



## 그래프 임베딩을 사용한 효과적인 상품 추천 방법

박성수 · 이건창\*  
성균관대학교 경영대학

# A noble approach to effective Recommender System using Graph Embedding

Sung Soo Park · Kun Chang Lee\*

SKK Business School, Sungkunkwan University

### [요 약]

본 연구는 그래프 임베딩 방법을 사용한 효율적인 상품 추천을 제시하는 것을 목적으로 한다. 제안된 방법은 구매자-상품 네트워크를 Skip-gram 기반의 그래프 임베딩인 DeepWalk로 학습하고, 구매자-상품의 유사도를 사용하여 상위 10개의 상품을 추천한다. 인터넷에 공개된 실제 상품 구매 이력 데이터로부터, 구매자 100명이 6개월간의 구매한 상품 2만개의 데이터를 사용한 실험으로 상품 추천 성능을 평가하였다. 실험결과 본 연구의 제안 방법이 이전의 협업 필터링에 비해 더 좋은 상품 추천 성능을 보였다. 특히, 상품 구매 빈도를 네트워크 간선의 가중치로 사용하는 경우 통계적으로 유의한 차이를 나타냈다. 본 연구의 결과는 온라인 쇼핑몰이나 마트, 백화점 등에서 보다 성능 높은 상품 추천을 할 수 있는 방안을 제시한다.

### [Abstract]

The purpose of this study is to propose efficient product recommendation using graph embedding method. The proposed method learns the buyer-product network with DeepWalk, which is a graph-based graph embedding, and recommends the top 10 products using the buyer-product similarity. From the actual product purchase history data published on the Internet, the product recommendation performance was evaluated by an experiment using 100 buyers' data of 20,000 products purchased for 6 months. Experimental results show that the proposed method has better product recommendation performance than previous collaborative filtering. Especially, when the frequency of purchasing goods is used as the weight of network trunk, statistically significant difference is shown. The result of this study suggests ways to recommend products with higher performance at online shopping, mart and department store.

**색인어** : 상품 추천, 그래프 임베딩, DeepWalk, 협업 필터링, 구매자-상품 유사도

**Key word** : Product recommendation, Graph embedding, DeepWalk, Collaborative filtering, Buyer-product similarity

<http://dx.doi.org/10.9728/dcs.2018.19.10.1919>



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Received** 05 September 2018; **Revised** 05 October 2018

**Accepted** 22 October 2018

\*Corresponding Author; Kun Chang Lee

**Tel:** +82-2-760-0505

**E-mail:** kunchanglee@gmail.com

## I. 서론

최근의 온라인상의 구매자가 인지하지 못하는 사이에, 거의 모든 웹 페이지의 한 부분에서 추천 광고나 온라인 상점의 상품 권유와 같은 추천 시스템을 거의 매일 접하게 된다. 갈수록 증가하는 다양한 상품과 서비스에 대한 정보 과부하 문제를 해결할 수 있는 방법으로 추천 시스템은 지난 수십 년 동안 많은 주목을 받았다. 2006년과 2009년 사이의 넷플릭스 경진대회(Netflix Prize)와 추천 시스템(Recommender Systems) 컨퍼런스에서 조직된 몇 가지 연구 과제는 기계 학습 및 데이터 마이닝 연구 분야의 수많은 연구자들로부터 주목 받았다[22].

추천 시스템은 정보 자원과 구매자를 조화하는 중요한 역할을 담당한다[1]. 지난 수년간 협업 필터링 기법, 내용 기반 기법, 하이브리드 기법 등 다양한 추천 알고리즘이 개발되었다[2]. 협업 필터링 방법은 구매자의 과거 행동뿐만 아니라 다른 유사한 구매자의 구매 의사 결정을 기반으로 모델링 한다. 내용 기반 방법은 비슷한 기능을 가진 새로운 상품을 추천하기 위해 상품의 중요한 기능 집합을 추출해서 모델링한다. 이 두 가지 유형의 방법을 혼합하는 방법을 하이브리드 방법이라 한다.

이러한 추천 방법은 어느 정도 효과적이지만, 이러한 접근법은 다음과 같은 문제가 있다. 기존 추천 알고리즘이 유사도 계산의 정확성을 높이는데 한계가 있다. 예를 들어, 고전적인 상품 기반 KNN 추천 알고리즘의 효율성은 KNN 그래프의 행렬 인수 분해는 시간과 비용이 높아 제한적인 유사도를 계산한다[3]. 따라서 추천의 성능을 높이기 위한 유사도 계산에 보다 효율적이고 성능 높은 방법을 확보하는 것은 중요한 문제이다.

