

Fad2Vec: 판매 데이터 임베딩 기반 패션 산업 단발성 유행 탐지

안승섭¹ · 진소연² · 이학연^{3*}

¹서울과학기술대학교 데이터사이언스학과 / (주)컬리 CLO 오피스

²서울과학기술대학교 데이터사이언스학과

³서울과학기술대학교 산업공학과

Fad2Vec: Detecting Fads in the Fashion Industry Based on Sales Embedding

Seungseop Ahn¹ · Soyeon Jin² · Hakyeon Lee³

¹Department of Data Science, Seoul National University of Science and Technology / CLO office, Kurly InC.

²Department of Data Science, Seoul National University of Science and Technology

³Department of Industrial Engineering, Seoul National University of Science and Technology

In the fashion industry, fads, a phenomenon in which demand for specific items surges in a short time and then disappears, are often observed. However, the conventional demand forecasting methods have limitations in predicting fads because it focuses on long-term trends. This paper presents a new approach called Fad2Vec, which adapts the Item2Vec approach to detect fads effectively. Hot periods composed of a hot point and a preceding period for each product are identified based on time-series sales data. Products are embedded into vectors representing the time points when they are temporarily fashionable. Then, they are clustered into a fad group where fashion items belong to the same cluster and show similar fad patterns. A case study of four famous fashion brands is provided to show how Fad2Vec works and verify its validity. The proposed Fad2Vec approach is expected to be practically utilized for the production planning of fast fashion companies.

Keywords: Fashion industry, Trend, Fad, Fad2Vec, Item2Vec, Word2Vec

1. 서론

Fad는 단발성 유행 및 짧은 주기로 판매량이 폭발적으로 성장하다가 사라지는 현상으로, 20세기 후반, 다원주의 및 인터넷과 모바일 기술의 발전으로 인해 소비자들의 니즈가 빠르고 다양하게 변화하며 등장한 개념이다(Solomon *et al.*, 2004). Fad가 등장하기 이전에는 대부분의 패션 기업들이 트렌드(trend)를 반영해 디자인 및 판매 전략을 수립하였다. 변화의 일반적인 방향을 뜻하는 트렌드는 인구 통계적 변화, 대중의 가치관

이나 태도, 생활양식 또는 기술 등의 변화 등으로 인해 발생하여 시장 전반에 광범위한 변화를 가져온다. 트렌드는 특히 시즌마다 새로운 유행이 등장하는 패션 산업에서 두드러지게 나타난다.

1980년대 중반까지 패션 산업의 트렌드는 리바이스 501 청바지, 화이트 셔츠 등과 같은 표준화된 스타일의 저비용 대량 생산되는 상품들이 주를 이루었다. 당시 소비자들은 스타일과 패션에 덜 민감했고 클래식한 스타일의 의류를 선호했다(Bhardwaj *et al.*, 2010). 하지만 이후 90년대부터 소비자들이

이 논문은 제1저자 안승섭의 석사학위논문을 발췌 및 수정하여 작성한 것임. 이 연구는 한국연구재단의 이공분야기초연구사업(기본연구)의 지원을 받아 수행되었음(NRF-2019R1F1A1057071 & NRF-2021R1F1A1045787).

* 연락저자 : 이학연 교수, 서울시 노원구 공릉로 232 서울과학기술대학교 프론티어관 607호, Tel : 02-970-6469, Fax : 02-974-2849,

E-mail : hylee@seoultech.ac.kr

2021년 12월 7일 접수; 2022년 1월 4일 수정본 접수; 2022년 1월 5일 게재 확정.

패션에 관심을 가지기 시작하면서 80년대의 클래식한 스타일의 의류에 대한 수요가 감소하기 시작했다(Bailey *et al.*, 2001). 변화된 트렌드는 20세기에 급격히 성장한 인터넷과 모바일이라는 신기술로 인해 실시간으로 전파되며 강한 파급력을 가지게 되었다(Blázquez *et al.*, 2014). 이에 따라 패션 트렌드의 주기가 급속히 짧아지면서 Fad의 중요성이 대두되기 시작했다. 80년대의 패션 트렌드는 봄과 여름(spring/summer: s/s) 및 가을과 겨울(fall/winter: f/w)을 바탕으로 개최되는 패션쇼 및 직물 박람회 등을 기반으로 하여, 1년 동안의 주기가 결정되었다. 하지만, 90년대 초반부터는 소비자들의 니즈에 따른 패션 의류의 다양성으로 인해 기존 시즌에 3~5개의 중간 시즌이 추가되었다(Barnes *et al.*, 2006). 이러한 흐름에 발맞춰 상품의 기획, 생산 및 유통과정을 통합하여 최신 유행을 반영한 저가의 의류를 최소 2주 정도의 짧은 주기로 대량생산 및 판매하는 Fad 대응 전략을 세운 패스트 패션(fast fashion) 브랜드가 등장했다. 대표적인 패스트 패션 브랜드로는 GAP, UNIQLO, ZARA, H&M 등이 있다.

패션 기업들이 공급 리드 타임(lead time)을 줄여 Fad에 대응하고 있지만, 다수의 재고 문제가 있어 수요 예측을 통한 Fad 대응 전략의 보완이 필요하다. 수요 예측은 고객 서비스의 향상과 재고 관리 및 생산 관리의 최적화를 위한 가장 중요한 작업이며, 부실한 수요 예측은 재고 부족 또는 높은 재고비용, 이로 인한 낮은 서비스 및 상품 품질 등 기업 경영에 치명적인 위험을 끼칠 수 있다(Rexhausen *et al.*, 2012). 또한 수요 예측을 전문가의 주관적인 예측에만 의존하게 되면 잘못된 결론을 내릴 위험이 있다(Tehrani *et al.*, 2016). 특히 Fad는 주기가 짧고 일정한 패턴을 보이지 않아 패션 산업의 수요 예측은 더욱 어려워지고 있다(Minner *et al.*, 2012). 이에 패션 산업의 수요 예측 성능 높이기 위한 목적으로 머신 러닝(machine learning) 기법을 적용하려는 시도들이 최근 활발히 이루어져 왔다. Thomassey and Fiordaliso(2006)는 K-Means 클러스터링과 의사결정나무(decision tree)를 이용하여 상품 판매량을 예측하였다. Frank *et al.*(2003)은 인공신경망(artificial neural network: ANN)을 바탕으로 미국 의류 회사의 판매 수량을 예측하고 그 성능을 지수평활법과 비교하였다. Loureiro *et al.*(2018)은 패션 소매회사의 판매 데이터를 바탕으로 선형회귀분석(linear regression)과 함께 앙상블(ensemble) 모델 및 딥러닝 모델을 활용하여 예측을 수행하였다. 그러나 이 연구들 역시 짧은 기간에 폭발적으로 유행했다가 장기간 잠복기를 거치고 다시 판매량이 급상승하는 단발성 유행인 Fad를 예측함에 있어서는 여전히 한계가 존재한다(Nenni *et al.*, 2013; Gutierrez *et al.*, 2008).

