

반복 구매제품의 재구매시기 예측을 위한 다층퍼셉트론(MLP) 모형과 순환신경망(RNN) 모형의 성능비교

송희석*

Comparison of Performance between MLP and RNN Model to Predict Purchase Timing for Repurchase Product

Hee Seok Song*

Abstract

Existing studies for recommender have focused on recommending an appropriate item based on the customer preference. However, it has not yet been studied actively to recommend purchase timing for the repurchase product despite of its importance. This study aims to propose MLP and RNN models based on the only simple purchase history data to predict the timing of customer repurchase and compare performances in the perspective of prediction accuracy and quality. As an experiment result, RNN model showed outstanding performance compared to MLP model. The proposed model can be used to develop CRM system which can offer SMS or app based promotion to the customer at the right time. This model also can be used to increase sales for repurchase product business by balancing the level of order as well as inducing repurchase of customer.

Keywords : Purchase Timing, Recommender, Neural Network, Recurrent Neural Network, Franchise Business, Repurchase Product

Received : 2016. 12. 03. Revised : 2017. 02. 06. Final Acceptance : 2017. 02. 08.

※ This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(No. 2015R1D1A1A09057672).

* Department of Global IT Business in Hannam University, 133 Ojungdong, Daedukgu, Daejeon City, 34430, Korea,
Tel : +82-42-629-8344, e-mail : hssong@hnu.kr

1. 서론

추천시스템은 학계 및 산업계에서 두루 관심이 집중되는 연구 분야로 대부분의 연구가 어떤 아이템을 추천할 것인가에 맞추어져 있다. 이에 비해 추천된 제품을 언제 추천하는 것이 적절한 것인지는 그 중요성에도 불구하고 아직 활발한 연구가 이루어지지 못하고 있다. 특히 음식이나 기호제품과 같이 빈번한 반복구매가 일어나는 아이템의 경우 재구매시기를 적절히 예상하여 적합한 프로모션을 제공할 수 있다면 해당 기업의 매출확대에 직접적으로 기여할 수 있는 방법이 된다. Sato et al.[2015]과 Zhao et al.[2014]의 연구에서도 반복 구매제품의 추천에 있어서 타이밍과 관련된 연구가 아직 초기단계에 있음을 지적하고 있다. 즉 고객의 과거 구매이력을 토대로 고객이 재주문할 시기를 예측하여 정확한 시기에 제품에 대한 프로모션을 보낸다면 경쟁회사로 고객의 유입을 막을 수 있을 뿐 아니라, 고객 주문이 부족할 때 선택적으로 가격할인 등의 쿠폰 프로모션을 실시함으로써 주문의 평준화에도 기여할 수 있다. 본 연구에서는 구매일자과 구매제품으로 구성된 간단한 구매이력 정보만을 활용하여 고객의 재구매시기를 예측하는 모형을 제안하는 것을 목표로 한다. 구매이력 정보만을 활용한 예측 모형은 반복구매가 빈번한 어떤 도메인에서도 적용 가능하기 때문에 범용성이 높다는 장점을 가진다. 제안된 모형은 모바일 SMS 또는 스마트폰용 앱을 기반으로 한 푸쉬 방식의 쿠폰 프로모션을 포함한 고객관계관리(CRM) 시스템을 구축하는 데 활용될 수 있다.

한편 최근 인공지능 바둑프로그램인 알파고의 성공으로 딥러닝에 대한 관심이 고조되고 있다. 딥러닝은 기존 신경망의 계층을 보다 깊이 설계한 것으로 자율주행, 문자인식, 자연어 처리 등에서 좋은 성능을 나타내고 있으나 아직 딥러닝을

기업경영에 응용한 연구는 부족한 상황이다. 이에 본 연구에서는 최근 주목받고 있는 딥러닝 기술 중 순환신경망 모형(RNN; Recurrent Neural Network)을 중심으로 고객관계관리 분야에 이를 적용하여 그 유용성을 살펴보기로 한다. 본 연구에서는 프랜차이즈를 포함한 배달음식 도메인에서 고객별 재구매시기를 예측하는 것을 목표로 하며, 연구모형에서 학습할 패턴은 주말 및 평일 구매형, 식사 및 간식 구매형 등의 패턴이 될 것이며 개인별로 학습된 패턴에 기반하여 적절한 구매 시기를 예측하게 된다. 구매 시기 예측에 활용할 데이터마이닝 기법으로는 대표적인 신경망 모형인 다층퍼셉트론(MLP; Multi Layer Perceptron)과 시간 축에 대해 의미를 연결하는 방식에서 탁월한 성능을 자랑하는 순환신경망(RNN; Recurrent Neural Network)을 활용하여 모형화 하고 그 성능을 비교하고자 한다.

본 연구의 구성은 다음과 같다. 먼저 제 2장에서는 구매 시기를 예측하기 위한 기존의 방법들과 그 성능을 판단하기 위해 관련 문헌들을 조사하였으며 제 3장에서는 대표적인 딥러닝기법 중 하나인 순환신경망에 대해 소개하고 학습방법과 그 한계를 살펴보기로 한다. 제 4장에서는 성능 평가를 위해 다층퍼셉트론과 순환신경망의 구조와 입력 자료를 설계하며, 제 5장에서는 실험에 사용한 데이터 셋과 성능평가 방법 및 실험결과를 요약한다. 마지막으로 제 6장에서는 제시한 모형을 고객관계 관리에 적용하기 위한 전략을 소개하고 향후 연구방향을 제시한다.

2. 기존연구

기존연구에서 재구매시기를 예측하는 접근법은 고객특성이나 구매행동과 같은 독립변수를 활용하는 접근법과 별도의 독립변수에 의존하지 않고 시계열 데이터만을 사용하여 예측하는 접근법

으로 크게 구분된다. 독립변수를 사용하여 재구매시기를 예측하는 접근법으로는 회귀분석이나 로짓분석 등을 활용하는 통계적 방법과 상황정보를 속성으로 활용하는 상황인지 추천이 있으며 시계열 데이터만을 활용하여 재구매시기를 예측하는 접근법으로는 시계열 모형, 추계적 확률모형, 시간간격 패턴마이닝 등 다양한 기법이 있다.

고객특성이나 구매행동과 같은 독립변수를 활용하여 재구매시기를 예측하는 대표적인 연구로는 Agrawal et al.[2011]의 연구와 Oh et al.[2014] 및 Gould and Dong[2000]의 연구가 있다. Agrawal et al.[2011]은 내구재에 대해 고객의 최적 구매시기를 결정하는 모형을 제시한 바 있다. 가구나 고가 전자제품의 경우, 처음에는 가격이 높지만 시간이 가면서 가격이 하락하는 경향이 있는 반면에 고객은 신제품을 남들보다 먼저 구매함으로써 얻는 만족감이 동시에 존재하기 때문에 가격과 선구매에 대한 만족이라는 혜택이 상충관계에 있게 된다. 이 연구에서는 가격과 고객의 선구매에 대한 혜택을 변수로 하여 시간에 따른 고객의 효용함수를 구성하여 최적의 구매시기를 결정하는 모형을 제안한 바 있다. 한편 Oh et al.[2014]은 TV 프로그램 추천시기에 대한 모형을 제시하고 추천기를 개발한 바 있는데 이 연구에서는 프로그램을 보기 시작한 시간, 프로그램을 시청한 시간이 사용자의 TV 프로그램에 대한 선호도에 영향을 미친다는 것을 발견하였다. TV 프로그램 시청 시작시간과 시청지속시간에 따른 TV 프로그램을 끝까지 볼 확률을 모형링하고 이를 사용자의 TV 프로그램의 선호도에 곱하여 시간에 따른 TV 프로그램 선호도를 구하였다. 이후 추천을 시도할 때 사용자가 현재 보고 있는 TV 프로그램 보다 시간에 따른 선호도가 높은 TV 프로그램이 존재할 때만 추천을 진행하였다. 상기 두 연구는 공히 추천시기 결정을 대상으로 하는 연구이지만 내구재와 TV 프로그램 등 다른

아이템을 대상으로 하고 있으며, 아이템의 특성과 추천 목적이 다르기 때문에 본 연구에서와 같이 음식이나 기호식품의 재구매시기 예측에는 직접적으로 이용하기가 어렵다.

독립변수를 활용하여 재구매시기를 예측하는 연구에는 생존분석을 활용하는 방법도 있다. Wang and Zhang[2013]은 생존분석에서 비례 위험 모형(Proportional hazards modeling)을 적용하여 추천시기를 포함하는 추천시스템을 제안한 바 있다. 생존분석 기반의 재구매시기 예측방법은 시간에 따른 구매(또는 비 구매) 확률에 대한 누적분포를 통계분포를 이용하여 가정하고, 실제 구매시기 데이터와의 차이에 대한 오류를 최소화하도록 분포에 사용된 파라미터를 예측한 후, 이 분포 함수에 근거하여 예측을 수행하는 방법으로 다양한 독립변수를 활용하여 재구매시기를 예측할 수 있다는 장점은 있지만, 과거 구매이력에 대한 시계열 패턴을 반영하지는 못한다는 단점이 있다. 또한 생존분석은 내구재의 재구매시기 예측에는 적합하나 음식이나 기호제품과 같은 짧은 구매주기 상품에 대한 재구매시기 예측에는 적합하지 않다.

