

이중 DNN을 이용한 가맹점 추천 시스템 (DoubleDNN)

칼리나 바야르체체¹, 나광택², 이주홍³
 인하대학교 컴퓨터공학과
 e-mail: ¹kb0422.bk@gmail.com
²kwangteakna@gmail.com
³juhong@inha.ac.kr

Merchant Recommender System using Double DNN

Bayartsetseg Kalina, KwangTek Na, Ju-Hong Lee
 Dept of Computer Engineering, Inha University

요 약

은행과 신용카드 업계에 있어, 고객의 다음 신용 카드 사용처(다음 방문 가맹점)를 예측할 수 있다면 고객의 라이프 스타일을 파악 할 수 있으며, 여러 프로모션과 비즈니스 기회를 포착할 수 있어 매출 증대를 꾀할 수 있다. 우리가 제안하는 모델은 고객이 다음에 방문할 가맹점을 예측/추천하는 것을 목표로 한다. 가맹점 방문과 같이 순차적으로 발생하는 이벤트에는 노이즈가 있을 수 있다. 이 노이즈를 제거하기 위해 두 개의 신경망을 이용한 DoubleDNN을 제안한다. 실험은 BC카드사의 데이터분포를 따르는 인공 생성된 신용카드 사용내역 데이터를 이용하였으며, DoubleDNN은 기존의 다른 추천 모델보다 좋은 성능을 보였다.

1. 서론

추천 시스템은 기계학습 분야 중 실제 비즈니스에서 가장 성공적으로 널리 사용되는 응용 프로그램 중 하나이다. 특히 Collaborative Filtering 알고리즘은 추천 시스템에서 사용되는 주요 알고리즘 중 하나이며 구매 순서에 따라 다음 항목을 예측한다. 최신의 모델들은 대부분 심층 신경망을 기반으로 작동한다.

은행과 신용카드 업계에 있어, 고객의 다음 신용 카드 사용처 (다음 가맹점)를 예측할 수 있다면 고객의 라이프 스타일을 파악 할 수 있으며, 여러 프로모션과 비즈니스 기회를 포착할 수 있어 매출 증대를 꾀할 수 있다. 우리 모델 (DoubleDNN)은 고객이 다음에 방문할 가맹점을 예측/추천하는 것을 목표로 한다. 우리의 결과는 DoubleDNN이 다른 최신의 DNN 추천 모델보다 우수하다는 것을 보여준다. 모델을 학습 및 검증하는데 사용된 데이터는 실제 BC카드사의 데이터의 분포를 따르는 인공 데이터이다.

본 연구의 주된 공헌은 다음과 같다:

1. 우리는 원 데이터의 contingency table과 Iterative Proportional Fitting (IPF) 알고리즘을 사용하여 신용카드 사용내역 인공데이터를 생성했다.
2. 본 논문에서는 데이터 상에 노이즈가 있는 경우 두 개의 DNN 모델을 이용하여 이를 상쇄시킬 수 있는 새로운 아키텍처(DoubleDNN)를 제안한다.

논문의 나머지 부분은 다음과 같이 구성된다. 먼저 2장에서는 본 논문의 동기가 된 다른 관련 연구를 소개한다. 3장에서는 제안 된 모델을 소개하며, 4장에서는 우리의 실험방법과 실험 결과를 제시한다. 5장에서는 연구 결과를 요약한다.

2. 관련 연구

2.1. Neural Collaborative Filtering

Neural Collaborative Filtering (NCF) [2]는 세 개의 신경망을 이용하여 상품을 추천한다.

첫 번째 신경망은 활성화함수로 identity 함수를 이용한 Generalized Matrix Factorization (GMF)이다. 두 번째 신경망은 ReLU를 활성화함수로 이용하는 Multi-Layer Perceptron (MLP)이다. 이 두 신경망의 역할은 고객(상품)을 임베딩 벡터로 변환시키는 것이다. 변환된 각 임베딩 벡터는 결합되어 Neural Matrix Factorization (NeuMF)라 불리는 마지막 신경망의 입력으로 들어간다. NeuMF는 사용자가 구매할 상품을 예측한다.

훈련은 예측 값 \hat{y}_{ui} 과 타겟 값 y_{ui} 사이의 손실을 최소화함으로써 수행된다.

$$L = \sum_{(u,i) \in YU Y^-} w_{ui}(y_{ui} - \hat{y}_{ui})^2 \quad (1)$$

여기서 Y 는 관찰 된 상호 작용의 집합을 나타내며,

Y^- 는 관찰되지 않은 상호 작용의 집합을 나타낸다. w_{ui} 는 학습 인스턴스(u, i)의 가중치를 나타내는 하이퍼매개 변수이다.

