

# 딥러닝 기반 고객 행동분석 B2B 추천 시스템

문혜영, 최윤종, 권예하, 최희선, 장혜인, 손종수†  
CJ올리브네트웍스

## Deep Learning Based Recommender System for B2B

### Marketplaces: Customer behavior analysis

Moon Hye-young, Choi Yun-jong, Kwon Yae-ha, Choi Hee-seon, Chang Hye-in, Sohn Jong-soo†  
CJ OliveNetworks

E-mail : hyeyoung.moon@cj.net yunjong.choi@cj.net yae-ha.kwon@cj.net heeseon.choi@cj.net  
hi.chang@cj.net jongsoo.sohn@cj.net

#### 요 약

최근 식품업계에서는 코로나19 장기화로 불확실성이 커지면서 B2C(기업과 소비자 간 거래)에 치우친 사업 구조 개편을 통해 매출 성장을 이루기 위한 전략들이 나오고 있다. 포화상태에 이른 B2C 시장에 비해 B2B(기업과 기업 간 거래) 시장은 성장 가능성이 높으며 특히 추천 시스템을 핵심 마케팅 기술 요소로 간주하여 적극적으로 활용하려는 시도가 증가하고 있는 추세다. 본 논문에서는 Node Embedding 기반의 Deep Learning 기법을 활용하여 구매 패턴에 따른 ‘상대적 사용자 세그먼트’를 수행하고, 분류정보를 활용한 동일 군집 내 인기 상품과 교차 상품을 추천하는 방법을 제안한다. 제안하는 방법은 B2B시스템에서의 고객 성향, 구매 패턴 등을 자동으로 종합하여 파악하고, 숨겨진 구매 유도 요소를 추출하며, 더 나아가 추가적인 개발 없이 고객에게 최적화 된 상품 추천을 해줄 수 있는 기술적 해결 방안을 제공한다.

#### 1. 서론

최근 B2B사업을 통해 수익구조를 다변화하는 기업이 늘고 있다. 특히 식품업계에서는 B2B 시장으로 사업을 확장하고 있는데 코로나19로 불확실성이 커짐에 따라 안정적 수익구조를 구축하는게 중요해졌기 때문이다. B2B 식자재 유통사업의 경우 대량으로 식자재를 납품해야 하기 때문에 거래 금액이 크고 일반적으로 마케팅 비용이 크게 들지 않아 수익률이 높다는 장점이 있다.

B2B 식자재 유통시장의 급속한 성장 속에서 능동적 마케팅 활동의 일환으로 고객추천 마케팅이 많은 관심을 받게 되었다. 그러나 B2B 식자재 유통 기업들은 전통적인 고객층을 지속적으로 관리 해오면서 데이터를 효과적으로 사용하기 보다는 관행적 거래방식이나 영업 담당자에게 의존하고 있는 상황이다. 이에 대다수의 B2B 공급업체들이 고객 추천에 대한 효과성을 증대시킬 수 있는 구체적인 방안을 찾는데 어려움을 겪고 있다.

본 논문에서는 너무 많은 품목을 취급하고 있는 B2B 식자재 유통업체에서 고객사가 관심을 갖거나 필요로 하는 품목은 수십에서 수백 개 밖에 되지 않는다는 점에 착안하여 고객사의 업종과 관심 상품에 따라 제한된 품목을 다양하게 추천하기 위해 Node Embedding 기법을 활용한 딥러닝 기반의 식자재 추천 방법을 제안한다. 제안하는 방법은 첫 번째로 고객, 주문, 상품 데이터를 기반으로 사용자 유사성 정의를 위한 관계 그래프를 학습하고 사용자 세그먼트를 추출하여 유의미한 고객군으로 재 분류한다. 두 번째로 인기 제품군과 교차 제품군으로 추천 식자재 후보를 결정하고 제품군-스코어 테이블을 기반으로 개인화된 필터링 랭킹 방법을 제시함으로써 고객별 적중률 높은 추천 상품을 노출한다. 이를 통해 고객사 각각의 맞춤형 주문 환경을 제공할 수 있게 된다.