한편 시계열 데이터만을 사용하여 고객의 재구매시기를 예측하는 연구 중 가장 활발하게 연구되는 분야로 주기적 패턴 마이닝(Periodic pattern mining), 순차패턴 마이닝(Sequential pattern mining), 그리고 순환패턴 마이닝(Cyclic pattern mining) 등이 있다. 이들은 시계열로 수집된 이벤트 데이터로부터 주기성이나 순차패턴 또는 순환패턴을 발견하는 데이터마이닝 기법들이다. 이들 기법 중 본 연구에서와 같이 순차패턴에서 이벤트 간 시간 간격을 발견하는데 초점을 둔 연구로는 Chiang et al.[2005], Chen and Huang[2005], Hu et al.[2009]의 연구가 있다. 이들 연구는 순차패턴 마이닝을 확장하여 다양한 소비재 제품 구매에 대한 반복구매 주기를 발견하는데 목표를 두고 있다.

시계열 데이터로부터 순차패턴 간격을 구하여 다음 구매 시기를 예측하는 이들 방법은 본 연구에서도 일부 활용이 가능하지만 고객 중 구매시점에 있어서 주기성이 뚜렷이 존재하지 않는 경우에는 유용성이 저하될 수 있는 방법이라 할 수 있다. 시계열 데이터를 활용한 재구매시기 예측에 활용될 수 있는 또 다른 방법으로 추세적 확률모형이 있다. 기존연구는 고객이 자사제품을 구매할지 하지 않을지 측면에서 확률모형 또는 수리모형을 구성한 연구가 대부분이지만 모든 고객이 언젠가는 자사의 물건을 구매한다고 가정하면(구매하지 않는 고객은 충분히 오랜 시간 후에 구매) 시간에 따른 입력 데이터를 바탕으로 마코프체인 모형과 같은 동적모형을 이용하여 구매 시기에 대한 예측이 가능해진다. Rendle et al.[2010]과 Wang et al.[2011]의 연구에서는 잠재요인(Latent factor) 모형과 마코프체인 모형을 결합하여 제품 구매 순서에 대한 시계열 데이터를 바탕으로 다음 구매가 예상되는 장바구니 제품을 예측하는 모형을 제시한 바 있다.

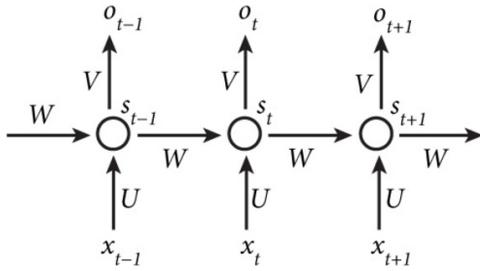
유비쿼터스 컴퓨팅의 출현과 함께 상황인지 추천시스템(Context-Aware Recommender)에 대한 연구도 활발히 진행되었으며 이들 연구에서도 추천시기를 포함한 제품 추천 방법을 제안하고 있다[Zhang et al., 2015]. 상황인지추천에서도 제품만이 아니라 추천시기와 추천 상황을 결정하는 기능을 포함하고 있으나 추천시기 결정이 도메인 특성에 따라 다르게 모형화된다는 점에서 반복 구매제품에 대한 범용 구매 시기 예측 모형을 개발하기 위한 본 연구와는 차이가 있다.

3. 순환신경망(RNN; Recurrent Neural Network)

3.1 순환신경망의 구조와 작동원리

최근 관심이 고조되고 있는 딥러닝은 주로 심

층 신경망을 지칭하며 기존의 신경망에 비해 더 많은 중간 계층을 사용하여 데이터에 대한 표현 능력을 향상시킨 신경망이다. 심층 신경망은 효과적인 학습방법이 존재하지 않는 단점이 있었으나 Hinton et al.[2006]의 연구에서 사전 학습이라는 학습방법이 제안되면서 이미지 인식, 자연언어처리 등의 분야에서 이를 적용한 연구가 급증하고 있다. 현재 딥러닝의 대표적인 모형으로는 회선신경망(CNN; Convolutional neural network) 모형과 순환신경망 모형 등이 있으며 이 중 순환신경망 모형은 시계열 데이터에 대한 학습과 추론에 적합한 것으로 알려져 있다. 순환신경망은 현재의 입력정보를 처리하는데 있어서 이전의 입력정보에 대한 이해를 바탕으로 처리가 이루어질 수 있도록 과거와 현재의 입력정보를 연결하여 처리하는 구조로 설계되어 있다. 따라서 순환신경망은 독립적인 입력변수를 학습 자료로 주로 사용하는 기존 신경망과는 달리 입력변수 간 서로 의존성이 존재할 것으로 예상되는 도메인에 주로 활용되는 모형이다. <Figure 1>은 전형적인 자연어문장 처리를 위한 순환신경망 구조를 보여주고 있다. <Figure 1>의 순환신경망은 입력층(x_{t-1}, x_t, x_{t+1}), 은닉층(s_{t-1}, s_t, s_{t+1})과 출력층(o_{t-1}, o_t, o_{t+1})으로 구성되며, 입력층과 은닉층을 연결하는 가중치 U, 은닉층과 출력층을 연결하는 가중치 V, 은닉층의 각 노드들을 연결하는 가중치 W로 구성되어 있다. 순환신경망이 기존 다층퍼셉트론 신경망과 다른 점은 이전정보에 대한 기억이 용이하도록 은닉층의 노드 간에 순차적으로 연결이 되어 있어 은닉층의 각 노드들이 네트워크 메모리 역할을 한다는 점이다. 즉 출력 값 o_t 는 은닉노드 s_t 에만 의존하지만 s_t 노드는 과거의 시간 스텝들에서 일어난 일들에 대한 정보를 전부 담고 있기 때문에 이전 입력정보에 대한 이해를 바탕으로 현재 입력정보의 처리가 가능한 것이다.



<Figure 1> Architecture of Recurrent Neural Network

<Figure 1>의 순환신경망 구조는 다음과 같이 수식으로 표현될 수 있다.

$$s_t = \tanh(Ux_t + Ws_{t-1}),$$

$$\text{where } \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

$$o_t = \text{softmax}(Vs_t),$$

$$\text{where } \text{softmax}(x) = \frac{e^x}{\sum e^x}$$

위 식에서 t시간의 은닉노드의 값은 t시간의 입력노드 값과 t-1시간의 은닉노드 값의 함수로 구성되며, t시간의 출력값은 t시간의 은닉노드 값의 함수로 구성된다. 은닉층 노드에 주로 적용되는 활성화 함수로는 -1에서 1사이의 값을 출력하는 하이퍼볼릭 탄젠트(Hyperbolic Tangent; tanh로 표기) 함수이며, 출력층 노드에 적용되는 활성화 함수로는 출력 값을 확률 값으로 변환하는 소프트맥스 함수가 주로 이용된다[Mulder et al., 2015].

순환신경망의 구조는 <Figure 1> 외에도 다양한 방식으로 구성할 수 있다. <Figure 1>은 전형적인 자연어문장 처리를 위한 순환신경망 구조로 매 시간 스텝마다 출력 값을 내지만(예; $[x_1, x_2, x_3] = [this, is, a], [o_1, o_2, o_3] = [is, a, book]$) 문장을 입력하면 긍정 또는 부정적인 감정을 추측하는 감성분석과 같은 응용분야에서는 굳이 출력 값으로 문장의 각 단어에 대한 다음 단어를 목표 값으

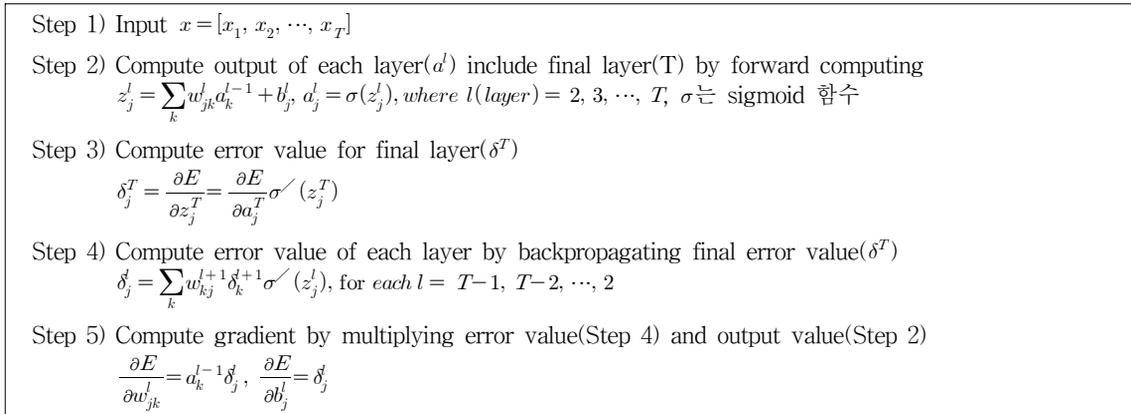
로 제시할 필요가 없으며, 긍정 또는 부정의 한 개의 최종 목표 값을 갖는 구조로 설계 할 수 있다.

