

카드 데이터 기반 심층 관광 추천 연구¹⁾

Card Transaction Data-based Deep Tourism Recommendation Study

홍민성 (Minsung Hong)

경희대학교 스마트관광연구소²⁾

김태경 (Taekyung Kim)

광운대학교 경영학부³⁾

정남호 (Namho Chung)

경희대학교 스마트관광원⁴⁾

〈 국문초록 〉

관광산업에서 발생하는 방대한 카드 거래 데이터는 관광객의 소비 행태와 패턴을 암시하는 중요한 자원이 되었다. 거래 데이터에 기반을 둔 스마트 서비스 시스템을 개발하는 것은 관광산업과 지식관리시스템 개발자들의 주요한 목표들 중 하나이다. 그러나 기존 추천 기법의 근간이 되어 온 평점을 활용하기 어렵다는 점은 시스템 설계자들이 학습 과정을 평가하기 어렵게 한다. 또한 시간적, 공간적, 인구통계학적 정보와 같이 추천 성과를 높일 수 있는 보조 요소들을 적절히 활용하는 방법도 어려운 상황이다. 이러한 문제들에 대하여 본 논문은 카드 거래 데이터를 기반으로 관광 서비스를 추천하는 새로운 방식인 CTDDTR을 제안한다. 먼저 Doc2Vec를 이용하여 시간성 선호도를 임베딩하여 관광객 그룹과 서비스 벡터로 데이터를 표현하였다. 다음 단계로 딥러닝 기술 중 하나인 다중 계층 퍼셉트론을 도입하여 얻어진 벡터와 관광 RDF로부터 도출한 보조 요소를 통합하여 심층 추천 모듈을 구성하였다. 추가로, 지식경영 분야의 RFM 분석 기법을 심층 추천 모듈에 도입하여 심층 신경망을 학습하는데 사용되는 평점을 생성함으로써 평점 부재 문제에 대응하였다. 제안한 CTDDTR의 추천 성능을 평가하기 위해 제주도에서 8년 동안 발생한 카드 거래 데이터를 사용하였고, 제안된 방법의 우수한 추천 성과와 보조 요소의 효과를 증명하였다.

주제어: 스마트 관광 추천, 심층 학습, 자원 기술 프레임워크, RFM 분석

1) 이 논문은 2019년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2019S1A3A2098438)

이 논문은 2022년 광운대학교 교내학술연구비 지원에 의해 연구되었음(2022-0142)

2) 제1저자, mshong.res@gmail.com

3) 교신저자, kimtk@kw.ac.kr

4) 제3저자, nhchung@khu.ac.kr

1. 서론

관광 산업은 세계 GDP의 약 10.5%를 차지하고, 14억 6천만 명의 관광객에게 서비스를 제공하는 거대한 사업 영역이다. 지식 시스템의 발전으로 여행객은 여행 목적지와 숙박, 레스토랑, 교통 또는 이벤트와 같은 관련 사항을 포함하여 많은 것들을 스스로 결정하는 추세다. 정보통신기술과 인터넷 서비스의 진보는 관광객이 보다 풍부한 정보에 쉽게 접근할 수 있는 길을 열었다고 평가된다(Hong & Jung, 2021b). 그러나 너무 많은 정보를 쉽게 접근할 수 있다는 사실은 한편 문제가 되기도 한다. 개인은 관심사에 맞는 관광지 정보나 서비스를 찾고 결정하는 일에 많은 시간을 들이고 책임을 스스로 지게 되고(Esmeli et al., 2020), 정보의 과부하 문제도 겪기 때문에 추천 시스템에 의존하는 경우가 점차 많아지고 있다. 이는 관광객의 명시적 피드백(평점)이나 암시적 피드백(클릭, 장바구니 등)을 이용하여 그들의 요구에 맞는 적절한 항목을 자동으로 제공하기 위한 시스템의 설계를 필요로 한다(Cai et al., 2018; Hong, 2021; Hong & Jung, 2022). 이러한 시스템은 사용 가능한 정보를 기반으로 관광객의 선호도를 예측하기 때문에 예측 정확도는 사용되는 정보에 따라 다를 수 있다(Thasal et al., 2018). 거래 결제 데이터를 활용할 경우 서비스 이용 사실에 따른 추천 시스템을 개발할 수 있는 장점은 있으나 평점 데이터를 얻기 곤란하다는 문제가 있어 예측 정확도가 떨어질 가능성도 있다(Sharifhosseini, 2019).

영화나 책과 같은 전통적인 상품을 제공하는 것과 달리 관광 서비스의 추천이 쉽지 않다는 사실이 반복적으로 보고되고 있다(예, Tan et al., 2014; Liu et al., 2014; Zhu et al., 2017; Hong & Jung, 2021a). 예를 들어, 관광 서비스는 보통 높은 가격, 많은 시간 소비뿐만 아니라 관광객과 관광 서비스 간의 상호작용 또한 상

대적으로 희소하다는 점에서 추천 정확도를 높이기 쉽지 않다(Rakesh et al., 2017; Zhao et al., 2021). 또한, 관광 서비스를 추천하는 것은 출발지, 목적지의 풍경, 여행 시즌 등의 상황에도 영향을 받는다. 시간 정보는 시간 경과에 따른 관광객 활동과 선호도의 변화를 이해하고 파악하는데 도움이 되며, 이러한 변화는 인위적인 효과나 환경 변화 요인에 의해 발생할 수 있다(Li et al., 2019). 친구의 추천이나 관광 유행, 그들이 속한 연령대에 따라 관광객이 좋아하는 여행 스타일이 다를 수 있는데 이는 다른 맥락 정보나 평가 사실 등을 폭넓게 활용해야 원하는 수준의 성능을 낼 수 있다는 점을 암시한다. 지식 경영 연구 분야에서 관광객을 대상으로 한 추천 시스템 개발 연구는 새로운 관심은 다양한 지식을 융합할 수 있는 유연한 추천 프레임워크를 요구하고 있다(Chen et al., 2020).

본 연구는 엄격히 개인 정보를 보호해야 하고 거래 데이터로부터 개인을 식별하는 것이 불가능하다는 가정을 기반으로 한다. 따라서 일반 데이터 보호 규정(GDPR : general data protection regulation)에 따라 익명 처리된 카드 거래 데이터를 제공받아 실험에 사용했다. 이와 같은 맥락에서 대규모 카드 거래 데이터를 활용한 관광객 그룹에 관광서비스를 제공하는 CTDDTR(card transaction data-based deep tourism recommendation)이라는 추천 방법을 제안하며, 이는 다음과 같이 두 가지로 구성된다. 첫 번째는 시간성을 고려하여 카드 거래 데이터를 모델링하고, 자연어 처리 기술인 Doc2Vec을 활용하여 관광 그룹과 관광 서비스를 고정된 크기의 실수 값을 갖는 벡터로 모델링한다. 다음으로 해당 결과에 대하여 관광 산업 관련하여 지식 그래프로 표현된 RDF(resource description framework)를 바탕으로 각종 보조 정보(예를 들어, 인구통계학 정보, 도착지 등)를 추가로 고려한다. 이후 심층 신경망 중 하나인 다중 계층 퍼셉트론(MLP : multi-layer perceptron)을 이용하

여 이들을 통합한다. 이러한 과정을 거쳐 관광 서비스에 대한 관광객 그룹의 선호도를 예측하게 된다. 평점 데이터 문제에 대응하기 위해 Sharifhosseini(2019)의 연구를 바탕으로 RFM 모델을 응용한 머신 러닝 접근법을 취하고 시간과 맥락, 신경망 학습 방법을 복합적으로 사용한다. 결과 평가를 위해 제주도에서 8년간 발생한 약 1,500만 건의 카드 거래 데이터를 사용했다. 성능 평가는 다양한 관점에서 최신 추천 알고리즘들과의 비교 실험을 하는 방식을 채택했다. 실험 결과를 통해 제안한 CTDDTR이 비교 추천 알고리즘보다 뛰어난 성능을 보이고, 각 보조 정보의 활용이 추천 성능을 높이는 데 긍정적인 영향을 미친다는 점을 알 수 있었다.

본 연구는 지식 저장소에 축적된 카드 거래 데이터를 추천 시스템 개발로 연결시킬 때 평점 데이터 결손이 가져올 문제를 해결하기 위하여 지식 경영 분야의 RFM 프레임워크 기반으로 생성된 평점 데이터를 활용하는 아이디어를 제공하고 이를 실험적으로 검증한다는 점에서 의의가 있다. 구체적으로 관광 영역에서 발생하는 정보 처리와 의사결정을 지원하는 새로운 방식을 탐구하고 이를 통한 응용프로그램 개발에 필요한 접근법을 보여준다는 점에서 연구 공헌을 찾을 수 있다. 실무적 측면에서는 익명성을 보장하면서도 카드 거래 데이터를 활용하는 방법에 관한 시사점을 제공하였고, 지식 그래프인 RDF를 관광 영역에서 활용하는 일에 대한 유효성 검증을 수행했다는 점에서 지식 데이터 활용이 필요한 근거를 제공하였다.

2. 기존문헌 연구

2.1. 관광 추천 시스템

지난 수십 년 동안 관광 산업은 대규모로 성장했으

며 매우 많은 관광 서비스가 물리적 또는 가상으로 제공되고 있다. 그러나 서비스의 수가 많을수록 관광객이 자신의 흥미에 맞는 적절한 서비스를 찾고 결정하기가 더 어려워지고 있으며, 이러한 관광객의 노력을 줄이기 위해 스마트 관광 분야에 추천 시스템을 적용하는 시도는 계속되고 있다(Guo et al., 2019; Hong & Jung, 2021a). 이같은 추천 시스템은 관광객의 명시적 피드백(평점)이나 암시적 피드백(클릭, 장바구니 내역 등)을 분석하여 다양한 관광 상품(관광 명소, 레스토랑, 박물관 등) 장보를 자동으로 제공한다(Chen et al., 2021; Hong & Jung, 2021b). Lu et al.(2015)과 Chaudhari and Thakkar(2019)의 연구에 따르면, 관광 추천 시스템은 콘텐츠 기반, 협업 필터링, 도메인 및 하이브리드 방식과 같은 네 가지 접근 방법으로 나뉜다. 일반적인 산업 도메인(예 : 영화)에서의 추천 시스템과 마찬가지로 콘텐츠 기반 접근 방식은 아이템을 추천하기 위해 아이템의 특성(예 : 장르, 감독, 개봉 연도 등)을 기반으로 유사한 아이템 추천하는 반면, 협업 필터링은 유사한 관심사를 갖는 사용자가 과거에 이용한 아이템을 추천한다. 도메인별 접근 방식은 다양한 보조 정보(예: 맥락 정보, 시간 정보, 위치 정보, 소셜 정보 등)를 활용하여 추천 품질을 향상시킨다. 마지막 하이브리드 방식은 앞서 언급한 각 방식의 단점을 극복하고 더 나은 추천 성능을 달성하기 위해 이들을 병렬 또는 순차적으로 결합한다(Esmail et al., 2020).