## 2. 관련연구

### 2.1 B2B 상품추천 관련 연구

B2B 추천시스템 관련연구[1]에 따르면 B2C 추천 시스템 구매자의 경우 일반적으로 일회성 상황을 위한 구매가 많기 때문에 새로운 품목을 제안하여 신규 고객의 유입을 늘리는 것이 중요하나 B2B 추천 시스템의 구매자는 정기적으로 동일 품목 재구매가 많기 때문에 구매자에게 대체 가능한 품목을 제안하므로서 수익성이 있는 상향 판매, 교차 판매가 촉진될 수 있도록 하는 것이 가장 중요해진다. 그러나 이러한 B2B 추천 시스템 특성을 분석하여 식자재 유통분야에서 맞춤형 추천 서비스를 제공하고자 한 연구는 미흡한 상태이다.

### 2.2 Node Embedding

Graph Embedding 방법에는 DeepWalk나 Node2vec, Graph2vec[2,3,4] 과 같은 임베딩 방법들이 있는데 이는 Document-Word Embedding 과 동일한 기술을 기초로 한다. Graph Embedding 의 근간이 되는 Document-Word Embedding 모델[5,6]은 문장, 문단, 심지어 전체 큰 문서와 같은 임의의 길이의 단어 시퀀스 표현을 학습 할 수

있는 PV-DBOW (paragraph vector-distributed bag of words, doc2vec skipgram)라고 하는 Skip-gram 모델로 특정 문단(=graph)의 임베딩이 그 문단에 등장하는 단어(=subgraphs)의 임베딩과 비슷해지는 것으로 유사한 단어(=subgraphs)을 많이 포함하고 있는 문단끼리 유사한 임베딩을 갖게 된다.

어떤 문단  $d$ 를 어떻게 임베딩할 것인가에 대한 함수는 어떤 문단  $d$ 를 이루고 있는  $w$ 의 sequence로 나타낼 수 있다. 그리고 이를 임베딩하는 Skip-gram 모델은 수식 1을 maximize하는 방향으로 학습한다.

$$\sum_{j=1}^{l_i} \log Pr(w_j|d_i) \quad (1)$$

$$Pr(w_j|d_i) =$$

$$\frac{\exp(\vec{d} \cdot \vec{w}_j)}{\sum_{w \in V} \exp(\vec{d} \cdot \vec{w})} \quad (2)$$

즉, 특정 문단의 임베딩이 그 문단에 등장하는 임베딩과 비슷해지는 것으로 분모는 모든 문단들의 단어를 의미하고, 분자는 해당 문단의 단어들을 의미한다. 따라서 유사한 단어들이 많이 포함된 문단끼리는 비슷한 임베딩을 갖게 된다.[5]

본 연구에서는 유사고객사 기반 관심품목을 다양하게 추천하기 위해 일자별 식자재 구매이력 데이터를 가지고 샘플링 전략에 따른 Node Embedding 모델을 적용하여 관계 그래프를 학습해 사용자 세그먼트를 추출한다.

## 3. 딥러닝 기반 고객 행동분석 B2B 추천 시스템

### 3.1 데이터 정의

과거 주문 이력 데이터를 수집하고 관계 그래프의 형태로 전처리 한다. 이를 위해 1년 동안 실제 B2B 폐쇄형 주문물에서 판매된 식자재 이력을 일자 별로 수집한다. 관계 그래프는 사용자-제품군 간의 관계성으로 표현되며, 해당 프로세스는 알고

리즘상에서 자동화되어 처리된다.

관계 그래프에서 ‘사용자’란 식자재를 주문하는 사업장을 의미하며, ‘제품군’이란 그림 1과 같이 제품의 한 단계 상위 개념으로, ‘\*\*\* 유기농 두부 (부침용) 300gx2입’이나 ‘\*\*\* 국산콩 두부 부침용 (210g)’과 같은 구체적인 제품의 명칭이 아닌 ‘부침용 두부’와 같은 포괄 명칭을 의미한다.