3.2 순환신경망의 학습

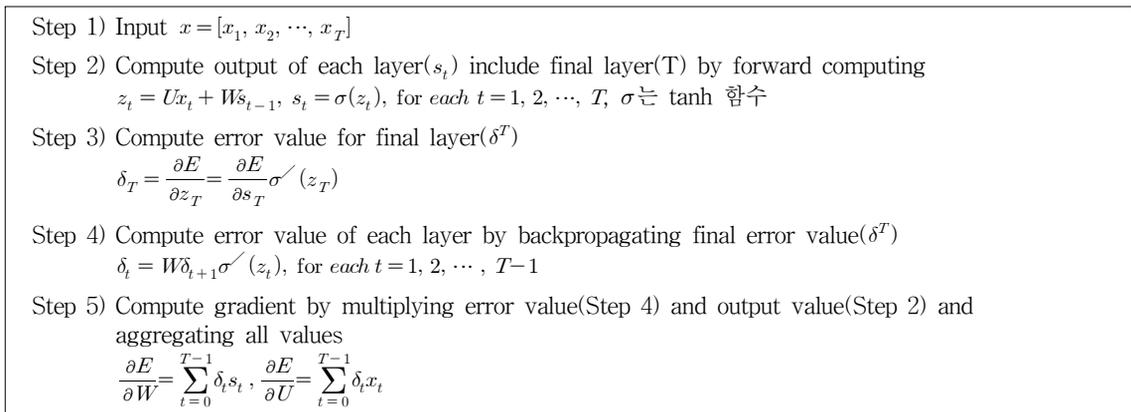
순환신경망을 학습하는 것은 기존의 신경망 학습과 유사하다. 순환신경망에서 학습이라 함은 기존 신경망과 같이 목표 값과 출력 값의 차이의 함수로 구성되는 손실함수(Loss function)를 최소화하는 연결 가중치 U, V, W를 발견하는 것으로 주로 확률적 경사하강법(SGD; Stochastic Gradient Descent)을 사용하여 학습을 진행한다. 확률적 경사하강법의 기본 아이디어는 모든 학습 데이터에 대해 반복적으로 학습하되 매 반복에서 오류를 줄이는 쪽으로 파라미터 값들을 조금씩 움직이는 것이다. 오류를 줄이는 방향은 손실함수의 경사(Gradient)로 주어진다. 확률적 경사하강법에서는 학습률(Learning rate)도 필요한데, 이는 매 반복에서 얼마만큼 파라미터 값에 대한 갱신이 이루어질지를 정해준다. 순환신경망 학습에서 많이 사용되는 손실함수는 교차 엔트로피(Cross-entropy) 함수로 학습 데이터가 N개가 있다면, 실제 목표 값 y_n 에 대한 네트워크의 예측 값 o_n 의 손실은 다음과 같은 식으로 계산된다.

$$E = -\frac{1}{N} \sum_{n \in N} y_n \log o_n$$

먼저 기존 다층퍼셉트론 신경망의 역전파 알고리즘을 이용하여 가중치 갱신에 필요한 경사(Gradient)를 구하는 방법을 살펴보면 <Figure 2>와 같다. 경사값은 손실함수를 가중치에 대해 편미분함으로써 구해질 수 있으나, 계산량을 줄이기 위해 <Figure 2>와 같은 알고리즘을 사용하고 있다. 먼저 입력 값에 대해 전방향 계산을 통해 각 계층 노드의 출력 값(a')을 구한다(Step 1, 2). 예측된 출력 값과 목표 값에 기반하여 최종출



<Figure 2> Procedure of Computing Gradient for Backpropagation Algorithm



<Figure 3> Procedure of Computing Gradient for BPTT Algorithm

력에 대한 오류값을 구한다(Step 3). 구해진 최종 출력에 대한 오류값을 역전파시킴으로써 모든 하위층 은닉노드의 오류값을 구한다(Step 4). 마지막으로 계층별로 구해진 오류값과 이전계층에서 전방향으로 계산된 활성화함수 값을 곱하여 가중치 갱신에 필요한 최종 경사(Gradient)를 계산한다.

<Figure 2>에서 제시한 다층퍼셉트론 신경망의 역전파 알고리즘과 달리 순환신경망에서는 기간별로 예측 값이 따로 존재하기 때문에 손실함수가 여러 기간에 걸쳐 계산된 오류값을 합산하여 만들어진다는 점과 다층퍼셉트론에 사용된 가중치는 계층별로 서로 다른 가중치를 사용하지만 순환신경망에 사용되는 가중치는 모든 계층에

서 동일한 가중치를 공유한다는 점 때문에 BPTT (Backpropagation Through Time)라는 약간 변형된 역전파 알고리즘을 사용하여 학습을 진행하게 된다. BPTT에서는 아래식과 같이 각 기간별로 산출된 경사 값을 더해 주는 방법으로 최종 경사 값을 구하여 경사하강을 진행하게 된다.

$$E = \sum_{t=1}^T E(t) \text{ 이므로, } \frac{\partial E}{\partial W} = \sum_{t=1}^T \frac{\partial E(t)}{\partial W}$$

결국 BPTT에서는 <Figure 3>과 같이 시간 진행에 따라 가중치를 공유하는 제약 식이 추가된다는 점과 각 기간별로 산출된 경사를 합산하여 최종 경사값을 얻는다는 점이 다르다.

BPTT 학습방법을 <Figure 1> 구조의 순환 신경망에 적용하면, T기의 오류 값을 계산한 후 T기의 가중치 갱신을 위한 경사를 구하고, 이 오류를 역방향으로 전파하여 T-1기, T-2기 등의 오류 값을 계산하여 T-1기, T-2기 등의 가중치 갱신을 위한 경사를 각 각 구하게 된다. 그러나 BPTT에서는 가중치를 공유하기 때문에 각 시기별 경사를 구하여 이를 합산한 경사 값을 가중치 갱신에 사용하게 된다.

3.3 순환신경망의 한계

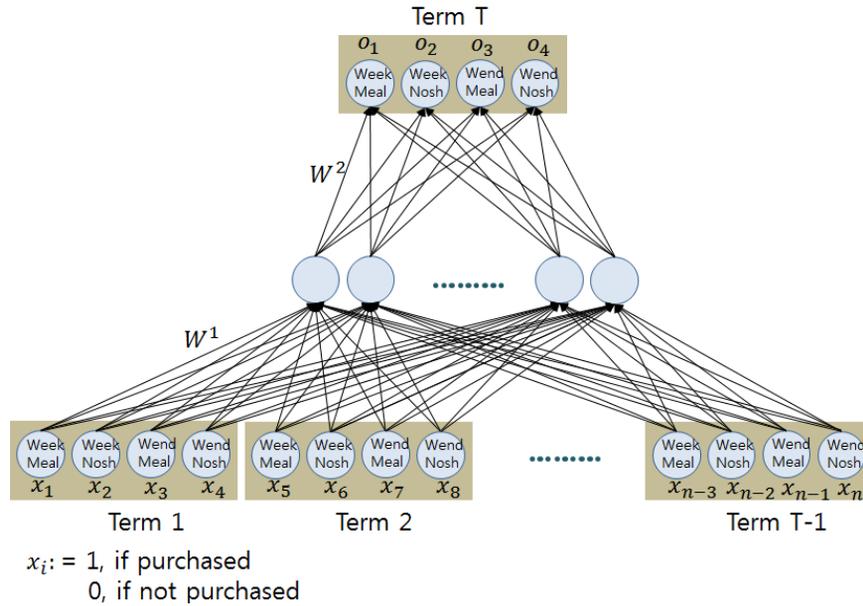
비록 순환신경망이 이전의 입력정보와 연계하여 현재의 입력정보를 처리하지만, 너무 오래전의 입력정보를 기억하는 데는 한계를 가진다. Bengio et al.[1994]에 의하면 순환신경망에서 긴 시퀀스의 입력정보를 BPTT로 학습할 경우 경사 사라짐(Vanishing gradient) 문제 또는 경사 폭발(Exploding gradient) 문제가 발생할 수 있다. 이는 먼 과거까지 학습을 하는 과정에서 반복되어 곱해지는 순환가중치 때문에 경사 값이 기하급수적으로 사라지거나 증가하기 때문에 발생하는 현상이다. 경사 사라짐 문제를 해결하기 위해서는 적당히 좋은 값으로 가중치를 초기화하는 방법, 정규화(Regularization) 방법, 활성화함수를 하이퍼볼릭탄젠트 또는 시그모이드 함수 대신 ReLU 함수를 사용하는 방법 등이 있다. 이 외에도 근본적인 해결책으로 LSTM(Long Short-Term Memory) 또는 GRU(Gated Recurrent Unit) 등 다양한 확장된 순환신경망 모형이 존재한다. LSTM은 1997년에 처음 제안되었고, 현재 자연어처리 분야에서 가장 널리 사용되는 모형 중 하나이며, GRU는 2014년에 발표된 LSTM의 간략화 버전이다. 두 모형 다 경사 사라짐 문제 해결을 위해 도입된 모형으로 효과적으로 긴 시퀀스를 처리하는 것으로 알려져 있다.

