data: A revolution that will transform supply chain design and management", *Journal of Business Logistics*, Vol.34, No.2, 2013, pp. 77-84.
- [32] Wang, Y., F. Caron, J. Vanthienen, L. Huang, and Y. Guo, "Acquiring logistics process intelligence: Methodology and an application for a chinese bulk port", *Expert Systems with Applications*, Vol.41, No.1, 2014, pp. 195-209.
- [33] Wu, L, X. Yue, A. Jin, and D. C. Yen, "Smart supply chain management: A review and implications for future research", *The International Journal of Logistics Management*, Vol.27, No.2, 2016, pp. 395-417.
- [34] Yoo, S., M. Cho, E. Kim, S. Kim, Y. Sim, D. Yoo, H. Hwang, and M. Song, "Assessment of hospital processes using a process mining technique: Outpatient process analysis at a tertiary hospital", *International Journal of Medical Informatics*, Vol.88, 2016, pp. 34-43.

Exploring the Prediction of Timely Stocking in Purchasing Process Using Process Mining and Deep Learning*

Youngsik Kang** · Hyunwoo Lee*** · Byoungsoo Kim****

Abstract

Applying predictive analytics to enterprise processes is an effective way to reduce operation costs and enhance productivity. Accordingly, the ability to predict business processes and performance indicators are regarded as a core capability. Recently, several works have predicted processes using deep learning in the form of recurrent neural networks (RNN). In particular, the approach of predicting the next step of activity using static or dynamic RNN has excellent results. However, few studies have given attention to applying deep learning in the form of dynamic RNN to predictions of process performance indicators. To fill this knowledge gap, the study developed an approach to using process mining and dynamic RNN. By utilizing actual data from a large domestic company, it has applied the suggested approach in estimating timely stocking in purchasing process, which is an important indicator of the process. The analytic methods and results of this study were presented and some implications and limitations are also discussed.

Keywords: *Deep Learning, Process Mining, Dynamic Recurrent Neural Networks, Purchasing Process, Logistics*

* This work was supported by Jungseok Logistics Foundation Grant

** Department of MIS, Myongji University

*** Department of MIS, Myongji University

**** Corresponding Author, School of Business, Yeungnam University

◎ 저 자 소 개 ◎



강 영 식 (yskang@mju.ac.kr)

고려대학교 산업공학과를 졸업하고, 포항공대에서 산업공학 석사학위(경영정보 전공)를 취득했다. 2008년 한국과학기술원에서 경영정보 공학박사를 취득했으며 현재 명지대학교 경영정보학과에 재직 중이다. 주요 관심분야는 프로세스 마이닝과 딥러닝, RPA(Robotic Process Automation)이다.



이 현 우 (hyunwwkd@gmail.com)

명지대학교 경영정보학과 학부과정에 재학 중이며 프로세스 마이닝과 RPA(Robotic Process Automation)를 활용하는 과제를 수행 중이다. 주요 관심분야는 RPA와 프로세스 마이닝, 머신러닝을 활용한 프로세스 자동화이다.



김 병 수 (kbsyu@yu.ac.kr)

현재 영남대학교 경영대학 경영학과 조교수로 재직 중이다. KAIST 경영대학에서 경영공학 박사 학위를 취득하였고, 지식경영연구센터, SK 텔레콤 IntelligenceLab, 삼성경제연구소 경영전략실, 서울여자대학교 경영학과에서 근무하였다. 주요 관심분야는 서비스 경영, 협업, 지식 생태계, IT 경영, 지식 경영 등이다.

논문접수일 : 2018년 11월 13일

게재확정일 : 2018년 12월 03일

1차 수정일 : 2018년 11월 30일