본 연구는 전통적인 협업 필터링 기반의 상품추천의 한계를 극복하고자 최근 활발한 연구가 이루어지는 딥 러닝 기반의 그래프 임베딩 방법을 적용하여 보다 정확도 높고 효과적인 상품 추천 방안을 제시하는 것을 목적으로 한다. 본 연구는 다음과 같은 연구 문제(Research Question)를 다룬다.

RQ: 상품-상품 네트워크에 대해 그래프 임베딩 기반의 상품 추천이 협업 필터링 방법의 상품 추천보다 성과가 좋을 것이다.

## II. 관련 연구

### 2-1 그래프 임베딩

그래프 임베딩의 목적은 고차원의 데이터 셋을 저차원의 벡터 공간에 임베딩하여 표현하는 것이다[4]. 그래프 임베딩과 관련한 많은 연구들이 수행되었으며, 각각의 접근법들은 서로 다른 특징과 효과를 가지고 있다[5]. 본 절에서는 전통적으로 연구되던 그래프 임베딩 방법부터 최근의 그래프 임베딩 방법까지를 고찰한다.

전통적인 그래프 임베딩 방법(Classical Graph Embedding Methods)은 원래 차원 축소(dimension reduction) 기법으로 연구가 진행되었다. 그래프는 일반적으로 이미지 데이터와 같은 데이터 셋으로 표시된 자질(Feature)로부터 생성된다. 그래프 임베딩의 대표적인 방법인 Isomap은 네트워크 상의 노드  $i$ 와  $j$

에 대한 연결 그래프 생성하고[6], 다차원 스케일링(MDS; Multi-Dimensional Scaling) 방법을 적용하여 노드  $i$ 에 대한 좌표 벡터  $u_i$ 를 얻는다. Isomap은 저차원 공간에 노드 쌍의 최단 거리를 거의 정확하게 보존하는 노드  $i$ 의 표현을 학습한다. Isomap은 노드 쌍의 최단 경로의 계산으로 인한 복잡성의 증가하는 것이 문제점으로 지적된다. 이를 극복하고자, 넓게 분리된 항목 사이의 쌍 방향 거리를 추정할 필요가 없도록 LLE(Locally linear embedding) 방법이 제안되었다[7]. LLE는 각 노드와 그 이웃이 여러 겹의 로컬 선형 경로 위에 있거나 그 근처에 놓여 있다고 가정한다. 부분적인 기하학적 구조를 특성화하도록 재구성하고, 저차원 공간에서 국소 선형 복원(locally linear reconstruction)을 기반으로 한 이웃 보존 매핑(neighborhood-preserving mapping)을 구성한다. 라플라시안 아이젠맵(LE; Laplacian eigenmaps)은 KNN을 사용하여 그래프를 구성하고[8], 히트 커널(Heat Kernel)을 사용하여 그래프에서 노드  $i$  및  $j$ 의 가중치  $W_{ij}$ 를 선택한다[9]. 또 다른 방법으로는 비선형 LE의 선형 근사법인 지역 보존 표현(LPP; locality preserving projection)이 제안되었다[10]. 이러한 그래프 임베딩 방법은 주로 그래프 재구성(graph reconstruction)을 목표로 한다. 또한, 그래프 임베딩은 대부분 간선 표현에 의해 인코딩된 노드들 사이의 인접도가 원래의 자질 공간에서 잘 정의된 자질 표현 데이터 셋으로 구성된 그래프에서 동작한다. 네트워크 재구성을 위해 학습된 그래프 임베딩 공간은 네트워크 추론(network inference)에 반드시 좋은 것은 아니다.

그래프 임베딩에 관한 최근 연구동향은 네트워크 구조를 보존하고 네트워크의 속성을 임베딩 공간에 저장하는 것이다. 네트워크 구조는 1차 구조(first-order structure)와 2차 구조(second-order structure) 및 커뮤니티 구조(community structure)와 같은 고차원 구조(higher-order structure)가 포함된다. 네트워크 구조는 서로 다른 세분성(granularities)으로 나타나는 서로 다른 그룹으로 분류될 수 있다. 최근의 그래프 임베딩은 이러한 네트워크 구조적인 특성을 유지하는 방법들이 많이 연구되고 있다. 그래프 임베딩에서 일반적으로 구현되는 네트워크 구조는 인접 구조(neighborhood structure), 상위 노드 근접성(high-order node proximity) 및 네트워크 커뮤니티(network communities)가 포함된다. 인접 구조 및 상위 노드 근접성 관련 그래프 임베딩의 그래프 임베딩 방법들은 DeepWalk로부터 시작되었다. DeepWalk는 노드의 인접 구조를 보존할 수 있는 네트워크의 노드 표현을 학습하기 위해 제안되었다[11].

