



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

공학석사 학위논문

시간 정보와 인기도 샘플링을
이용한 세션 기반 추천 모델

A Session-based Recommendation Model
using Temporal Information and
Popularity-based Sampling

2018년 12월

서울대학교 대학원

전기·정보공학부

이희원

시간 정보와 인기도 샘플링을 이용한 세션 기반 추천 모델

지도교수 심 규 석

이 논문을 공학석사 학위논문으로 제출함
2018년 11월

서울대학교 대학원
전기·정보공학부
이 희 원

이희원의 공학석사 학위논문을 인준함
2018년 12월

위 원 장 김태환 (인)

부위원장 심규석 (인)

위 원 홍성수 (인)

초 록

세션 기반 추천모델은 작은 규모의 인터넷 쇼핑 사이트에서 옷을 추천하거나 뉴스 사이트에서 관심 뉴스를, 동영상 제공 사이트에서 영상을 추천하는 경우에 주로 사용된다. 즉, 세션 기반 추천모델은 처음 접속해보는 사이트 혹은 로그인을 하지 않아 과거 접속 정보가 없는 경우에 추천을 하는 모델이다. 익명의 사용자와 상품 간의 상호작용에 대한 현재의 짧은 세션 정보만을 바탕으로 추천을 하기 때문에 상품 간의 유사도를 바탕으로 추천하는 모델이 과거에는 많이 사용되었다. 하지만 이와 같은 상품 기반 유사도 추천 모델은 마지막에 선택한 상품만을 바탕으로 추천을 한다는 단점이 있었다. 이러한 문제를 해결하기 위해 최근에 순환 신경망 기반의 모델이 제안되었고 좋은 성능을 보여주고 있다. 그렇지만 순환 신경망을 기반으로 한 기존 연구도 상품을 선택한 순서만을 이용해 추천을 하기 때문에 시간 정보를 활용하지 못하는 단점이 있다.

본 논문에서는 순환 신경망 기반 모델에 월, 요일, 시간, 분, 초의 시간 정보를 추가로 반영하여, 계절, 요일 등의 변화에 따른 사용자의 선호도 변화를 반영한 모델을 제안한다. 그리고 로그 기반의 인기도 샘플링 방법과 모델을 학습시키는 손실 함수에 가중치로직을 반영하여 모델의 성능을 높일 수 있는 방법을 제안한다. 또한 제안하는 모델의 성능을 검증하기 위해 실생활 데이터를 이용하여 실험을 진행하였다.

주요어 : 추천 시스템, 순환 신경망 모델, 인기도 기반 샘플링, GRU

학 번 : 2017-21079

목 차

제 1 장 서론	1
제 1 절 연구의 배경 및 내용	1
제 2 장 관련 연구	3
제 1 절 세션 기반 추천 시스템과 순환신경망	3
제 2 절 배치 사이즈와 인기도 기반 샘플링	8
제 3 절 베이지안 개인화 순위	10
제 3 장 제안하는 세션 기반 추천 모델	13
제 1 절 시간 정보를 반영한 추천 모델 구조	13
제 2 절 최적화	18
제 4 장 실험	21
제 1 절 실험 환경 및 실험 데이터	21
제 2 절 성능 측정 지표 및 실험 설정	23
제 3 절 실험 결과 및 분석	24
제 5 장 결론	28
참고문헌	29
Abstract	31

표 목 차

[표 1] 실험 데이터 클릭수 통계정보	18
[표 2] 실험 데이터 기본정보	22
[표 3] 인기도 기반 샘플링과 로그 기반 인기도 샘플링 성능 비교	24
[표 4] 기존 연구와 제안 모델의 Recall@20 성능 비교	26

그 림 목 차

[그림 1] 상품 기반 협업 필터링 과정	3
[그림 2] 순환 신경망 구조	5
[그림 3] 게이트 순환 유닛 구조	6
[그림 4] 제안하는 세션 기반 추천 모델 구조	13
[그림 5] 피드포워드 층의 구조	16
[그림 6] 인기도 기반 샘플링과 로그 기반 인기도 샘플링의 예	19
[그림 7] 인기도 기반 샘플링과 로그 기반 인기도 샘플링 성능 비교	24

제 1 장 서론

제 1 절 연구의 배경 및 내용

우리는 일상생활에서 잘 인지하지 못하는 순간에도 다양한 추천 정보를 제공받으며 살아가고 있다. 작은 규모의 인터넷 쇼핑 사이트에서 옷을 고르거나 뉴스 사이트에서 뉴스를 보고 동영상 제공 사이트에서 영상을 보는 순간순간마다 추천을 받고 있다. 그리고 이와 같은 사이트는 일반적으로 사용자가 로그인을 하지 않은 상태에서 검색을 하게 되는데 이 짧은 기간 동안 사용자의 행동 정보를 세션이라고 말한다. 즉, 본 논문에서 설명하는 세션 기반 추천 모델은 처음 접속해보는 사이트 혹은 로그인을 하지 않아 과거 접속 정보가 없는 경우에 추천을 하는 모델이다.

세션 기반 추천 모델은 세션의 짧은 정보만을 바탕으로 상품을 추천하기 때문에 과거에는 상품 간의 유사도를 바탕으로 추천하는 연구[1]가 많이 있었다. 하지만 상품 간의 유사도를 바탕으로 추천하는 경우에는 마지막에 선택한 상품과 유사한 상품만을 추천하는 단점이 있었다. 이러한 문제를 해결하기 위해 최근에는 상품 선택의 순서를 반영한 순환 신경망(recurrent neural network: RNN) 기반의 연구들[2, 3, 4, 5]이 많이 진행되었고, 성능 면에서도 많은 발전을 이루었다. 그렇지만 기존 순환 신경망 기반의 연구들도 시간의 순서에 대한 고려만 있고 시간 정보가 가지고 있는 계절, 요일, 시간대에 따른 사용자의 선호도 변화를 고려하지 않고 있다.

본 논문에서는 순환 신경망 기반 모델에 월, 요일, 시간, 분, 초의 시간 정보(time feature)를 추가로 반영하여, 시간에 따른 사용자의 선호도 변화를 반영한 모델을 제안한다. 그리고 로그 기반의 인기도 샘플링 방법(popularity-based sampling)과 모델을 학습시키는 손실 함수에 가중치 로직을 반영하여 모델의 성능을 높일 수 있는 방법을 제안한다. 먼저 2장에서는 관련 연구 및 게이트 순환 유닛(gated recurrent unit: GRU)[6]에 대한 설명을 포함하여, 베이저안 개인화 랭킹(bayesian

personalized ranking: BPR)[9]과 인기도 기반 샘플링에 대하여 소개한다. 3장에서는 본 논문에서 제안하는 시간 정보와 인기도 샘플링을 이용한 세션 기반 추천 모델에 대해 소개한다. 4장에서는 제안한 모델과 시간 정보를 활용하지 않은 기존 연구 모델의 성능을 비교하고 실험 결과를 분석한다. 마지막으로 5장에서는 내용을 정리하고 결론을 맺는다.

제 2 장 관련 연구

제 1 절 세션 기반 추천 시스템과 순환신경망

세션 기반 추천 시스템은 현재의 짧은 세션에서 익명의 사용자가 선택한 상품 정보의 리스트와 상품을 선택한 시간 정보만을 바탕으로 추천이 이루어진다. 과거에는 대표적으로 상품 기반 협업 필터링(item-based collaborative filtering)[1]과 콘텐츠 기반 추천 모델(content-based recommendation)[7]이 주로 사용되어 왔다. 최근에는 순환신경망 모델의 단점을 보완한 게이트 순환 유닛(gated recurrent unit: GRU)[6]이 세션 기반 추천 모델로 좋은 성능을 보이고 있다. 그래서 2장에서는 앞에서 소개한 기존 추천 모델과 게이트 순환 유닛에 대해 소개하고자 한다.