Fad는 일정한 주기 또는 경향에 따라 출현하는 것이 아니므로 시계열 데이터의 추세를 바탕으로 사전에 예측하는 것은 거의 불가능하기 때문에, 예측보다는 빠르게 그 출현을 탐지하고 유연하게 대응하는 접근 방식이 요구된다. 따라서 본 연구는 패션 산업에서 Fad를 효과적으로 탐지할 수 있는 새로운 방법으로써, Item2Vec을 차용한 Fad2Vec을 제시한다. Item2Vec은

소비자별 구매 이력을 참고하여, 상품의 선호도를 반영해 상품을 임베딩(embedding)한 후, 각각의 상품 벡터의 유사도를 계산해 상품을 추천해 주는 방법이다(Barkan *et al.*, 2016). Fad2Vec은 패션 상품의 판매 데이터에서 상품별로 특정 기준에 부합하는 유행 시점들을 추출해 시점별로 유행한 상품 벡터를 생성하여 임베딩한 후, 벡터의 유사도를 계산해 단발성 유행 패턴을 예측한다. 비슷한 유행 패턴을 보이는 상품들은 높은 벡터 유사도를 가지므로, 군집화를 통해 하나의 상품의 유행이 시작되었을 때, 같은 군집 내 상품들이 조만간 함께 유행하게 될 것임을 파악할 수 있다.

본 연구의 구성은 다음과 같다. 제2장에서는 머신러닝을 위한 패션 상품 판매 예측 연구 및 Item2Vec 관련 연구를 살펴본다. 제3장에서는 Fad2Vec의 기본 개념과 프레임워크를 설명한다. 제4장에서는 국내 유명 패션 기업 4개 브랜드 판매 데이터에 Fad2Vec을 적용하고 검증한다. 마지막으로 제5장에서는 연구의 의의 및 향후 연구 방향을 포함한 결론을 제시한다.

2. Item2Vec

ANN을 통한 자연어 처리(natural language processing: NLP) 기법이 급속하게 발전하고 있다(Collobert *et al.*, 2008; Barkan *et al.*, 2017). 최근 번역 및 추천 시스템 등에 널리 활용되고 있는 Word2Vec은 저차원 공간에 단어를 분산 표현하는 모델이다(Mikolov *et al.*, 2013). 문장 내 중심단어와 주변 단어 사이의 관계를 이용하여 임베딩하기 때문에 벡터에 단어의 의미를 담을 수 있다는 것이 장점이다. 기존의 원-핫 인코딩(one-hot encoding) 방법은 단순히 단어의 나열을 벡터로 나타낸 것으로, 각 벡터의 차원은 단순히 해당 단어가 벡터 내에서 가지는 인덱스를 표현할 뿐이다. 또한 말뭉치 내 단어의 개수가 벡터의 차원이 되므로, 희소벡터(sparse vector)화 되며, 벡터에 단어의 의미를 포함한 다양한 정보를 담지 못한다는 단점이 있다. 따라서 의미를 내포하는 벡터의 형태로 단어를 표현하는 임베딩 벡터를 주로 사용하게 된다. 임베딩 벡터는 훈련 데이터로 학습을 통해 계산되는데, 이때 쓰이는 학습 방법이 Word2Vec이다. Word2Vec은 문서에서 비슷한 위치에 등장하는 단어들은 비슷한 의미를 가진다는 분포가설(distributional hypothesis)에 기반한다(Mikolov *et al.*, 2013).

Word2Vec은 CBOW(continuous bag-of-words), Skip-gram 2가지 방식으로 학습을 진행한다. CBOW는 주변단어를 통해 중심단어를 예측하는 방법으로 임베딩 벡터를 생성한다. 중심단어를 예측하기 위해 중심단어를 기준으로 앞, 뒤 몇 개의 단어를 불지를 결정하는 범위를 윈도우 사이즈(window size)라고 한다. 윈도우 사이즈가 결정되면 CBOW는 슬라이딩 윈도우(sliding window) 방식으로 주변 단어와 중심 단어 선택을 바꿔가며 학습 데이터 세트를 생성한다. 이후 중심단어의 임베딩 벡터와 주변 단어의 임베딩 벡터들의 유사도가 높아지도록,

주변단어와 중심단어가 아닌 단어들의 임베딩 벡터의 유사도는 낮아지도록 학습이 진행된다. Skip-gram은 CBOW와 반대로 중심단어에서 주변단어를 예측하며 임베딩 벡터를 생성하는 방법이다. 중심단어가 주어질 때 주변단어가 등장할 확률을 최대한으로 만들어주도록 임베딩 벡터값을 조정해 주게 된다. 이를 통해 중심단어의 임베딩 벡터와 주변단어의 임베딩 벡터의 유사도가 높아지게 된다. CBOW나 Skip-gram으로 학습된 단어 임베딩 벡터는 단어와 단어 사이의 관계적 맥락에 따른 의미를 포함하게 된다.

Item2Vec은 상품 간의 유사성을 생성하기 위해 Word2Vec의 단어 학습 방법을 사용하여 아이템, 즉 상품을 임베딩하는 방법이다. Word2Vec의 문장 내에서 단어를 학습하는 과정은 Item2Vec에서 장바구니 내에서 상품을 학습시키는 과정으로 볼 수 있다. 이 과정에서 장바구니 내에서 구매한 상품의 시간적 배열은 무시한다. 이를 통해 Item2Vec은 상품 구매 데이터에서 상품의 선호도를 반영하여 임베딩 할 수 있게 되며, 고객들의 장바구니에 자주 같이 등장했던 상품들은 비슷한 선호도가 반영된 임베딩 벡터값을 갖게 된다. 이를 통해 상품 선호도의 유사성을 반영하여 유사 상품 추천이 가능하다. Item2Vec은 학습 방식에서 Word2Vec과 차이를 보인다. Word2Vec은 윈도우 사이즈를 설정하여, 범위 내에 존재하는 단어들의 임베딩 벡터값의 내적을 크게 만드는 방식으로 학습을 진행하는데 반해, Item2Vec은 윈도우 사이즈를 최대화하여 사용자가 구매한 상품목록에 담긴 모든 상품이 유사한 임베딩 벡터를 갖도록 학습을 진행한다.