DeepWalk는 짧은 무작위 걷기(Random Walk)에 나타나는 노드의 분포가 자연 언어로 된 단어의 분포와 유사하다는 것을 발견했다. 이러한 발견에 영감을 얻어 널리 사용되는 단어 표현 학습 모델인 Skip-Gram 모델을 DeepWalk에 적용하여 노드의 표현을 학습한다[12]. 구체적으로 DeepWalk는 그림 1.과 같이 일련의 걷기 순서(Walking Sequence)를 생성하기 위해 네트워크 상에서 절단된 랜덤 워크(Truncated Random Walk)를 채택한다. Skip-Gram에 따라 각 걷기 순서  $s = \{v_1, v_2, \dots, v_s\}$ 에 대해

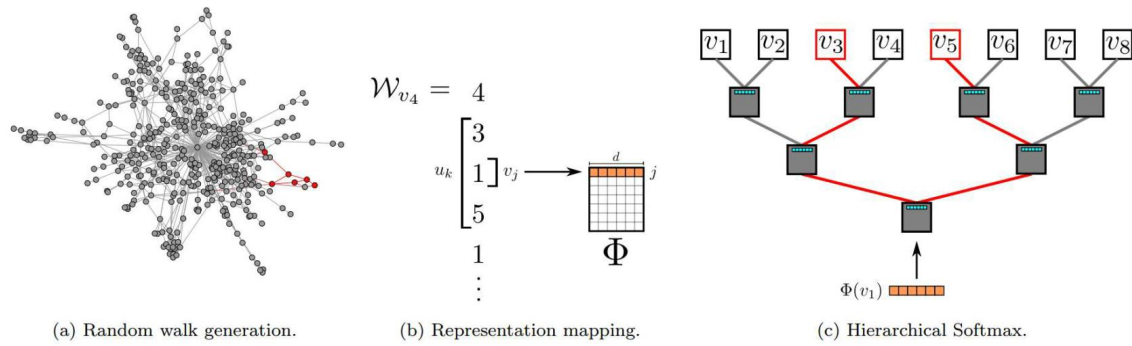


그림 1. DeepWalk 개요[11]  
Fig. 1. Overview of DeepWalk[11]

Deep walk는 식(1)과 같이 걷기 순서에서 노드  $v_i$ 의 이웃 노드의 확률을 최대화하는 것을 목표로 한다.

$$\max_{\Phi} \Pr(\{v_{i-w}, \dots, v_{i+w}\} \setminus v_i | \Phi(v_i)) = \prod_{j=i-w, j \neq i}^{i+w} \Pr(v_j | \Phi(v_i)) \quad (1)$$

여기서  $w$ 는 윈도우 크기이고,  $\Phi(v_i)$ 는 현재의  $v_i$  표현이며,  $v_{i-w}, \dots, v_{i+w} \setminus v_i$ 는  $v_i$ 와 연결된 지역 맥락 노드(local context nodes)이다. 마지막으로 계층적 소프트 맥스(hierarchical soft-max)를 사용하여 임베딩을 효율적으로 추론한다[13].

많은 네트워크 임베딩 방법은 잠재적인 저차원 공간에서 인접 구조, 고차원 근접 및 커뮤니티 구조를 포함하여 노드의 로컬 구조를 보존하는 것을 목표로 하는 다양한 연구들이 수행되고 새로운 방법들이 제시되면서 네트워크 추론에 더 많은 관심을 기울이고 있다. 또한 최근 그래프 임베딩 방법은 주로 소셜 네트워크(social networks), 생물 네트워크(biology networks) 및 전자 상거래 네트워크(e-commerce networks)와 같이 자연스럽게 형성된 네트워크에서 적용되는 것을 목표로 한다. 하지만, 많은 그래프 임베딩 연구들은 현재까지 연구용 데이터에만 적용이 되었고, 그래프 임베딩의 효과를 실제 비즈니스에 적용하기 위한 추가적인 연구들이 필요한 시점이다.

### 2-2 상품 추천

추천 시스템의 초기 개념은 다양하고 많은 양의 문서들을 필터링 해서 자신의 선호도에 적합한 문서를 찾기위한 방법으로 제안되었다[14]. 이후 Amazon.com과 같은 전자 상거래 기업들을 중심으로 추천 시스템을 적용하여 온라인에 게시된 수많은 상품과 구매자를 연결하여, 상품 판매의 효과 증대와 고객 충성도를 높이는 것을 목적으로 사용되었다[15]. 추천 시스템의 정의는 구매자의 충분한 경험 없이 대안에 대한 선택을 돕도록 구매자에게 상품을 제안해 어떤 상품을 구매해야 할지 의사 결정을 지원하는 시스템이라 한다[16].

