

# 이커머스 구매전환을 향상을 위한 지능형 고객 관리 시스템

강다빈(\*), 오승민(\*), 홍서영(\*), 신슬비(\*), 임상순(\*)

(\*) 성결대학교 컴퓨터공학과, db021129@naver.com, osm189214@gmail.com, redzerotwo02@naver.com, seulbi0108@naver.com, lssgood80@gmail.com

## 1. 연구 배경

최근 코로나로 인해 온라인 활동이 중요해지면서 이커머스 시장이 더욱 커지고 있다. 이에 따라 이커머스 업체 간 경쟁이 심화되고 고객유치와 구매전환율이 중요해지고 있다[1]. 따라서, 기존 고객들의 물품 구매 패턴 및 물품 구매 만족도를 분석하고 이에 기반을 둔 지능형 고객 상태 관리 시스템의 필요성이 증가하고 있다.

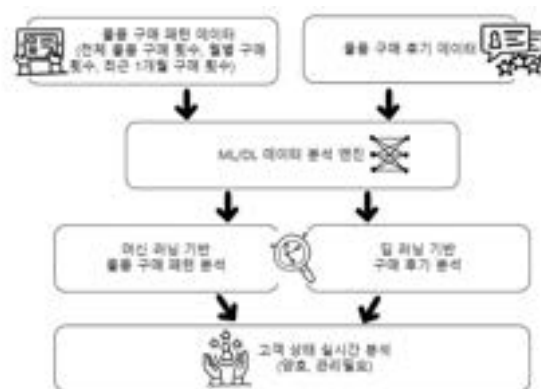
본 논문에서 제안하는 지능형 고객 상태관리 시스템은 해당 이커머스를 사용하는 고객들의 전체 물품 구매 횟수, 월평균 물품 구매 횟수, 최근 1개월 물품 구매 횟수를 분석하고 최근 1개월 물품 구매 횟수가 월평균 물품 구매 횟수보다 특정 임계치(Threshold) 이하로 내려가는 경우 해당 고객의 구매 만족도를 분석하고 관리자에게 ‘양호’, ‘관리 필요’ 등의 분류된 고객 상태를 실시간으로 전달한다. 또한, 물품구매횟수와 긴밀한 상관관계가 있는 것은 물품 구매 평점 및 후기 정보이다. 따라서, 본 연구의 핵심이 되는 고객 상태 분류 기법은 딥 러닝 기반 리뷰 감성 분석을 활용하여 해당 고객의 제품 구매 만족도를 정확히 분석하고자 한다.

딥 러닝을 활용하여 평점과 리뷰데이터 분석을 하면 소비자 이탈의 근원을 해결할 수 있다. 고객들이 해당 이커머스에서 구매한 상품의 만족도를 나타낼 수 있는 평점 및 리뷰 데이터에 대한 긍정, 부정의 비율을 상세 분석하고 소비자들의 추후 행동을 파악한다. 그 후 고객 리뷰 패턴 분석 정보에 기반을 두어 맞춤형 이벤트 프로모션을 소비자들에게 진행한다면 고객 유치와 구매전환율 상승을 성공적으로 해결할 수 있다. 따라서, 본 고에서는 실시간 이커머스 고객의 상태관리 지원을 위한 평점 및 감성 리뷰 분석 서비스를 개발하고자 한다.

## 2. 연구 내용

### 2-1. 이커머스 지능형 고객 관리 시스템 구조도

(그림 1) 지능형 고객 관리 시스템 내부 구성도



### 2-2. CNN 및 LSTM 기반 리뷰 감성 분석

본 연구에서는 고객유치와 구매전환율 상승을 위한 평점 분석 및 파악을 위해 다음과 같이 구현하였다. 리뷰의 내용이 긍정인지 부정인지 알아보기 위한 첫 번째 지표인 리뷰의 평점을 이용한다. 리뷰 및 평점을 크롤링하여 평점의 표를 만든다. 4점 이상의 평점은 긍정, 2점 이하의 평점은 부정으로 레이블링을 진행한다. 또한, 상품의 전반적인 만족도를 파악하기 위해 평균 평점을 구한다. 판매자 및 회사는 평점 비율과 평균을 통해 소비자들의 만족도를 확인할 수 있다. 리뷰는 평점과 더불어 소비자의 변화와 의도를 잘 파악할 수 있는 지표이다. 리뷰를 토큰화하여 딥 러닝을 통해 스스로 긍정 부정 리뷰를 확률로 나타내는 서비스를 다음과 같이 구현하였다.