2.2 Deep Neural Networks for YouTube Recommendations

YouTube Recommender System [1]은 사용자가 동영상 상을 시청한 후 그 사용자가 다음에 시청할만한 동영상을 추천하는 것을 목표로 한다.

[1]은 두 개의 심층망으로 구성된다. 하나는 후보 생성을 담당하고, 다른 하나는 생성된 후보의 순위를 매기는데 사용된다. 후보 생성 신경망은 사용자의 동영상 시청내역을 입력으로 받아 사용자가 시청할 만한 수백만 건의 동영상을 찾아낸다. 순위 신경망은 동영상과 사용자의 관계를 설명할 수 있는 더 많은 특징을 사용하여 학습된다. 순위 신경망의 입력은 후보 생성 신경망의 출력이다. 즉 후보 생성 신경망이 생성해낸 수백만 건의 동영상 중에서 수백 개의 동영상을 추출하여 순위를 매긴다.

2.3. 연구동기

2.1 장과 2.2 장에서 설명했던 것처럼 DNNs는 다음 상품추천을 위해 사용자와 상품과의 관계를 성공적으로 파악한다. YouTube Recommender System [1]에서는 다음 상품의 확률을 계산하기 위해 식 (2)를 사용한다.

$$P(i_t | (i_{t-1}, i_{t-2}, \dots, i_{t-k}), u_j) \quad (2)$$

하지만 순차적으로 발생(동영상 시청)하는 이벤트에는 노이즈가 있을 수 있다. 즉 i_{t-1} 과 i_t 의 연관 관계가 적어 다음 상품(가맹점) 추천 결과에 안 좋은 영향을 끼칠 수 있다. 본 논문에서는 위와 같은 상황을 극복하기 위한 DoubleDNN 모델을 제시한다.

3. 제안 방법

가맹점 추천 문제는 분류문제로 해석할 수 있다. 여기서 레이블은 고객이 다음에 방문할 가맹점이 된다. Softmax 계층은 출력의 확률을 계산하는데 사용되며 이는 다음에 방문할 가맹점의 확률이 된다.

본 논문에서는 두 개의 심층 신경망을 사용한다. 첫 번째 신경망은 주어진 고객과 가맹점 벡터에 대하여 해당 고객이 다음에 방문할 가맹점들의 확률을 계산한다. 여기서 가맹점 벡터는 t-1 시점과 t-2 시점의 가맹점 벡터의 평균을 이용한다. 고객 벡터와 가맹점 벡터는 word2vec를 이용해 생성한다.

$$P(i_t | i_{t-1}, i_{t-2}, u_j) \quad (3)$$

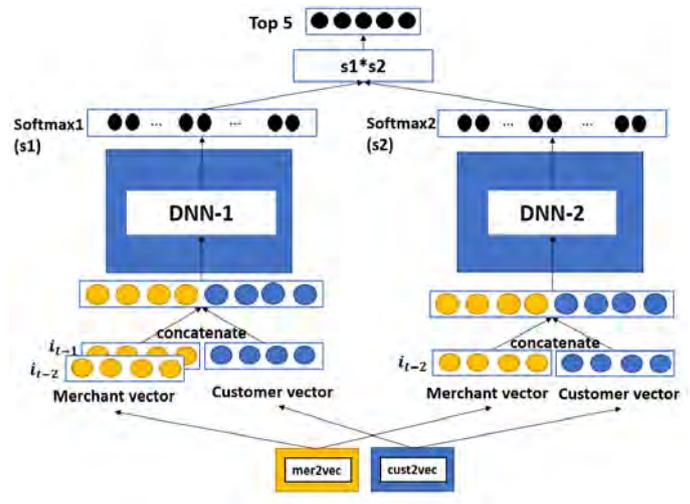
2.2절에서 설명한 것처럼 가맹점 i_{t-1} 은 노이즈로 작용

할 수 있다. 따라서 우리는 두 번째 심층 신경망을 이용하여 이를 상쇄시킨다. 이 신경망은 t-2 시점의 가맹점 벡터와 해당 고객의 벡터를 이용해 t-1 시점의 가맹점들의 확률을 계산한다.

$$P(i_{t-1} | i_{t-2}, u_j) \quad (4)$$

두 신경망의 출력인 가맹점 확률 벡터를 이용해 i_t 과 i_{t-1} 의 결합 확률 분포를 계산한다.

$$P(i_t, i_{t-1} | i_{t-2}, u_j) = P(i_t | i_{t-1}, i_{t-2}, u_j) * P(i_{t-1} | i_{t-2}, u_j) \quad (5)$$



<그림 1> DoubleDNN 모델

그림 1은 DoubleDNN 모델을 보여준다.