추천 알고리즘은 크게 세 가지로 크게 나누어진다. 협업 필터링은 가장 성숙되고 널리 사용되는 방법으로[17], 구매자가 입력한 상품의 평점, 태그 등의 명시적인 관련성 지표(relevance indications) 또는 읽기, 다운로드와 같은 구매자의 행위로 부터 계산된 암시적 관련성 지표를 기반으로 구매자-구매자 또는 상품-상품 간 유사성(Similarity)를 인식하는 것이다[18]. 일반적인 구매자 기반 협업 필터링 알고리즘은 먼저 공통 상품의 중복되는 상호 작용이나 비슷한 등급을 기준으로 유사한 구매자를 식별한다. 그런 다음 유사한 구매자의 선호도를 기반으로 추천 사항을 만든다. 상품 기반 추천 알고리즘은 상품 간 유사성을 분석 한 다음 이러한 유사한 상품을 사용하여 추천 할 상품 집합을 식별한다. 또 다른 방법인 내용기반 필터링(Content-based filtering)은 구매자가 관심 있어 하는 상품의 내용을 기반으로 추천을 수행한다. 즉, 내용기반 필터링 추천 시스템은 구매자의 선호도가 높았던 상품들에 대한 공통적인 특성을 생성하고, 이를 상품 프로파일과 비교하여, 여러 상품들 중에서 구매자 선호도 특성과 가장 유사한 특성을 갖는 상품을 구매자에게 추천하는 방법이다[19]. 하이브리드 추천 시스템은 협업 필터링과 내용 기반 필터링 등의 다양한 추천 기술을 결합하여 더 적은 단점과 더 나은 추천 성능을 얻는 방법이다[17].

다양한 추천 시스템 중 협업 필터링이 구매자의 성향과 유사한 상품을 추천하는 방법으로 실무에서도 우수성을 인정받아 여러 기업에 적용되고, 학계를 중심으로 관련된 연구가 다양하게 진행되고 있다[20]. 일반적으로 구매자 기반 협업 필터링을 이용한 목표 구매자에 대한 추천 목록 생성 절차는 다음과 같다.

- 1단계 : 구매자의 구매이력이나 평점(Rating)을 기반으로 구매자-상품 행렬(Buyer-product matrix)을 구성한다.
- 2단계 : 같은 상품에 대한 평점을 기준으로 다른 구매자들의 평점과 비교하여 구매자간 유사도를 구하고, 이를 바탕으로 인접 집단(Neighbor group)을 구성한다. 이 단계에서 인접 집단을 파악하기 위한 유사도 계산 방법으로는 코사인 유사도(Cosine similarity)나 피어슨 상관계수(Pearson Correlation Coefficient) 등이 사용된다.

- 3단계 : 인접 집단을 기반으로 상품들에 대한 예상 선호도를 계산하여 추천 목록을 생성한다.

1단계에서 생성되는 구매자-상품 행렬은 그래프를 표현하는 인접행렬과 같은 형태이다. 구매자-상품 행렬은 고차원이며 희소성이 높은 행렬로 이것을 그래프 임베딩 방법으로 저차원의 벡터로 변환해서 보다 효율적인 상품 추천이 가능하다.

### III. 연구 방법

#### 3-1 실험 절차

본 연구는 기존의 협업 필터링 추천 시스템에 그래프 임베딩을 결합하여 추천의 성능을 높이는 것을 목적으로 그림 2과 같은 실험 절차로 연구한다.

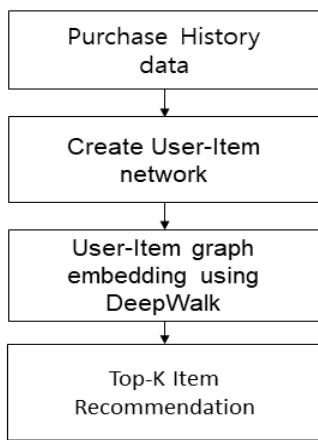


그림 2. 그래프 임베딩 기반 상품 추천 블록 다이어그램  
 Fig. 2. The block diagram of product recommendation based on graph embedding

- 1단계: 본 연구의 목적에 따라 실험용 데이터를 준비한다. 연구 목적을 확인할 수 있는 데이터를 Kaggle에 공개된 구매 이력 데이터<sup>1)</sup>를 다운로드 받았다. 총 구매 이력 데이터는 약 3억 5천만 건으로, 실험을 위하여 무작위 추출한 구매고객 100명에 대한 2012년 11월부터 2013년 4월까지 6개월 간의 구매이력 약 4만 4천 건을 추출하였다. 재구매, 중복구매를 제외하고 데이터 셋에 포함된 상품의 고유개수는 20,945개 이다.