상품 기반 협업 필터링은 M 개의 세션과 N 개의 상품, 그리고 사용자가 상품을 선택했는지 여부를 그림 1과 같이 입력으로 받아 상품을 추천한다. 현재 사용자 세션 u_a 의 마지막 선택 상품이 i_i 일 때, i_i 와 i_1 부터 i_n 까지의 코사인 유사도(cosine similarity)를 각각 연산한 후 가장 높은 값을 나타내는 Top-N개의 상품을 추천하는 것이 목표이다. 코사인 유사도는 두 개의 상품 i 와 j 간의 상관관계를 식 (2-1)을 통해 계산한다.

$$sim(i, j) = \frac{\sum_{u \in U} (R_{u,i} - \bar{R}_i)(R_{u,j} - \bar{R}_j)}{\sqrt{\sum_{u \in U} (R_{u,i} - \bar{R}_i)^2} \sqrt{\sum_{u \in U} (R_{u,j} - \bar{R}_j)^2}} \quad \text{식 (2-1)}$$

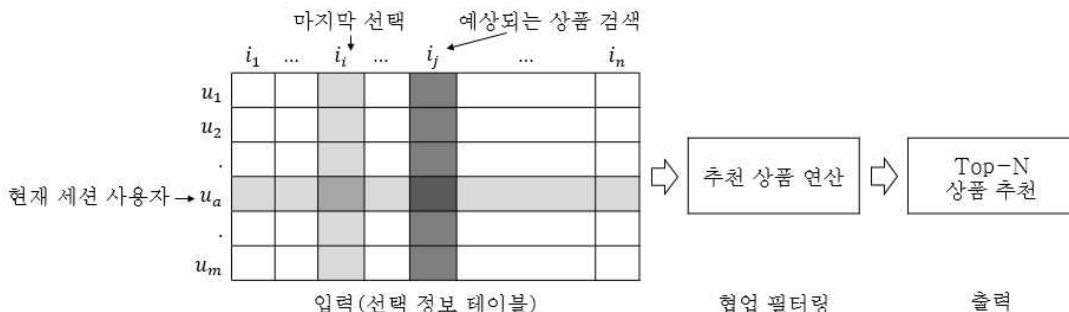


그림 1. 상품 기반 협업 필터링 과정

여기서 $R_{u,i}$ 는 세션 u 에서 상품 i 가 선택되었을 시 1의 값을, 선택되지 않았을 시 0의 값을 나타내는 변수이다. \bar{R}_i 는 i 번째 상품의 평균값을 나타낸다. 상품 기반 협업 필터링은 굉장히 널리 쓰이는 모델이지만 데이터가 충분하지 않은 상태에서는 추천이 부정확한 콜드 스타트(cold start) 문제와 상품이 많아지면 입력 데이터를 저장하기 어려운 확장성(scalability) 문제, 대부분의 데이터가 0을 나타내는 희소성(sparsity) 문제를 가지고 있다. 그중 세션 기반 추천에서 치명적인 문제는 마지막에 선택한 상품만을 기준으로만 추천을 한다는 점이다.

콘텐츠 기반 추천 모델은 상품의 콘텐츠 정보의 유사도를 기준으로 추천을 하는 모델이다. 상품을 설명하는 정보에서 핵심 키워드를 추출하고 그 키워드의 유사도를 계산한 후 Top-N개의 상품을 추천하는 것이 목표이다. 상품 정보에서 키워드를 추출하는 방법으로 TF-IDF(term frequency - inverse document frequency)를 사용한다. TF는 특정 단어가 문서 내에서 얼마나 자주 등장하는지를 나타내는 수치로 값이 클수록 문서에서 중요하다고 판단할 수 있다. 하지만 단어 자체가 모든 문서에서 자주 등장하는 경우에는 그 단어가 흔하게 사용된다고 판단할 수 있다. 이를 문서빈도(document frequency: DF)라고 하며 이 값의 역수를 역문서빈도(inverse document frequency: IDF)라고 한다. TF-IDF는 TF와 IDF를 곱한 값을 나타내고 식 (2-2)와 같이 계산한다.

$$W_{t,d} = tf_{t,d} \times \log\left(\frac{N}{df_t}\right) \quad \text{식 (2-2)}$$

여기서 t 는 단어에 대한 인덱스, d 는 문서에 대한 인덱스를 의미하며, $tf_{t,d}$ 는 문서 d 에서 t 가 반복되는 수를 나타낸다. df_t 는 t 가 포함된 문서의 수를 나타내고 N 은 총 문서의 수를 나타낸다. TF-IDF를 통해 상품 정보에서 키워드 집합을 도출하면, 상품 정보 간의 유사도를 자카드 유사도(jaccard similarity)를 이용하여 계산한다. 자카드 유사도 값은 식 (2-3)과 같이 계산한다.

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \quad \text{식 (2-3)}$$

콘텐츠 기반 추천은 콜드 스타트 문제가 없다는 장점이 있으나, 이 역시 마지막에 선택한 상품을 기준으로 추천을 한다는 단점을 보완하지는 못한다.

순환신경망(recurrent neural network: RNN) 모델은 음성, 문자 등 순차적으로 등장하는 데이터 처리에 용이한 모델이다. 그래서 시퀀스 길이에 관계없이 입력과 출력을 받아들일 수 있는 네트워크 구조로 되어 있다. 즉, 다양하고 유연하게 구조를 생성할 수 있는 점이 장점이다. 순환신경망의 구조는 그림 2와 같다. x_t 는 입력, h_t 는 은닉 상태(hidden state), y_t 는 출력을 나타낸다. 은닉 상태 벡터는 식 (2-4)로 계산되고 출력은 식 (2-5)로 계산된다.

$$h_t = \tanh(W_{hh}h_{t-1} + W_{xh}x_t + b_h) \quad \text{식 (2-4)}$$

$$y_t = W_{hy}h_t + b_y \quad \text{식 (2-5)}$$

여기서 W_{hh} , W_{xh} , W_{hy} 는 각 입력과 은닉 상태 벡터의 차원을 동일하게 변환해 주는 변환 행렬이며, b_h , b_y 는 과대적합(overfitting)을 방지하는 항이다. 활성화함수(activation function) f로는 tanh를 사용한다. 기본 순환신경망의 구조는 순차적 데이터 처리에 용이한 장점이 있지만 내부의 유닛

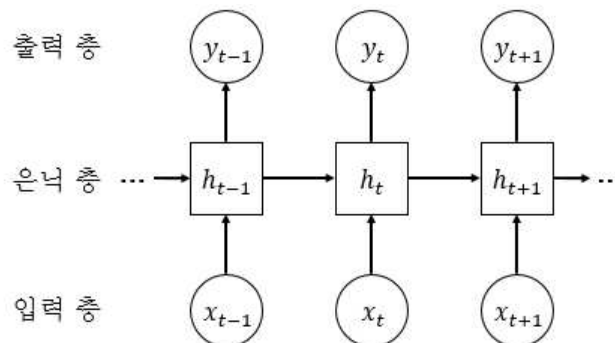


그림 2. 순환 신경망 구조

이 많아지면 경사 하강법(gradient descent)으로 학습이 잘되지 않는, 경사 값이 사라지는(vanishing gradient) 문제가 발생한다는 단점이 있다.