Item2Vec 기법은 여러 도메인에 성공적으로 적용되어 왔다. Barkan and Koenigstein(2016)은 Microsoft Xbox 음악 제공서비스에서 사용자가 재생한 음악의 아티스트 데이터와 Microsoft Store의 사용자 장바구니 데이터를 이용하여 Item2Vec 방법을 검증하였다. 그 결과, 유사한 장르를 가진 아티스트들과 비슷한 품목 및 제조사의 상품끼리 높은 유사도의 임베딩 값을 가지는 것을 확인하였다. Sadeghian *et al.*(2019)은 Item2Vec 방법을 호텔 산업에 적용하였다. 사용자가 호텔을 조회한 데이터와 호텔의 서비스 데이터, 지리정보를 결합한 벡터를 이용해

호텔의 임베딩 벡터를 생성하였다. 그 결과 비슷한 정보를 가진 호텔끼리 유사도가 높게 임베딩 되었으며, 사용자들이 향후 조회할 호텔을 예측하는 성능도 우수하게 나타났다. Book2vec은 도서의 내용을 확인하기 어려운 경우, 사용자가 책을 읽은 데이터만으로 Item2Vec 방법을 통해 도서를 임베딩하는 방법이다(Anvari *et al.*, 2018). Googlereads 사이트의 데이터로 Item2Vec 방법을 활용해 도서들을 임베딩하였다. 그 결과, 같은 저자 및 시리즈 도서들에 비슷한 임베딩 벡터값을 할당해 도서의 추천 시스템에 유용하게 활용될 수 있음을 확인하였다. 또한 세계적인 음악 스트리밍 서비스 기업은 Spotify는 음악 플레이리스트에 Item2Vec 알고리즘을 적용하여 고객들에게 음악 추천 서비스를 제공하고 있다(Anderson *et al.*, 2020). 본 연구에서는 Item2Vec을 차용 및 변형하여 패션 산업에서의 Fad 예측을 위한 Fad2Vec 방법론을 제시한다.

3. Fad2Vec

Fad2Vec은 단어의 의미를 학습시키는 Word2Vec의 방법론에 기반하여, 상품의 유사성을 학습시키는 Item2Vec을 차용한 방법론이다. Fad2Vec은 상품의 판매 데이터로부터 찾아낸 시계열 패턴을 학습시키는 것이 아닌 상품의 유행 시점을 학습시키는 모델로, 임베딩된 벡터들은 상품 간 유행 시점이 얼마나 비슷한지를 나타낸다. 비슷한 임베딩 값을 가지는 상품들은 비슷한 시기에 유행하는 상품 집합, 즉 Fad 군집으로 볼 수 있다. 따라서 특정 상품의 갑작스러운 유행이 시작되었음을 감지하였다면, Fad 군집 내에 속하는 다른 상품들도 빠른 시일 내에 유행이 시작될 가능성이 높으며, 이에 따라 상품 디자인 및 생산 계획을 수립함으로써 Fad에 선제적으로 대응이 가능하다. <Table 1>은 Word2Vec, Item2Vec과 비교를 통한 Fad2Vec의 기본 개념을 설명한 표이며, <Figure 1>은 Fad2Vec의 프레임워크를 도식화한 것이다.

Fad2Vec의 절차는 다음과 같다. 먼저 판매 데이터에서 상품별 유행구간(hot period)을 도출한다. 유행구간은 특정한 조건

Table 1. Comparison among Word2Vec, Item2Vec, and Fad2Vec

Method	Word2Vec		Item2Vec		Fad2Vec	
Objective of learning	Learning the semantic meaning of words		Learning the product similarity		Learning the popular period	
Meaning of vectors	Meaning of words		Preference of products		Similarity of popular periods	
Application fields	Word and sentence classification		Product recommendation		Production planing	
Types of data	Word vectors of each sentence		User's shopping cart vectors		Products vectors that were popular at the time	
Data form	Sentence	Word	User	Product	Year/Month	Product
	1	cat, dog, ...	U1	P1, P2, ...	19/01	P1, P2, ...
	2	coke, tea, ...	U2	P1, P3, ...	19/02	P1, P3, ...
	3	pen, note, ...	U3	P10, P11, ...	19/03	P10, P11, ...

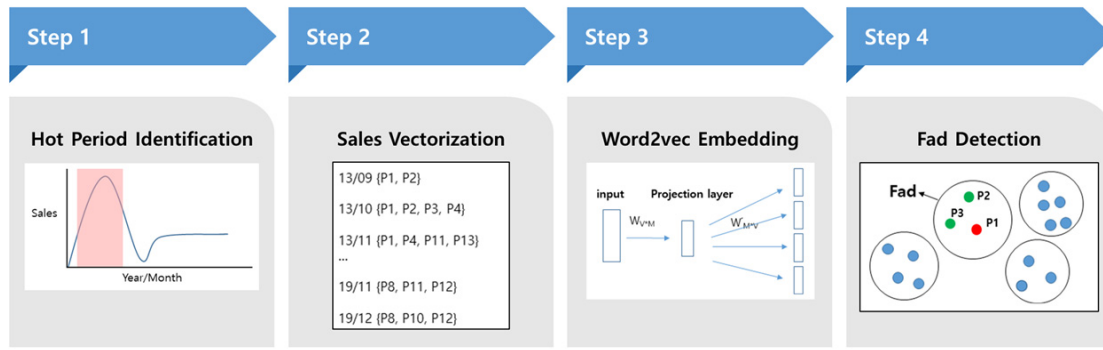


Figure 1. Framework of Fad2Vec

을 만족하는 시점들로 이루어져 있는데, 유행구간은 제품 판매 주기에서 유행구간은 Fad를 정의하는 기준이 되기 때문에 Fad2Vec의 성능을 좌우하는 가장 중요한 개념이라고 볼 수 있다. 유행구간은 구간 내 판매량 최고점을 의미하는 단기고점(hot point)과 선행구간(preceding period) 2가지 개념으로 이루어진다. 본 연구에서는 단기고점을 1) 해당 시점의 판매량이 판매량의 3개월 이동평균보다 크고 2) 전 기간 최고 판매량의 10%보다 크며 3) 판매량의 증가추세가 끝나는 지점이라는 3가지 조건을 모두 만족하는 시점으로 정의하였다. 선행구간은 유행구간 내 최고점을 찍기까지의 상승구간을 의미한다. 본 연구에서는 선행구간을 모든 시점 간 기울기가 양인 구간에서 판매량이 1 이상인 시점부터 단기고점 직전 시점까지로 정의하였다.