2단계: 거래 데이터 셋에서 구매고객과 상품에 대한 구매 빈도, 구매 수량, 월 평균 구매 수량에 대한 변수를 생성한다. 생성된 데이터의 예시는 표 1, 표 2과 같다.

표 1. 상품 구매 수량 데이터 셋 예시

Table. 1. Example of quantity of product purchase data set

Customer	Product Code	QTY
108664900	4_10_1010_105150050	1
108664900	4_10_1011_104300040	1
108664900	4_29_2904_101530010	16
1127710454	214_26_2620_101370010	1
1127710454	214_26_2620_1072652878	7

표 2. 월 평균 상품 구매 수량 데이터 셋 예시

Table. 2. Example of Monthly average purchase quantity data set

Customer	Product Code	QTY
108664900	4_10_1010_105150050	0.166
108664900	4_41_4107_102113020	0.333
108664900	4_41_4107_102400020	0.5
1127710454	214_26_2620_1072652878	1.5
1127710454	214_29_2903_105100050	2.5
1127710454	214_32_3201_105010050	0.333

- 3단계: 2단계의 데이터 셋은 그래프 임베딩의 입력 데이터로 사용한다. 예를 들어, 그래프 임베딩 관점에서는 구매 고객과 상품 네트워크의 간선 가중치로 볼 수 있다. 그래프 임베딩 방법은 Perozzi가 제안한 DeepWalk를 사용한다. 각 노드 당 랜덤 워크 횟수는 10회, 워크의 길이는 80회, 생성할 벡터 차원의 수는 128차원으로 설정한다.

- 4단계: 생성된 그래프 임베딩된 자질로부터 구매고객-상품 간의 코사인 유사도를 계산하고, 협업 필터링에 입력하여 상품 추천을 실행한다. 입력데이터를 7:3으로 학습 데이터 셋과 검증 셋으로 분할하고, 학습 데이터 셋으로 만들어진 추천모형으로 검증 데이터 셋에 대한 상위 10개의 상품을 추천한다. 이후 검증 데이터 셋에서 추천 성공 여부에 대한 결과를 판단한다.

#### 3-2. 성능 측정

본 연구의 추천 시스템의 성능 평가는 Top-N 추천 성능 측정 방법을 사용한다. 이 방식은 예를 들어 상위 10개의 추천 상품의 목록을 제공하고, 구매자가 선호할 것으로 기대되는 상품이 추천되었는가를 평가하는 기법이다. 즉, 구매자가 상품을 선택하는 것을 얼마나 잘 지원하는가를 평가하게 된다[21]. Top-N 추천 성능 측정은 정확도(Precision), 재현율(Recall), F 척도(F-measure)의 지표가 사용된다. 본 연구의 추천 시스템의 성능 평가는 정확도와 재현율, 그리고 정확도와 재현율의 기하 평균인 F1 척도를 사용한다. 정확도, 재현율, F1척도의 공식은 각각 식 (2), (3), (4)와 같다.

1) <https://www.kaggle.com/c/acquire-valued-shoppers-challenge/data>

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1 = \frac{2 \times (Precision + Recall)}{(Precision \times Recall)} \quad (4)$$

여기서 TP(True Positive)는 상품 추천된 상품을 구매한 건수이고, FP(False Positive)는 추천된 상품을 구매하지 않은 건수이다. TN(True Negative)는 추천되지 않은 상품을 미구매한 건수이고, FN (False Negative)은 추천되지 않은 상품을 구매한 건수이다.

#### IV. 연구 결과

실험의 결과는 표 3에서 표 5와 같다. 구매 빈도를 사용한 상품추천의 결과(표 3)는 그래프 임베딩 기반의 상품 추천의 F1이 4.99에서 5.44로 나타났고, 그래프 임베딩을 적용하지 않은 상품추천은 0.31에서 1.91로 나타났다. 두 방법 간의 추천 정확도 차이는 3.18에서 5.14가 나오고 있으며, 두 방법 간의 유의성 차이가 유의한 것(p<.0001)으로 나타났다.

구매 수량을 사용한 상품 추천의 결과(표 4)는 그래프 임베딩을 사용하지 않은 상품 추천의 F1이 0.03에서 0.83이고, 그래프 임베딩을 사용한 상품 추천이 0.03에서 1을 보였다. 전반적으로 그래프 임베딩을 사용한 상품 추천의 성과가 약간 더 높지만 추천 상품의 수가 두 개인 경우는 그래프 임베딩을 사용한 추천이 더 낮았다. 두 방법 간의 정확도 차이는 -0.04에서 0.15로 나타났지만 그래프 임베딩을 사용한 상품 추천은 추천 상품의 수가 증가할수록 추천의 정확도가 증가하는 것을 보인다. 두 방법 간 추천 정확도의 유의하지 않은 것으로 나타났다(p=0.667).