활성 함수로 hyperbolic tangent를 사용하였으며, 손실 함수로는 멀티 클래스 분류 문제에 많이 사용되는 Categorical cross-entropy를 이용한다. categorical cross-entropy는 수식 (6)과 같이 모델 확률 분포 q와 타겟 확률 분포 p간의 거리를 측정한다.

$$H(p, q) = - \sum_{i=1}^N p_i \log q_i \quad (6)$$

본 논문에서 제안하는 모델의 모델 확률 분포 q_i 는 softmax 계층에서 계산된다. Softmax 함수는 다음과 같다.

$$s_i = \frac{e^{out_i}}{\sum_{j=1}^N e^{out_j}} \quad (7)$$

여기서 out_i 은 출력 계층의 i 번째 output 이다.

4. 실험

4.1 인공데이터 생성

N개의 패턴으로 데이터를 표현하는 일반적인 방법은 contingency table을 이용하는 것이다. 우리는 contingency table이 N개 패턴의 빈도 분포를 보여 줄 수 있도록 구성하였다. 각각의 패턴에 4가지 범주 속성을 생성하였고 각 범주 속성은 고객 (12 그룹), 가맹점 (15 그룹), 거래 날짜 (12 범주), 금액 (6 범주)로 구성된다.

반응변수가 따로 없고 변수들의 범주에 따라 관측된 도수들로 이루어진 contingency table 형식의 데이터에 변수들간의 독립성이나 조건부 독립성 등의 관계를 탐색하는데 사용되는 로그 선형 모델을 사용했다.

본 논문에서는 BC카드사의 데이터분포를 따르는 신용카드 사용내역 데이터를 만들기 위해 iterative Proportional Fitting (IPF) [4] 알고리즘을 사용했다.

4.2 실험 데이터

모델의 학습과 검증을 위해 신용카드 사용내역의 합성

	고객 수	가맹점 개수	데이터 건수
신용카드 데이터 1	4,427	6,959	159,169

데이터를 사용하였다. 데이터에 대한 자세한 정보는 표 1과 같다.

<표 1> 실험 데이터 요약

각각의 고객이 마지막에 방문한 가맹점을 검증 셋으로 사용되었고 나머지 데이터는 훈련 셋으로 사용되었다.

4.3 평가 지표

각 고객에 대해 top-5 가맹점을 추천한다. 그리고 이에 대한 Precision, Recall, F1-score, Normalized Discounted Cumulative Gain(NDCG)를 평가 지표로 사용하였다. 각각의 수식은 다음과 같다.

$$precision@5 = \frac{\sum_{u \in U} |Y_u \cap \hat{Y}_u|}{5 * |U|} \tag{8}$$

$$recall@5 = \frac{\sum_{u \in U} |Y_u \cap \hat{Y}_u|}{|\hat{U}|} \tag{9}$$

$$F1 - score@5 = \frac{2 * precision * recall}{precision + recall} \tag{10}$$

$$NDCG@5 = \frac{DCG@5}{IDCG@5} = \frac{\sum_{j=1}^5 \frac{r_j}{\log_2(j+1)}}{\sum_{j=1}^5 \frac{\hat{r}_j}{\log_2(j+1)}} \tag{11}$$

여기서 U는 고객 집합이며, Y는 타겟 가맹점의 집합이

다. \hat{Y} 은 모델이 추천한 가맹점 집합이며, |U| 는 고객 수이다. r_j 는 j^{th} 가맹점에 대한 점수, \hat{r}_j 은 j^{th} 가맹점에 랭킹 점수이다.

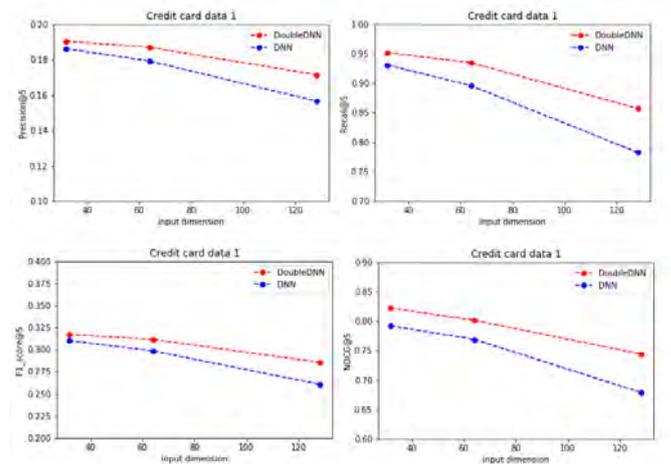
4.4 실험 결과

4.4.1 DNN 모델

가맹점 임베딩 벡터들의 평균 벡터와 고객 임베딩 벡터를 인풋으로 사용하여 고객이 다음에 이용할 가맹점을 예측한다. 가맹점 벡터의 평균은 시점 t-3에서 t-1까지의 가맹점 벡터의 평균을 이용한다. 삼층 레이어를 사용하였고 각 레이어의 노드 개수는 256, 512, 256개이다. 활성화함수(activation function)로 hyperbolic tangent 함수, 손실 함수(loss function)로 수식 (6) cross-entropy를 이용하였다.