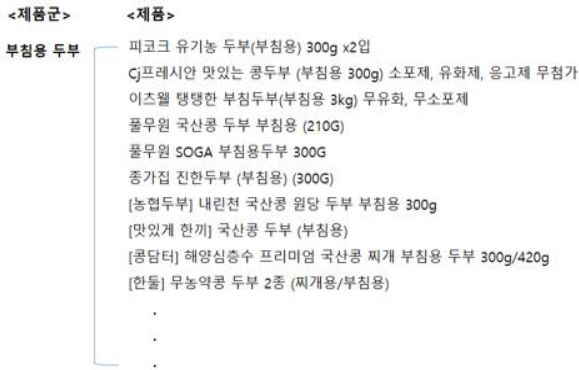


그림1 제품군-제품 용어 규정

### 3.2 관계 그래프 학습

정의된 관계 그래프의 각 Node가 d차원의 벡터로 임베딩 되어 학습이 진행된다. 그림 2와 같이 비슷한 구매 이력을 가진 사용자끼리는 공간 상에서 임베딩 값이 가까이 위치하게 되고, 상이한 구매 이력을 가진 사용자끼리는 임베딩 값이 멀게 위치하도록 학습이 진행된다.

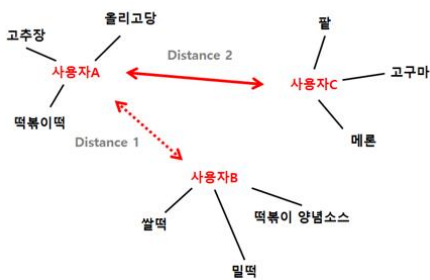


그림 2 관계 그래프 학습 시각적 예시

### 3.3 사용자 세그먼트 추출

사용자 세그먼트는 3.2에서 학습된 임베딩 정보를 바탕으로 판단된다. 사용자는 고유의 임베딩으로 표현되어 공간 상의 특정 위치에 존재하게 되며, 상대적인 관계 속에서 유사성이 결정된다. 즉,

사용자A는 ‘품질보다 가격을 중시하는 고객’과 같이 고정된 세그먼트에 속하는 것이 아니라 ‘사용자C보다는 사용자B와 구매 경향성이 유사한 사용자’와 같은 형태로 정의되는 것이다.

학습한 모델 추론 시, 특정 사용자A가 입력으로 들어왔을 때, 특정 사용자A를 ‘기준 사용자’라고 정의할 수 있다. 이러한 기준 사용자A와 나머지 N-1의 사용자들을 1대1로 비교하여 가장 유사한 사용자부터 가장 비 유사한 사용자까지 순서를 매길 수 있다.

### 3.4 추천 상품 후보 선정

최종 추천 제품은 인기 제품과 교차 제품의 두 가지 형태로 제공된다. 이를 위해 사전에 인기 제품군과 교차 제품군을 선정한다.

인기 제품군은 기준 사용자가 구매한 제품군 내에서 아직 구매한 이력이 없는 인기 제품군을 의미하며 교차 제품군은 기준 사용자가 구매하지 않은 제품군 내에서의 인기 제품군을 의미한다.

기준 사용자가 최근 k개월 구매한 이력이 있는 제품군에 속하는 제품들은 인기 제품군으로 추천되며 이는 표 1에서의 추천 후보 제품군 #1에 해당한다. 다음으로 기준 사용자가 k개월 구매한 이력이 없고 N개의 유사 사용자는 구매한 이력이 있는 제품군을 교차 제품군으로 선정하며 이는 표 1에서 추천 후보 제품군 #2에 해당한다.