월 평균 구매 수량을 사용한 상품 추천 결과는 표 5와 같다. 그래프 임베딩을 사용하지 않은 상품 추천의 F1이 0.03에서 0.44이고, 그래프 임베딩을 사용한 상품 추천이 0.07에서 0.54를 보였다. 전반적으로 그래프 임베딩을 사용한 상품 추천의 성과가 약간 더 높지만 추천 상품의 수가 세 개인 경우는 그래프 임베딩을 사용한 추천이 더 낮았다. 두 방법 간의 정확도 차이는 -0.01에서 0.14로 나타났지만, 그래프 임베딩을 사용한 상품 추천은 추천 상품의 수가 증가할수록 추천의 정확도를 평가하는 지표인 F1이 증가하는 패턴을 보인다. 두 방법 간 추천 정확도의 차이는 유의하지 않다(p=0.374).

표 3. 구매 빈도를 사용한 상품 추천 비교

Table 3. Compare product recommendations using purchase frequency

Cut off	Traditional CF (1)			Graph Embedding CF (2)			Diff F1 (2)-(1)
	Precision	Recall	F1	Precision	Recall	F1	
1	11.39	0.16	0.31	<b>16.25</b>	<b>3.27</b>	<b>5.44</b>	5.14
2	<b>11.39</b>	0.31	0.60	10.63	<b>3.41</b>	<b>5.16</b>	4.57
3	<b>10.12</b>	0.41	0.79	8.75	<b>3.56</b>	<b>5.06</b>	4.27
4	<b>8.86</b>	0.48	0.92	7.81	<b>3.72</b>	<b>5.04</b>	4.12
5	<b>8.35</b>	0.56	1.05	7.75	<b>3.94</b>	<b>5.22</b>	4.17
6	<b>8.01</b>	0.65	1.21	6.88	<b>4.05</b>	<b>5.10</b>	3.89
7	<b>7.78</b>	0.74	1.34	6.25	<b>4.13</b>	<b>4.97</b>	3.63
8	<b>8.07</b>	0.94	1.69	6.25	<b>4.43</b>	<b>5.19</b>	3.50
9	<b>7.88</b>	1.03	1.82	5.69	<b>4.47</b>	<b>5.01</b>	3.18
10	<b>7.59</b>	1.10	1.92	5.75	<b>4.65</b>	<b>5.14</b>	3.22
Prob	0.0001						

표 4. 구매 수량을 사용한 상품 추천 비교

Table 4. Product Recommendations Using Purchased Quantities

Cut off	Traditional CF (1)			Graph Embedding CF (2)			Diff F1 (2)-(1)
	Precision	Recall	F1	Precision	Recall	F1	
1	1.00	0.01	0.03	1.00	<b>0.02</b>	0.03	0.01
2	<b>3.00</b>	<b>0.10</b>	<b>0.19</b>	2.50	0.07	0.14	-0.04
3	3.00	0.15	0.29	<b>3.33</b>	0.19	<b>0.37</b>	0.08
4	3.50	0.26	0.48	<b>3.50</b>	<b>0.26</b>	<b>0.48</b>	0.00
5	3.80	0.35	0.65	<b>4.20</b>	<b>0.38</b>	<b>0.70</b>	0.05
6	3.33	0.37	0.67	<b>3.67</b>	<b>0.40</b>	<b>0.72</b>	0.05
7	3.14	0.40	0.71	<b>3.43</b>	<b>0.44</b>	<b>0.77</b>	0.06
8	3.00	0.43	0.76	<b>3.63</b>	<b>0.51</b>	<b>0.90</b>	0.14
9	2.89	0.47	0.80	<b>3.44</b>	<b>0.55</b>	<b>0.94</b>	0.14
10	2.80	0.51	0.86	<b>3.40</b>	<b>0.59</b>	<b>1.00</b>	0.15
Prob	0.667						