<Figure 2>는 판매 그래프에 본 연구에서 정의한 유행구간을 적용한 예시이다. 14/01(14년 01월)은 3개월 이동평균 값보다 크고 전 기간 최고 판매량의 10%보다 크므로 단기고점으로 볼 수 있다. 판매가 발생한 13/11부터 13/12, 14/01은 기울기가 모두 양인 것을 확인할 수 있다. 따라서 13/11부터 단기고점인 14/01의 직전인 13/12까지가 선행구간으로 볼 수 있다. 단기고점과 선행구간을 합해, 13/11부터 14/01까지를 유행구간으로 최종 설정한다.

상품별 유행구간을 정의한 후, 각 시점에 유행한 상품만을 추출하여 벡터를 만든다. 이 벡터를 Word2Vec의 Skip-gram으로 학습시킨다. Skip-gram을 사용하는 이유는 단어의 의미를 학습하는 작업에서 Skip-gram이 CBOW보다 더 좋은 성능을 보이며(Mikolov *et al.*, 2013), 네거티브 샘플링(negative sampling)을 활용할 때 자주 등장하지 않는 상품들까지 정확한 임베딩 값을 할당할 수 있기 때문이다(Barkan *et al.*, 2016). 이 과정을 통해 같은 시기에 유행했던 상품들은 유사한 임베딩 벡터값을 갖도록 학습된다. 비슷한 임베딩 벡터값을 갖는 상품들의 군집을 Fad로 정의한다. Fad 군집 내 어떤 상품이 갑작스럽게 유행하게 되면, 해당 상품의 벡터와 유사도가 높은 상위 N개 상품의 생산 및 판매 계획을 조정함으로써 Fad에 대응할 수 있다.

4. 실험 및 검증

4.1 데이터 및 전처리

본 연구에서는 Fad2Vec의 유효성을 검증하기 위해 10년간(2011년 01월~2019년 12월) 월별 국내 패션 기업 L사 4개 브랜드

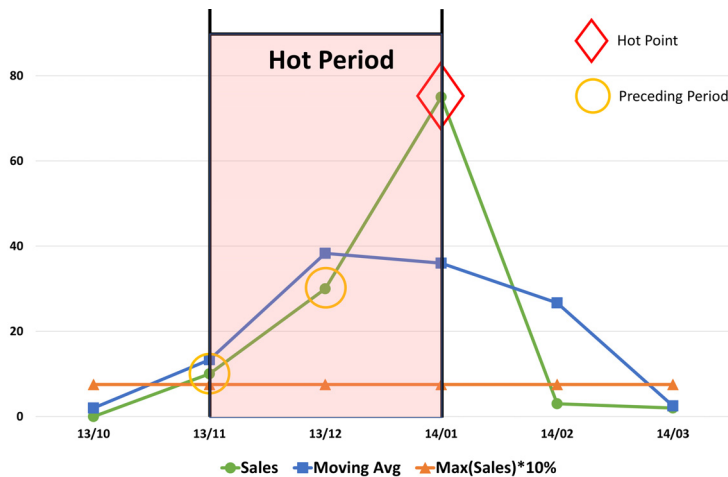


Figure 2. Example of a Hot Period

Table 2. Product Features and Codes

Feature	Attributes
Brand	HZ, DM, MA, JN (4 brands)
Category	Jumper, T-shirts, Half coat, Long trench coat, Trench coat, Casual shirts, Pants, Muffler, Pull over, Jacket, Gloves, Long coat, Living, Tie, Stationary, Perfume, Accessories, Shirts, Sweater, Shoes, Sets etc. (64 categories)
Season	Spring, Summer, Fall, Winter, First half, Second half, All (7 periods)
Color	107 colors

드의 신사복 및 영캐주얼 판매 데이터를 사용하였다. 판매 데이터는 판매연월, 상품코드, 판매수량으로 구성되어 있다. 이 중 상품코드는 6개의 정보를 포함하는 일련의 알파벳 코드로 이루어져 있다. 상품코드는 브랜드, 품목, 생산연도, 시즌, 일련번호, 색상 정보로 구성되어 있다. 브랜드는 총 4개의 브랜드로 구성되어 있으며, 품목은 64개 품목, 시즌은 봄, 여름, 상반기, 가을, 겨울, 하반기, 무관 총 7개의 기간으로 색상은 107개 색상으로 구분되어 있다. 상품코드의 주요 정보는 <Table 2>와 같다.

상품코드로 구별된 상품들을 주요 특성 중심으로 통합하기 위해서 107개의 상세 색상을 22개의 큰 범주의 색상코드로 통합하였다. 또한 하프 코트와 롱 트렌치 코트, 트렌치 코트 등 세부 카테고리를 코트라는 큰 범주의 카테고리로 통합하고 생활용품, 문구류 등 비의류 품목을 제외하여 최종적으로 13개 품목으로 정리하였다. 또한 상품의 특성과는 무관한 생산연도 코드와 일련번호 코드를 제거하여, 총 11자리 코드를 7자리 코드로 축소 후 통합하였다. 그 결과 1,445개 상품 유형이 분석 대상으로 도출되었으며, 통합된 코드에 따라 상품별 월 판매량도 통합하였다.

한편, 상품코드를 통합하였음에도 기간 내 판매실적이 저조해 분석하기 어려운 상품이 존재하였다. 따라서 본 연구에서는 두 가지 조건을 통해 분석 대상이 될 상품을 선정하였다. 최근에 출시되어 유행에 민감한 상품의 경우 그렇지 않은 상품에 비해 판매 기간이 짧고, 이러한 데이터가 모델 학습을 방해한다고 판단하여 1) 전체 기간 108개월 중 15% 이상 판매가 발생한 상품 혹은 2) 연속 6개월 동안 판매된 상품을 대상으로 분석 대상이 될 상품을 선정하였다. 그 결과, 최종적으로 590개 상품이 분석 대상으로 선정되었다.

<Figure 3>은 분석 대상 상품 중 특정 상품(DMSHBBK)의 월별 판매량 추이를 나타낸 그래프이다. 특정 시점에서 판매량이 급격하게 상승했다가 급감하고 오랜 시간 후 다시 급상승하는 것을 볼 수 있으며, 판매량이 급격히 상승하는 구간이 유행구간으로 정의된 것을 확인할 수 있다. 본 연구에서는 계절성이 없고 비선형 패턴을 띠는 Fad 상품들을 학습 데이터로 선정하고 판매량이 급격히 상승하는 시점들을 추출하여 분석을 수행하였다.