4. 반복 구매제품의 재구매시기 예측모형 구성

본 연구에서는 반복 구매제품의 재구매시기 예측을 위해 대표적인 신경망모형인 다층퍼셉트론(MLP; Multi Layer Perceptron)과 최근 딥러닝으로 각광 받고 있는 순환신경망(RNN; Recurrent Neural Network)을 활용하여 각각 모형을 구성하고 예측 성능을 비교하고자 한다. 본 연구에서는 고객별로 시간에 따른 반복구매 패턴이 존재한다는 것을 전제한다. 다층퍼셉트론 신경망모형은 이전 정보를 기억하는 메커니즘이 없기 때문에 본 연구에서 목표로 하는 반복 구매제품의 재구매시기를 예측할 때 고객별 구매빈도나 간격 패턴만을 학습하여 다음 구매 시기를 예측할 가능성이 높다. 그러나 순환신경망의 경우 고객별 구매빈도나 간격 패턴 외에도 시계열적 변화에 대한 학습이 가능하기 때문에 점점 구매 간격이 짧아지거나 느려지는 패턴도 학습이 가능할 것이다.

4.1 신경망 구조 및 입력자료 설계

먼저 반복구매제품의 재구매시기 예측을 위해 제안한 다층퍼셉트론의 구조는 <Figure 4>와 같다. 예측에 사용할 다층퍼셉트론 모형은 3계층 구조의 신경망으로 입력노드, 출력노드, 은닉노드로 구성된다. 각 입력노드는 주차별로 평일식사, 평일간식, 주말식사, 주말간식의 4개 노드를 포함하며 총 T-1기까지의 구매이력을 활용하여 T기의 구매여부를 예측하는 방식으로 설계했기 때문에 (T-1)×4개의 입력노드로 구성된다. 출력노드는 T기의 구매패턴에 해당하는 것으로 평일식사, 평일간식, 주말식사, 주말간식의 4개 노드로 구성된다. 각각의 입력노드는 해당 주차에 해당 목적으로 구매한 바가 있으면 1, 그렇지 않으면 0으로 코딩하기로 한다.



<Figure 4> Architecture of MLP and Input Data Format

한편 재구매시기 예측을 위한 효과적인 순환 신경망 모델을 설계하기 위해서는 시간 의존적인 특성을 유지하도록 시계열적 자료를 입력하는 것이 필요하다. 따라서 구매패턴이란 개념을 <Table 1>과 같이 정의하여 기간별 구매패턴을 입력 자료로 활용하기로 한다. <Table 1>에는 총 16가지의 구매패턴이 정의되어 있는데 각 패턴은 4자리의 0 또는 1의 값을 갖는 디짓(Digit)으로 구성된다.

본 연구에서 제안한 순환신경망은 아래식과 같이 특정 구매패턴이 나타날 확률로 표현되며,

이전 기간의 구매패턴을 알고 있을 때 다음 기간의 구매패턴이 나타날 확률의 곱으로 계산되며 순환신경망의 구조는 <Figure 5>와 같다.

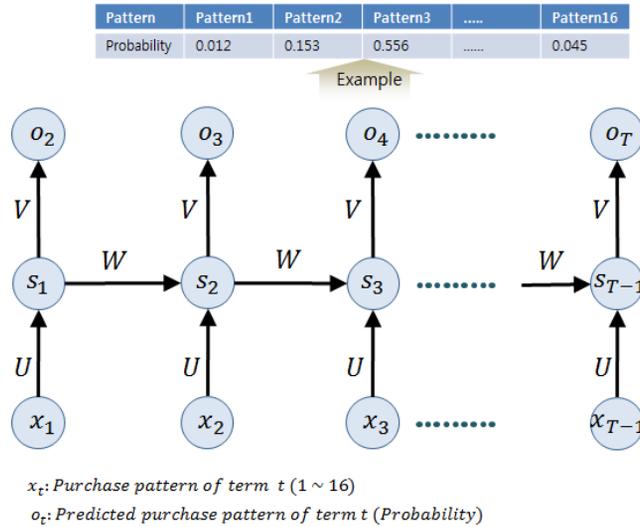
$$P(b_1, \dots, b_T) = \prod_{i=1}^T P(b_i | b_1, \dots, b_{i-1}),$$

where b_i : i 기의 구매 패턴

본 연구에서 구매패턴은 다음 기간에 구매가 예상되는 고객에게 제공될 프로모션의 종류를 고려하여 결정하였다. 즉 본 연구는 배달음식 도

<Table 1> Definition of Purchase Pattern

NO	Patterns	NO	Patterns
1	0-0-0-0 No purchase	9	1-0-0-0 Weekday meal purchase
2	0-0-0-1 Weekend nosh purchase	10	1-0-0-1 Weekday meal+Weekend nosh
3	0-0-1-0 Weekend meal purchase	11	1-0-1-0 Meal purchase
4	0-0-1-1 Weekend purchase	12	1-0-1-1 Weekday Meal+Weekend
5	0-1-0-0 Weekday nosh purchase	13	1-1-0-0 Weekday purchase
6	0-1-0-1 Nosh purchase	14	1-1-0-1 Weekday+Weekend nosh
7	0-1-1-0 Weekday nosh+Weekend meal	15	1-1-1-0 Weekday+Weekend meal
8	0-1-1-1 Weekday nosh+Weekend	16	1-1-1-1 Always purchase



(Figure 5) Architecture of RNN and Input Data Format

메인의 재구매시기 예측을 목표로 하고 있기 때문에 위 식에서 각 기간(Period)은 주 단위로 설정하고 구매패턴은 주중, 주말로 일차 구분하고 점심 또는 저녁 같은 식사용으로 구매하는 지 주간 또는 야간의 간식용으로 구매하는 지로 세분하여 구매패턴을 구성하였다. 이는 주문 평준화(Balancing)를 위해 주문이 뜬한 주중에 오퍼할 프로모션과 주문이 몰리는 주말에 오퍼할 프로모션이 다를 수 있으며, 적합한 칼로리의 메뉴를 오퍼하기 위해 식사용인지 간식용인지를 구분할 필요가 있기 때문이다. 순환신경망의 입력 값은 주차별 구매패턴이 되며, 학습된 순환신경망모형을 사용하여 예측한 결과 값인 출력값 o_i 는 16 가지 구매패턴에 대한 확률분포로 주어지며 아래 식과 같이 가장 큰 확률을 갖는 인덱스를 최종 구매패턴으로 선택한다.

$$o_i = \operatorname{argmax}(k_1, \dots, k_{16})$$

where k_i : predicted probability of purchase pattern i

순환신경망을 학습하는 것은 목표 값에 가까운 값을 출력하도록 연결가중치를 결정하는 것으

로 경사 사라짐 문제를 회피하기 위해서는 연결가중치에 대한 적절한 초기 값 설정이 필요하다. Glorot and Bengio[2010] 연구에서는 경사 사라짐 문제를 피하기 위한 초기화 방법으로 n 이 이전 계층(Layer)에서부터 들어오는 연결 수라고 할 때 $[-1/\sqrt{n}, 1/\sqrt{n}]$ 구간에서 랜덤으로 정하는 것이 좋다고 하였다. 따라서 본 연구에서도 이 방식을 적용하여 연결가중치에 대한 초기화를 수행하기로 한다. 손실함수는 교차 엔트로피 손실함수를 사용한다. 특정 고객의 구매패턴을 하나의 학습 데이터로 입력하고, 총 손실 값은 매 시간 스텝마다의 손실의 총 합으로 계산한다. 학습 알고리즘은 앞에서 제시된 확률적 경사하강법(Stochastic Gradient Descent)에 기반한 BPTT 알고리즘을 적용한다.