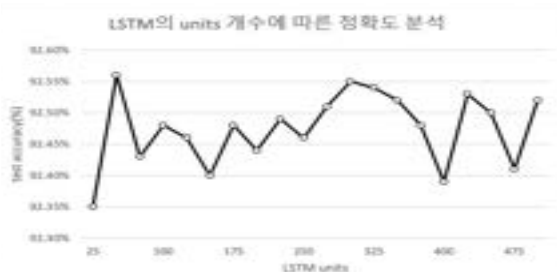
평점을 이용하여 긍정 부정 리뷰를 분류하고 토큰화하여 형태소 단위로 분리한다. 토큰화된 글자를 등장 빈도수에 따라 나열한다. 이후 빈도수대로 토큰화된 글자의 긍정과 부정을 CNN[2]과 LSTM[2]을 이용하여 순차적

으로 학습시킨다. 본 모델에서는 최종적으로 sigmoid 함수를 이용하여 데이터를 긍정 및 부정으로 분류한다. 일반적으로 CNN은 필터와의 합성곱 연산을 이용해 주로 이미지 또는 시계열 데이터에서 특성을 추출하기 위하여 사용되는 신경망으로, 자연어 처리에서 단어 벡터의 각 부분의 특징을 추출하기 위해 사용하기도 한다. LSTM은 RNN을 보완하기 위해 나온 것으로 Forget Gate에서 불필요한 정보를 제거하고, Input Gate에서는 저장할 정보를 기억한다. 이를 반복하고 업데이트하며 Output Gate에서 정확도가 높은 정보를 출력한다. 제안한 시스템에서는 이 과정을 통해 리뷰를 토근화시켜 긍정, 부정 여부와 정확도를 출력한다.

### 3. 실험 및 결과

본 실험에서는 Conv1d 계층과 Pooling 계층은 고정시키고 LSTM units의 개수에 따른 정확도를 분석한다. 25 units부터 500 units까지 25units 간격으로 데이터를 분석하였으며, (그림 2)는 이에 대한 분석한 결과이다.

(그림 2) LSTM의 units 개수에 따른 정확도 분석



대체로 92.55% 정도가 가장 높은 정확도임을 알 수 있다. 이를 통해 높은 정확도를 얻기 위해 unit의 개수는 50, 300, 425, 500으로 설정하는 것이 적합하다. 또한, STM의 unit 개수를 250으로 설정했을 때 layer의 1부터 5까지의 정확도를 측정하였으며, (그림 3)은 분석 결과를 나타낸다.

실험 결과를 바탕으로 LSTM layer가 3개일 때 가장 높은 정확도인 92.57%의 값을 가지는 것을 알 수 있다. 즉, LSTM의 unit 개수를 50, 300, 425, 500 단위로 변경하는 환경에서 연속으로 3개의 LSTM 층을 쌓을 때 (250 units 기준) 가장 정확도가 높음을 알 수 있다.

(그림 3) LSTM의 layer 개수에 따른 정확도 분석



### 4. 기대효과 및 활용

평점 및 리뷰 분석 데이터를 활용하여 고객 구매 전환율 향상과 맞춤 프로모션을 이용한 고객 유치의 기대효과를 볼 수 있다. 딥 러닝을 이용한 기술로 판매자 및 회사는 소비자 관리를 효과적으로 실행하며 상품의 근본적인 문제점을 통찰할 수 있다. 이 점이 앞으로의 소비 동태를 파악하는 지표가 될 것으로 예상된다.

### 5. 결론

본 논문에서는 지능형 고객 상태 관리 시스템의 구조를 제안하고 고객 물품 후기 분석을 위한 CNN 및 LSTM 기반 감성 분석 프레임워크를 제안하였다. 평점을 긍정과 부정으로 구분하였고 평점을 기반으로 CNN의 Conv1D와 MaxPooling1D과 LSTM을 통해 리뷰의 긍정, 부정 여부를 학습시켰다. 이후 LSTM의 unit 개수와 layer 개수에 변화를 주어 정확도를 높이는 실험을 진행하였다. 향후 부정적 단어와 긍정적 단어를 정제된 형태로 빈도수에 따라 나열하여 판매자가 명시적으로 알 수 있도록 발전시켜 활용하고자 한다.

### 참고 문헌

- [1] 장정재, "코로나19 시대 부산의 '비대면 마케팅' 활성화 방안 - 언택트 소비 증가에 대응하는 '라이브 커머스(live commerce)'에 주목", BDI 정책포커스(2020),1-16.
- [2] T. N. Sainath, O. Vinyals, A. Senior and H. Sak, "Convolutional, Long Short-Term Memory, fully connected Deep Neural Networks", IEEE ICASSP 2015