관광 여정 추천의 경우 하이브리드 방식이 적절하게 사용된다. 예를 들어 Kotiloglu et al.(2017)은 후보지 선호도와 데이터 기반 협업 필터링 방법을 적절히 섞는 방식을 취한다. 구체적으로 살펴보면, 이는 여러 날에 걸친 여행 기간의 여행 일정을 개인화하는 방법으로 협업 필터링에 기반을 둔 FFTS(filter-first, tour-second) 프레임워크로 요약될 수 있다. 즉, 관광객이 선택한 필수 방문지 후보가 주어지면 이에 대한 선

호도를 추정하고, 만족도를 최대화할 것으로 예상되는 선택 가능한 방문지를 추천하는데 이때 선택 가능한 방문지는 사용자의 소셜 미디어 및 온라인 데이터에 기반을 둔 아이템 기반 협업 필터링으로 정해진다. Kotiloglu et al.(2017)은 이때 관광객의 복잡한 선택을 처리하기 위해 반복 타부 검색 알고리즘(iterated tabu search algorithm)을 이용하였다. 또한 콘텐츠 기반, 협업 필터링 및 지식 기반 솔루션을 결합한 하이브리드 접근 방식도 고려될 수 있다(Pessemier et al., 2017). 이는 개인과 단체 여행객을 위한 관광지 추천 시스템을 개발할 때 활용될 수 있는 방법으로 사용자 프로필과 개인 관심사, 다음 목적지에 대한 특정 요구 사항 등, 개인의 선호를 비롯한 선택지들을 복합적으로 고려하여 관광지를 추천하자는 것이다.

개별적인 추천 알고리즘이나 하이브리드 방식 모두 평점과 같은 명시적 피드백은 추천 후보지를 선택할 때 고려할 수 있는 가치 있는 정보로 간주된다. 그러나 실제로 관광지에 방문하여 구매 의사결정을 한 흔적을 보여주는 거래 데이터(transaction data)에서는 평점을 발견하기가 쉽지 않으며 이를 「평점 희소성 문제(rating scarcity problem)」로 간주된다. 평점과 같은 명시적 피드백에 기반을 둔 관광 추천 시스템은 평점 희소성 문제로 인해 관광객 선호도가 충분히 수집되지 않을 수 있다. 희소성이란 사용자별 평점의 양을 의미하는데, 예를 들어 관광객은 1년에 한두 번만 호텔을 예약하는 경우가 많으며, 이는 해당 관광객의 선호도를 분석하는데 데이터가 충분히 수집되기 어렵다는 것을 의미한다. 평점 희소성 문제를 해결하기 위해 Al-Ghossein et al.(2018)은 호텔 추천을 위해 식당이나 관광지와 같은 다른 도메인의 피드백을 활용하는 「교차 도메인 추천 시스템」을 제안했다. 일반적으로 관광객이 방문할 목적지를 선택한 뒤에 호텔을 검색한다고 가정했으며, 목적지 선택이 호텔 선택에 중요한 요

소를 고려하여 위치 기반 소셜 네트워크 데이터를 바탕으로 관광객의 이동 패턴을 학습하였다. 추천은 이와 같은 이동 패턴을 기반으로 이루어진다.

2.2. 심층 학습 기반 관광 추천

심층 학습(deep learning)은 정보 검색 및 추천 시스템에 적용되어 우수한 효과를 입증하고 있으나(Zhang et al., 2019), 상품(Katarya & Arora, 2020), 영화(Tahmasebi et al. 2021), 음악(Fessahaye et al., 2019)과 같은 다른 산업 영역에 비해 관광 추천에 이를 적용한 사례는 상대적으로 찾기 어렵다. 확인된 사례들은 심층 학습 기술을 이용하여 사용자들의 피드백을 직접적으로 학습하기보다 고객 리뷰에서 정보를 추출하거나(예, Wang, 2020) 관광 사진 분류(예, Fudholi et al., 2021), 상위 레벨의 시각 정보 추출(예, Zhao et al., 2021) 등 간접적인 서비스 개발에 국한된다.

한편, 일부 연구자들은 심층 학습에 기반을 둔 「다중 기준 추천 방법」을 제안하였다. Shambour(2021)는 순방향 신경망(feed-forward neural network)과 오토 인코더(autoencoder) 기술에 기반을 둔 다중 기준 추천 기법인 AEMC(autoencoder-based multi-criteria recommendation)를 제안했다. 각 기준별 평점을 독립적으로 예측하기 위해 각각에 대한 오토인코더 처리단위를 구성하고, 예측한 평점의 산술 평균을 최종 예측 평점으로 이용하는 방식이다. 여행지 리뷰 플랫폼인 TripAdvisor로부터 수집한 다중 기준(호텔 서비스, 위치, 가격, 청결도)에 대한 평가 점수를 학습하여 호텔을 추천할 때의 예측 정확도를 평가하고 그 결과가 우수성함을 주장하였다. Nassar et al.(2020)은 DMCF(deep multi-criteria collaborative filtering)를 제안했는데 이는 다중 기준 데이터와 두 개의 심층 인공 신경망들을 이용하여 평점을 개별적으로 예측하는 방법이다. 구체적으로

TripAdvisor의 호텔 리뷰 데이터에서 DMCF의 성능을 실험하였고 다중 기준의 결합으로 추천 성능이 크게 개선됨을 확인했다.

추천 시스템 개발에 주로 도입되는 행렬 인수분해(matrix factorization)방법과 마찬가지로 단어 임베딩(word embedding) 기술은 입력 요소를 학습하여 저차원의 벡터 공간으로 표현하기 위해 사용된다(Ozsoy, 2016). 단어 임베딩은 빅데이터에 포함된 문자 데이터를 효과적으로 처리하기 위해 사용되며 자연어 처리 기술에서 중요한 위치를 차지하고 있다(Musto et al., 2015; Park & Liu, 2020; Back & Chung, 2020). Musto et al.(2015)은 콘텐츠 기반 추천에서 잠재 의미 색인(latent semantic indexing)과 무작위 색인(random indexing), Word2Vec을 실증적으로 비교했다. 실험 결과에서는 Word2Vec이 MovieLens 및 DBbook 데이터 세트에서 더 우수한 성능을 보였다. Back & Chung(2020)은 Word2Vec 기반 소셜 관계 마이닝을 사용하여 멀티미디어 콘텐츠를 추천하는 알고리즘을 제안했다. Word2Vec 기법을 적용하여 변환된 감정 단어의 벡터를 이용하여 사용자 간의 유사도 및 신뢰 관계를 구하고, 이를 기반으로 멀티미디어 콘텐츠를 추천한 것이다. 또한 자연어 처리 기술을 응용하여 사용자 선호도를 직접적으로 나타내는 상품 이용 내역을 문장으로, 상품을 단어로 간주하고 사용자와 상품에 대한 벡터 학습에 단어 임베딩 기술을 사용한 경우도 있다(Ozsoy, 2016; Esmeli et al., 2020; Vuong et al., 2021). 이 방법은 평가 점수와 같은 명시적 피드백이 없는 경우에도 다양한 추천 도메인에 쉽게 적용할 수 있는 것으로 보인다.

단어 임베딩 방법을 응용하여 사용자 선호도를 활용한 Ozsoy(2016)의 접근법은 평가 점수가 주어지지 않는 카드 거래 데이터를 활용하는 경우에도 적지 않은 시사점을 준다. 따라서 본 연구는 Ozsoy(2016)와 마찬가지로 단어 임베딩 접근법의 Doc2Vec 기술을 적용

한다. 여기에서 나아가 지식 그래프를 활용하여, 관광 산업 RDF에서 도출한 보조 요소들을 통합하고 관광객 그룹의 관광 서비스에 대한 선호도를 예측하기 위해 인공 신경망 중 하나인 MLP를 적용한다. 제안한 기법의 공정한 평가를 위해 RFM 기반 추천 알고리즘(Sharifhosseini, 2019), 최신 심층 학습 기반 추천 알고리즘인 Shambour(2021)의 AEMC, Nassar et al.(2020)의 DMCF과 및 Liu et al.(2018)의 시간성을 반영한 세션 기반 추천 모델의 성능을 함께 비교한다. 여기에 더해 CTDDTR의 여러 변형을 평가하고 제안한 기술의 주요 구성 모듈의 효능도 검증한다.

3. 시간성 선호도 임베딩

3.1. 카드 거래 데이터 분석

본 절에서는 Le and Mikilov(2014)가 제안한 Doc2Vec을 기반으로 관광객과 서비스 항목 간의 상호작용을 벡터로 표현하는 시간성 선호도 임베딩(TE : temporal preference embedding) 접근에 대하여 상술한다. 관광객의 서비스 이용 내역은 단어 임베딩 시의 문서처럼 다루어지고 관광 서비스는 단어처럼 다룬다. 즉, 각 내역은 관광 서비스라는 단어 요소들이 시간 순으로 나열된 것과 같이 표현된 데이터로 취급된다. Misztal- Radecka et al.(2021)가 사용자와 클릭한 상품을 문서 및 단어로 간주하는 것과 유사하게, 카드 거래 데이터에 포함된 관광 그룹 및 서비스는 TE에서 문서 및 단어로 간주한 것이다. 이때, 문서 내 단락은 특정 시간 간격으로 구분하여 시간 요소가 반영되도록 조정된다. 이러한 방법은 시간 간격이 모델링에 큰 영향을 주기 때문에 사용하는 데이터 세트의 시간적 특징을 고려하여 설정해야 한다. 따라서 우리는 관광 산업과 관련된 카드 거래 데이터만

사용하기 위해, 숙박과 음식점, 관광 명소 등 관광 산업과 관련된 데이터만 통계적으로 추출하였다. Zheng et al.(2018)이 시간적 역학을 활용하기 위해 시간 간격으로 관광객 수를 분석한 것과 유사하게 본 연구는 <그림 1>과 <그림 2>와 같이 월별로 카드 거래 건수를 분석하였다.

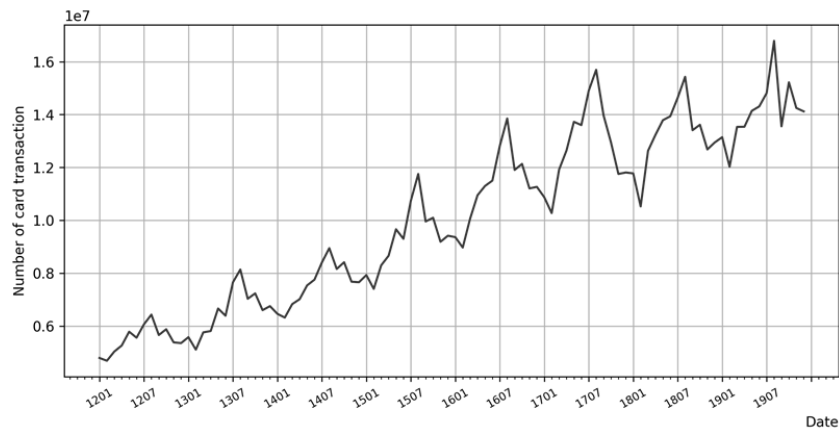
<그림 1>로부터 관광 산업 관련 카드 거래 수가 8년 동안 주기적, 점진적으로 증가했음을 알 수 있다. 유사한 주기는 <그림 2>의 왼쪽에서도 관찰된다. 데이터는 제주도에서 추출된 것으로 이때 관광 성수기와 비수기에 따라 반복적 특성이 드러난다. 또한 <그

림 2>의 오른쪽에서 볼 수 있듯이 5월-6월과 11월-12월을 제외하고 인접 월간 격차의 대부분은 크게 다르다. 따라서 관광서비스 소비는 계절적 영향을 받으며 카드거래 경향은 월별로 시간에 민감함을 유추할 수 있다. 이를 다음과 같은 규칙으로 표현한다.