4.4.2. DoubleDNN 모델

네트워크 1 (DNN-1)의 구조는 삼층 레이어를 사용하였고 각 레이어의 노드 개수는 256, 512, 256개이다. 네트워크 2 (DNN-2)는 이층 레이어를 사용하였고 노드 개수는 각각 256, 512이다. 활성화함수(activation function)와 손실 함수(loss function)는 DNN 모델과 동일하다.



<그림 2> 실험 결과

그림 2는 본 논문에서 제안하는 모델(DoubleDNN)이 다른 최신 DNN 모델 보다 성능이 우수하다는 것을 보여준다. 가맹점 벡터의 차원이 32, 64, 128로 변함에 따라 DoubleDNN은 항상 DNN보다 성능이 우수하다.

5. 결론

본 논문에서는 가맹점 방문과 같이 순차적으로 발생하는 이벤트에는 노이즈가 있어서 제거하기 위해 두 개의 신경망을 이용한 DoubleDNN 모델을 제안하였다. 실험 결과는 제안한 모델은 최신 DNN 모델에 비해 더 좋은 성능을 보이는 것을 확인하였다.

가맹점 벡터의 차원에 따라 평가 지표들이 바뀌었다. 임베딩 벡터의 차원이 커지면 오버피팅이 발생하고 추천 모델의 성능이 저하될 수 있는 것을 알아냈다.

향후 연구에서는 장기간의 순차적 종속성을 포착하고

시계열과 같은 다른 정보를 모델에 통합할 수 있는 Recurrent Neural Network(RNN)과 같은 다른 신경망을 사용해서 연구할 예정이다.

감사의 글

본 연구는 2018년도 산학협력 기술개발사업(첫걸음) "고객 마케팅을 위한 딥러닝 기반의 고객 세분화 및 추천알고리즘 개발" (과제번호-S2599930) 지원으로 수행되었습니다. 연구수행을 위해 필요한 카드 사용 데이터의 접근에 도움을 주신 BC카드사에 감사드립니다.

참고문헌

[1] Paul Covington, Jay Adams and Emre Sargin. "Deep Neural Networks for YouTube Recommendations", In RecSys'16 Proceedings of the 10th ACM Conference on Recommender Systems, September 2016. pp.191-198

[2] Xiangnam He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, Tat-Seng Chua. "Neural Collaborative Filtering", In WWW'17 Proceedings of the 26th International Conference on World Wide Web, April 2017. pp.173-182.

[3] Hélène Massam, Jinnan Liu and Adrian Dobra. "A Conjugate Prior for Discrete Hierarchical Log-linear models", In The Annals of Statistics Vol.37, No.6A, 2009. pp.3431-3467.

[4] W. Edwards Deming and Frederick F.Stephan. "On a Least Squares adjustment of a sampled frequency table when the expected marginal totals are known", In The Annals of Mathematical Statistics. Vol.35, 1940. pp.615-630.

[5] Hao Wu, Yue Ning, Prithwish Chakraborty, Jilles Vreeken, Nikolaj Tatti, Naren Ramakrishnan. "Generating Realistic Synthetic Population Datasets", arXiv:1602.06844v3 [cs.DB] 25 Feb 2016.

[6] Agresti, Alan. "Categorical Data Analysis", Third Edition. Wiley. 2012.

[7] Suvash Sedhain, Aditya Menon, Scott Sanner, Lexing Xie. "AutoRec: Autoencoders Meet Collaborative", In WWW'15 Companion Proceedings of the 24th International Conference on World Wide Web, May 2015. pp.111-112.

[8] Aaron van den Oord, Sander Dieleman, Benjamin Schrauwen. "Deep content-based music recommendation", In NIPS'13 Proceedings of the 26th International Conference on Neural Information Processing Systems. Vol.2, December 2013. pp.2643-2651.

[9] Hao Wang, Naiyan Wang and Dit-Yan Yeung. "Collaborative Deep Learning for Recommender

Systems", In KDD'15 Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 2015. pp.1235-1244.

[10] Tomas Mikolov, Kai Chen, Greg Corrado and Jeffrey Dean. "Efficient Estimation of Word Representations in Vector Space", arXiv:1301.3781v3 [cs.CL] 7 Sep 2013.

[11] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado and Jeffrey Dean. "Distributed Representations of Words and Phrases and their Compositionality", In NIPS'13 Proceedings of the 26th International Conference on Neural Information Processing Systems Vol.2, December 2013. pp.3111-3119.