그래서 본 논문에서는 경사 값이 사라지는 문제를 효과적으로 개선한 모델, LSTM(long short-term memory)[8]과 게이트 순환 유닛(GRU) 중 모델 파라미터를 적게 사용하여 효과적인 성능을 나타낸 게이트 순환 유닛을 사용한다. 게이트 순환 유닛의 구조는 그림 3과 같이 구성되어 있다. 여기서 z 는 업데이트 게이트(update gate)를 의미하며 얼마만큼의 정보를 업데이트 할지를 나타낸다. 계산은 식 (2-6)과 같이 현재 시점 t 의 입력 x_t 와 이전 은닉 상태 h_{t-1} 로 계산되며, W_z, U_z 는 벡터의 차원을 동일하게 변환해주는 변환 행렬이다. 활성화함수 σ 는 시그모이드를 사용한다.

$$z_t = \sigma(W_z x_t + U_z h_{t-1}) \quad \text{식 (2-6)}$$

r 은 리셋 게이트(reset gate)를 의미하며 계산은 식 (2-7)과 같이 한다.

$$r_t = \sigma(W_r x_t + U_r h_{t-1}) \quad \text{식 (2-7)}$$

다음으로 기억에 관한 처리 과정은, 현재 시점 t 에 기억할 정보를 식 (2-8)을 통해 계산한다.

$$\tilde{h}_t = \tanh(W_h x_t + U_h (r_t \odot h_{t-1})) \quad \text{식 (2-8)}$$

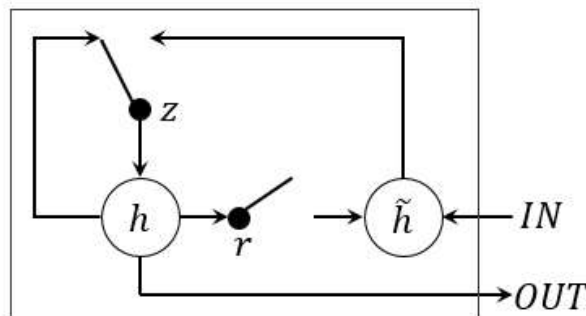


그림 3. 게이트 순환 유닛 구조

식 (2-8)은 현재 시점의 정보 x_t 와 과거정보 h_{t-1} 을 얼마만큼 반영할지 r_t 를 이용하여 계산한다. 리셋 게이트의 활성화함수는 시그모이드이므로 0~1 사이의 값을 갖게 되는데, 0은 정보를 잊고 1은 정보를 기억한다는 의미를 갖는다. 여기서 \odot 는 원소 단위 곱을 의미해서 정보의 기억량을 각각의 곱으로 계산한다. 마지막으로 다음 상태를 업데이트 하는 계산은 식 (2-9)와 같다.

$$h_t = (1 - z_t)h_{t-1} + z_t\tilde{h}_t \quad \text{식 (2-9)}$$

과거정보 h_{t-1} 와 현재정보 \tilde{h}_t 를 업데이트 게이트를 이용해서 선형 보간 방법으로 계산한다.

제 2 절 배치 사이즈와 인기도 기반 샘플링

신경망 모델의 학습 과정은 손실 함수의 값을 최소화(Optimization) 시키기 위해 모델 변수(parameter)를 업데이트해가는 과정이다. 그리고 이 과정을 일회 처리할 때 사용하는 데이터의 양을 배치 사이즈(batch size)라고 한다. 즉, 신경망 모델의 학습은 하나의 배치가 처리될 때마다 그 결과를 바탕으로 손실 함수 값을 계산하고 모델 변수를 업데이트 하는 과정을 통해 진행한다.

배치 사이즈는 신경망 모델을 학습할 때 매우 중요한 역할을 하는데, 배치 사이즈의 크기를 어떻게 설정하느냐에 따라 장단점이 달라진다. 하나의 배치를 전체 데이터셋으로 구성하는 경우, 전체 데이터를 반영하여 모델 변수를 업데이트하게 되어 정확한 손실 함수 값을 반영할 수 있다는 장점이 있다. 단점으로는 데이터셋이 클 경우 한번 수행하는데 소요 되는 시간이 매우 커 전역 최적값(global optima)을 구하지 못할 수 있다. 또한 메모리 측면에서도 데이터를 처리할 수 있는 메모리 제약으로 GPU 병렬 연산 처리 시 에러가 발생할 수 있다. 이와 반대로 배치 사이즈를 1로 사용하는 경우에는 수행시간이 매우 빨라 학습이 빠르게 진행된다는 장점이 있다. 단점은 데이터 전체의 특성을 반영하기가 힘들어 매 배치 처리마다 성능이 좋을 것이라는 보장을 할 수 없다. 그래서 많은 양의 데이터를 학습시키는 모델에서는 이 두 가지 경우의 장점을 조합한 미니 배치(mini-batch)를 사용한다. 적게는 수십에서 많게는 수백의 데이터를 하나의 배치로 구성하여 학습함으로써 메모리 제약 안에서 빠르게 학습하고 전체 데이터의 특성을 고루 반영할 수 있도록 한다.

본 논문에서는 미니 배치에 사용하는 데이터를 생성할 때, 인기도 기반 샘플링(popularity-based sampling)[4]을 개선하여 사용한다. 인기도 기반 샘플링은 전체 데이터셋에서 상품이 선택된 누적 분포를 이용하여, 전체 사용자가 많이 선택한 상품일수록 자주 추출되도록 하는 샘플링 기법이다. 인기도 기반 샘플링을 활용한 모델 학습은, 목표 상품(target item)과 샘플링으로 추출된 네거티브 샘플(negative sample)의 점수를 비교하는 방식으로 진행한다. 이때, 학습은 목표 상품의 점수가 네거티브

샘플의 점수보다 높아지는 방향으로 이루어지는데, 이미 점수가 낮은 상품보다는 점수가 높은 상품을 위주로 네거티브 샘플링 해야 학습이 효율적으로 이루어지게 된다. 이는 이미 점수가 크게 벌어진 상품의 경우는 이미 목표 상품과 서로 잘 구분되어 성능을 향상시킬 요소가 적지만, 점수가 높은 네거티브 샘플은 학습을 통해 성능을 높일 요소가 많기 때문이다. [4]에서는 일반적으로 인기도가 높은 상품의 점수가 높아지는 경향이 있다는 점에 활용하여 i 번째 상품에 대한 선택 수를 $supp_i$ 로 정의하고 $supp_i^\alpha$ ($0 \leq \alpha \leq 1$)에 비례하는 확률 분포를 생성하여 샘플링 하였다. α 값이 0에 가까울 경우 균등 분포(uniform distribution)와 유사한 확률 분포가 되고 α 가 1에 가까울 경우 인기도 비율 그대로를 확률로 사용한다는 의미이다.

제 3 절 베이지안 개인화 순위

베이지안 개인화 순위(bayesian personalized ranking: BPR)[9]는 쌍위 순위 손실 함수로써, 이는 사용자가 선택한 목표 상품의 점수가 네거티브 샘플(목표 상품 이외의 상품들 중 샘플링한 상품들의 집합)의 점수보다 높아지는 방향으로 학습되게 유도하는 역할을 한다. 계산은 식 (2-10)과 같이 계산되며, 목표 상품과 네거티브 샘플간의 점수를 비교하는 평균 손실 함수이다. 음의 값으로 함수가 구성되어, 목표 상품과 네거티브 샘플과의 점수 차가 커질수록 전체적인 손실 값을 낮추는 역할을 한다.

$$L_{bpr} = -\frac{1}{N_S} \cdot \sum_{j=1}^{N_S} \log(\sigma(r_i - r_j)) \quad \text{식 (2-10)}$$

여기에서 N_S 는 네거티브 샘플의 크기를 나타내고, r_i 는 아이템 i 에 대한 점수를 의미하며, i 와 j 는 각 목표 상품과 네거티브 샘플의 인덱스를 의미한다. 베이지안 개인화 순위 함수 BPR은 학습을 진행할수록 경사 값이 사라지는(vanishing gradient) 문제가 발생한다는 단점이 있다. 쌍위 손실 값의 평균을 사용하기 때문에 학습이 진행될수록 다수의 상관성이 없는 샘플들이 기울기의 변화를 약화시키는 역할을 하게 되기 때문이다. BPR의 기울기 값을 구하기 위해 미분으로 값을 계산하면, 식 (2-11)과 식 (2-12)를 이용하여 식 (2-13)과 같이 도출된다.

$$\begin{aligned} \frac{d}{dx} \sigma(x) &= \frac{d}{dx} \frac{1}{1+e^{-x}} = \frac{e^{-x}}{(1+e^{-x})^2} = \frac{1+e^{-x}-1}{(1+e^{-x})^2} \\ &= \frac{1}{1+e^{-x}} - \frac{1}{(1+e^{-x})^2} = \sigma(x)(1-\sigma(x)) \quad \text{식 (2-11)} \end{aligned}$$

$$\frac{d}{dx} \log f(x) = \frac{f'(x)}{f(x)} \quad \text{식 (2-12)}$$

$$\frac{dL_{bpr}}{dr_i} = -\frac{1}{N_S} \sum_{j=1}^{N_S} (1-\sigma(r_i - r_j)) \quad \text{식 (2-13)}$$

식 (2-13)을 보면 $r_i \gg r_j$ 의 값을 나타내는 다수의 샘플들이 $1 - \sigma(r_i - r_j)$ 에 의해 0에 가까운 값을 갖게 되고 기울기 값의 변화가 작아 값을 업데이트하는데 어려움을 갖게 된다. 그래서 [4]에서는 BPR의 단점을 보완한 BPR-max 손실 함수를 제안하였다.