4.2 모형의 구현

순환신경망 모형을 코드로 구현하기 위해서는 기존에 존재하는 다양한 라이브러리를 이용하는 것이 효과적이다. 최근에 구글에서 오픈소스로 공개한 인공지능 라이브러리인 텐서플로(Tensor-

flow), 뉴욕대(NYU)의 토치(Torch), 캘리포니아 대 버클리의 카페(Caffe), 몬트리올대의 시애노(Theano) 등이 기계학습 관련 오픈 소스 라이브러리의 대표적인 예이다. 본 연구에서는 기계학습 분야에서 비교적 폭 넓게 활용되고 있는 시애노(Theano) 라이브러리[Bastien et al., 2012, <http://deeplearning.net/software/theano/tutorial>]를 기반으로 순환신경망 모형을 구현하고자 한다. 시애노는 다양한 플랫폼 환경의 상용서비스 개발 언어로 폭 넓게 사용되고 있는 파이썬(Python) 언어 기반의 라이브러리이다. 파이썬과 시애노를 기반으로 모형을 구축하면 상용 서비스를 위한 시스템 구축 시 실험에 사용한 모형과 코드를 그대로 재활용할 수 있기 때문에 상용화를 위한 비용과 시간을 줄일 수 있다는 장점이 있다. 시애노 사용의 또 다른 장점은 경사를 구하기 위한 `grad()` 함수를 제공하기 때문에 복잡한 미분식을 직접 구현할 필요가 없어서 실험의 신속 정확성을 기할 수 있다. 본 실험에서 사용된 운영체제는 윈도우7이다.

5. 성능 평가

5.1 실험용 데이터셋

본 연구에서는 랜덤으로 발생된 인공 데이터를 이용하여 실험용 데이터셋을 구성하기로 한다. 입력자료의 패턴(생성타입)을 몇 가지로 정의하여 생성타입별로 인공 데이터를 발생시킨 후 다음기간의 구매패턴이 무엇이 될지를 예측하는 방법을 택하게 되면, 각 신경망 별로 생성타입별 구매패턴 예측성능을 파악할 수 있다는 장점이 있다. 이는 실제 구매이력 데이터를 입력자료로 사용했을 경우 파악하기 어려운 부분이다. 실험용 데이터 셋을 구성하기 위해 다음과 같은 방법으로 여덟 가지의 생성타입을 정의하고 타입별 100개의 랜덤 레코드를 발생시켜 총 800명에 대한 8주간의 구매이력 데이터 셋을 마련하기로 한다. 실험을 위해 데이터를 무작위로 구분하여 60%는 학습용으로 40%는 테스트용 데이터로 구분하고 실험을 진행한다.

〈Table 2〉 Generation Types and Generating Methods for Input

Generation TYPE	Customer Type	Generating methods
T1	Rare purchase customer	- Randomly assign 0 or 1 for 32 buckets - 10% probability for 1
T2	Loyal customer	- Randomly assign 0 or 1 for 32 buckets - 40% probability for 1
T3	Increasing purchase	- 10% probability of 1 for 1~2 weeks - 20% probability of 1 for 3~4 weeks - 30% probability of 1 for 5~6 weeks - 40% probability of 1 for 7~8 weeks
T4	Decreasing purchase	- 40% probability of 1 for 1~2 weeks - 30% probability of 1 for 3~4 weeks - 20% probability of 1 for 5~6 weeks - 10% probability of 1 for 7~8 weeks
T5	Weekday meal purchase	- 40% probability of 1 for (4n+1)th buckets - 10% probability of 1 for the other buckets
T6	Weekday nosh purchase	- 40% probability of 1 for (4n+2)th buckets - 10% probability of 1 for the other buckets
T7	Weekend meal purchase	- 40% probability of 1 for (4n+3)th buckets - 10% probability of 1 for the other buckets
T8	Weekend nosh purchase	- 40% probability of 1 for (4n+4)th buckets - 10% probability of 1 for the other buckets

5.2 성능 평가지표 및 평가방법

일반적으로 추천시스템의 성능은 예측정확도, 추천품질, 커버리지, 참신성 등 다양한 측면에서 평가가 가능하다. 본 연구에서는 신경망의 예측정확도를 구하기 위해 실제 고객이 T기에 구매한 패턴(y_T)과 신경망을 통해 T기에 예측된 구매패턴(o_T)이 일치하면 정분류로 판단하여 정분류율을 예측정확도로 사용하기로 한다. 한편, 본 연구에서 제안한 추천시스템은 최소 한번 이상 구매할 것으로 예측된 고객(<Table 1>에서 구매패턴 1을 제외한 나머지 패턴)을 대상으로 적합한 시기에 적절한 프로모션을 제공하는 형태로 추천이 이루어진다. 즉 T기에 한건도 구매하지 않을 것으로 예상되는 고객에게는 별도의 프로모션이 제공되지 않는다. 일반적으로 구매 데이터는 희소성(Sparsity) 특성으로 인해 한 건도 구매하지 않은 고객 데이터가 많이 존재하게 되며, 구매를 하지 않을 것으로 예측된 고객이 실제 구매하지 않은 경우에도 정분류로 계산하여 구매정확도는 높아지는 현상이 생기게 된다. 따라서 예측정확도 외에 추천품질을 판단하는 추가 지표가 필요하다. 반복구매패턴 예측연구에서는 특정 기간 동안 최소 한 번 이상 구매하는 것으로 예측된 고객 중 실제 얼마나 정확히 구매패턴을 맞추었는지가 중요한 추천품질지표가 될 수 있다. 이는 최소 한 번 이상 구매하는 것으로 예측된 고객을 대상으로 프로모션 문자를 발송하게 되기 때문이다. 또한 최소 한 번 이상 구매할 것으로 예측된 고객 중 실제 한 건도 구매하지 않은 고객의 경우, 해당 프로모션 문자를 스팸으로 인식할 가능성이 높다. 따라서 본 연구에서는 전자를 정밀도(Precision), 후자를 스팸률(Spam)로 정의하여 추천품질을 평가하는 척도로 사용하기로 한다. 아래 식은 정밀도와 스팸률을 측정하기 위한 식이다.

A : 기간 T에 최소 한 번 이상 구매할 것으로 예측된 고객집합(구매패턴 1 외의 패턴으로 예측된 고객)

B : 기간 T에 예측된 구매패턴과 실제 구매패턴이 일치하는 고객집합

C : 실제 기간 T에 구매하지 않은 고객집합(실제 구매패턴이 구매패턴 1(0-0-0-0)인 고객)

$$\text{정밀도} = \frac{\|A \cap B\|}{\|A\|}$$

$$\text{스팸률} = \frac{\|A \cap C\|}{\|A\|}$$

5.3 실험결과

5.3.1 구매패턴 예측정확도 평가

<Table 3>은 다층퍼셉트론(MLP) 신경망에 의한 예측결과로 여덟번째 주의 구매패턴에 대한 예측정확도를 생생타입별로 보여주고 있다. 예측에 사용된 신경망 모형은 입력, 은닉, 출력층으로 구성된 3계층 구조의 신경망으로 28개의 입력노드와 4개의 출력노드로 구성하였다. 각각의 입력노드는 주차별로 평일식사, 평일간식, 주말식사, 주말간식의 4개 노드를 포함하며 총 7주간의 구매패턴을 활용했기 때문에 28개의 입력노드로 구성된다. 출력노드는 8주차의 구매패턴에 해당하는 것으로 평일식사, 평일간식, 주말식사, 주말간식의 4개 노드로 구성된다. 각각의 입력노드는 해당 주차에 해당 목적으로 구매한 바가 있으면 1, 그렇지 않으면 0으로 코딩하였으며, 출력노드의 출력 값에 대한 해석 또한 0.5를 기준으로 이보다 큰 값이 출력되면 1, 작거나 같은 값이 출력되면 0으로 변환하여 예측정확도를 평가하였다. 실험에 사용된 신경망 모형은 과잉적합(Overfitting)을 완화하기 위해 정규화 항목(Regularization term)을 손실함수에 추가하여 프로그래밍하였으며, 학습률(Learning rate)은 0.01로 설정하여