| 사용자  | 부침용 두부 | 떡국 떡   | 콩나물       | 순두부   | 어묵  |
|------|--------|--------|-----------|-------|-----|
| 사용자C | 콩나물    | 드레싱    | 떡볶이 떡     |       |     |
| 사용자E | 순두부    | 어묵     | 농축 육수 베이스 | 고추장   |     |
| 사용자F | 어묵     | 떡볶이 떡  | 곡물가루      |       |     |
| 사용자L | 어묵     | 부침용 두부 | 전처리 농산물   | 떡볶이 떡 | 드레싱 |
| 사용자M | 부침용 두부 | 콩나물    | 어묵        | 떡볶이 떡 |     |
| 사용자N | 순두부    | 떡국 떡   | 농축 육수 베이스 |       |     |

추천 후보 제품군 #1  
추천 후보 제품군 #2

표 1 추천 후보 제품군 추출

추천 후보 제품군이 결정됨과 동시에 표 2와 같이 제품군-제품-스코어 테이블을 가지게 되며 여기서 스코어란 유사 사용자 N개에 대한 해당 제품의 공통 노출 빈도를 의미한다. N개 유사 사용

자 전체에서 제품A를 최근 k개월간 구매한 이력이 있다면 스코어는 N이 되고 N개 유사 사용자 중에 제품A를 최근 k개월간 구매한 이력이 한 건도 없다면 스코어는 0이 된다.

| 교차 제품군 - 교차제품 - 스코어 |                |     | 인기 제품군 - 인기제품 - 스코어 |                 |     |
|---------------------|----------------|-----|---------------------|-----------------|-----|
| 제품군                 | 제품             | 스코어 | 제품군                 | 제품              | 스코어 |
| 드레싱                 | ***** 유자레몬 드레싱 | 2   | 부침용 두부              | *** 부침두부        | 6   |
| 드레싱                 | *** 코를솔로 드레싱   | 4   | 부침용 두부              | *** 진한두부(부침용)   | 3   |
| 떡볶이 떡               | *** 밀떡         | 13  | 떡국 떡                | ***** 쌀떡국 3kg   | 2   |
| 고추장                 | **** 고추장       | 3   | 콩나물                 | *** 국산 콩나물 800g | 7   |
| 고추장                 | *** 태양초 고추장    | 0   | 콩나물                 | ** **** 콩나물     | 3   |
| 고추장                 | ** *** 태양초 고추장 | 14  | 순두부                 | ** 냉장 순두부 400g  | 11  |
| 고추장                 | *** 고추장        | 7   | 순두부                 | ** 고소한순두부 400g  | 0   |
| 곡물가루                | **** 새싹보리 분말   | 4   | 어묵                  | *** 어묵 4종       | 8   |
| 전처리 농산물             | ***** 신선야채     | 10  |                     |                 |     |

표 2 제품군-제품-스코어

### 3.5 랭킹

제품군-제품-스코어를 기반으로 추천 시스템 설계자의 추천 전략에 따라 상품의 랭킹을 적용할 수 있으며 다양한 파라미터를 적용할 수 있다. 적용 가능한 파라미터로는 인기 제품과 교차 제품 중 어떤 제품들을 추천으로 제공할지를 위한 Category 설정, ‘유사 사용자 N개’ 설정, 제품 후보군 결정을 위한 ‘최근 k개월의 구매 이력’ 설정, ‘스코어 상위 N개의 제품’을 추천 후보로 결정하기 위한 Top N 지수, 최종 추천 제품을 결정하기 위한 스코어 기반 Item Sample Weight 설정 등이 있다. 이 외에도 동적으로 파라미터를 추가하여 추천 전략을 결정할 수 있다.

## 4. B2B 상품 추천 결과

### 4.1 사용자 세그먼트 추출 결과

Node Embedding기반 관계 그래프 학습결과로 앞서 언급한 사용자 세그먼트는 사용자 유사성을 나타내고 그 결과 식자재를 주문하는 사업장의 세그먼트 추출결과는 표 3과 같이 실행 예시를 기재하였다.