BPR-max는 네거티브 샘플중 가장 큰 값을 나타내는 r_{\max} 보다 목표 상품의 값이 더 크게 나타날 확률을 최대화 시키는 것을 목표로 한다. 이를 조건부 확률로 나타내면 식 (2-14)와 같다.

$$P(r_i > r_{\max}) = \sum_{j=1}^{N_s} P(r_i > r_j | r_j = r_{\max}) P(r_j = r_{\max}) \quad \text{식 (2-14)}$$

$P(r_i > r_j)$ 는 기존 BPR 함수의 $\sigma(r_i - r_j)$ 와 유사한 확률을 나타내고 $P(r_j = r_{\max})$ 는 softmax 함수와 유사한 확률을 나타낸다고 가정하여 BPR-max 손실 함수를 식 (2-15)와 같이 정의하였다.

$$L_{bpr - \max} = -\log \sum_{j=1}^{N_s} s_j \sigma(r_i - r_j) \quad \text{식 (2-15)}$$

여기서 s_j 는 j 상품에 대한 softmax 값을 의미한다. softmax 함수를 도입함으로써 점수가 높은 네거티브 샘플 위주로 값이 업데이트될 수 있도록 하는 장점이 생긴다. 점수가 낮은 네거티브 상품은 이미 목표 상품과 잘 구분이 되고 있어 더 이상 업데이트가 될 필요가 없기 때문에 BPR-max가 전반적으로 좋은 성능을 보이게 된다. BPR-max 손실 함수는 r_i 목표 상품 점수가 낮을 때 전반적으로 파라미터 값들이 업데이트 되도록 하고 r_i 값이 높아질수록 높은 점수를 가진 네거티브 샘플에 집중해서 파라미터 값들이 업데이트 되도록 도와준다. BPR-max 손실 함수는 추가로 과대적합(overfitting)을 방지하고 학습이 원활이 이뤄질 수 있도록 l_2 정규화(l_2 regularization) 항을 추가하였다. 정규화 항목을 추가할 때에도 softmax 값을 동일하게 반영하고 λ 정규화 파라미터를 활

용하여 식 (2-16)과 같이 최종 함수를 적용하였다.

$$L_{bpr-\max} = -\log \sum_{j=1}^{N_s} s_j \sigma(r_i - r_j) + \lambda \sum_{j=1}^{N_s} s_j r_j^2 \quad \text{식 (2-16)}$$

본 논문 모델에서 적용한 손실 함수에 대한 자세한 구성이나 인기도 샘플링을 개선한 방법에 대한 상세한 설명은 3장에서 설명한다.

제 3 장 제안하는 세션 기반 추천 모델

본 연구에서 제안하는 모델은 [4]에서 제안한 세션 기반 추천 모델을 기반으로 하고 있다. 기본 구조는 동일하게 사용하고 있으나, 추천이 발생하는 시점의 시간 특성 정보(time feature)를 추출하고 이 정보를 추천에 이용할 수 있도록 모델을 구성하였다. 이를 통해 시간의 변화가 추천에 반영될 수 있도록 하였다. 또한 데이터가 특정 상품에 지나치게 편향되어 선택되는 현상(skewed data)이 있을 때 인기도 샘플링이 성능이 좋지 않아 이를 보완할 수 있도록 로그 기반의 인기도 샘플링을 반영하였다. [4]에서 제안된 BPR-max 손실 함수는 학습이 진행됨에 따라 학습 속도가 느려진다는 단점이 있어 가중치를 부여할 수 있도록 손실 함수를 개선하여 반영하였다.

제 1 절 시간 정보를 반영한 추천 모델 구조

본 논문에서는 그림 4에서와 같이 네트워크 구조를 구성한다. 시간 특성 정보(time feature)를 추출하고 이 정보를 게이트 순환 유닛 층(GRU layer)의 결과 벡터와 연결(concatenation)하도록 구성한다. 그리고 해당 벡터를 피드포워드 층(feedforward layer)을 거쳐 추천 결과를 상품

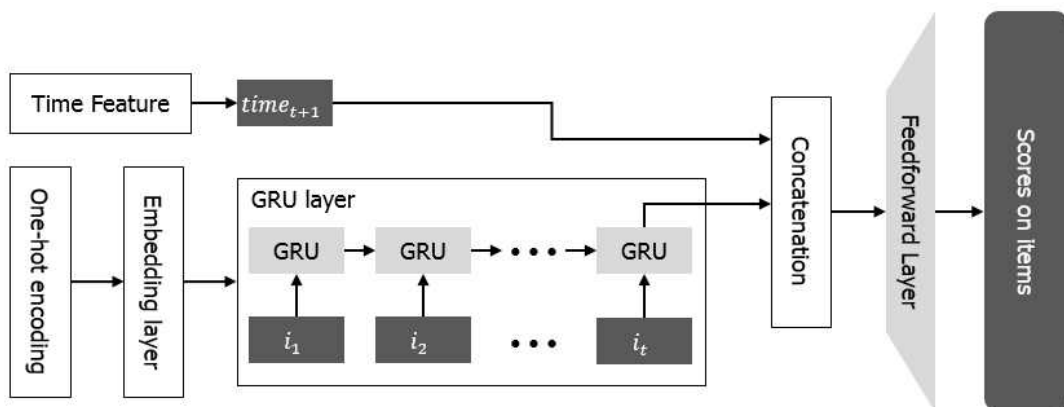


그림 4. 제안하는 세션 기반 추천 모델 구조

별 점수로 도출하도록 구성한다. 본 논문에 적용된 각 층에 대해서는 상세히 나누어 설명하도록 하겠다.

1. 원 핫 인코딩(one-hot encoding)

원 핫 인코딩이란 텍스트 기반의 데이터를 숫자 기반의 벡터로 변환하는데 사용하는 방법론이다. 학습 모델에서 학습에 필요한 모든 상품의 개수를 N 으로 정의했을 때 N 차원의 벡터로 각각의 상품을 표현할 수 있다. 상품을 순서대로 정렬하고 해당 상품의 순서에 해당하는 위치의 차원 값만 1로 표시하고 나머지는 0의 값을 갖는 N 차원의 벡터가 상품을 대표하는 벡터 값이 된다. 예를 들어 총 상품이 5라고 했을 때, 첫 번째 상품은 [10000] 두 번째 상품은 [01000]과 같이 표현될 수 있다. 원 핫 인코딩은 학습에 용이하게 구조를 변환하고 처리과정도 단순하다는 장점을 가지고 있다. 하지만 비슷한 특성을 가지는 상품 간에 상관성을 표현하지 못한다는 단점을 가지고 있다. 상품을 추천하는 시스템 모델 입장에서는 비슷한 유형의 상품들이 비슷한 벡터로 표현될 수 있다면 보다 좋은 성능을 나타낼 수 있을 것이다.