4.2 실험 결과

상품별 유행구간을 추출하여 시점별 유행한 상품 벡터를 생성하였다. <Table 3>은 상품 벡터의 일부를 예시로 나타낸 것이다. 생성된 벡터는 Word2Vec의 Skip-gram으로 학습하였다. Word2Vec은 파이썬 오픈 라이브러리인 Gensim 라이브러리를 사용하였다. 학습 데이터에 한 번 이상 등장한 상품을 대상으로 은닉층은 100차원, 네거티브 샘플링은 5개로 학습은 100번 반복하도록 설정하였다. 윈도우 사이즈는 각 벡터를 한 번에 포함할 수 있도록 최대치인 9,999로 설정하였다.

Table 3. Example of Product Vectors

Year/Month	Product code
11/03	DMPAABI, HZTSDBK, ...
11/04	DMSHANV, MATRDNV, ...
...	...
19/12	DMSHBBK, MASHBGR, ...

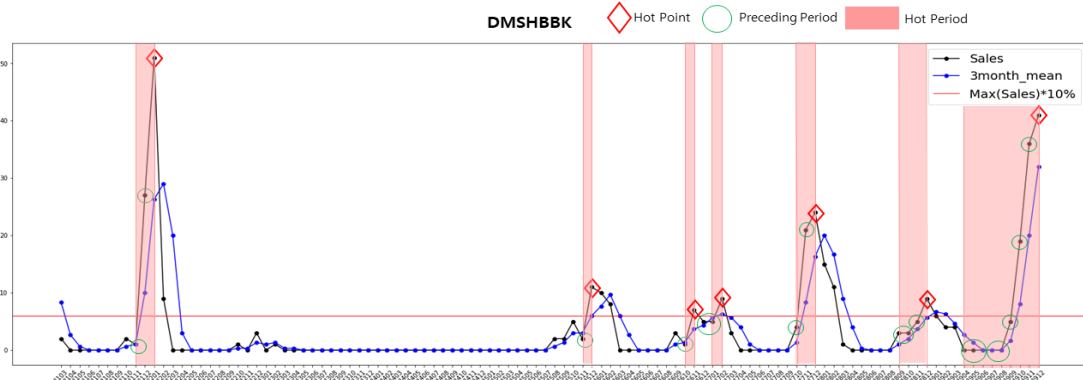


Figure 3. Example of Detecting Hot Periods from Sales Data

임베딩된 벡터의 값은 해당 상품의 유행 시점을 나타낸다. <Figure 4>는 t-SNE(t-distributed stochastic neighbor embedding)를 이용하여 상품별 임베딩 결과를 2차원 평면에 나타낸 것이다. t-SNE는 고차원 데이터를 쌍별 유사성 행렬로 변환하여 2차원에 축소하는 시각화 방법이다(Van der Maaten *et al.*, 2008). <Figure 4>를 보면 여러 Fad 군집이 나타난 것을 확인할 수 있으며, Fad는 하나의 상품으로 나타나는 것이 아니라 여러 품목의 상품 집합으로 이루어져 있음을 알 수 있다. Fad 내 특정 상품이 갑자기 유행하게 되면 상품과 코사인 유사도를 통해 유사한 상품 벡터 상위 N개를 Fad 군집 내에서 찾는다. 이는 학습 기간 내 유행 시기가 가장 비슷하게 겹친 상품을 탐색하는 것을 의미한다.

<Figure 5>는 <Figure 4>의 그래프에서 표시된 영역을 확대한 것이다. 예를 들어, 임베딩된 상품 벡터들 중 “DMSBBK”라는 상품이 갑작스럽게 판매량이 상승하면 “DMSHBBK”와 임베딩 벡터값이 가장 비슷한 상품 벡터를 코사인 유사도 기반으로 탐색한다. “DMSHBBK”와 가장 비슷한 벡터값을 갖는 상품인 상위 5개 제품인 “MASHBGR”, “JNSHBWT”, “DMPABBL”, “HZJUBBL”, “HZPABBK”는 유행 시기가 비슷한 상품들로 “DMSHBBK”의 유행과 맞추어 상품 기획과 판매 계획을 조정할 수 있다.

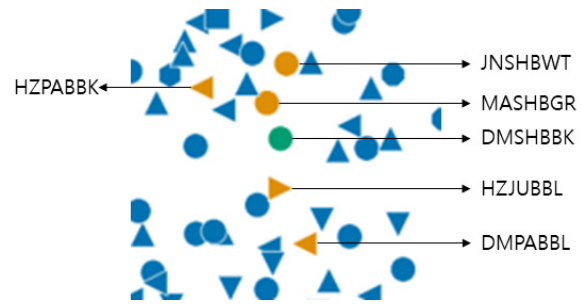


Figure 5. Fad Group of “DMSHBBK”

4.3 검증

본 연구에서는 Fad 탐지를 위한 Fad2Vec 기법의 타당성을 검증하기 위해 두 가지 방법으로 검증을 수행하였다. 첫 번째로 저 특정 상품이 유행했을 때 3개월 이내에 같은 Fad 군집에 속하는 상품이 실제로 같이 유행했는지, 즉 적중률(hit ratio)을 측정하였다. 먼저 3개월 검증 기간의 유행을 예측하기 위한 입력 데이터로써 3개월의 첫 달을 사용하며, 이를 기준연월이라고 한다. Fad2Vec 모델은 기준연월 전 N년 동안, 각 시점에 유행했던 상품만 추출해 학습시킨다. 그 후, 학습된 모델에 기준연월에 유행한 상품을 입력하여 코사인 유사도 기반 상위 N개 상품만을 도출한다. 본 연구에서는 기준연월 전 6년(72개월) 동안 유행했던 상품을 추출하여 모델을 학습시켰고 학습된 모