<Table 3> MLP-Prediction Accuracy

Models		Data	Total Acc	Prediction accuracy by generation type							
				T1	T2	T3	T4	T5	T6	T7	T8
1	Hid = 100 Epo = 5,000	Train	0.435	0.667	0.250	0.150	0.583	0.433	0.517	0.400	0.483
		Test	0.347	0.525	0.100	0.125	0.400	0.375	0.350	0.325	0.575
2	Hid = 100 Epo = 10,000	Train	0.429	0.650	0.267	0.117	0.550	0.433	0.517	0.400	0.500
		Test	0.334	0.500	0.100	0.125	0.425	0.350	0.350	0.325	0.500
3	Hid = 100 Epo = 15,000	Train	0.438	0.650	0.333	0.117	0.550	0.433	0.483	0.383	0.550
		Test	0.331	0.500	0.075	0.125	0.425	0.350	0.325	0.350	0.500
4	Hid = 100 Epo = 20,000	Train	0.438	0.583	0.350	0.167	0.533	0.483	0.483	0.383	0.517
		Test	0.303	0.500	0.075	0.125	0.375	0.300	0.250	0.425	0.375
5	Hid = 200 Epo = 15,000	Train	0.442	0.617	0.350	0.133	0.517	0.467	0.467	0.400	0.583
		Test	0.334	0.525	0.150	0.100	0.350	0.350	0.300	0.400	0.500
6	Hid = 300 Epo = 15,000	Train	0.446	0.550	0.467	0.167	0.550	0.367	0.500	0.417	0.550
		Test	0.309	0.575	0.075	0.150	0.350	0.350	0.275	0.300	0.400
7	Hid = 500 Epo = 15,000	Train	0.481	0.583	0.517	0.167	0.550	0.517	0.517	0.417	0.583
		Test	0.331	0.550	0.075	0.125	0.375	0.325	0.375	0.350	0.475
8	Hid = 800 Epo = 15,000	Train	0.498	0.583	0.550	0.183	0.617	0.517	0.533	0.500	0.500
		Test	0.341	0.550	0.125	0.150	0.300	0.375	0.375	0.375	0.475
9	Hid = 800 Epo = 20,000	Train	0.517	0.517	0.567	0.217	0.600	0.583	0.567	0.600	0.483
		Test	0.319	0.500	0.100	0.200	0.250	0.350	0.350	0.425	0.375

<Table 4> RNN-Prediction Accuracy

Models		Data	Total Acc	Prediction accuracy by generation type							
				T1	T2	T3	T4	T5	T6	T7	T8
1	Hid = 20 Epo = 40 Trunc = 3	Train	0.423	0.633	0.283	0.117	0.683	0.367	0.517	0.300	0.483
		Test	0.400	0.600	0.050	0.125	0.600	0.350	0.400	0.500	0.575
2	Hid = 20 Epo = 40 Trunc = 4	Train	0.442	0.617	0.317	0.133	0.683	0.400	0.533	0.333	0.517
		Test	0.394	0.600	0.025	0.150	0.600	0.400	0.400	0.425	0.550
3	Hid = 20 Epo = 50 Trunc = 3	Train	0.423	0.633	0.250	0.133	0.683	0.367	0.517	0.333	0.467
		Test	0.406	0.600	0.075	0.125	0.575	0.400	0.400	0.500	0.575
4	Hid = 20 Epo = 60 Trunc = 3	Train	0.444	0.633	0.333	0.167	0.683	0.400	0.517	0.350	0.467
		Test	0.397	0.575	0.075	0.150	0.575	0.350	0.400	0.450	0.600
5	Hid = 25 Epo = 70 Trunc = 3	Train	0.435	0.633	0.283	0.167	0.600	0.417	0.517	0.350	0.517
		Test	0.388	0.600	0.075	0.125	0.575	0.375	0.400	0.375	0.575
6	Hid = 25 Epo = 90 Trunc = 3	Train	0.496	0.633	0.383	0.317	0.617	0.517	0.517	0.467	0.517
		Test	0.384	0.600	0.075	0.150	0.550	0.350	0.375	0.450	0.525
7	Hid = 30 Epo = 80 Trunc = 3	Train	0.490	0.650	0.400	0.283	0.667	0.450	0.550	0.383	0.533
		Test	0.409	0.600	0.100	0.150	0.575	0.400	0.400	0.475	0.575
8	Hid = 35 Epo = 90 Trunc = 3	Train	0.475	0.633	0.433	0.183	0.667	0.450	0.517	0.383	0.533
		Test	0.391	0.600	0.075	0.125	0.575	0.425	0.375	0.425	0.525
9	Hid = 35 Epo = 50 Trunc = 3	Train	0.479	0.633	0.383	0.283	0.700	0.433	0.533	0.383	0.483
		Test	0.384	0.600	0.050	0.100	0.575	0.425	0.375	0.400	0.550
10	Hid = 30 Epo = 80 Trunc = 4	Train	0.508	0.600	0.517	0.317	0.667	0.483	0.517	0.500	0.467
		Test	0.359	0.575	0.100	0.100	0.450	0.400	0.325	0.450	0.475

학습을 진행하였다. 다층퍼셉트론모형의 구매패턴에 대한 예측정확도 평가결과, 은닉노드의 수와 훈련 반복횟수가 많을수록 훈련데이터의 예측정확도는 높았으나, 테스트 데이터에서의 예측정확도는 과잉적합으로 인해 낮아지는 결과를 보여주었다(모형 9). 테스트 데이터에 적용한 결과, 가장 높은 예측정확도를 나타낸 모형은 모형 1(은닉 = 100, 반복 = 5,000)로 예측정확도가 34.7%로 나타났다. 생성타입별로 테스트 데이터의 예측정확도를 살펴보면, T1과 같이 구매빈도가 상대적으로 높은 생성타입에서 예측정확도가 높게 나타났으며 T2와 같이 구매빈도가 희박한 데이터에서는 예측정확도가 낮게 나타나는 것을 확인할 수 있다.

<Table 4>는 순환신경망(RNN)에 의한 구매패턴 예측결과로 예측정확도를 생성타입별로 보여주고 있다. 실험에 사용된 순환신경망모형은 초기 학습률을 0.005로 설정하였으며, 손실함수의 값이 직전 반복 때 보다 높아지면 학습률을 반으로 줄여나가는 방식으로 학습을 진행하였다. 다층퍼셉트론 신경망에서와 같이 각 예측은 1주에서 7주까지의 구매패턴 데이터를 입력하고, 마지막 8주차($t = 8$)의 예측정확도를 계산하였다. BPTT에서는 가중치 갱신을 위한 경사도 계산 시, 어느 단계까지 경사정보를 전파할 것인지를 제한할 수 있는데 이때 사용되는 파라미터가 BPTT Truncate이다. 이때 설정된 BPTT Truncate 기간까지만 거슬러 올라가면서 경사도 값을 계산하게 되며, 결과적으로 이 기간까지의 구매패턴만을 t 기의 구매를 예측하는데 반영하게 된다. 본 실험에서는 다양한 기간에 걸쳐 경사도를 계산해 보았으나 대체로 최근 3주 전까지 구매패턴을 반영하여 경사도를 계산한 경우(BPTT Truncate = 3)에서 예측 정확도가 높게 나타났다. 제시한 순환신경망을 테스트 데이터에 적용한 예측정확도는 모형 7(은닉 = 30, 반복

= 80, 역전파 = 3)에서 40.9%(0.409)로 가장 높은 정확도를 보여주었다. 결국 다층퍼셉트론보다는 순환신경망모형의 예측정확도가 우수함을 알 수 있었다. 순환신경망 예측결과에서 생성타입별로 예측정확도를 비교하면, 다층퍼셉트론 신경망에서와 같이 구매빈도가 상대적으로 높은 생성타입에서 예측정확도가 높게 나타났다. 생성타입 측면에서 다층퍼셉트론과 순환신경망의 예측결과를 비교하면, T4의 경우 다층퍼셉트론보다는 순환신경망을 활용할 경우 예측정확도가 현저히 높게 나타났다. T4는 시간에 따라 점진적으로 감소하는 구매패턴으로 이는 순환신경망이 시간에 의존하는 데이터를 다층퍼셉트론보다 잘 구분하여 예측할 수 있음을 보여주고 있다.