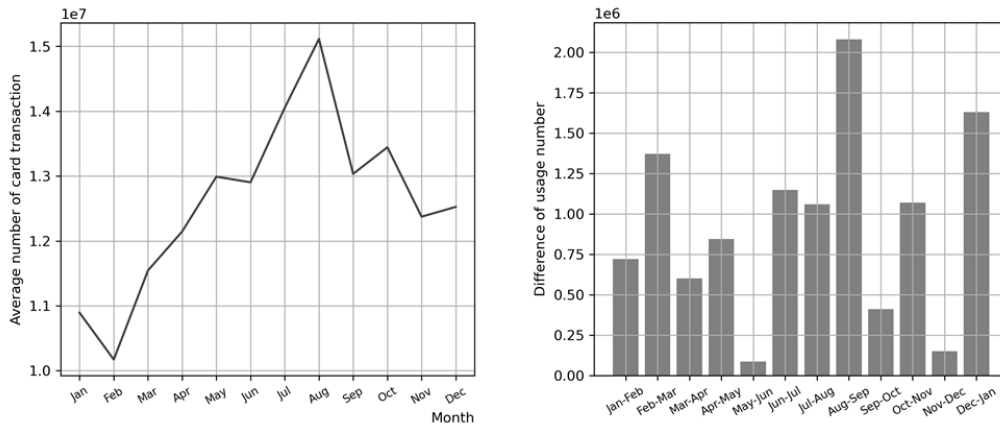
규칙 1: 관광 서비스 소비에는 계절적 경향이 있으며 시간이 지남에 따라 카드 거래 증가와 함께 반복된다.

규칙 2: 이러한 경향은 시간에 따라 월별 또는 계절별로 구분할 수 있다.

단어 임베딩 기술을 자연어 처리가 아닌 다른 맥락에 적용하기 위해 위와 같은 규칙 명제의 타당성이 확



<그림 1> 8년간의 관광 분야의 월별 카드 거래 수



<그림 2> 월별 평균 거래 수와 거래 수 변화량

보될 필요가 있으며, 현재 분석을 진행하는 제주도 관광 데이터 셋은 이를 충족한다고 볼 수 있다.

3.2. 시간적 선호도 학습

TE의 시간적 선호도 학습 방식은 개월별 또는 계절별로 여러 단락으로 구분하여 사용자 그룹과 서비스 벡터를 학습하는 절차를 취한다. 보다 구체적으로 <그림 3>은 TE의 입력 데이터가 처리되는 방식을 보여준다. 여기에서 userG는 관광객 그룹을, ser은 관광 서비스를 나타내는 첫머리 글자들이다.

```

단계 1. 그룹과 서비스를 문서 라벨과 단어로 설정:
(userG1, userG2, ...)
{ser1, ser2, ser3, ser4, ser5,ser6, ser7, ser8, ser9, ser10, ...}

단계 2. 시간적 거래 데이터를 그룹별로 생성:
userG1: ser1, ser2, ser3, ser4, ser5, ser2, ser6, ser7, ser10, ser5, ser8, ser9
userG2: ser3, ser4, ser6, ser8, ser10, ser1, ser3, ser7, ser4, ser4, ser7, ser9, ser2

단계 3. 시간 간격에 따라 그룹별 거래 데이터를 문단으로 분할:
userG1:
[ser1, ser2, ser3, ser4, ser5]
[ser2, ser6, ser7, ser10]
[ser5, ser8, ser9]
userG1:
[ser3, ser4, ser6, ser8, ser10]
[ser1, ser3, ser7]
[ser4, ser4, ser7, ser9, ser2]

단계 4. Doc2Vec 입력 데이터 생성 (문단):
para1 (userG1, period1) / [ser1, ser2, ser3, ser4, ser5, (userG1, period1)]
para2 (userG1, period2) / [ser2, ser6, ser7, ser10, (userG1, period2)]
para3 (userG1, period3) / [ser5, ser8, ser9, (userG1, period3)]
para4 (userG2, period1) / [ser3, ser4, ser6, ser8, ser10, (userG2, period1)]
para5 (userG2, period2) / [ser1, ser3, ser7, (userG2, period2)]
para6 (userG2, period3) / [ser4, ser4, ser7, ser9, ser2, (userG2, period3)]

단계 5. Doc2Vec 모델을 학습하고 그룹과 서비스의 표현 벡터 생성:
userG1: [float 1, float 2, ..., float n]
userG2: [float 2, float 2, ..., float n]
ser1: [float 1, float 2, ..., float n]
ser2: [float 1, float 2, ..., float n]
    
```

<그림 3> 카드 거래 데이터 예시와 처리 절차

시간적 요인의 영향을 반영하기 위해 <그림 3>과 같이 카드 거래 데이터를 전처리하고 Doc2Vec에 전달한다. 먼저 첫 번째 단계와 같이 문서 레이블 및 단어로써 사용자 그룹과 서비스의 ID를 설정한다. 사용자 그룹이 사용하는 서비스는 해당 그룹의 상호 작용으로써 문서를 구성하기 위해 시간 순으로 정렬된다. 다음으로 문서에 포함된 서비스 ID는 지정된 기간별(즉, 3.1절에서 분석한 월별 또는 계절별)로 문단으로 분리된다. <그림 3>의 네 번째 단계에서 볼 수 있듯이 각 단락의 마지막에 사용자 그룹 및 시간 간격 ID를 각

문장의 태그로써 추가한다. 가공된 데이터는 두 개의 Doc2Vec 모델, PV-DM(paragraph vector-distributed memory)와 PV-DBOW(paragraph vector -distributed bag of words)에 입력되어 관광객 그룹과 서비스 아이템에 대한 실수 값을 갖는 고정된 크기의 표현된 벡터로 변환 된다.

4. 심층 관광 추천

김태경 등(2021)은 관광 지식 그래프를 RDF로 도출하였다. 이를 바탕으로 앞에서 언급한 TE에 관광 RDF를 보조 요소로 융합하여 서비스에 대한 그룹의 선호도를 예측한다. 이러한 절차를 「심층 관광 추천(DR: deep tourism recommendation)」으로 명명하고 구체적인 내용을 설명한다.

4.1. RFM 분석을 통한 평점 생성

TE 데이터를 평점으로 매핑하기 위해서 RFM(recency, frequency and monetary) 분석 프레임워크를 사용한다. 고객의 충성도를 간접적으로 알아보기 위해 마케팅 분야에서 RFM에 기반을 둔 측정치를 활용한다(Sharifhosseini, 2019). 구매 결정 이후 높은 평점을 부여한다는 것은 관광 서비스를 만족스럽게 체험했다는 사실이나 향후 재구매 의도가 높다는 뜻으로 받아들여지기 때문에 RFM과 평점 사이에는 비례적 관계가 있다고 가정한다(Awangga et al., 2019). 일반적으로 심층 신경망을 통해 선호도를 예측하려면 정답 데이터를 포함하는 데이터 세트를 이용하여 모델을 학습하지만, 카드 거래 데이터에는 평점이 없기 때문에 지식 경영 분야에서 사용되는 RFM 분석을 이용하여 평점을 추정하여 정답 데이터로 모델을 훈련하고 평가하

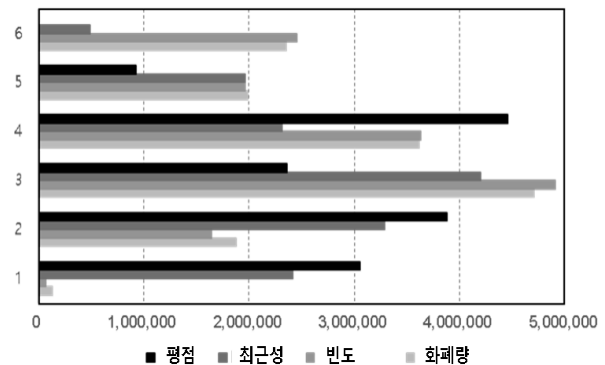
는 것이다. 구체적으로 고객이 자주, 최근에 더 많은 돈을 특정 서비스에 소비한 경우, 해당 서비스에 대한 고객의 충성도 또는 선호도, 평점이 높은 것으로 해석할 수 있다. 본 연구에서 각 요인은 다음과 같이 최신성(R: recency), 빈도(F: frequency), 화폐량(M: monetary)으로 정의된다.

- **최신성(R)**은 $R = M + (12 \times (Y - Y_b))$ 로 계산되며, 여기서 Y_b 와 Y 는 데이터에 포함된 거래의 가장 빠른 연도와 각 사용자 그룹의 해당 거래 연도를 나타낸다.
- **빈도(F)**는 관광객 그룹당 거래의 수를 의미한다.
- **화폐량(M)**은 관광객 그룹당 총 거래 금액을 나타낸다.

최신성의 경우 데이터 세트가 2012년부터 2019년까지 발생한 카드 결제를 포함하기 때문에 $Y_b = 2012$ 로 설정한다. 결과적으로 세 가지 요소가 더 높은 경우, 사용자 그룹이 서비스에 대한 선호도가 더 높다고 해석할 수 있다. 세 가지 요소(즉, R과 F, M)를 결합하기 위해 중요도에 따른 가중치를 Sharifhosseini(2019)의 연구에 근거하여 각기 1과 2, 4로 적용했다.

RFM 분석을 이용하여 각 거래 데이터에서의 서비스에 대한 관광객 그룹의 평점을 생성한다. 먼저 각 요소별로 모든 거래 데이터를 k-평균 군집화(clustering) 기법을 적용하여 4개의 그룹으로 분류한다. 그리고 각 클러스터마다 요소의 평균값 5개를 구해서 모든 거래 데이터를 각기 6개의 그룹으로 재분류하고 1부터 6까지의 정수를 할당한다. 즉, 한 거래 데이터의 빈도 요소에 할당된 점수가 6인 경우, 6 미만인 다른 거래들보다 해당 거래가 더 자주 발생했음을 의미하게 된다. 다음으로 거래별로 할당된 3개 요소의 점수에 가중치를 곱했으며 이에 3차원 k-평균 군

집화를 적용하고 3개의 요소를 함께 고려하여 모든 거래 데이터를 5개의 군집으로 다시 분류한다. 그런 다음 각 군집에 속한 거래 데이터들의 가중치가 적용된 요소의 합계의 평균값을 계산하고, 평균값이 높은 순으로 군집별로 평점을 5부터 1까지 할당한다. 즉, 하나의 군집에 속한 모든 거래 데이터는 동일한 평점을 갖게 된다. 결과적으로 평점 5를 할당 받은 거래 데이터의 관광객 그룹은 최근에 더 많은 비용을 들여 서비스를 자주 이용한다는 것을 나타내고, 해당 그룹이 해당 서비스를 선호한다는 것을 의미한다. <그림 4>는 RFM 분석으로 평점을 할당한 거래 데이터의 분포를 보여준다. 의도한 바와 같이 모든 거래 데이터는 세 가지 요소별로 1에서 6까지 다양하게 분포되고, 각 요소의 가중치에 따라 평점이 1에서 5까지 적절하게 부여되었음을 알 수 있다.