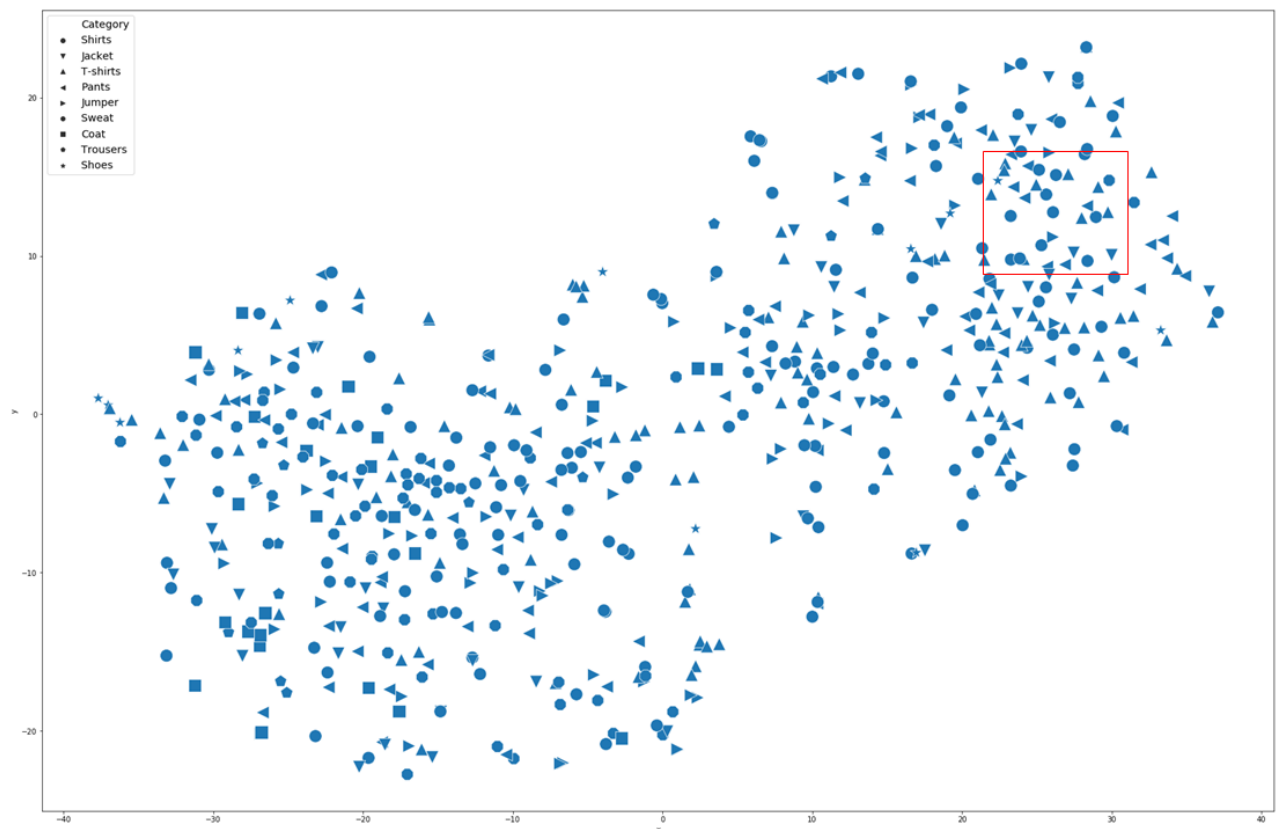


Figure 4. Visualization of Product Embedding Using t-SNE

델을 이용하여 코사인 유사도 기반 상위 5개 상품을 추출하였다. 적중률은 모델에서 입력된 상품과 같이 유행할 것으로 예측된 상품의 수를 실제 유행한 상품의 수로 나눈 것으로 계산된다. 이를 통해 기준연월에 유행한 모든 상품의 적중률이 구해지면 이들의 평균값을 해당 기준연월의 적중률로 정의한다.

위의 방법을 통해 17년 3월부터 19년 9월까지의 적중률을 계산하였다. 평균 적중률은 0.695이며, 기간별 적중률에 대한 세부 사항은 부록 A에 첨부하였다. 이는 Fad2Vec을 통해 3개월 내에 유행할 것으로 예측된 상품 10개 중 7개가 실제로 유행했다는 것을 의미하며, 따라서 Fad2Vec을 통한 Fad 탐지 결