표 5. 월 평균 구매 수량을 사용한 상품 추천 비교

Table 5. Product Recommendations Using Average Monthly Purchase Quantities

Cut off	Traditional CF (1)			Graph Embedding CF (2)			Diff F1 (2)-(1)
	Precision	Recall	F1	Precision	Recall	F1	
1	1.00	0.02	0.03	<b>2.00</b>	<b>0.03</b>	<b>0.07</b>	0.03
2	1.50	0.05	0.10	<b>2.00</b>	<b>0.06</b>	<b>0.12</b>	0.02
3	2.00	0.11	<b>0.21</b>	<b>2.00</b>	0.11	0.20	-0.01
4	2.00	0.14	0.27	<b>2.50</b>	<b>0.18</b>	<b>0.33</b>	0.06
5	1.60	0.14	0.27	<b>2.20</b>	<b>0.19</b>	<b>0.34</b>	0.08
6	1.67	0.17	0.31	<b>2.00</b>	<b>0.20</b>	<b>0.37</b>	0.06
7	1.43	0.17	0.31	<b>1.71</b>	<b>0.20</b>	<b>0.36</b>	0.06
8	1.50	0.20	0.36	<b>1.63</b>	<b>0.22</b>	<b>0.39</b>	0.03
9	1.33	0.20	0.35	<b>1.78</b>	<b>0.29</b>	<b>0.49</b>	0.14
10	1.40	0.26	0.44	<b>1.80</b>	<b>0.32</b>	<b>0.54</b>	0.10
Prob	0.374						

## V. 결 론

본 연구의 결과 그래프 임베딩을 사용한 상품 추천은 전반적으로 기존의 그래프 임베딩을 사용하지 않은 상품 추천에 비해 약간 우수하거나 뚜렷하게 높은 추천 성능을 나타낸다. 특히, 구매 빈도를 사용한 그래프 임베딩 기반의 상품 추천은 이전의 상품 추천에 비해 상대적으로 유의한 상품 추천 정확도를 보여 그래프 임베딩 기반의 상품 추천이 효과적인 성과를 나타내는 것을 확인 하였다. 이것은 구매고객-상품간의 관계 매트릭스를 그래프 임베딩이 더 효과적으로 자질화하기 때문이다.

본 연구의 의의는 다음과 같다. 첫째, 최근 컴퓨터 공학 분야에서 활발한 연구가 이루어지는 그래프 임베딩을 경영학 관점에서 상품 추천과 연계해 그 효과를 실증하였다. 일부의 연구가 영화 추천 등에 활용이 되었지만, 일반적인 소매 상품 추천에 그래프 임베딩을 적용한 연구는 거의 찾아볼 수 없다. 둘째, 단순 차원 축소를 목적으로 하는 이전의 그래프 임베딩 방법에 비해 그래프의 노드나 간선을 추론하는 것에 목적을 둔 DeepWalk를 상품 추천에 적용하였다. 연구를 통해 그래프를 한정된 벡터 공간에 임베딩 할 때 더 높은 성능과 효과를 보였다. 셋째, 모든 데이터는 그래프 형태로 표현이 가능하고 이를 활용한 연구는 네트워크를 처리하는 시간의 한계로 인해 제한적인 연구가 수행되었다. 하지만, 본 연구에서 제안된 그래프 임베딩 방법으로 그래프를 제한된 벡터 공간에 표현하게 되어 보다 큰 빅 데이터를 활용한 연구에 대한 방향을 제시하였다.

또한, 실무적으로 본 연구가 제안한 방법을 기반으로 보다 효과적이고 정확도 높은 상품 추천으로 기업의 성과를 향상시킬 수 있다. 다양한 상품들이 거래되는 온라인 마켓에서 보다 정확한 상품 추천을 하는 것은 중요한 문제이고 이를 극복하기 위한 새로운 방법을 본 연구에서 제시하였다.

본 연구의 한계는 다음과 같다. 첫째, 구매자 100명에 대한 2만 여개의 상품 데이터로 한정하여 실험을 수행하였다. 추후 연구를 통해 보다 많은 구매자와 상품이 포함된 빅 데이터를 실험해 제안된 방법의 성능을 측정해야 한다. 둘째, 보다 다양한 그래프 임베딩 방법을 사용하여 보다 성능 높은 상품 추천 방법을 찾아야 한다. 최근의 연구는 DeepWalk 외에도 Node2vec, LINE 등의 다양한 그래프 임베딩 기법들이 제안되고 있다. 이러한 그래프 임베딩 방법들이 상품추천, 고객이탈 등의 다양한 경영 문제에 어울리도록 하는 연구가 필요하다. 셋째, 구매고객-상품-유통사-제조사로 이어지는 네트워크에 대한 연구가 필요하다. 이러한 네트워크는 네트워크의 노드가 서로 다른 속성을 가지는 이질적 네트워크로 분석을 위한 새로운 방법이 필요하다.

## 감사의 글

이 성과는 2017년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. 2017R1A2B4010956).