<그림 4> RFM 분석에 따른 거래 데이터 분포

4.2. 심층 관광 추천

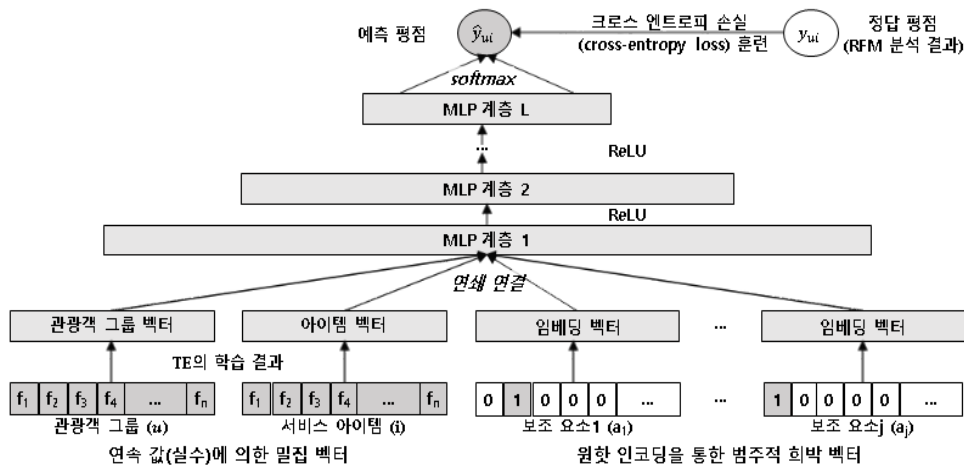
본 절에서는 앞서 설명한 TE와 RFM 분석의 결과, 그리고 관광 RDF에서 도출한 보조 요소를 통합하여 관광객 그룹의 서비스에 대한 선호도를 예측하는 심층 관광 추천(DR) 모듈에 대해서 설명한다. <그림 5>는 관광객 그룹과 서비스 항목의 표현 벡터와 각 보조 요

소를 통합하는 MLP 신경망 구조를 도식화 한 것이다.

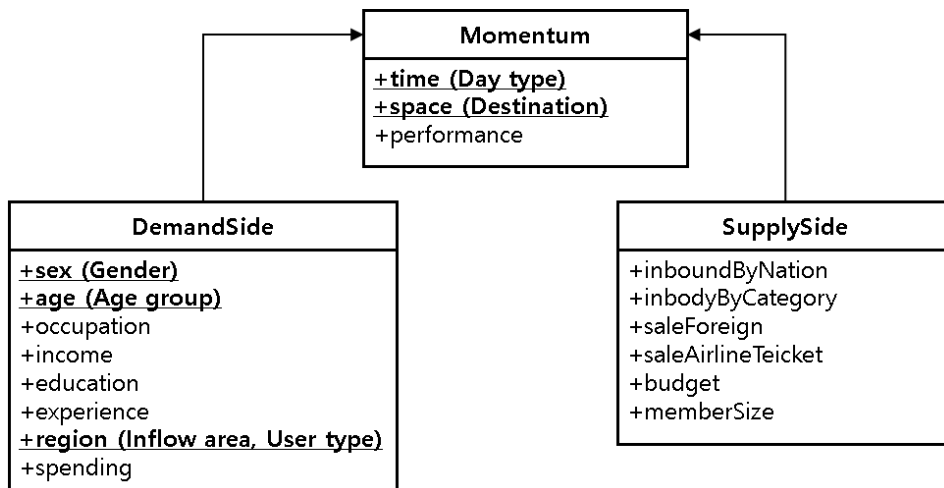
활성화 함수에는 여러 선택지가 있으나, 예비 실험 결과 더 좋은 성능을 보인 ReLU(rectifier unit)를 이용한다. 실제 ReLU는 타 함수에 비해 학습 데이터에 너무 맞추어져 일반 상황에서 성능이 떨어지는 과적합화 될 가능성이 적기 때문에 선호된다(He et al., 2017). 또한 MLP의 네트워크 구조는 타워 패턴을 따라 설계했다. 타워 패턴은 아래 계층이 가장 넓고, 위로 쌓이는 연속 계층에는 더 적은 수의 뉴런을 포함한다. He

et al.(2016)에 따르면, 타워 패턴은 상위 계층에 소수의 은닉 유닛을 사용하여 데이터의 추상적인 특징을 잘 학습할 수 있게 한다. 단순성과 사용 용이성으로 인해 점점 많이 쓰이고 있는 소프트맥스(softmax)와 크로스 엔트로피 손실(cross-entropy loss)을 결합한 소프트맥스 손실(softmax loss)을 목적함수로 사용한다.

고객의 다양한 정보, 비정형 데이터를 자원으로 묶어 효율적으로 분석하기 위한 방법 중 하나로써 RDF 기반의 정보 시스템의 도입은 지속하여 강조되어 왔



<그림 5> 심층 관광 추천의 MLP 신경망 구조



<그림 6> 관광 산업 RDF의 UML(김태경 등, 2021)과 보조 요인

다(김태경 등, 2021). 본 연구에서는 김태경 등(2021)의 연구에서 도출한 관광 RDF와 카드 거래 데이터 상의 교차 특성 요소들을 추천에 활용한다. <그림 6>은 김태경 등(2021)이 정리한 관광 산업 도메인의 RDF의 일부를 UML로 도식화 한 것으로, 굵게 밑줄 친 특성들이 카드 거래 데이터와 교차된 요소(괄호 내 나열)들로 MLP의 선호도 학습에 활용된다. 보조 요인으로 <그림 6>의 괄호 안에 나열된 6개의 범주형 항목들이 있어 <그림 5>의 오른쪽 아래와 같이 이들을 임베딩 계층에서 저차원의 조밀한 실수 값을 갖는 벡터로 변환한다. 각 요소들에 대한 상세 정보는 5.1절의 데이터 세트 서술에서 기술한다.

요약하면 제안된 CTDDTR의 추천 절차는 다음과 같다.

- 1) TE에서 Doc2Vec를 이용하여 관광객 그룹과 서비스 아이템의 벡터를 학습한다.
- 2) 평점을 생성할 때 RFM 모델을 이용하고, 관광 RDF를 통해 보조 요소들을 선택한다.
- 3) 학습한 그룹과 아이템의 표현 벡터와, 보조 요소들의 정보를 통합하고, 얻어진 평점을 기반으로 MLP의 선호도 예측 모델을 학습한다.
- 4) 관광객 그룹의 서비스 선호도를 예측하여, 선호도가 높은 서비스 N개를 추천한다.

본 연구에서 제안하는 CTDDTR은 개인정보가 보호되어야 하는 일반적인 산업 환경에서의 추천 알고리즘으로써 관광객 그룹에 서비스를 제공하지만, 관광객 개별 거래 데이터가 사용 가능한 경우 개인화된 추천 서비스에 쉽게 적용할 수 있다.

5. 실험 및 논의

5.1. 실험 설계

제안한 CTDDTR의 효율성과 보조 요인들의 추천 성능에 대한 영향을 평가하기 위해, 한국의 주요 카드사 중 하나인 S카드의 카드 거래 데이터를 사용했다. 한국의 최대 관광지인 제주도를 목표 지역으로 선정하고, 여기서 발생한 약 1,900만 건의 거래내역을 확보했다. 관광객의 개인 정보는 통계적으로 처리되어 관광객 그룹을 관광객의 성별, 연령대, 주거지역, 거래일 유형으로 분류함으로써 개인 정보 보호를 위해 노력했다. 결과적으로 1,260개의 사용자 그룹을 대상으로 추천 성능을 평가하였다. 또한 관광과 관련 없는 서비스에서 발생한 카드 거래는 한국표준산업코드¹⁾(KSIC : Korea standard industry code)를 통해 제거했다. 관광 관련 업종(예 : 숙박업과 요식업, 동물원 등)과 무관한 KSIC에 해당하는 거래 데이터를 걸러낸 결과, <표 1>에 나열된 바와 같이 약 1,500만개의 거래 데이터에 413개의 관광서비스가 포함되었다. 정리된 데이터에서는 제주시와 서귀포시에 국한하여 거래가 발생되었고 79개의 목적지를 포함하며, 거래일 유형(day type: 주중, 주말, 주중 공휴일, 주말 공휴일), 고객 유형(user type: 관광객 및 거주자), 18개의 국내 고객 유입 지역(inflow area), 7개의 연령대(age group) 및 2개의 성별(gender)과 같은 5개 보조 요인들이 확인되었다. 개월별 및 계절별로 각각 241,640(724,920)개와 191,388(574,164)개의 TE 시퀀스를 생성하여 Doc2Vec 모델로 임베딩하였다.

추천 성능을 알아보기 위해 5분절 교차검증법(5-fold cross-validation)을 수행하였고, 평균 절대 에러, F1-점수, 평균 상호 순위점수, 정규화된 할인 누적 이

1) KSIC: http://kssc.kostat.go.kr/ksscNew_web/ekssc/main/main.do

〈표 1〉 카드 거래 데이터의 통계 정보

| 특성 | 수량 |
|--------------------------------|------------|
| 관광 관련 카드 거래 | 14,673,210 |
| 관광객 그룹 | 1,260 |
| 관광 서비스 | 413 |
| 거래 발생지 (제주시와 서귀포시 행정단위) | 79 |
| 거래일 유형(주중, 주말, 주중 공휴일, 주말 공휴일) | 5 |
| 고객 유형 (관광객 또는 거주자) | 2 |
| 국내 고객 유입 지역 | 18 |
| 연령대 | 7 |
| 성별 | 2 |

〈표 2〉 서비스 이용 수에 따른 관광객 그룹의 분포

| 서비스 수 | 5개 미만 | 10개 미만 | 20개 미만 | 30개 미만 | 40개 미만 | 50개 미만 | 50개 이상 |
|---------|-------|--------|--------|--------|--------|--------|--------|
| 퍼센트 (%) | 33.31 | 19.25 | 18.78 | 10.34 | 6.71 | 4.20 | 7.41 |

익과 같이 다양한 지표의 평균값으로 비교 알고리즘들을 평가하였다. 다음은 각각의 지표에 대한 간략한 설명이다. 본 연구에서는 상위 N개의 서비스 추천의 성능을 이들 성능 지표들에 근거하여 평가했다.