5.3.2 구매패턴 예측품질 평가-정밀도와 스팸률

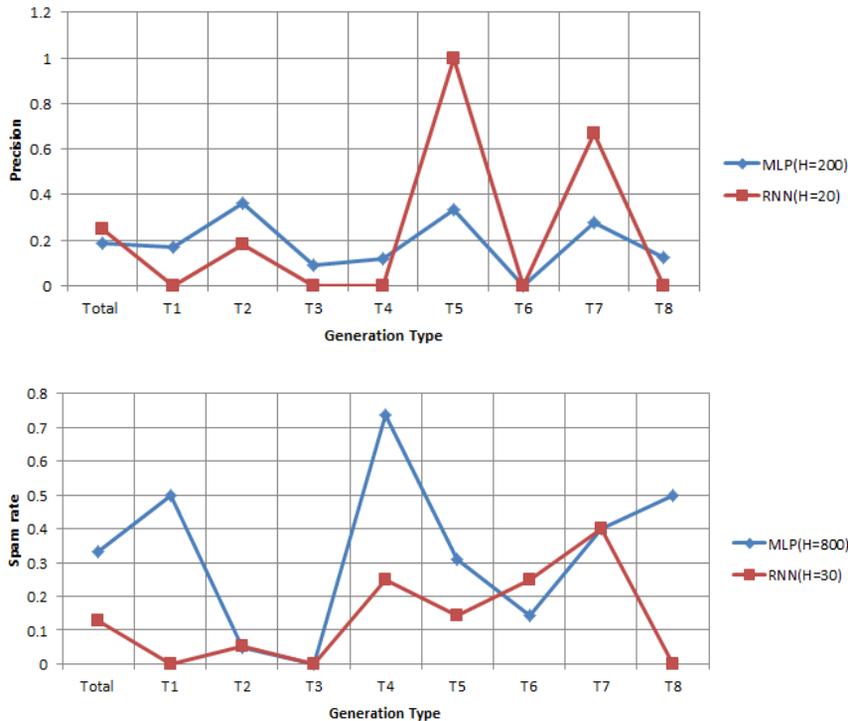
<Table 5>와 <Table 6>은 각각 다층퍼셉트론(MLP)신경망과 순환신경망(RNN)의 여덟 번째 주의 구매패턴에 대한 예측결과를 정밀도와 스팸률 측면에서 생성타입별로 보여주고 있다. 순환신경망 예측결과에서 정밀도와 스팸률에 숫자가 나타나지 않은 셀은 해당 생성타입에서 최소한 번이상 구매가 예측되는 고객이 한 명도 없어서 정밀도와 스팸률을 계산할 수 없는 셀이다. 다층퍼셉트론모형 중에서는 모형 5에서 가장 높은 정밀도(0.189)를 보여주었고, 모형 8에서 가장 낮은 스팸률(0.333)을 보여주었다. 순환신경망 모형 중에서는 모형 3의 정밀도(0.250)가 가장 우수하였고, 모형 7의 스팸률(0.128)이 가장 우수한 것으로 나타났다. 다층퍼셉트론과 순환신경망의 예측품질을 비교하면 정밀도 측면에서는 순환신경망의 품질이 다층퍼셉트론신경망에 비해 약간 우수한 것으로 나타났으나 스팸률 측면에서는 절반에 가까운 매우 우수한 성능을 보여주었다. <Figure 6>은 우수 모형에 대한 예측품질을 비교하여 보여주고 있다.

〈Table 5〉 MLP-Prediction Quality

Models		Data	Total Acc	Precision and spam rate by generation type							
				T1	T2	T3	T4	T5	T6	T7	T8
1	Hid = 100 Epo = 5,000	Prec	0.140	0.000	0.250	0.167	0.000	0.400	0.000	0.133	0.000
		Spam	0.439	1.000	0.083	0.167	0.800	0.400	0.500	0.533	0.000
2	Hid = 100 Epo = 10,000	Prec	0.160	0.000	0.231	0.143	0.143	0.286	0.000	0.211	0.000
		Spam	0.440	1.000	0.077	0.143	0.643	0.429	0.333	0.526	0.600
3	Hid = 100 Epo = 15,000	Prec	0.169	0.143	0.182	0.167	0.133	0.333	0.000	0.211	0.143
		Spam	0.434	0.714	0.091	0.167	0.600	0.444	0.333	0.474	0.571
4	Hid = 100 Epo = 20,000	Prec	0.157	0.111	0.125	0.100	0.150	0.200	0.000	0.304	0.133
		Spam	0.426	0.556	0.062	0.100	0.600	0.500	0.500	0.391	0.667
5	Hid = 200 Epo = 15,000	Prec	0.189	0.167	0.364	0.091	0.118	0.333	0.000	0.278	0.125
		Spam	0.422	0.667	0.000	0.182	0.706	0.444	0.400	0.444	0.500
6	Hid = 300 Epo = 15,000	Prec	0.142	0.250	0.154	0.133	0.111	0.250	0.000	0.143	0.167
		Spam	0.398	0.375	0.077	0.067	0.667	0.375	0.417	0.476	0.556
7	Hid = 500 Epo = 15,000	Prec	0.152	0.167	0.100	0.091	0.133	0.308	0.000	0.182	0.182
		Spam	0.362	0.500	0.050	0.091	0.733	0.462	0.143	0.409	0.545
8	Hid = 800 Epo = 15,000	Prec	0.162	0.167	0.190	0.077	0.105	0.308	0.000	0.200	0.167
		Spam	0.333	0.500	0.048	0.000	0.737	0.308	0.143	0.400	0.500
9	Hid = 800 Epo = 20,000	Prec	0.159	0.200	0.083	0.167	0.091	0.308	0.000	0.273	0.158
		Spam	0.348	0.600	0.000	0.000	0.727	0.385	0.200	0.364	0.579

〈Table 6〉 RNN-Prediction Quality

Models		Data	Total Acc	Precision and spam rate by generation type							
				T1	T2	T3	T4	T5	T6	T7	T8
1	Hid = 20 Epo = 40 Trunc = 3	Prec	0.167	-	0.091	0.000	-	0.000	-	0.667	0.000
		Spam	0.167	-	0.091	0.000	-	1.000	-	0.333	0.000
2	Hid = 20 Epo = 40 Trunc = 4	Prec	0.207	-	0.071	0.250	0.000	1.000	-	0.375	0.000
		Spam	0.241	-	0.071	0.000	0.000	0.000	-	0.625	1.000
3	Hid = 20 Epo = 50 Trunc = 3	Prec	0.250	-	0.182	0.000	0.000	1.000	-	0.667	-
		Spam	0.150	-	0.091	0.000	0.333	0.000	-	0.333	-
4	Hid = 20 Epo = 60 Trunc = 3	Prec	0.159	0.000	0.154	0.182	0.000	0.167	0.000	0.200	0.333
		Spam	0.182	1.000	0.077	0.091	0.333	0.333	0.000	0.400	0.000
5	Hid = 25 Epo = 70 Trunc = 3	Prec	0.167	-	0.125	0.000	0.000	0.286	0.500	0.273	0.000
		Spam	0.250	-	0.062	0.000	0.333	0.286	0.500	0.636	0.000
6	Hid = 25 Epo = 90 Trunc = 3	Prec	0.140	-	0.133	0.091	0.143	0.167	0.250	0.250	0.000
		Spam	0.228	-	0.067	0.000	0.429	0.333	0.500	0.375	0.333
7	Hid = 30 Epo = 80 Trunc = 3	Prec	0.191	0.000	0.158	0.167	0.000	0.286	0.250	0.400	0.000
		Spam	0.128	0.000	0.053	0.000	0.250	0.143	0.250	0.400	0.000
8	Hid = 35 Epo = 90 Trunc = 3	Prec	0.137	-	0.105	0.000	0.000	0.429	0.000	0.222	0.000
		Spam	0.196	-	0.053	0.000	0.200	0.143	1.000	0.444	1.000
9	Hid = 35 Epo = 50 Trunc = 3	Prec	0.154	-	0.071	0.000	0.000	0.500	0.250	0.267	0.000
		Spam	0.250	-	0.071	0.091	0.500	0.000	0.500	0.467	0.500
10	Hid = 30 Epo = 80 Trunc = 4	Prec	0.123	0.000	0.200	0.000	0.000	0.333	0.000	0.308	0.000
		Spam	0.284	1.000	0.100	0.071	0.462	0.167	0.500	0.385	0.500



〈Figure 6〉 Prediction Quality Comparison

보다 세부적으로 생성타입별로 예측품질을 비교하면 다음과 같다. 먼저 정밀도 측면에서 가장 우수한 다층퍼셉트론모형 5와 순환신경망모형 3을 비교해 보면, 다층퍼셉트론의 경우 T2(0.364)와 T5(0.333)에서 정밀도가 우수했으나, 순환신경망의 경우 T7(0.667)에서 우수한 정밀도를 보여주었다. 스팸률 측면에서 가장 우수한 다층퍼셉트론 모형 8과 순환신경망 모형 7을 비교해 보면, 다층퍼셉트론의 경우 T4(0.737)에서 가장 스팸률이 높았으나 순환신경망의 경우 T7(0.400)에서 가장 높은 스팸률을 보여주었다.