2. 임베딩(embedding)

임베딩은 미리 정의된 차원에서 연속형의 값을 갖는 벡터 값으로 표현된다. 일반적으로 20~200 정도의 작은 차원으로 구성하고 원 핫 인코딩과는 달리 모든 차원이 값을 가지고 있다. 그리고 해당 차원 값들의 조합으로 상품의 속성을 표현한다. 임베딩의 차원 값들은 머신 러닝을 통해 학습하고, 학습을 통해 비슷한 속성의 상품들이 비슷한 벡터 값을 나타내게 된다. 즉, 상품의 속성이 얼마나 유사한지 벡터간의 거리를 통해 유사도를 측정할 수 있게 된다. 벡터의 유사도를 이용한 일반화(generalization)된 학습이 가능하다는 점은 세션 기반 추천 시스템과 같이 짧은 세션의 학습 데이터를 이용해서 학습을 진행해야하는 경우에 유용하다. 또한 임베딩을 사용하면 차원의 저주(curse of dimensionality)라 불리는 고차원 벡터에서 학습이 되지 않는 문제가 발생하지 않는 장점이

있다. 상품의 총수 N 이 너무 큰 경우 원 핫 인코딩만으로는 학습이 되지 않는 문제가 발생할 수 있다.

3. GRU와 시간 특성 정보(time feature) 연결

그림 4는 시간 특성 정보를 활용하여 추천하는 모델 구조를 나타낸다. 시작 시점 1에서 현재시간 t 시간까지 선택한 상품들의 이력을 i_1, i_2, \dots, i_t 로 표현할 때, GRU를 거쳐 은닉 정보(hidden representation)가 생성된다. 이 은닉 정보 벡터와 시간 $t+1$ 에 만들어진 시간 특성 정보 $time_{t+1}$ 벡터를 연결(concatenation)하여 피드포워드 층(feedforward layer)의 입력으로 생성한다. 이는 사용자가 새로운 상품을 선택하는 시점에 새로운 상품 추천이 이루어질 수 있도록 구조를 반영한 것이다.

시간 특성 정보는 월, 요일, 시간, 분, 초로 구분한 5가지 실수형 데이터로 생성하였다. 특성 정보는 식 (3-1) ~ 식 (3-5)를 통해 계산한다.

$$time_{month} = (month - 5.5) / 11 \quad \text{식 (3-1)}$$

$$time_{week} = (week - 3) / 6 \quad \text{식 (3-2)}$$

$$time_{hour} = (hour - 11.5) / 23 \quad \text{식 (3-3)}$$

$$time_{min} = (min - 29.5) / 59 \quad \text{식 (3-4)}$$

$$time_{sec} = (sec - 29.5) / 59 \quad \text{식 (3-5)}$$

여기서 $month, week, hour, min, sec$ 값은 정수 값이며 $month$ 는 0~11까지의 정수를, $week$ 는 0~6의 정수, $hour$ 는 0~23의 정수, min 는 0~59의 정수, sec 는 0~59의 정수 값을 가진다. 모든 시간 특성 정보가 정규화(normalization) 될 수 있도록 중심을 0으로 두고 구간이 1이 될 수 있도록 계산하였다. 위의 식을 거쳐 생성된 시간 특성 정보는 5차원의 시간 특성 정보 벡터가 되고 이 벡터를 100차원의 GRU 은닉 정보 벡터와 연결하여 105차원의 벡터가 생성된다. GRU 은닉 정보 벡터는 하이퍼파라미터(hyperparameter)로 모델 구성에 따라 변경 가능하지만, 본 논문에서는 100으로 설정하여 실험한다.

3. 피드포워드 층(feedforward layer)

피드포워드 층은 GRU의 은닉 정보 벡터와 시간 특성 정보가 연결된 벡터를 입력 받아 각각의 상품을 나타내는 N 차원의 벡터로 변환하는 역할을 한다. 그림 5는 완전 연결 층(fully connected layer)라고도 불리는데, 입력 단 벡터의 각 차원 값을 출력 단에 얼마나 반영할지 가중치(weight)를 곱해서 모두 더한 값을 출력으로 나타낸다.

$$y_j = \sum_{i=1}^n w_{ij}x_i, \text{ for } j = 1, \dots, N \quad \text{식 (3-6)}$$

식 (3-6)과 같이 계산을 하며, n 은 입력단의 차원을 나타내고 j 는 1에서 N 까지의 값으로 출력 단 각 차원 값을 계산한다. w_{ij} 는 학습을 통해 반영되는 가중치 값이다.

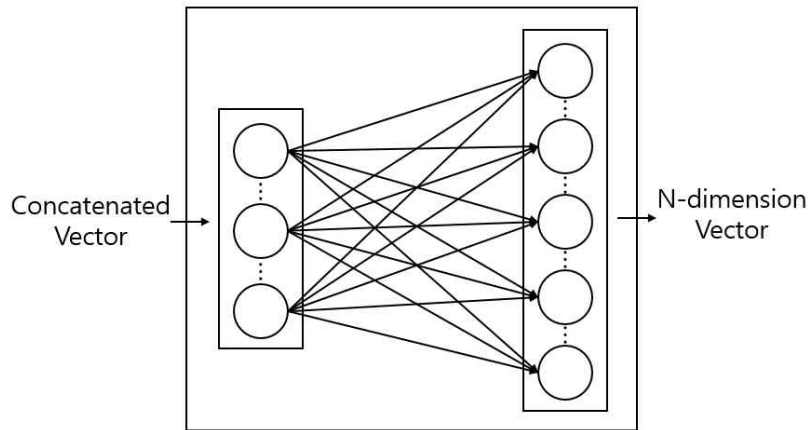


그림 5. 피드포워드 층의 구조

4. 소프트맥스(softmax) 함수와 점수 계산

소프트맥스(softmax) 함수는 점수를 모두 합한 결과가 1이 되고 각각의 점수는 0에서 1사이의 값을 갖도록 변환시키는 함수이다. 즉, 전체의 합이 1이기 때문에 소프트맥스 결과 값은 확률로 볼 수 있다. 소프트맥

스 함수 계산은 식 (3-7)과 같이 지수로 표현되는데, 이는 입력 값이 음수일 경우 음의 확률로 계산되는 것을 방지하기 위함이다.

$$S(y_i) = \frac{e^{y_i}}{\sum_{j=1}^N e^{y_j}}, \text{ for } i = 1, \dots, N \quad \text{식 (3-7)}$$

분모는 지수로 변환된 입력 값들을 모두 더한 값으로 전체를 나타내고 분자는 i 번째 지수로 변환된 입력 값을 나타낸다. 그러므로 모든 소프트맥스의 값 $S(y_i)$ 는 0에서 1의 값을 나타내고 합이 1이 되도록 구성됨을 알 수 있다. 본 논문에서 각 상품의 점수는 소프트맥스로 도출된 확률 값을 그대로 이용하여 표현하였다. 이 점수는 손실 함수의 입력 값으로 사용한다.

제 2 절 최적화

1. 로그 기반 인기도 샘플링

[4]에서 적용된 인기도 기반 샘플링은 특정 상품 위주로 데이터 쏠림 현상(data skew)이 있을 경우 좋은 성능을 발휘하지 못하는 문제가 있다. 본 논문에서는 표 1에서 보는 것과 같이 3가지의 데이터셋을 실험 데이터로 사용하는데, RSC15와 LFM은 평균과 중위수 값의 차이가 매우 크고 표준편차의 값이 큰 값을 보이고 있어 데이터 쏠림 현상 큰 것을 알 수 있다. (표 1의 $supp$ 는 기존 인기도 샘플링 통계 값을 나타내고, $\log(supp)$ 은 본 논문에서 적용한 로그 기반 샘플링의 통계 값을 나타낸다.) 그래서 일반적인 인기도 샘플링을 사용할 경우 균등 분포 샘플링이 더 좋은 성능을 보였다. 이는 네거티브 샘플링이 특정 상품 위주로만 선택되어 학습이 전체 상품 대상으로 이뤄지지 않는다는 것을 알 수 있다. 이에 인기도 기반 샘플링을 로그 스케일 변환을 적용하여 샘플링하도록 모델을 개선하여 적용하였다. 그리고 2장 2절에서 설명한 α 값을 사용하는 로직은 로그 기반 인기도 샘플링에서도 동일하게 적용하였다. 수식은 식 (3-9)과 같다.