5.1.1. 평균 절대 에러(mean absolute error: MAE)

N개의 테스트 데이터가 주어지고 r_{ui} 과 \hat{r}_{ui} 이 정답 평점과 예측된 평점이라 할 때, 둘의 차이(에러)의 평균을 의미한다. 따라서 MAE가 낮을수록 해당 추천 알고리즘의 성능이 더 좋다.

$$MAE = \frac{1}{N} \sum_{u,i} |r_{ui} - \hat{r}_{ui}|$$

5.1.2. F1-score

U개의 관광객 그룹에 대한 추천의 정확도(P)와 재현율(R)이 주어졌을 때 아래와 같이 정의되며, 높을수록 예측 성능이 좋다고 판단한다. 이때 관련된 아이템은 관광객이 실제 이용한 서비스를 의미한다.

$$F_1 = 2 \times P \times R / (P + R)$$

$$P = \frac{1}{U} \sum_u \frac{\text{추천된 관련아이템 수}}{\text{추천된 아이템 수}},$$

$$R = \frac{1}{U} \sum_u \frac{\text{추천된 관련아이템 수}}{\text{관련된아이템 수}}$$

5.1.3. 평균 상호 순위점수(mean reciprocal rank: MRR)

하나의 아이템에 초점을 맞춘다. MRR는 U개의 관광객 그룹이 주어지고 k_u 가 관광객 그룹 u를 위해 N개의 아이템을 제공 시 관련된 아이템의 위치를 나타낼 때, 아래와 같이 정의되며 높을수록 추천 성능이 좋음을 의미한다. MRR 단독으로는 좋은 평가 지표가 아닐 수 있지만 <표 2>에서 볼 수 있듯이 관광 산업에서의 관광객이 이용하는 서비스는 소수이기 때문에 고려해볼 가치가 있다. 또한 5개 미만의 서비스를 사용하는 사용자 그룹이 전체의 30% 이상을 차지하는 것을 볼 수 있다.

$$MRR(U) = \frac{1}{U} \sum_{u \in U} \frac{1}{k_u}$$

5.1.4. 정규화된 할인 누적 이익(normalized discounted cumulative gain: NDCG)

일부 아이템이 다른 아이템보다 관련성이 높다는 사실에 근거한다. 즉, 관련성이 높은 아이템 항목이 중간 정도의 관련된 또는 관련성이 없는 항목보다 상위에 배치되어야 하고 아이템의 위치에 비례하여 관련성 점수에 페널티를 준다. r_i 이 i 위치에서의 결과의 관련 정도를 나타내고, R_k 이 상위 k 까지의 관련성에 따라 정렬된 관련된 아이템의 목록을 나타낼 때, 다음과 같이 정의된다. 결과적으로, 해당 측정치가 높을수록 상위 k 개의 아이템 추천의 성능이 좋다고 평가할 수 있다.

$$NDCG_k = DCG_k / IDCG_k$$

$$DCG_k = \sum_{i=1}^k \frac{2^{r_i} - 1}{\log_2(i + 1)},$$

$$IDCG_k = \sum_{i=1}^{|R_k|} \frac{2^{r_i} - 1}{\log_2(i + 1)}$$

5.2. 비교 추천 알고리즘

본 절에서는 제안한 방법과 성능 비교를 위한 추천 알고리즘 7가지(Ozsoy, 2016; Liu et al., 2018; Sharifhosseini, 2019; Nassar et al., 2020; Shambour, 2021)를 간단히 설명한다. 아래에서 설명하는 모든 알고리즘은 Tensorflow와 Keras 2.4.0을 사용하여 Python으로 구현되었으며, 동일한 실험 환경에서 평가하였다.

카드 결제 데이터 세트에는 고객의 평점 정보가 없기 때문에 카드 거래 데이터에 직접 적용하는 것은 불가능하다. 따라서 Sharifhosseini(2019)는 RFM 분석을 활용하여 관광객 그룹의 서비스에 대한 평점을 생성하고 협업 필터링 기법에 기반을 둔 GSVD와 SVD를 제안하였다. 본 연구에서는 위에서 제시한 RFM 분석

을 이용한 결과로 평가하기 때문에, GSVD_{RFM}와 SVD_{RFM}로 표기한다. 또한 Ozsoy(2016)가 제안한 Word2Vec 기반의 내용 기반(KNI_{W2V}), 협업 필터링(NN_{W2V}), 하이브리드 추천(KIU_{W2V})을 고려하며, 비교의 공정성을 위해 Doc2Vec를 활용한 방식(KNI_{D2V}, NN_{D2V}, KIU_{D2V})도 구현하여 평가하였다. Word2Vec의 경우 Skip-gram 모델이, Doc2Vec의 경우 PV-DM 모델이 좋은 성능을 보였기 때문에, 해당 모델에 기반을 둔 성능 평가 결과와 비교했다. 이 외에도 최신 심층 학습 기반의 추천 알고리즘으로써 2.2절에서 언급한 AEMC와 DMCF, 그리고 시간성에 기반을 둔 세션 기반 추천 알고리즘인 STAMP(short-term attention/memory priority model)와 비교 평가했다(Liu et al., 2018). 각 알고리즘별로 연구 논문을 참조하여 매개변수들을 설정하였으며, 데이터 세트에 맞추어 수정하여 구현하였다. 앞의 두 심층 학습 기반 추천 알고리즘들은 RFM 분석을 이용하여 학습하기 때문에, AEMC_{RFM}와 DMCF_{RFM}로 표기한다. 반면 사용자의 현재와 일반적 관심을 고려하는 STAMP의 경우, RFM 분석으로 생성한 평점의 유효성을 확인을 위한 대조 알고리즘으로써, 해당 분석 기법을 적용하지 않았으며, 세션의 시간은 가장 좋은 성능을 보인 개월 단위로 정의하였다.

추가적으로 제안한 방식의 주요 파트가 추천 성능에 미치는 영향을 파악하기 위해 CTDDTR의 변형을 고려하였다, Doc2Vec 모델을 구별하기 위해 제안된 방식의 이름 뒤에 Doc2Vec 모델에 대한 접미사 dm(PV-DM)과 db(PV-DBOW)를 추가하였다. 또한 TE에서 시간적 요소를 고려하지 않는 변형은 CTDDTRo(CTDDTR without temporality)로 명명하였다. 예를 들어 CTDDTRodm은 PV-DM 기반 TE에서 시간성을 고려하지 않는 방식을 나타내며 각각의 모형은 접미사들로 구분될 수 있다.

5.3. CTDDTR의 변형의 성능 비교 분석

본 절에서는 제안한 알고리즘의 최적의 설정 값을 찾기 위해 다양한 설정 값에서의 CTDDTR와 CTDDTRo를 평가했으며, 보다 일반적인 결과를 얻기 위해 5번 테스트를 진행하여 평균 및 오차범위를 표기하였다. Doc2Vec의 여러 매개변수가 CTDDTR에서 추천 성능에 영향을 미칠 수 있으므로, 매개 변수들이 CTDDTR의 성능에 어떻게, 그리고 얼마나 영향을 미치는지를 추가로 확인하였다. 이때 「최소 빈도수(minimum word count)」를 이용해 이용내역에 포함된 관광 서비스의 빈도수가 설정 미만인 경우 제거했다. 추가적으로 추천 시스템에서 데이터는 일반적으로 매우 희소하며 몇 번만 관찰되는 아이템 항목을 포함하기 때문에, 정보 손실을 방지하기 위해 이 매개변수를 1로 설정했다. 「창 크기(window size)」는 Doc2Vec에서 각 서비스의 벡터를 학습하면서 고려되는 서비스들의 최대 간격을 설정하며, 「학습 횟수」는 Doc2Vec 모델 학습의 반복 횟수를 지정한다. 마지막으로 「벡터 크기」는 PE

로부터 반환되는 관광객 그룹과 서비스 벡터의 크기를 의미하며, DR에 직접적으로 입력되기 때문에 성능에 영향을 미친다.

<표 3>은 각 변형 모델별 MAE 성능과 가장 우수한 모델의 성능 개선율을 나열한다. 문서(즉, 관광객의 서비스 이용 내역) 측면에서는 시간성을 고려한 CTDDTR이 더 나은 성능을 보였으며, 이는 사용자 그룹과 서비스 항목을 모델링할 때, 시간적 요소를 반영하는 TE에서 결과치가 더 긍정적이었다. Doc2Vec 모델 측면에서는 PV-DM이 PV-DBOW 보다 전반적으로 나은 성능을 보였고, 굵은 폰트와 음영으로 표시된 바와 같이 벡터 크기를 15로 설정한 CTDDTRdm이 가장 우수한 성능을 보였다.

앞서 언급한 바와 같이 벡터 크기 외에도 TE에는 추천의 성능에 영향을 미치는 여러 설정 값이 존재하며, <표 4>는 각 설정에 따른 PTE의 MAE 측정치를 나열한 것이다. 3.1절에서 분석한 결과를 바탕으로 시간 간격을 주와 개월, 계절 단위로 변경하여 실험하였고, 개월별로 설정한 경우 가장 우수한 성능을 보였다.

<표 3> CTDDTR 변형별 MAE 성능 비교

| 모형 | 벡터크기 | MAE(오차) | 개선율 | 모형 | 벡터크기 | MAE(오차) | 개선율 |
|-----------|------|-----------------|-------|-----------|------|------------------------|------|
| CTDDTRodb | 10 | 0.6992 (±0.070) | 16.2% | CTDDTRodb | 10 | 0.6348 (±0.049) | 7.7% |
| | 15 | 0.6862 (±0.057) | 14.6% | | 15 | 0.6125 (±0.030) | 4.3% |
| | 20 | 0.7265 (±0.083) | 19.3% | | 20 | 0.6456 (±0.041) | 9.2% |
| CTDDTRodm | 10 | 0.6768 (±0.35) | 13.4% | CTDDTRdm | 10 | 0.6016 (±0.031) | 2.6% |
| | 15 | 0.6676 (±0.045) | 12.2% | | 15 | 0.5862 (±0.023) | - |
| | 20 | 0.6952 (±0.045) | 15.7% | | 20 | 0.6232 (±0.024) | 5.9% |

※ 음영으로 가장 우수한 성능을 보인 모형을 표시함, 개선율은 비교대상보다 최대 성능 모델이 개선된 정도를 표시함

<표 4> PTE 설정에 따른 CTDDTRdm 성능 변화

| 시간 간격 | MAE | 오차범위 | 최소 빈도 | MAE | 오차범위 | 학습 횟수 | MAE | 오차범위 |
|--------|--------|-------|-------|--------|-------|-------|--------|-------|
| week | 0.7905 | 0.030 | 1 | 0.5862 | 0.022 | 2 | 0.7596 | 0.121 |
| month | 0.5862 | 0.021 | 3 | 0.7977 | 0.130 | 5 | 0.5862 | 0.021 |
| season | 0.6016 | 0.027 | 5 | 0.9892 | 0.197 | 10 | 0.5961 | 0.020 |

※ 음영으로 가장 우수한 성능을 보인 모형을 표시함

또한 최소 빈도수는 앞서 언급한 바와 같이 정보 손실을 줄일 수 있도록 1로 설정하는 것이 바람직함을 알 수 있다. 최소 빈도수를 높게 설정할 경우, 거의 이용되지 않은 일부 서비스 항목들이 Doc2Vec 모델을 훈련할 때 제거되어 심지어 추천되지 않는다는 것을 관찰했다. 이는 추천 연구 분야에서 잘 알려진 아이템 콜드 스타트(item cold-start) 문제이다. 또한 5보다 적은 학습 횟수에서는 높은 MAE를 보였고 이후부터는 거의 유사한 결과를 얻었다. 따라서 학습의 효율성을 위해 훈련을 5번 반복하는 것으로 설정함이 바람직하다. 이외에 MLP에서 통합되는 보조 요소의 임베딩 벡터 크기에 따른 성능은 <표 5>에서 볼 수 있듯이 20의 크기가 가장 적절한 것으로 드러났다.