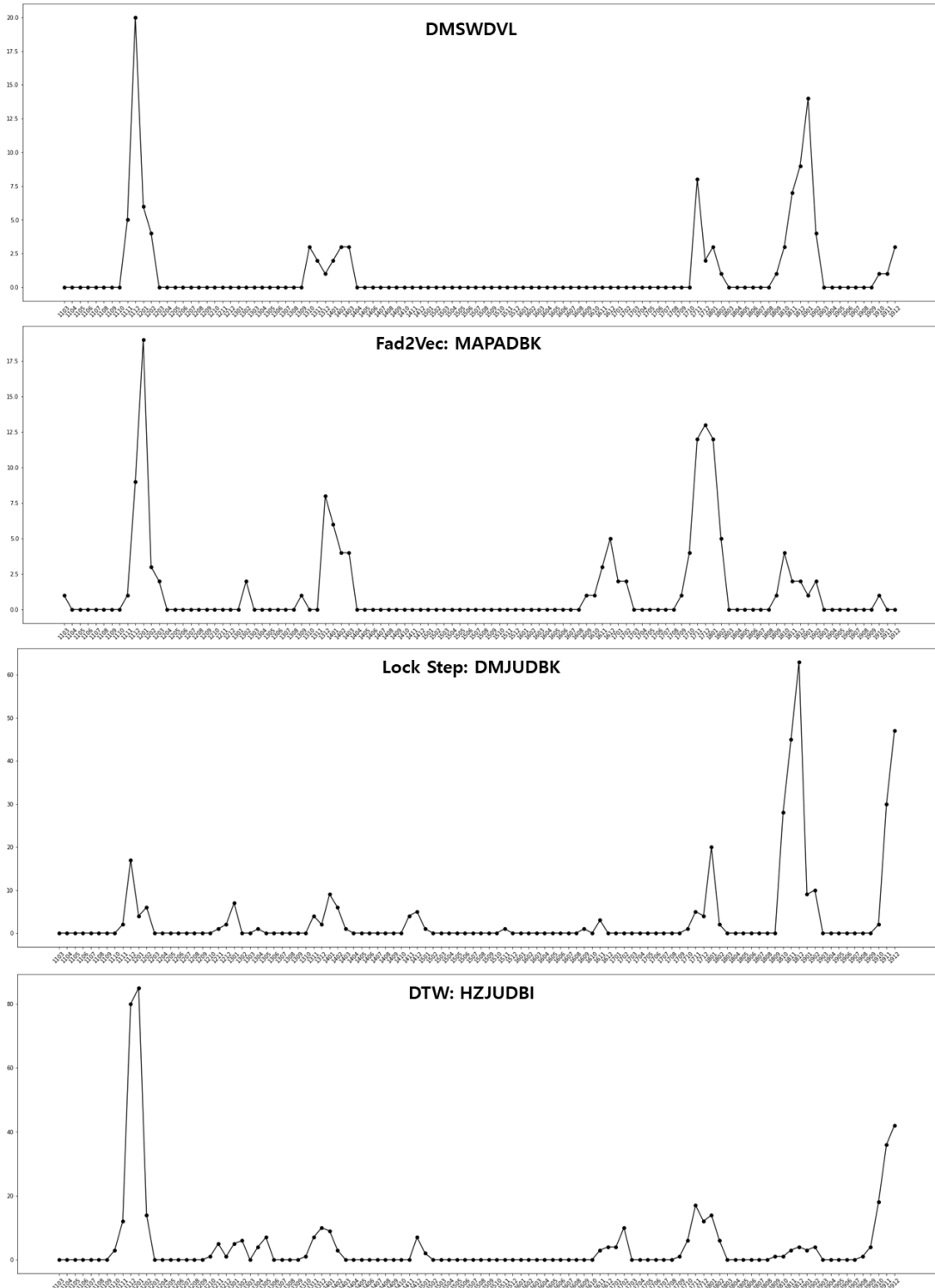


Figure 6. Comparison among Fad2Vec, Lock Step, and DTW

과의 활용도가 매우 우수하다고 할 수 있다.

두 번째로는 Fad2Vec과 기존의 시계열 유사도 측정법 적용 결과를 시각적으로 비교하였다. 비교대상 시계열 유사도 측정법으로는 신호들의 파장 유사성 분석 및 음성인식 분야에서 사용되는 Lock-step 방식과 동적시간왜곡(dynamic time warping: DTW) 방식의 유클리디안(euclidean) 거리 측정법을 사용하였다. Lock-step은 n번째 점을 순서대로 비교하는 방식이며, DTW는 고점과 저점을 기준으로 1:N 시점으로 비교가 가능한 방식이다.

<Figure 6>은 예시로써 “DMSWDVL” 상품과 가장 유사한 판매 패턴을 보이는 상품을 추출한 결과를 나타낸 것이다. Fad2Vec, Lock-step, DTW에서 “DMSWDVL”와 가장 유사한 상품으로 각각 “MAPADBK”, “DMJUDBK”, “HZJUDBI”가 도출되었다. 실제 “DMSWDVL”과 가장 유사한 Fad 패턴을 보이는 것은 Fad2Vec을 통해 도출된 “MAPADBK”로, 이로부터 Fad2Vec이 불규칙한 시계열 데이터의 패턴을 식별함에 있어 매우 효과적임을 확인할 수 있다.

5. 결론

본 연구에서는 Item2Vec 방법론을 차용한 Fad2Vec 방법론을 제시하였다. 2011년 01월부터 2019년 12월까지의 판매 데이터에서 시점별 유행 상품을 추출하여, 벡터를 만든 후 Item2Vec 방식으로 학습시켜, 동일 시점에 유행한 상품들이 높은 유사도를 가지는 임베딩 벡터값을 갖도록 하였다. 적중률 측정 및 기존의 시계열 유사성 측정기법을 비교한 결과, Fad2Vec이 단발성 유행을 탐지하는 데 있어 매우 효과적임을 확인하였다.

본 연구는 기존의 통계적 모델로는 예측이 어려웠던 Fad를 탐지하고 대응하기 위한 효과적인 접근 방법을 제시하였다는 측면에서 의의를 가진다. 판매량 예측을 위해 일정한 계절성을 가지거나 과거 판매 데이터가 충분히 확보되어야 하는 기존 방법에 비해서 Fad2Vec은 판매 주기가 매우 불규칙적이고 과거 판매 데이터가 불충분하더라도 유사한 주기를 가진 상품을 제안함으로써 Fad에 대응할 수 있다. 따라서 Fad2Vec이 Fad 대응 전략에 필요한 정량적이고 객관적인 예측 데이터를 제공함으로써 패스트 패션 기업의 상품 기획 및 생산 계획에 유용하게 활용될 수 있을 것으로 기대된다.

그러나 본 연구에는 향후 연구에서 극복해야 할 몇 가지 한계점이 있다. 첫째, 판매 데이터에 존재하는 상품코드만 활용하여 상품의 모든 특성을 상품 벡터에 반영하지 못했다. 본 연구에서는 브랜드, 품목, 시즌 및 색상만을 활용했으나 실제 의류 및 액세서리에는 핏과 소재 등 다양한 특성들이 존재한다. 따라서 패션 아이템의 구매 요인으로 작용하는 특성들을 추가한 상품 벡터를 생성한다면 좀 더 세밀하게 상품의 유행 주기를 탐색할 수 있을 것으로 보인다. 둘째, 유행구간인 유행구간을 지정하는 과정에서 3개월 이동평균과 최고 판매량의 10%

를 조건으로 사용했다. 그러나 카테고리별로 다른 조건을 적용하는 등의 더 정교한 방식의 유행구간 추출방식이 필요할 것으로 보인다. 셋째, 본 연구에서는 각 제품의 월별 판매량 그래프에서 보이는 계절성을 고려하지 않았다. 따라서 시계열 분석기법을 활용하여 계절성을 제거한 후 분석 및 검증을 수행하고, 수정 전 Fad2Vec과 비교하는 방식으로 Fad2Vec의 성능을 다각도로 검증할 필요가 있다.