## 참고문헌

- [1] Deshpande, M., and Karypis, G., "Item-Based Top- N Recommendation Algorithms," *ACM Transactions on Information Systems*, Vol. 22, No. 1, pp. 143-177, 2004.
- [2] Bobadilla, J., Ortega, F., Hernando, A., and Gutiérrez, A., "Recommender Systems Survey," *Knowledge-Based Systems*, Vol. 46, pp. 109-132, 2013.
- [3] Rendle, S., and Schmidt-Thieme, L., "Pairwise Interaction Tensor Factorization for Personalized Tag Recommendation," in *Proceedings of the Third ACM International Conference on Web Search and Data Mining*, New York, USA: ACM, pp. 81-90, 2010.
- [4] Yan, S., Xu, D., Zhang, B., Zhang, H. j., Yang, Q., and Lin, S. 2007. "Graph Embedding and Extensions: A General Framework for Dimensionality Reduction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 29, No. 1, pp. 40-51, 2007.
- [5] Fu, Y., & Ma, Y. "Graph embedding for pattern analysis", *Springer Science & Business Media*, 2012.
- [6] Tenenbaum, J. B., Silva, V. de, and Langford, J. C., "A Global Geometric Framework for Nonlinear Dimensionality Reduction", *American Association for the Advancement of Science*, pp. 2319-2323, 2000.
- [7] Roweis, S. T., and Saul, L. K., "Nonlinear Dimensionality Reduction by Locally Linear Embedding", *American Association for the Advancement of Science*, pp. 2323-2326, 2000.
- [8] Belkin, M., and Niyogi, P., "Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering," in *Advances in Neural Information Processing Systems* 14, MIT Press, pp. 585-591, 2002.
- [9] Berline, N., Getzler, E., & Vergne, M., "Heat kernels and Dirac operators", *Springer Science & Business Media*, 2003.
- [10] He, X., and Niyogi, P., "Locality Preserving Projections," in *Advances in Neural Information Processing Systems* 16, MIT Press, pp. 153-160, 2004.
- [11] Perozzi, B., Al-Rfou, R., and Skiena, S., "DeepWalk: Online Learning of Social Representations," in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, USA: ACM, pp. 701-710, 2014.
- [12] Mikolov, T., Chen, K., Corrado, G., and Dean, J., "Efficient Estimation of Word Representations in Vector Space", In *Proceedings of Workshop at ICLR*, pp. 1-12, 2013
- [13] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J., "Distributed representations of words and phrases

and their compositionality", In *Advances in neural information processing systems*, pp. 3111-3119, 2013.

[14] Goldberg, D., Nichols, D., Oki, B. M., and Terry, D., "Using Collaborative Filtering to Weave an Information Tapestry," *Commun. ACM*, New York, NY, USA: ACM, pp. 61-70, 1992.

[15] Kohavi, R., and Provost, F. 2001. "Applications of Data Mining to Electronic Commerce," *Data Mining and Knowledge Discovery*, Vol. 5, No.1-2, pp. 5-10.

[16] Bobadilla, J., Ortega, F., Hernando, A., and Gutiérrez, A., "Recommender Systems Survey," *Knowledge-Based Systems*, Vol. 46, pp. 109-132, 2013.

[17] Burke, R., "Hybrid Recommender Systems: Survey and Experiments," *User Modeling and User-Adapted Interaction*, Vol. 12, No. 4, pp. 331-370, 2002.

[18] Burke, R., "Evaluating the Dynamic Properties of Recommendation Algorithms," In *Proceedings of the Fourth ACM Conference on Recommender Systems*, New York, USA: ACM, pp. 225-228., 2010.

[19] Pazzani, M. J., & Billsus, D., "Content-based recommendation systems", In *the adaptive web*, pp. 325-341, Springer, Berlin, Heidelberg, 2007.

[20] Schafer, J. B., Konstan, J. A., & Riedl, J., "E-commerce recommendation applications", *Data mining and knowledge discovery*, Vol. 5, No.1-2, pp. 115-153, 2001.

[21] Hofmann, T., "Collaborative Filtering via Gaussian Probabilistic Latent Semantic Analysis," in *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval*, New York, USA: ACM, pp. 259-266, 2003.

[22] Jeong, S. Y., Kim, H. J., "A Recommender System Using Factorization Machine." *Journal of Digital Contents Society*, Vol. 18, No. 4, pp. 707-712, 2017.



**이건창(Kun Chang Lee)**

1984년 : 카이스트 경영과학과  
(공학석사-의사결정지원)  
1988년 : 카이스트 경영과학과  
(공학박사-인공지능)

1988년 ~ 현재: 경기대학교 경영정보학과 교수 역임 및 현재  
성균관대학교 경영대학 교수  
※관심분야 : 창의성과학, 인공지능, 헬스 인포매틱스,  
빅데이터 분석, 감성분석



**박성수(Sung Soo Park)**

2010년 : 성균관대학교 경영대학원  
(경영학석사)

2010년~현 재: 성균관대학교 경영대 박사과정  
※관심분야 : 인공지능(Artificial Intelligence), 빅데이터(Big Data), 감성분석(Sentiment Analysis) 등