4.2절에서 언급한 바와 같이 DR 모듈은 관광객 그룹과 서비스 아이템의 정보뿐만 아니라 관광 RDF에 기반을 두어 카드 거래 데이터로부터 도출된 여러 보조 요소를 통합하여 서비스 선호도를 예측한다. RDF <표 6>은 각 보조 요소별 CTDDTRdm의 MAE를 나열한 것이다. 결과적으로 “거래일 유형”과 “목적지”, “유

입 지역”이 MAE 측면에서 제안한 추천 알고리즘에 더 긍정적인 영향을 미친다는 것을 알 수 있다. 이는 추천 성능을 향상시키기 위해 시공간적 보조 요인이 중요함을 강조한다. Chaudhari and Thakkar(2019)의 연구 결과와 유사하게, “연령대”와 “성별”과 같은 인구 통계정보 또한 CTDDTRdm의 성능에 긍정적인 영향을 주었다. 그러나 “사용자 유형”은 부정적인 영향을 미쳤다. 비즈니스 관점에서 이러한 분류가 사용자 선호도 모델링과 추천 서비스에 큰 영향을 미치지 않을 수 있음을 의미한다. 이는 “사용자 유형”보다 더 주요한 요소, 가령 “유입 지역”, “연령대”, “성별”에서 이미 정보로 반영되었기 때문으로 추정된다.

5.4. 알고리즘 성능 비교

<표 7>에서 볼 수 있듯이, 실험 결과는 제안한 방식이 모든 성능 지표에서 다른 모델을 능가한다는 것을 보여준다. MAE와 MRR에서, CTDDTR은 평균 28.5%와 15.5%의 개선으로 다른 접근 방식보다 우수하였으

<표 5> MLP 임베딩 벡터 크기에 따른 CTDDTRdm 성능 변화

| 임베딩 크기 | 10 | 20 | 30 | 40 | 50 |
|--------|--------|--------|--------|--------|--------|
| MAE | 0.7155 | 0.5862 | 0.6080 | 0.6701 | 0.7292 |
| 오차범위 | 0.027 | 0.023 | 0.021 | 0.269 | 0.272 |

※ 음영으로 가장 우수한 성능을 보인 모형을 표시함

<표 6> 보조 요소에 따른 CTDDTRdm 성능 변화

| 보조요소 | MAE | 오차범위 | 순위 |
|--------|--------|--------|----|
| 없음 | 0.5862 | 0.023 | |
| 거래일 유형 | 0.5346 | 0.0160 | 1 |
| 목적지 | 0.5353 | 0.0267 | 2 |
| 유입 지역 | 0.5498 | 0.0301 | 3 |
| 연령대 | 0.5650 | 0.0329 | 4 |
| 성별 | 0.5709 | 0.0724 | 5 |
| 관광객 유형 | 0.5945 | 0.1814 | 6 |

〈표 7〉 비교 알고리즘 간 성능 비교 결과

| Methods | 성능 평가 결과 | | | | CTDDTRdm의 성능 개선율 | | | | MRR -NDCG |
|---------------------|----------|--------|--------|--------|------------------|-------|-------|-------|--------------|
| | MAE | F1s | MRR | NDCG | MAE | F1s | MRR | NDCG | |
| GSVD _{RFM} | 0.8729 | 0.5992 | 0.6080 | 0.3796 | 39.2% | 41.0% | 24.2% | 55.2% | 0.2285 |
| SVD _{RFM} | 0.9167 | 0.5103 | 0.5459 | 0.3307 | 42.1% | 65.6% | 38.4% | 78.2% | 0.2152 |
| KN _{W2V} | 0.8513 | 0.6080 | 0.6201 | 0.4794 | 37.6% | 39.0% | 21.8% | 22.9% | 0.1407 |
| NN _{W2V} | 0.7897 | 0.6172 | 0.6261 | 0.4902 | 32.8% | 36.9% | 20.6% | 20.2% | 0.1360 |
| KIU _{W2V} | 0.6923 | 0.6916 | 0.6346 | 0.4914 | 23.3% | 22.2% | 19.0% | 19.9% | 0.1432 |
| KN _{D2V} | 0.7813 | 0.6330 | 0.6521 | 0.5060 | 32.0% | 33.5% | 15.8% | 16.4% | 0.1461 |
| NN _{D2V} | 0.7397 | 0.6572 | 0.6661 | 0.5123 | 28.2% | 28.6% | 13.4% | 15.0% | 0.1538 |
| KIU _{D2V} | 0.6832 | 0.6742 | 0.6746 | 0.5305 | 22.3% | 25.4% | 12.0% | 11.1% | 0.1442 |
| STAMP | - | 0.7947 | 0.7527 | 0.5564 | - | 6.4% | 0.4% | 5.9% | 0.1963 |
| AESC _{RFM} | 0.6042 | 0.7667 | 0.7480 | 0.5196 | 12.1% | 10.2% | 1.0% | 13.4% | 0.2284 |
| DSCF _{RFM} | 0.6267 | 0.7437 | 0.7262 | 0.4930 | 15.3% | 13.6% | 4.0% | 19.5% | 0.2332 |
| CTDDTRdm | 0.5310 | 0.8452 | 0.7554 | 0.5892 | - | - | - | - | 0.1662 |

※ 음영으로 가장 우수한 성능을 보인 모형을 표시함, 개선율은 비교대상보다 최대 성능 모델이 개선된 정도를 표시함. 기울임은 비교 대상 중의 가장 우수한 성능을 모형을 표시함.

며, F1-score와 NDCG에서도 평균 29.3%와 25.2%의 개선율을 보였다. NDCG는 추천 연구에서 많이 사용되는 상위 10개의 아이템을 추천을 대상으로 한다.

조금 구체적으로 살펴보자. Word2Vec 기반 접근 방식(즉, KN_{W2V}와 NN_{W2V}, KIU_{W2V})을 고려할 때 CTDDTRdm은 모든 지표에서 평균적으로 약 26.4% 우수했다. Ozsoy(2016)가 제안한 알고리즘은 Word2Vec 모델을 훈련하고 추천 모듈에서 사용자 그룹 및 항목 유사성을 사용하기 위해 사용자 그룹 및 서비스 항목으로 시작하는 문서를 구성하며, 관광객 그룹과 서비스 항목을 단어로써 동일한 개념으로 간주한다. 이러한 방식은 서비스 항목 간의 불균형을 만든다. 예를 들어, 서비스 항목을 시간 순으로 정렬할 때 사용자 그룹과 최근 소비한 서비스 사이의 거리는 과거에 이용된 서비스보다 훨씬 멀어지고, CTDDTRdm보다 낮은 성능을 보이는 것으로 예상된다.

단어 임베딩 도입에 따른 이와 같은 불공정함을 배제하기 위해, Ozsoy(2016)의 알고리즘에 Doc2Vec를 적

용한 KN_{D2V}, NN_{D2V} 및 KIU_{D2V}와 비교했다. 그 결과 기존의 방식(KN_{W2V}와 NN_{W2V}, KIU_{W2V})보다 제안된 방법은 더 나은 성능을 보였다. 이러한 결과는 Doc2Vec 모델이 관광객 그룹을 단어가 아닌 문서로 간주하고 이용된 서비스 간의 불균형이 완화되기 때문으로 예상된다. 그러나 CTDDTRdm은 여전히 해당 추천 알고리즘을 능가하며 심지어 CTDDTRdm조차도 그보다 낮은 MAE를 보여준다(<표 6> 참조). 이러한 결과는 TE에서 고려한 시간적 요인과 DR에서 도입한 MLP가 추천 성능에 긍정적인 영향을 미친다는 것을 보여준다.

다음으로 최신 인공신경망 기반 접근 방식인 AEMC_{RFM} 및 DMCF_{RFM} 과 비교 평가한 결과이다. CTDDTRdm은 MRR에서 평균 2.5%로 약간 더 나은 반면 MAE와 F1-score, NDCG에서 13.7%와 11.9%, 16.4%의 개선으로 두 가지 접근 방식보다 더 나은 성능을 보였다. MAE의 계산 오류 정도와 달리 MRR과 NDCG는 실제 적용 측면에서 top-k 권장 항목을 측정한다. MRR은 1순위 항목만 고려하고 NDCG는 추천 항목의 순서를 반영하

며, 이러한 지표의 용도와 관련하여 제안한 CTDDTRdm 방법이 top-k 항목 권장 사항에서 AEMC_{RFM} 및 DMCF_{RFM} 보다 더 정확하고 관련성 있는 항목을 제공하며, 실제 응용서비스 관점에서 더 안정적이고 우수할 것이라고 기대할 수 있다.

또한 시간성에 기반을 둔 세션 추천 모델인 STAMP 과 함께 성능을 비교 평가한다. 5.2절에서 언급한 바와 같이 STAMP는 RFM 방식을 적용하지 않기 때문에 F1-score와 MRR, NDCG의 3가지 척도만 평가한다. STAMP는 두 가지 DNN 기반 접근 방식보다 더 나은 성능을 보여주지만 F1-score 및 NDCG에서 CTDDTRdm에 비해 결과가 약간 나쁘다. 이러한 결과는 제안된 TE가 세션 기반 모델처럼 잘 작동하며, MLP 기반 DR은 RFM 방식을 도입하여 1에서 5까지 척도화된 사용자 그룹의 선호도를 적절하게 예측할 수 있음을 의미한다.

마지막으로 <표 7>에 나열된 MRR-NDCG 측면에서 성능을 비교한다. 두 지표의 용도에 따르면 MRR-NDCG이 높을수록 서비스 1개 추천과 10개 항목 추천 간의 성능 격차가 더 커진다. 실제로 추천 시스템은 일반적으로 사용자 그룹에 여러 항목을 제공하고 결과 그룹은 더 많은 선택의 기회를 갖는다. 일반적으로 관광객은 추천된 관광 서비스를 이용 후에 만족하지 않으면 해당 추천 서비스를 다시 이용하지 않으려하기 때문에, 최상위로 제공되는 추천의 정확도가 중요하다. 이와 관련하여 CTDDTRdm과 Word2Vec 및 Doc2Vec 기반 접근 방식은 DNN 기반 접근 방식보다 상위 1위 항목 권장 사항과 상위 10위 항목 권장 사항 간의 격차가 더 작은 것으로 나타났다. 이러한 결과는 RFM 분석에 의해 생성된 평점이 DNN 모델의 훈련에서 노이즈가 될 수 있기 때문이다. 반면에 암시적 피드백 기반 STAMP는 0.1963의 측정치를 가지고, CTDDTRdm은 세션 기반 모델 STAMP 보다 낮은 차이를 보였다.