참고문헌

- Anderson, A., Maystre, L., Anderson, I., Mehrotra, R., and Lalmas, M. (2020), Algorithmic Effects on the Diversity of Consumption on Spotify, In *Proceedings of The Web Conference*, 2020, 2155-2165.
- Anvari, S. and Amirkhani, H. (2018), Book2Vec: Representing Books in Vector Space Without Using the Contents. In *2018 8th International Conference on Computer and Knowledge Engineering (ICCKE)*, IEEE, 176-182.
- Bailey, T. (2001), Organizational innovation in the apparel industry, *Industrial Relations*, **32**(1), 30-48.
- Barkan, O. (2017), Bayesian neural word embedding, In *Proceedings of the AAAI Conference on Artificial Intelligence*, **31**(1), 3135-3143.
- Barkan, O. and Koenigstein, N. (2016), Item2Vec: Neural item embedding for collaborative filtering. In *2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP)*, IEEE, 1-6.
- Barnes, L., Lea-Greenwood, G., Tyler, D., Heeley, J., and Bhamra, T. (2006), Supply chain influences on new product development in fashion clothing, *Journal of Fashion Marketing and Management: An International Journal*, **10**(3), 316-328.
- Bhardwaj, V. and Fairhurst, A. (2010), Fast fashion: Response to changes in the fashion industry, *The International Review of Retail, Distribution and Consumer Research*, **20**(1), 165-173.
- Blázquez, M. (2014), Fashion shopping in multichannel retail: The role of technology in enhancing the customer experience, *International Journal of Electronic Commerce*, **18**(4), 97-116.
- Collobert, R. and Weston, J. (2008), A unified architecture for natural language processing: Deep neural networks with multitask learning, In *Proceedings of the 25th International Conference on Machine Learning*, 160-167.
- Frank, C., Garg, A., Sztandera, L., and Raheja, A. (2003), Forecasting women's apparel sales using mathematical modeling, *International Journal of Clothing Science and Technology*, **15**(2), 107-125.
- Gutierrez, R. S., Solis, A. O., and Mukhopadhyay, S. (2008), Lumpy demand forecasting using neural networks, *International Journal of Production Economics*, **111**(2), 409-420.
- Loureiro, A. L. D., Miguéis, V. L., and da Silva, L. F. M. (2018), Exploring the use of deep neural networks for sales forecasting in fashion retail, *Decision Support Systems*, **114**, 81-93.
- Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013), Efficient estimation of word representations in vector space, arXiv preprint arXiv:1301.3781.
- Minner, S. and Kiesmüller, G. P. (2012), Dynamic product acquisition in closed loop supply chains, *International Journal of Production Research*, **50**(11), 2836-2851.
- Nenni, M. E., Giustiniano, L., and Pirollo, L. (2013), Demand forecasting

- in the fashion industry: a review, *International Journal of Engineering Business Management*, **5**, 37.
- Papalexopoulos, A. D. and Hesterberg, T. C. (1990), A regression-based approach to short-term system load forecasting, *IEEE Transactions on Power Systems*, **5**(4), 1535-1547.
- Rexhausen, D., Pibernik, R., and Kaiser, G. (2012), Customer-facing supply chain practices: The impact of demand and distribution management on supply chain success, *Journal of Operations Management*, **30**(4), 269-281.
- Sadeghian, A., Minaee, S., Partalas, I., Li, X., Wang, D. Z., and Cowan, B. (2019), Hotel2vec: Learning Attribute-Aware Hotel Embeddings with Self-Supervision, arXiv preprint arXiv:1910.03943.
- Solomon, M. R. and Rabolt, N. J. (2004), *Consumer behavior: In fashion*. Prentice Hall.
- Tehrani, A. F. and Ahrens, D. (2016), Enhanced predictive models for purchasing in the fashion field by using kernel machine regression equipped with ordinal logistic regression, *Journal of Retailing and Consumer Services*, **32**, 131-138.
- Thomassey, S. and Fiordaliso, A. (2006), A hybrid sales forecasting system based on clustering and decision trees, *Decision Support Systems*, **42**(1), 408-421.
- Van der Maaten, L. and Hinton, G. (2008), Visualizing data using t-SNE, *Journal of Machine Learning Research*, **9**(11).

저자소개

안승섭: 서울과학기술대학교 경영학과에서 학사학위를 취득하고 서울과학기술대학교 데이터사이언스학과에서 석사학위를 취득하였다. 연구 분야는 로지스틱스 애널리틱스, 비즈니스 애널리틱스 등이다.

진소연: 서울과학기술대학교 컴퓨터공학과에서 학사학위를 취득하였으며, 현재 서울과학기술대학교 데이터사이언스학과에 재학 중이다. 관심 분야는 텍스트 분석, 비즈니스 애널리틱스, 네트워크 분석 등이다.

이학연: 이학연 교수는 서울대학교 산업공학과에서 학사학위를 취득하였으며, 동대학원에서 박사학위를 받았다. 현재 서울과학기술대학교 산업공학과 교수로 재직 중이다. 주요 연구 분야는 기술예측, 이노베이션 애널리틱스, 디지털 혁신 전략 등이다.

<Appendix A> Hit ratio

Year/Month	Training period	Validation period	Hit ratio
17/03	11/03~17/02	17/03, 17/04, 17/05	0.74
17/04	11/04~17/03	17/04, 17/05, 17/06	0.81
17/05	11/05~17/04	17/05, 17/06, 17/07	0.83
17/06	11/06~17/05	17/06, 17/07, 17/08	0.75
17/07	11/07~17/06	17/07, 17/08, 17/09	0.47
17/08	11/08~17/07	17/08, 17/09, 17/10	0.58
17/09	11/09~17/08	17/09, 17/10, 17/11	0.69
17/10	11/10~17/09	17/10, 17/11, 17/12	0.77
17/11	11/11~17/10	17/11, 17/12, 18/01	0.84
17/12	11/12~17/11	17/12, 18/01, 18/02	0.84
18/01	12/01~17/12	18/01, 18/02, 18/03	0.62
18/02	12/02~18/01	18/02, 18/03, 18/04	0.5
18/03	12/03~18/02	18/03, 18/04, 18/05	0.69
18/04	12/04~18/03	18/04, 18/05, 18/06	0.79
18/05	12/05~18/04	18/05, 18/06, 18/07	0.77
18/06	12/06~18/05	18/06, 18/07, 18/08	0.77
18/07	12/07~18/06	18/07, 18/08, 18/09	0.55
18/08	12/08~18/07	18/08, 18/09, 18/10	0.59
18/09	12/09~18/08	18/09, 18/10, 18/11	0.66
18/10	12/10~18/09	18/10, 18/11, 18/12	0.78
18/11	12/11~18/10	18/11, 18/12, 19/01	0.77
18/12	12/12~18/11	18/12, 19/01, 19/02	0.72
19/01	13/01~18/12	19/01, 19/02, 19/03	0.62
19/02	13/02~19/01	19/02, 19/03, 19/04	0.57
19/03	13/03~19/02	19/03, 19/04, 19/05	0.66
19/04	13/04~19/03	19/04, 19/05, 19/06	0.77
19/05	13/05~19/04	19/05, 19/06, 19/07	0.76
19/06	13/06~19/05	19/06, 19/07, 19/08	0.72
19/07	13/07~19/06	19/07, 19/08, 19/09	0.55
19/08	13/08~19/07	19/08, 19/09, 19/10	0.59
19/09	13/09~19/08	19/09, 19/10, 19/11	0.72
19/10	13/10~19/09	19/10, 19/11, 19/12	0.75
Average			0.695