| 기준 사업장    | 유사사업장(n=5)   |
|-----------|--|
| **산후조리원   | ***산후조리원 / **산후조리원 / **미래여성병원 / ***산부인과 / **산후조리원 강서점  |
| **남부노인복지관 | **종합사회복지관 / ***복지회 / **종합사회복지관 / ***봉사회 / **시 노인복지관    |
| ***지역아동센터 | **지역아동센터 / **홀스쿨아동센터 / **멘토지역아동센터 / **잠종은아동센터 / ***공부방 |
| **커피      | ****커피 / **가정카페 / **메이커리 / **대학교 재과커피과 / **커피복합문화센터점   |

표 3 사용자 세그먼트 추출 결과 예시

## 4.2 B2B 온라인 주문물 추천 결과

본 논문에서 제시한 Node Embedding 모델기반의 B2B 추천 기법은 실제 B2B 온라인몰 테스트를 통해 실제 서비스 환경에서 추천 결과를 검증하였다. 프리미엄 식자재를 주로 주문하는 사업장, 가성비 좋은 상품을 주로 주문하는 사업장, 대용량 상품을 주로 주문하는 고객 등 개별 군집 특성에 맞는 추천 상품결과를 확인하였다. 결과 예시는 표4와 같다.

| 기준 사업장     | 제품군     | 추천 상품               |
|------------|---------|---------------------|
| ** 구청      | 콩나물     | *** 콩나물(짬뽕 4kg)     |
| ***** 어린이집 | 콩나물     | ** 콩나물(친환경_G마크 1kg) |
| ** 회사 구내식당 | 킹크랩/랍스터 | 랍스터(40미 4.5kg)      |
| ** 실버타운    | 킹크랩/랍스터 | 랍스터(40미 4.5kg)      |

표 4 상품 추천 결과 예시

구청 구내식당에서는 단가 경쟁력 있는 콩나물이 추천되었고 어린이집의 경우에는 친환경 콩나물이 추천되었다. 또한 기존 사업장 분류가 오피스로 구분되어있는 프리미엄 구내식당이 실버타운과 유사사업장으로서 분류되어 고급 식자재를 추천 받았다.

## 5. 결론 및 향후 연구

본 논문에서는 Node Embedding기법을 활용하여 딥러닝기반의 식자재 추천 방법을 제안하고 검증하였다. 사용자 세그먼트 추출 결과를 확인하였으며 추천 서비스 도메인의 특성과, 그 특성을 고려한 설계를 진행하여 추천시스템을 구축하고 실제 서비스에 적용하여 사업장 별 적중률 높은 추천 상품이 노출됨을 확인하였다.

앞으로 본 연구는 추천 효과의 정량평가를 목표로 연구될 예정이다. B2C 추천의 경우, 상품의 구매가 매출의 확대로 연결되어 추천 상품의 구매율에 연관된 지표를 평가지표로 사용한다. B2B 추천은 사용자의 구매가 지정된 예산 범위에서 이루어지는 경우가 대부분으로 매출 이익 또는 거래의 유지 기간 등의 평가지표 개발이 필요하다. A/B 테스트 등을 활용하여 효과적인 평가 지표를 찾는 방향으로 연구할 예정이다.

## [참고문헌]

- [1] Zhang, Xuirui and Wang, Hengshan (2005) "Study on Recommender Systems for Business-To-Business Electronic Commerce," Communications of the IIMA: Vol. 5 : Iss. 4 , Article 8.
- [2] B. Perozzi, R. Al-Rfou, and S. Skiena. DeepWalk: Online learning of social representations. In KDD, 2014.
- [3] Grover, Aditya, and Jure Leskovec. "node2vec: Scalable feature learning for networks." Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2016.
- [4] A. Narayanan, M. Chandramohan, R. Venkatesan, L. Chen, Y. Liu, S. Jaiswal, graph2vec: Learning Distributed Representations of Graphs (2017).
- [5] Le, Quoc, and Tomas Mikolov. "Distributed representations of sentences and documents." Proceedings of the 31st International Conference on Machine Learning (ICML-14). 2014.
- [6] Mikolov, Tomas, et al. "Efficient estimation of word representations in vector space." arXiv preprint arXiv:1301.3781 (2013).