이러한 결과는 RFM 분석 기반의 등급 생성이 유효한 방식임을 의미하고, 단일 또는 다중 서비스 아이템 추천서비스에서의 CTDDTRdm의 효과성을 증명한다. 또 다른 원인으로 시간성의 고려에 의한 차이일 수 있다. CTDDTRdm과 STAMP의 경우 서비스 이용의 순차적 연속성을 데이터 모델링 과정에서 반영하지만, DNN 기반의 두 추천 알고리즘은 이를 간과한다. 이러한 관광 서비스 이용의 연속성은 관광객이 일일 관광 계획에 따라 여러 관광 서비스를 순차적으로 이용하기 때문에 관광 추천에 있어 매우 중요하다. 결과적으로 이러한 결과로부터 CTDDTRdm이 평가 점수가 없는 거래 데이터 사용에 특화된 성능을 보인다는 것을 암시한다.

최영제 등(2020)은 도서추천 시스템을 연구하기 위해 거래 데이터를 사용했다. 이들의 연구는 SAR(simple algorithm for recommendation)을 바탕으로 단어 임베딩 절차를 거쳤다는 점에서 본 연구와 접근법이 유사하다. 최영제 등(2020)은 연구 결과로 요인 가중치 적용에 있어 반복 구매가 선호로 이어지지 않는다는 점을 고려해야 한다는 결론에 이르렀지만, 시간 요인을 세부적으로 고려하지는 않았다. 정영진, 조운호(2017) 역시 도서구입 추천 시스템을 대상으로 연구를 수행했다는 점에서 관광 서비스 도메인을 주제로 한 본 연구와 차별화된다. 이들은 협업 필터링을 바탕으로 BCBCF(book contents-based collaborative filtering)을 제안하였는데 사용자 프로필과 토픽 모델링의 결과를 하이브리드 하였다. 본 연구에서 맥락 정보를 반영하기 위해 지식 그래프를 사용한 것에 비해 정영진, 조운호(2017)는 텍스트 마이닝의 결과 및 사용자 관련 정보에서 맥락 정보를 얻는 것으로 보인다.

모바일 인터넷과 스마트 기기의 보급은 스마트 관광을 수행하기 유리한 환경을 제공하지만 정보 과부화 문제를 야기하기도 하는데 추천 시스템은 이와 같

은 상황에서 기업의 가치를 높일 수 있는 유효한 수단이 될 수 있다(문현실 등, 2020). 본 연구는 대량으로 수집되며 사용자의 실제 구매 활동을 반영하는 카드 데이터를 대상으로 했다는 점에서 기존의 연구 시도들과 차별화된다. 또한 RFM 분석 모델을 적용하되 단어 임베딩 기법과 RDF 지식 그래프 및 딥 러닝 방법을 복합적으로 적용하여 사용자 평점이 누락된 경우에도 선호도와 요인 정보 등이 반영되도록 했다는 점에서 새로운 시도로 받아들여진다.

6. 결론 및 향후 연구

글로벌 관광 산업이 급속도로 성장하고 있으며, 관광 상품을 제공하여 관광객을 지원하기 위한 추천 시스템이 활발히 연구 및 산업계에 융합되고 있다. 최근에는 관광 소비 스타일과 패턴을 포함하는 카드 거래 데이터가 추천 시스템에 적용되고 있다. 그러나 기존 관광 서비스 추천 연구에서는 카드 거래 데이터가 개인정보 보호로 인해 실제 응용서비스에서 이용 가능한 정보가 제한된다는 점을 충분히 고려하지 못했다. 또한 관광 서비스 선택에 중요한 역할을 미치는 요소들에 대한 반영도 중요하며, 더 정확한 추천을 위해 시간에 따른 관광객의 흥미 변화도 고려될 필요가 있다.

본 연구에서는 GDPR을 고려한 상황에서 관광객 그룹에게 관광 서비스를 추천하기 위해 카드 거래 데이터를 기반으로 하는 CTDDTR이라는 새로운 방법을 제안했다. 제안하는 방법은 시간적 사용자 그룹의 선호도를 고려한 사용자 그룹 및 서비스 항목의 벡터를 학습하는 TE와 학습된 벡터와 보조 요소를 융합하는 MLP에 기반을 둔 DR의 두 가지 모듈로 구성된다. 특히 Doc2Vec 기법을 기반으로 하는 선호도 모델링에 시간적 요인을 적극적으로 적용하기 위해 TE 방식으

로 순차적인 항목으로 구성된 사용자 그룹의 상호작용을 문서로 구성하고 일정 기간별로 문장으로 분할하여 Doc2Vec 모델을 학습한다. 또한 카드 데이터에 존재하지 않는 MLP의 학습을 위한 결과 값(즉, 평점)을 생성하기 위하여 사용되는 RFM 분석을 적용하였다(Sharifhosseini, 2019).

한국의 가장 유명한 관광지 중 하나인 제주에서 발생한 실제 카드 거래 데이터에 대한 실험 결과, 제안된 방법이 최신 DNN 및 세션 기반의 추천 알고리즘을 포함한 다양한 추천 접근 방식보다 우수한 것으로 나타났다. 특히 제안된 방법은 상위 1개 항목 추천과 상위 10개 항목 추천 관점에서 안정적이고 강력한 성능을 보였다. 또한, RFM 분석으로 얻은 평가 점수를 사용하여 MLP 모델을 훈련했다. 이는 CTDDTR 방법이 명시적 피드백(즉, 평가 점수)의 부재를 극복할 수 있는 능력을 가지고 있으며 카드 거래 데이터를 기반으로 실제 응용 프로그램에 쉽게 적용될 수 있음을 강조한다. 또한, 관광 산업 RDF로 도출한 보조 요소들을 통합한 성능 평가를 통해, 관광 서비스 추천에 있어 해당 보조 요소들의 중요도를 밝혔으며 제안하는 방법이 이러한 요소를 아이템 추천 과정에 효과적으로 반영할 수 있음을 보여주었다.

이러한 월등한 성능에도 불구하고 본 연구에는 향후 더 고찰해야 할 세 가지 한계가 남아있다. 실험에서는 GDPR 정책을 고려하여 개인 정보를 제한하고 익명성을 강화하기 위한 전처리 과정을 거쳤기 때문에 실제 개인을 위한 아이템 추천 서비스에 많은 제한이 뒤따르게 된다. 개인 정보를 보호하면서도 개인화 추천이 가능하도록 제도와 데이터 처리 방법에 대한 후속 연구가 뒤따를 필요가 있다. 또한 CTDDTR의 범용성을 고려하여 간단한 MLP 네트워크를 사용했다는 것도 한계점으로 지적된다. 향후에는 GRU, LSTM, RNN과 같은 보다 정교한 인공지능망을 조사하고 각

각의 장단점을 신중하게 고려하여 더 개선할 여지를 찾을 수 있을 것으로 기대한다. 동시에 암시적 피드백에 기반을 둔 DNN 기반 추천 알고리즘도 추가적으로 비교 분석될 필요가 있다. 예를 들어, 거래 발생에 RFM 분석을 도입하여 평점을 생성하지 않고, 세션 기반 알고리즘인 STAMP과 같이 상호작용의 여부로 간주하여 DNN 모델을 학습하는 접근도 고려할 수 있다. 마지막으로 본 연구는 다양한 산업 분야와 관련된 카드 거래 데이터가 있음에도, 초기 연구로서 스마트 관광 분야에 집중하였다. 향후에는 다양한 산업 분야 전반에 걸쳐 제안한 기법을 적용하고, 나아가 데이터 제공 업체(data provider)인 카드사와 협력하여 가명 처리된 금융 데이터 기반 금융 플랫폼에 이식하여 실제 서비스 환경에서 실증하고 부가가치를 창출에 기여할 것이다.

본 연구에서는 사용자 그룹의 선호도를 정확하게 예측하기 위해 카드 거래 데이터로부터 시간성을 고려한 사용자 그룹-서비스 항목의 상호작용을 모델링하고 MLP 네트워크에서 다양한 요소를 통합하는 데 중점을 두었다. 결과적으로 이 연구는 관광객의 카드 거래 데이터 분석에 기반을 두어 관광객 그룹이 선호할 서비스를 더 정확히 예측하고 관광 서비스 목록을 제공하여 관광객의 의사결정(관광 일정 계획 등)을 촉진하는 결과를 제공함으로써 스마트관광 시스템 개발 연구에 기여한다. 또한, 지식 그래프 데이터를 고려함으로써 보다 많은 온톨로지 데이터가 활용될 필요성을 환기했다는 점에서 의의가 있다.

〈참고문헌〉

[국내 문헌]