데이터	타입	평균	중위수	최대값	표준편차
RSC15 ①	$supp$	674.12	63.00	84278.00	2434.16
	$\log(supp)$	9.40	5.98	16.36	11.25
LFM ②	$supp$	384.56	42.00	588928.00	3373.29
	$\log(supp)$	8.59	5.39	19.17	11.72
BK ③	$supp$	23.89	11	7984	85.92
	$\log(supp)$	4.58	3.46	12.96	6.42

표 1. 실험 데이터 클릭수 통계정보

① <http://2015.recsyschallenge.com>

② <http://last.fm>

③ <https://snap.stanford.edu/data/loc-brightkite.html>

$$(\log_2(supp + 1))^\alpha$$

식 (3-8)

그림 6은 인기도 기반 샘플링과 로그 기반 인기도 샘플링을 적용했을 때 샘플링 될 확률을 예를 들어 설명한 것이다. 총 상품이 3개라고 가정하고, 상품1이 총 10000회 클릭, 상품2는 100회, 상품3은 1회 클릭 되었다고 했을 때, α 의 변화에 따라 아래와 같은 확률로 샘플링이 됨을 알 수 있다.

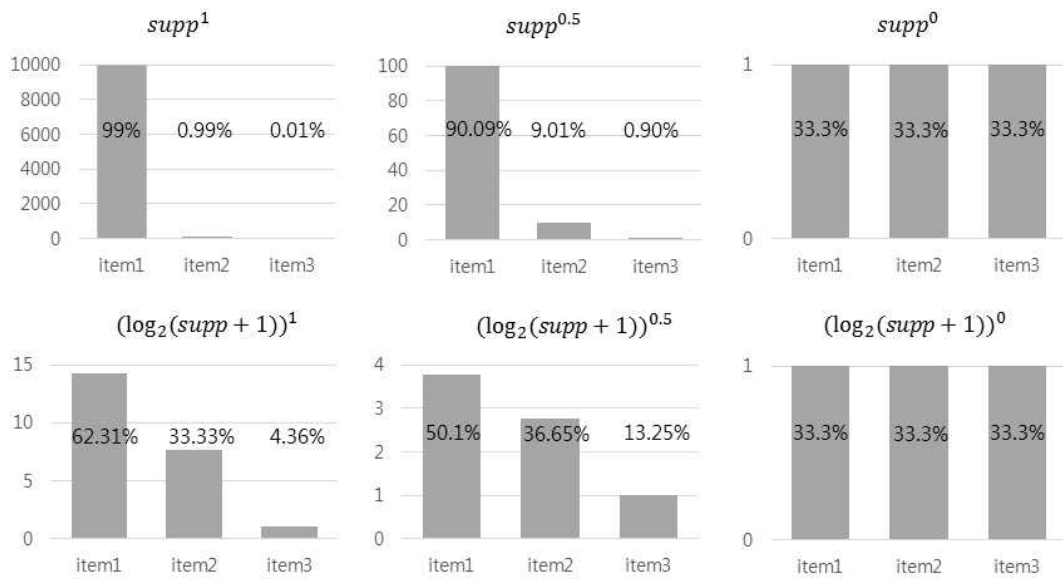


그림 6. 인기도 기반 샘플링과 로그 기반 인기도 샘플링의 예

2. BPR-max 가중치 적용

식 (2-16)의 BPR-max 손실 함수에는 네거티브 샘플의 점수가 가장 큰 값일 확률을 나타내는 소프트맥스 값 s_j 를 곱해주는 항이 있는데, 이 값이 학습이 진행됨에 따라 영향도가 낮아지고 그로 인해 학습 진행이 느려지는 현상이 발생한다. 그래서 학습 초기에는 소프트맥스 값을 그대로 유지해서 학습하고 학습이 진행될수록 소프트맥스 값에 가중치를 반영하도록 구성하였다. 식 (3-9)에서 α 가 새로 반영한 가중치 항이다.

$$s_j = \frac{e^{r_i \cdot \alpha}}{\sum_{j=1}^N e^{r_j \cdot \alpha}}, \text{ where } \alpha = \left\lfloor \frac{\text{epoch_idx}}{k} \right\rfloor + 1 \quad \text{식 (3-9)}$$

여기서 α 를 계산할 때 에폭(epoch)의 횟수를 나타내는 epoch_idx 값과 하이퍼파라미터 k 을 사용한다. k 는 α 값이 3이하가 되도록 하는 값 중 가장 좋은 성능이 나타나는 값을 사용하였다.

제 4 장 실험

제 1 절 실험 환경 및 실험 데이터

실험에 사용된 컴퓨터는 Intel Core(TM) i7-6700 CPU @ 3.4GHz에 메인메모리 4GB 용량이 장착된 머신이다. 그래픽 카드는 GeForce GTX980를 사용하였다. 실험에서 사용한 데이터는 총 3종류로 다음과 같다.

RSC15: Recsys Challenge 2015에서 제공한 데이터셋으로, 인터넷 구매 사이트의 클릭 정보이다. 세션, 상품 번호, 시간으로 구성되었다.

LFM: LastFM은 사용자가 실행한 노래의 아티스트, 실행 시간으로 구성된 데이터셋이다. 실험 시 사용자의 정보를 없애고 1일 이상의 접속 정보가 없을 시 새로운 세션으로 구분하여 세션 기반 데이터로 변환하여 구성하였다.

BK: Brightkite, 위치 기반 소셜 네트워크 서비스 데이터로 사용자, 장소, 시간으로 구성된 데이터셋이다. 실험 시에는 사용자 정보를 없애고 10일 이상 접속 정보가 없을 시 새로운 세션으로 구분하여 세션 기반 데이터로 구성하였다.

기존 연구[4] : 익명의 사용자 세션 아이디와 상품 아이디 정보, 상품 클릭 이벤트 시간 데이터를 사용하는 모델이다. 시간 데이터는 클릭 이벤트의 순서를 정렬하는 용도로만 사용하고 원 핫 인코딩을 통해 상품 정보를 표현하고 임베딩 레이어를 거쳐 게이트 순환 유닛, 피드포워드 레이어를 순차적으로 연결한 모델이다.

제안 모델 : [4]에서 제안한 모델에 시간 특성 정보를 추가로 반영하여 계절 변화, 요일의 변화(주중, 주말), 시간 대 등의 변화를 상품 추천에 활용한 모델이다. [4]에서 제안한 인기도 기반 샘플링과 BPR-max 손실 함수를 로그 기반 인기도 샘플링과 가중치를 반영한 BPR-max 손실 함수로 학습 효율을 개선한 모델이다.

데이터	학습 데이터		테스트 데이터	
	세션수	클릭수	세션수	클릭수
RSC15	7,966,257	31,637,239	15,324	71,222
LFM	1,000,642	18,530,958	5,605	277,301
BK	9,394	580,902	2,074	13,621

표 2. 실험 데이터 기본정보

표 2는 실험에 사용된 3가지 데이터 셋의 세션과 클릭 수를 보여주는 데이터로 데이터의 특성을 확인할 수 있다. 데이터는 학습 데이터와 테스트 데이터를 구분하여 표현하였다.

제 2 절 성능 측정 지표 및 실험 설정

세션 기반 추천 모델은 매 추천 수행 시, 한 화면에 보이는 정해진 숫자의 추천 리스트를 매회 계산하여 보여준다. 그리고 이 때 제시된 추천 리스트가 사용자에게 적합한 상품인지 여부를 판단하는 것이 성능 평가에 가장 중요한 요소이다. 본 논문에서는 이 목적에 부합하는 성능지표로 Recall@20을 사용하였다.

$$Recall@20 = \frac{\text{목표 상품이 Top20에 들어있는 횟수}}{\text{총 추천 횟수}} \quad \text{식 (4-1)}$$

매번 추천이 진행 될 때마다 점수가 높은 순으로 상품을 나열하여 순위를 매기는데, 현재시간 t 기준으로 사용자가 $t+1$ 시점에 선택한 상품이 추천 리스트 상위 20위 안에 들어있는지 여부를 카운트하여 식 (4-1)과 같이 계산한다.