6. 결론 및 시사점

본 연구에서는 구매일자와 구매제품으로 구성된 간단한 구매이력 정보만을 활용하여 고객의 재구매시기를 예측하는 신경망 기반의 모형을

제안하고 성능평가를 통해 반복구매 시기예측 응용 분야에서 순환신경망 모형의 우수성을 확인하였다. 프랜차이즈를 포함한 배달음식 비즈니스에서 성공의 두 축은 원가 절감과 매출확대이다. 이 중 원가절감은 경쟁이 치열할수록 원가 압박이 심하기 때문에 원가절감을 통한 이익확보에는 한계가 존재한다. 매출확대 측면에서는 양적으로 트래픽을 높이는 것 보다는 충성고객 확보를 통해 반복구매를 유도하는 것이 질적인 측면에서 바람직하다. 이와 더불어 매출확대를 위해 중요한 것은 주문의 평균화이다. 이는 일 주문수량의 변동으로 과다 주문 시, 곁품에 따른 판매기회의 상실이 우려되고, 과소 주문 시 고정비로 인한 이익 감소가 우려되기 때문이다. 결국 프랜차이즈를 포함한 배달음식 비즈니스의 성공 열쇠는 일 주문수량의 평균화와 반복구매 고객의 확대를 통한 매출 확대임을 알 수 있다. 주문

수량의 평균화와 동시에 경쟁사로의 고객 유출을 방지하고 반복구매를 촉진하기 위한 방법은 정확한 시기에 적절한 제품을 추천하는 것이다. 이를 위해 본 연구에서는 고객의 과거 구매이력을 토대로 고객이 재주문할 시기를 예측하는 순환신경망 기반의 개인화된 구매패턴 예측모형을 제안하였다. 제안한 모형을 바탕으로 주문수량이 부족할 때 해당시기에 구매할 것으로 예측된 고객에게 프로모션을 제공한다면 주문의 평균화가 가능하게 되며 매출확대 효과를 얻을 수 있다. 또한 치열한 경쟁상황에서 고객이 해당 제품과 서비스를 필요로 할 때 적절한 프로모션을 제공함으로써 경쟁회사로의 고객 유출을 방지하고자사로의 반복구매를 유도하게 되는 효과를 얻을 수 있다. 반복구매를 유도하기 위한 효과적인 프로모션 전략으로 평일과 주말의 제품별 고객 선호도 또는 주문수량에 기반하여 다른 프로모션을 제공하는 방법을 고려할 수 있으며, 식사용인지 간식용인지에 따라 적합한 영양 칼로리의 제품을 프로모션 할 수 있을 것이다. 또한 반복구매 주기가 짧은 고객은 비용보다는 신제품 등 제품 특성을 중시하는가 하면 반복구매 주기가 긴 고객은 비용을 중시 여기는 경향이 있다는 기존 연구를 토대로 [Bayus, 1991] 반복구매 주기에 따라 신제품이나 기존제품에 대한 다양한 프로모션을 제공할 수도 있을 것이다.

순환신경망은 언어모형에 주로 적용되고 있으나, 고객관계 관리와 같은 경영분야에 순환신경망 모형이 적용된 사례는 찾아보기 힘들다. 언어모형의 경우 문장의 패턴이 다양하고 어휘의 수가 많아 수행성능에 관한 이슈가 많은 반면 고객관계관리 분야의 경우 특징의 수가 상대적으로 작아 성능측면에서는 언어모형 보다 순환신경망의 적용이 용이하다. 그러나 언어모형이 공통의 문법을 가지는 다양한 문장들을 사례로 학습하는데 비해 고객 데이터의 경우, 고객마다 공통의 패

턴이 존재하는 것은 아니기 때문에 예측정확도나 예측품질에 있어서는 언어모형에 비해 뒤질 수밖에 없다. 또한 고객 구매데이터의 경우, 원자료에 0이 많이 나타나는 등 희소성(Sparsity)이 높은 것도 예측 정확도를 저하시키는 원인으로 작용한다. 따라서 이러한 문제를 극복하기 위해 기간별 구매량과 같은 추가적인 특징을 유도하여 예측에 사용하는 전략이 필요하다. 그럼에도 불구하고 반복구매시기 예측과 같이 시간에 의존하는 고객 구매패턴을 예측하는데 있어서는 기존 다층퍼셉트론 기법에 비해 순환신경망이 우수한 기법임을 알 수 있다.

향후 연구로써 고객별로 제공된 프로모션에 대한 반응이 피드백 되어 누적되면 보다 맞춤형된 프로모션을 설계할 수 있을 뿐 아니라 프로모션 반응률을 본 연구에서 제안한 예측모형의 입력 자료로 활용함으로써 예측 품질을 높일 수 있을 것이다. 이와 관련하여 Sato et al.[2015]는 그로서리 소매상을 대상으로 아이템에 대한 고객의 선호도만이 아니라, 아이템별 가격할인에 따른 민감도와 고객별 가격할인에 따른 민감도 정보를 결합하여 추천을 하는 맞춤형 베이시안 추천기(Bayesian Personalized Ranking : BPR)를 제안한 바 있다. 한편 본 연구에서는 인공적으로 생성된 데이터를 이용하여 실험을 진행하였으나 향후 실제 구매이력 데이터를 사용하여 제안한 모형의 예측 품질을 평가하고 추가적인 특징을 유도하여 데이터 희박성 문제를 극복하기 위한 노력도 필요할 것이다. 또한 다양한 아키텍처의 신경망을 설계하여 성능을 비교하는 것도 보다 높은 예측 품질을 확보하는데 필요할 것이다.

References

- [1] Agrawal, R., Ieong, S., and Velu, R., Timing When to Buy, ACM Conference on Informa-

- tion and Knowledge Management (CIKM), 2011.
- [2] Bastien, F., Lamblin, P., Pascanu, R., Bergstra, J., Goodfellow, I., Bergeron, A., Bouchard, N., Warde-Farley, D., and Bengio, Y., Theano : new features and speed improvements, NIPS 2012 deep learning workshop, 2012.
- [3] Bayus, B., L., “The Consumer Durable Replacement Buyer”, *Journal of Marketing*, Vol. 55, No. 1, 1991, pp. 42-51.
- [4] Bengio, Y., Simard, P., and Frasconi, P., “Learning long-term dependencies with gradient descent is difficult”, *IEEE Transactions on Neural Networks*, Vol. 5, No. 2, 1994, pp. 157-166.
- [5] Chen, Y. L. and Huang, T. C. K., “Discovering fuzzy time-interval sequential patterns in sequence databases”, *IEEE Syst. Trans. Man Cybernet Part B*, Vol. 35, No. 5, 2005, pp. 959-972.
- [6] Chiang, D. A., Lee, S. L., Chen, C. C., and Wang, M. H., “Mining interval sequential patterns”, *International Journal of Intelligent System*, Vol. 20, No. 3, 2005, pp. 359-373.
- [7] Glorot, X. and Bengio, Y., Understanding the difficulty of training deep feedforward neural networks, Proceedings of the International Conference on Artificial Intelligence and Statistics(AISTATS'10), 2010.
- [8] Gould, B. W. and Dong, D., “The Decision of When to Buy a Frequently Purchased Good : A Multi-Period Probit Model”, *Journal of Agricultural and Resource Economics*, Vol. 25, No. 2, 2000, pp. 636-652.
- [9] Hinton, G. E., Osindero, S., and Teh, Y., “A fast learning algorithm for deep belief nets”, *Neural Computation*, Vol. 18, 2006, pp. 1527-1554.
- [10] Hu, Y. H., Huang, T. C., Yang, H. R., and Chen, Y. L., “On mining multi-time-interval sequential patterns”, *Data Knowledge Engineering*, Vol. 68, No. 10, 2009, pp. 1112-1127.
- [11] Mulder, W. D., Bethard, S., and Moens, M.-F., “A survey on the application of recurrent neural networks to statistical language modeling”, *Computer Speech and Language*, Vol. 30, No. 1, 2015, pp. 61-98.
- [12] Oh, J., Kim, S., Kim, J., and Yu, H., “When to recommend : A new issue on TV show recommendation”, *Information Sciences*, Vol. 280, No. 1, 2014, pp. 261-274.
- [13] Rendle, S., Freudenthaler, C., and Schmidt-Thieme, L., Factorizing personalized markov chains for next-basket recommendation, In WWW Conference, 2010, pp. 811-820.
- [14] Sato, M., Izumo, H., and Sonoda, T., Discount Sensitive Recommender System for Retail Business, Proceedings of the 3rd Workshop on Emotions and Personality in Personalized Systems, 2015, pp. 33-40.
- [15] Wang, J., Sarwar, B., and Sundaresan, N., Utilizing related products for post-purchase recommendation in e-commerce, Proceedings of the fifth ACM conference on Recommender systems, 2011, pp. 329-332.
- [16] Zhao, G., Lee, M. L., and Wynne, H., Utilizing Purchase Intervals in Latent Clusters for Product Recommendation, Proceedings of the 8th Workshop on Social Network Mining and Analysis (SNAKDD'14), 2014, pp. 1-9.

■ 저자소개



송희석

고려대학교 경영학과 학사, 한국과학기술원 경영과학과 석사 및 경영공학과 박사학위를 취득하였으며 대우전자, 대우정보시스템 기술연구소 근무를 거쳐

현재는 한남대학교 글로벌IT경영학과 교수로 재직 중이다. 관심분야는 CRM과 Data Mining, 소셜 네트워크, 유비쿼터스 비즈니스, 비즈니스모델 등이며 주요연구 결과는 Knowledge-based systems, Expert systems with applications, Artificial Intelligence Review 등의 학술지에 발표하였다.