1. 김태경, 구철모, 정남호 (2021). 스마트관광에서 웹 메타 데이터의 의미와 활용 방안. *관광연구*, 35(3), 5-21.
2. 문현실, 임진혁, 김도연, 조윤희 (2020). 시각 정보를 활용한 딥러닝 기반 추천 시스템. *지식경영연구*, 21(3), 27-44.
3. 정영진, 조윤희 (2017). 온라인 구매 행태를 고려한 토픽 모델링 기반 도서 추천. *지식경영연구*, 18(4), 97-118.
4. 최영제, 문현실, 조윤희 (2020). 트랜잭션 기반 추천 시스템에서 워드 임베딩을 통한 도메인 지식 반영. *지식경영연구*, 21(1), 117-136.
5. Al-Ghossein, M., Abdessalem, T., & Barre, A. (2018). Cross-domain recommendation in the hotel sector. *In Proceedings of the Workshop on Recommenders in Tourism*, Vancouver, Canada, 1-6.
6. Awangga, R. M., Pane, S. F., & Wijayanti, D. A. (2019). GURILEM: A novel design of customer rating model using K-Means and RFM. *EMITTER International Journal of Engineering Technology*, 7(2), 404-422.
7. Baek, J. W., & Chung, K. Y. (2020). Multimedia recommendation using Word2Vec-based social relationship mining. *Multimedia Tools and Applications*, 80(26), 1-17.
8. Cai, G., Lee, K., & Lee, I. (2018). Itinerary recommender system with semantic trajectory pattern mining from geo-tagged photos. *Expert Systems with Applications*, 94, 32-40.
9. Chaudhari, K., & Thakkar, A. (2019). A comprehensive survey on travel recommender systems. *Archives of Computational Methods in Engineering*, 27(5), 1-27.
10. Chen, L., Wu, Z., Cao, J., Zhu, G., & Ge, Y. (2020). Travel recommendation via fusing multiauxiliary information into matrix factorization. *ACM Transactions on Intelligent Systems and Technology*, 11(2), 1-24.
11. Chen, L., Yang, W., Li, K., & Li, K. (2021). Distributed matrix factorization based on fast optimization for implicit feedback recommendation. *Journal of Intelligent Information Systems*, 56(1), 49-72.
12. Esmaeili, L., Mardani, S., Golpayegani, S. A. H., & Madar, Z. Z. (2020). A novel tourism recommender system in the context of social commerce. *Expert Systems with Applications*, 149, 113301.
13. Esmeli, R., Bader-El-Den, M., & Abdullahi, H. (2020). Using Word2Vec recommendation for improved purchase prediction. *In Proceedings of the 2020 International Joint Conference on Neural Networks*, Glasgow, United Kingdom, 1-8.
14. Fessahaye, F., Perez, L., Zhan, T., Zhang, R., Fossier, C., Markarian, R., & Oh, P. Y. (2019). T-RECSYS: A novel music recommendation system using deep learning. *In Proceedings of the IEEE International Conference on Consumer Electronics*, Las Vegas, NV, USA, 1-6.
15. Fudholi, D. H., Rani, S., Arifin, D. M., & Satyatama, M. R. (2021). Deep learning-based mobile tourism recommender system. *Scientific Journal of Informatics*, 8(1), 111-118.
16. Guo, L., Liang, J., Zhu, Y., Luo, Y., Sun, L., & Zheng, X. (2019). Collaborative filtering recommendation based on trust and emotion. *Journal of Intelligent Information Systems*, 53(1), 113-135.
17. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 770-778.
18. He, X., Liao, L., Zhang, H., Nie, L., Hu, X., & Chua, T. (2017). Neural collaborative filtering. *In Proceedings of the 26th International Conference on World Wide Web*, Perth, Australia, 173-182.
19. Hong, M. (2021). Decrease and conquer-based parallel tensor factorization for diversity and real-time of multi-criteria recommendation. *Information Sciences*, 562, 259-278.
20. Hong, M., & Jung, J. J. (2021a). Multi-criteria tensor model consolidating spatial and temporal information for tourism recommendation. *Journal of Ambient*

- Intelligence and Smart Environments*, 13(1), 5–19.
21. Hong, M., & Jung, J. J. (2021b). Multi-criteria tensor model for tourism recommender systems. *Expert Systems with Applications*, 170, 114537.
 22. Hong, M., & Jung, J. J. (2022). Sentiment aware tensor model for multi-criteria recommendation. *Applied Intelligence*, 1–20, <https://doi.org/10.1007/s10489-022-03267-z>
 23. Katarya, R., & Arora, Y. (2020). CAPSMF: A novel product recommender system using deep learning based text analysis model. *Multimedia Tools and Applications*, 79(47), 35927–35948.
 24. Kotiloglu, S., Lappas, T., Pelechris, K., & Repoussis, P. P. (2017). Personalized multi-period tour recommendations. *Tourism Management*, 62, 76–88.
 25. Le, Q. V., & Mikolov, T. (2014). Distributed representations of sentences and documents. *In Proceedings of the 31th International Conference on Machine Learning*, Beijing, China, 1188–1196.
 26. Li, M., Wu, H., & Zhang, H. (2019). Matrix factorization for personalized recommendation with implicit feedback and temporal information in social ecommerce networks. *IEEE Access*, 7, 141268–141276.
 27. Liu, Q., Chen, E., Xiong, H., Ge, Y., Li, Z., & Wu, X. (2014). A cocktail approach for travel package recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 26(2), 278–293.
 28. Liu, Q., Zeng, Y., Mokhosi, R., & Zhang, H. (2018). STAMP: Short-term attention/memory priority model for session-based recommendation. *In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, London, UK, 1831–1839.
 29. Lu, J., Wu, D., Mao, M., Wang, W., & Zhang, G. (2015). Recommender system application developments: A survey. *Decision Support Systems*, 74, 12–32.
 30. Misztal-Radecka, J., Indurkha, B., & SmywinskiPohl, A. (2021). Meta-User2Vec model for addressing the user and item cold-start problem in recommender systems. *User Modeling and User-Adapted Interaction*, 31(2), 261–286.
 31. Musto, C., Semeraro, G., De Gemmis, M., & Lops, P. (2015). Word embedding techniques for content-based recommender systems: An empirical evaluation. *In Proceedings of the 9th ACM Conference on Recommender Systems (RecSys)*, Vienna, Austria.
 32. Nassar, N., Jafar, A., & Rahhal, Y. (2020). A novel deep multi-criteria collaborative filtering model for recommendation system. *Knowledge-Based Systems*, 187, 104811.
 33. Ozsoy, M. G. (2016). From word embeddings to item recommendation. *CoRR*, abs/1601.01356
 34. Park, S. T., & Liu, C. (2020). A study on topic models using Lda and word2vec in travel route recommendation: Focus on convergence travel and tours reviews. *Personal and Ubiquitous Computing*, 26, 1–17.
 35. Pessemier, T. D., Dhondt, J., & Martens, L. (2017). Hybrid group recommendations for a travel service. *Multimedia Tools and Applications*, 76(2), 2787–2811.
 36. Rakesh, V., Jadhav, N., Kotov, A., & Reddy, C. K. (2017). Probabilistic social sequential model for tour recommendation. *In Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, Cambridge, United Kingdom, 631–640.
 37. Shambour, Q. (2021). A deep learning based algorithm for multi-criteria recommender systems. *Knowledge-Based Systems*, 211, 106545.
 38. Sharifhosseini, A. (2019). A case study for presenting bank recommender systems based on bon card transaction data. *In Proceedings of the 9th International Conference on Computer and Knowledge Engineering*, Iran, 72–77.
 39. Tahmasebi, H., Ravanmehr, R., & Mohamadrezaei, R. (2021). Social movie recommender system based on deep autoencoder network using Twitter data. *Neural Computing and Applications*, 33(5), 1607–1623.
 40. Tan, C., Liu, Q., Chen, E., Xiong, H., & Wu, X. (2014). Object-oriented travel package recommendation. *ACM Transactions on Intelligent Systems and Technology*, 5(3), 1–26.
 41. Thasal, R., Yelkar, S., Tare, A., & Gaikwad, S. (2018). Information retrieval and de-duplication for tourism recommender system. *International Research Journal of*

- Engineering and Technology*, 5(03), 1683–1687.
42. Vuong Nguyen, L., Nguyen, T. H., Jung, J. J., & Camacho, D. (2021). Extending collaborative filtering recommendation using word embedding: A hybrid approach. *Concurrency and Computation: Practice and Experience*, e6232. <https://doi.org/10.1002/cpe.6232>
 43. Wang, M. (2020). Applying internet information technology combined with deep learning to tourism collaborative recommendation system. *Plos One*, 15(12), e0240656.
 44. Zhang, S., Yao, L., Sun, A., & Tay, Y. (2019). Deep learning based recommender system: A survey and new perspectives. *ACM Computing Surveys*, 52(1), 1–38.
 45. Zhao, P., Xu, C., Liu, Y., Sheng, V. S., Zheng, K., Xiong, H., & Zhou, X. (2021). Photo2Trip: Exploiting visual contents in geo-tagged photos for personalized tour recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 33(4), 1708–1721.
 46. Zheng, X., Luo, Y., Sun, L., Zhang, J., & Chen, F. (2018). A tourism destination recommender system using users' sentiment and temporal dynamics. *Journal of Intelligent Information Systems*, 51(3), 557–578.
 47. Zhu, G., Cao, J., Li, C., & Wu, Z. (2017). A recommendation engine for travel products based on topic sequential patterns. *Multimedia Tools and Applications*, 76(16), 17595–17612.

저 자 소 개



홍 민 성 (Minsung Hong)

현재 경희대학교 스마트관광연구소의 연구 교수로 재직 중이다. 중앙대학교에서 컴퓨터공학 박사 학위를 취득하였고, 서부노르웨이연구소에서 박사후연구원을 역임하였다. 주요 관심분야는 추천시스템, 인공지능, 빅데이터, 데이터마이닝, 자연어처리 기술 등이다. 지금까지 Information Sciences, Expert Systems with Applications, Electronic Commerce Research and Applications, Journal of Ambient Intelligence and Humanized Computing 등 주요 학술지에 논문을 발표하였다.



김 태 경 (Taekyung Kim)

현재 광운대학교 경영대학 경영학부 교수로 재직 중이다. 서울대학교에서 경영학 박사 학위를 취득하였다. 주요 관심분야는 지식경영시스템, 지식공유, 지능형 정보기술 등이다. 지금까지 Information Systems Research, Journal of Management Information Systems, International Journal of Information Management 등 주요 학술지에 논문을 발표하였다.



정 남 호 (Namho Chung)

현재 경희대학교 호텔관광 경영대학 학장과 스마트관광원 교수를 겸임하고 있으며, 스마트관광연구소의 소장이다. 그는 영국의 University of Surrey의 호텔 및 관광 경영 대학에서 방문 연구원으로 재직하였다. 현재 대한민국 교육부와 한국연구재단이 지원하는 SSK대형 센터사업인 스마트관광도시 프로젝트와 BK21 Four 교육사업단을 이끌고 있다. 주요 관심분야는 여행 행동, 정보 검색 및 의사 결정, 목적지 마케팅, 지식 관리 및 목적지 관리 조직을 위한 정보 시스템 개발이다. 또한 탁월한 업적으로 경희대학교 명예의 전당에 이름을 올렸다.

〈 Abstract 〉

Card Transaction Data-based Deep Tourism Recommendation Study

Minsung Hong^{*}, Taekyung Kim^{**}, Namho Chung^{***}

The massive card transaction data generated in the tourism industry has become an important resource that implies tourist consumption behaviors and patterns. Based on the transaction data, developing a smart service system becomes one of major goals in both tourism businesses and knowledge management system developer communities. However, the lack of rating scores, which is the basis of traditional recommendation techniques, makes it hard for system designers to evaluate a learning process. In addition, other auxiliary factors such as temporal, spatial, and demographic information are needed to increase the performance of a recommendation system; but, gathering those are not easy in the card transaction context. In this paper, we introduce CTDDTR, a novel approach using card transaction data to recommend tourism services. It consists of two main components: i) Temporal preference Embedding (TE) represents tourist groups and services into vectors through Doc2Vec. And ii) Deep tourism Recommendation (DR) integrates the vectors and the auxiliary factors from a tourism RDF (resource description framework) through MLP (multi-layer perceptron) to provide services to tourist groups. In addition, we adopt RFM analysis from the field of knowledge management to generate explicit feedback (i.e., rating scores) used in the DR part. To evaluate CTDDTR, the card transactions data that happened over eight years on Jeju island is used. Experimental results demonstrate that the proposed method is more positive in effectiveness and efficacies.

Key Words: Smart tourism recommendation, Deep learning, Resource description framework, RFM analysis

* Smart Tourism Research Center, Kyung-Hee University

** Division of Business Administration, Kwangwoon University

*** Smart Tourism Education Platform, Kyung-Hee University