실험 설정과 관련해서 경사 하강 최적화 알고리즘으로는(gradient descent optimization algorithm) Adagrad를 사용하고 게이트 순환 유닛은 단층에 유닛 사이즈를 100으로 설정하였다. [4]에서 최고의 성능을 보였던 설정을 동일하게 사용하여 샘플 크기를 2048로 설정하였다.

제 3 절 실험 결과 및 분석

본 논문에서 두 가지 실험을 비교하였다. 첫 번째로 로그 기반 인기도 샘플링과 기존 인기도 기반 샘플링의 실험 결과를 비교하였다. 두 번째로 [4]의 기존 모델의 성능과 제안 모델의 성능을 데이터와 모델을 바꿔가며 결과를 비교하였다.

1. 샘플링 성능 비교

[4]에서 제안한 인기도 기반 샘플링과 본 논문에서 제안하는 로그 기반 인기도 샘플링의 성능을 비교하기 위해 2장 2절에서 설명한 α 값을 변경해 가며 실험을 진행하였다. 데이터는 RSC15를 사용하였고, 그림 7과 표 3은 동일한 실험을 통해 도출된 결과를 나타낸다. 여기서 $supp$ 는 인기도 기반 샘플링을 $log(supp)$ 는 로그 기반 인기도 샘플링을 나타낸다.

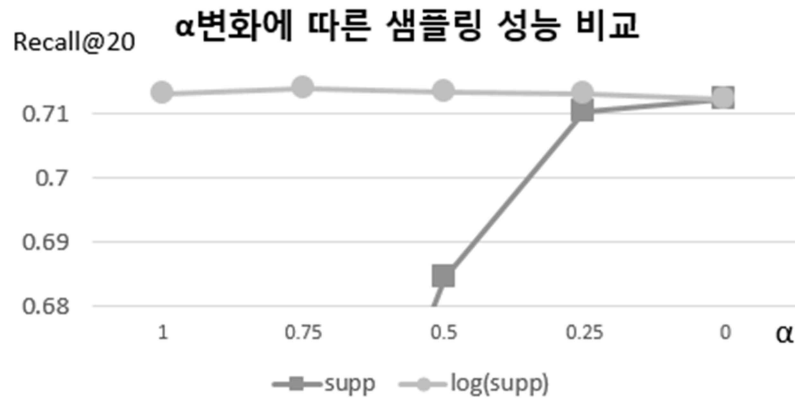


그림 7. 인기도 기반 샘플링과 로그 기반 인기도 샘플링 성능 비교

α설정 \ 샘플링	α = 1	α = 0.75	α = 0.5	α = 0.25	α = 0
$supp$	0.55518	0.62799	0.68449	0.71023	0.71219
$log(supp)$	0.71314	0.71395	0.71356	0.71315	0.71219

표 3. 인기도 기반 샘플링과 로그 기반 인기도 샘플링 성능 비교

실험 결과를 보면 인기도 기반 샘플링은 $\alpha = 0$ 의 값에서 가장 좋은 성능을 보여주고 로그 기반 인기도 샘플링에서는 $\alpha = 0.75$ 에서 가장 좋은 성능을 보여준다. 그리고 $\alpha = 0$ 일 때, 인기도 기반 샘플링과 로그 기반 인기도 샘플링이 동일한 값을 나타낸다. 그 이유는 $\alpha = 0$ 일 때는 균등 분포를 나타내기 때문이다. RSC15 데이터의 경우는 데이터 슢림 현상이 있어 [4]에서 제안한 인기도 기반 샘플링이 좋은 성능을 나타내지 않고, 오히려 균등 분포 샘플링을 사용했을 때 가장 좋은 성능을 보이는 것을 확인할 수 있다. 반면 로그 기반 인기도 샘플링은 α 값이 어떤 값을 갖더라도 균등 분포를 사용했을 때보다 좋은 성능을 보여준다는 점이 장점이다.

2. 모델 성능 비교

표 4는 RSC15, LFM, BK 세 개의 데이터셋을 대상으로 기존 연구 [4]와 본 논문에서 제안하는 시간 특성 데이터를 추가한 모델, 기존 연구에 로그 기반 인기도 샘플링만 추가한 모델, 기존 연구에 BPR-max 가중치 로직만을 추가한 모델, 마지막으로 시간 특성 데이터와 로그 기반 샘플링, BPR-max 가중치를 모두 적용한 모델, 총 5가지 모델의 성능을 실험한 결과이다. 여기서 RSC15 데이터셋은 15에폭, LFM은 20에폭, BK는 100에폭을 수행하는데, 매 에폭이 완료됐을 때 성능을 측정하고 그중 가장 높은 수치를 나타낸 값을 결과로 사용하였다. 각 모델 별 실험은 총 5차례 반복 수행한 결과의 평균을 최종 결과 값으로 사용하였다. 각 모델의 성능 비교는 Recall@20 결과를 이용하였으며, 기존 연구 결과를 제외한 나머지 모델들은 기존 모델에서의 성능보다 얼마만큼 성능이 향상되었는지 향상률을 함께 보여준다.

시간 특성 데이터를 추가하고 로그 기반 인기도 샘플링, BPR-max 가중치 모두를 반영한 최종 제안 모델이 모든 데이터셋에서 가장 좋은 성능을 보였다. 그리고 각각의 모델이 기존 연구보다 좋은 성능을 보임으로써, 본 논문에서 제안한 모델과 기법들이 모두 성능 향상에 기여하고 있다고 해석할 수 있다. RSC15의 경우는 기존 연구 모델에서 이미

모델	RSC15	LFM	BK
기존 연구[4]	0.71208	0.51400	0.25949
시간 특성 데이터 추가	0.71312 (+0.15%)	0.52232 (+1.62%)	0.30553 (+17.71%)
로그 기반 인기도 샘플링 추가	0.71390 (+0.26%)	0.52851 (+2.82%)	0.28054 (+8.11%)
BPR-max 가중치 추가	0.71281 (+0.10%)	0.51989 (+1.15%)	0.26681 (+2.82%)
최종 제안 모델	0.71437 (+0.32%)	0.53286 (+3.67%)	0.33430 (+28.83%)

표 4. 기존 연구와 제안 모델의 Recall@20 성능 비교

높은 수치의 성능을 보여주고 있어 큰 성능 향상을 확인할 수는 없었지만, 매 학습 결과가 기존 연구보다 좋은 성능을 보임을 확인할 수 있었다. LFM도 RSC15에 비해 높은 성능 향상률을 보였지만, BK 데이터셋에서 특히 좋은 성능을 보여주었다. 이는 BK 데이터셋이 시간 정보에 따른 사용자의 선택이 많은 영향을 받고 있음을 알 수 있다. BK 데이터셋은 시간과 위치 정보를 가지고 있는데, 식사 시간에는 식당을 추천하고 저녁 시간대에는 숙박 시설을 추천하는 것이 합리적이듯 시간 정보가 사용자의 선택에 큰 영향을 주는 것이다. 그러므로 시간의 변화에 따라 추천 상품의 변화가 큰 데이터에 대해서 제안 모델이 높은 성능을 보일 수 있음을 알 수 있다.

제 5 장 결론

본 논문에서는 세션 기반 추천 모델의 성능을 높이기 위해 시간 특성 데이터를 반영한 모델을 제시하였고, 로그 기반 인기도 샘플링 방법과 BPR-max 가중치 반영 방법을 제시하였다. 그리고 제시한 모델은 세션 기반 추천의 대표적인 모델 [4]와 실생활 데이터를 이용한 실험을 통해 성능 검증을 하였다. 제안 모델에서 제시한 시간 특성 데이터를 반영한 모델, 샘플링 기법과 BPR-max 가중치 기법을 개별로 적용했을 때 성능 개선 효과를 각각 비교하였다. 특히 3가지 다른 타입의 데이터셋을 바탕으로 실험을 진행하였는데, 시간대에 따른 사용자의 선호도가 많이 변하는 데이터셋에서 가장 좋은 성능을 보인 것을 확인하였다.

본 논문에서 제안한 기법들은 세션 기반 추천 모델뿐만 아니라 다른 유형의 추천 모델에서도 쉽게 적용해볼 수 있는 유용한 기법들이다. 본 논문은 Recsys Challenge에서 제공하는 것과 같이 제한적인 정보의 데이터셋을 기반으로 모델을 제안하였지만, 상품설명, 상품 리뷰, 상품 이미지 등을 추가로 제공하는 모델, [11], [12]와 같은 경우에서도 부분 모듈로 구성하여 적용해 볼 수 있는 구조이다. 본 논문에서 제안한 기법들은 앞으로 다양한 분야에 성능 향상을 위해 적용해 볼 수 있을 것으로 생각된다.

참 고 문 헌

- [1] Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. Item-based collaborative filtering recommendation algorithms. WWW, 2001
- [2] Hidasi, B., Karatzoglou, A., Baltrunas, L., & Tikk, D. Session-based recommendations with recurrent neural networks. ICLR, 2016
- [3] Twardowski, B. Modelling contextual information in session-aware recommender systems with neural networks. 10th ACM Conference on Recommender Systems, 2016
- [4] Hidasi, B., & Karatzoglou, A. Recurrent Neural Networks with Top-k Gains for Session-based Recommendations. arXiv preprint arXiv:1706.03847, 2017
- [5] Li, J., Ren, P., Chen, Z., Ren, Z., Lian, T., & Ma, J. Neural Attentive Session-based Recommendation. CIKM 2017
- [6] Bahdanau, D., Cho, K., & Bengio, Y. Neural machine translation by jointly learning to align and translate. arXiv:1409.0473, 2014.
- [7] Jeffrey D. Ullman Anand Rajaraman, Jure Leskovec. Mining of massive datasets, 2013.
URL: <http://infolab.stanford.edu/~ullman/mmds/ch9.pdf>

- [8] Hochreiter, Sepp, and Jürgen Schmidhuber. Long short-term memory. *Neural computation* 9.8, 1735–1780, 1997
- [9] Rendle, S., Freudenthaler, C., Gantner, Z., & Schmidt-Thieme, L. BPR: Bayesian personalized ranking from implicit feedback. *AUAI Press*, 2009.
- [10] Ko, Y. J., Maystre, L., & Grossglauser, M. Collaborative recurrent neural networks for dynamic recommender systems. *JMLR Vol. 63, No. EPFL-CONF-222477*, 2016.
- [11] Zhang, Y., Ai, Q., Chen, X., & Croft, W. B. Joint representation learning for top-n recommendation with heterogeneous information sources. pp. 1449–1458, *ACM* 2017.
- [12] Gu, Y., Zhao, B., Hardtke, D., & Sun, Y. (2016, April). Learning global term weights for content-based recommender systems. pp. 391–400, *WWW* 2017.
- [13] Pazzani, Michael J., and Daniel Billsus. Content-based recommendation systems. *The adaptive web*. Springer, Berlin, Heidelberg, 2007. 325–341.

Abstract

A Session-based Recommendation Model using Temporal Information and Popularity-based Sampling

Heewon Lee

Electrical and Computer Engineering

The Graduate School

Seoul National University

Session-based recommendation models are used to recommend items on various platforms, such as clothes on a small Internet shopping site, news on a news site, and videos on a video streaming site. Session-based recommendation models are useful when there is no past access information of an user. Such cases occur when an user first visits a site or does not log in. Traditional approaches utilize similarity-based models for recommendation since there is only the current short session information about the interaction between anonymous users and the products. However, such a similarity-based recommendation model has a drawback in that the recommendation is

based on only the last selected product. To alleviate the problem, recurrent neural network based models have recently been proposed and show good performances. Still, the existing works based on recurrent neural networks cannot utilize time feature information because they consider only the order of item selections.

In this paper, we propose a model that reflects users' preference changes over time, such as the seasonal changes or day of the week, by additionally reflecting the time information in the granularities of month, day, hour, minute and second to the neural network based model. In addition, we propose methods to increase the performance of the model by introducing a new loss function and a log based popularity sampling method. Furthermore, we show the effectiveness of our proposed model by conducting experiments with real-life data.

**keywords : recommendation system, recurrent neural network,
popularity based sampling, gated recurrent unit**

Student Number : 2017-21079

감사의 글

지난 2년 동안 무사히 석사과정을 마칠 수 있도록 도움주신 모든 분들께 감사의 인사를 드리고 싶습니다. 먼저 저의 지도 교수님이신 심규석 교수님께 감사의 말씀을 드립니다. 바쁘신 와중에도 연구에 있어 부족한 점, 보완할 점에 대해 꼼꼼히 방향을 제시해주시고 인생을 살아가는데 있어 중요한 마음가짐에 대해서도 자식처럼 아끼며 지도해주셔서 감사드립니다. 교수님의 열정적이고 성실하신 모습을 보며 많이 배울 수 있었고 연구실에서 보낸 시간은 저에게 큰 행운이었습니다. 그리고 바쁘신 와중에 귀한 시간을 내주시고 논문 심사위원을 맡아주신 김태환 교수님과 홍성수 교수님께도 진심으로 감사의 인사를 드립니다.

KDD 연구실 분들께도 감사의 인사드립니다. 연구실 첫 회식 자리에서 유머로 맞아주었던 진현씨, 연구실 소개 시간에 많은 도움을 주어 연구실에 들어오는데 큰 힘이 되었던 윤재씨와 광호씨, 방장으로써 든든한 모습의 우환씨, 새 방장으로써 굿은일도 마지않던 성웅씨, 적극적으로 먼저 다가와 도움을 주던 한준씨, 차분하게 꼼꼼하게 도와주는 수용씨, 옆자리에서 든든하게 힘이 돼주었던 대영씨, 유쾌함 긍정적인 마인드를 뽐어내던 동희씨, 진지함과 박학다식을 뽐내던 장혁씨, 수업도 함께 많이 듣고 많은 나이차에도 친구처럼 대해준 멋쟁이 동기 영준씨, 마지막 한학기 밖에 함께하지 못해 아쉬운 성욱씨와 영빈씨 모두 함께 해주어 감사합니다. 연구실에서 좋은 인연을 만들어가는 거 같아 정말 행복합니다.

SK하이닉스 분들께도 감사 인사드립니다. 학교에서 세미나가 있을 때 잊지 않고 저를 챙겨주시고 많은 조언해주셨던 송창록 전무님, 항상 밝은 웃음으로 맞아주셨던 이한주 상무님, 중간 중간 공부 잘하고 있는지 챙겨주셨던 강도욱 팀장님, 저의 빈자리를 채우느라 고생하신 품질시스템의 하성해 수석님, 박경석 수석님, 추정연 수석님, 공재홍 수석님, 김창준 수석님, 전유진 책임님, 전미경 책임님, 방진석 책임님, 진연하 책임님, 신은미 책임, 탁우영 책임, 김명환 선임, 김윤지 선임, 유진 선임, 장진주 선임, 최관용 선임, 홍선우 선임, 김상진 선임, 모든 회사가족 여러

분들에게 감사드립니다. 그리고 저를 지지해주시고 아낌없는 조언을 해주셨던 송재영 전 팀장님께도 감사 인사드립니다.

언제나 응원을 아끼지 않았던 저의 가족들에게도 감사드립니다. 항상 따뜻하게 웃어주시며 챙겨주시는 장인어른, 매주 바쁘신 와중에도 찾아와 주셔서 와이프 건강을 챙겨주시는 장모님, 언니와 형부한테 너무 잘하는 착한 우리 처제 모두 감사드립니다. 그리고 항상 저와 며느리 건강 챙겨주시고 밑반찬, 요리 할 것 없이 가득 챙겨다주시는 우리 어머니, 아버지, 집안에 경조사 굿은일 모두 도맡아 잘해주는 우리 형과 형수님 모두 감사드립니다. 마지막으로 튼튼이 키운다고 고생하고 있는 우리 와이프 박현경, 입덧 때문에 고생이 많은데도 옆에서 많이 도와주고 힘이 되어 주어 감사합니다. 우리 가족 모두 사랑합니다.

2018. 11. 이희원 드림.