

Route matching delivery recommendation system using text similarity

Jeongeun Song*, Yoon-Ah Song*

*Student, Graduate School of information, Yonsei University, Seoul, Korea

*Student, Graduate School of information, Yonsei University, Seoul, Korea

[Abstract]

In this paper, we propose an algorithm that enables near-field delivery at a faster and lowest cost to meet the growing demand for delivery services. The algorithm proposed in this study involves subway passengers (shipper) in logistics movement as delivery sources. At this time, the passenger may select a delivery logistics matching subway route. And from the perspective of the service user, it is possible to select a delivery man whose route matches. At this time, the delivery source recommendation is carried out in a text similarity measurement method that combines TF-IDF&N-gram and BERT. Therefore, unlike the existing delivery system, two-way selection is supported in a man-to-man method between consumers and delivery man. Both cost minimization and delivery period reduction can be guaranteed in that passengers on board are involved in logistics movement. In addition, since special skills are not required in terms of transportation, it is also meaningful in that it can provide opportunities for economic participation to workers whose job positions have been reduced.

▶ **Key words:** route recommendation, route matching, text similarity, short-distance delivery, subway delivery, platform service

[요 약]

본 연구에서는 급증하는 배송 서비스 수요에 맞춰 더 신속하고 최저 비용으로 근거리 배송을 가능하게 하는 알고리즘을 제안하고자 한다. 본 연구에서 제안하는 알고리즘에서는 배송원으로 지하철 승객을 물류 이동에 참여시킨다. 이때 승객은 이동 경로와 일치하는 배송 물류를 선택할 수 있다. 그리고 서비스 이용자의 입장에서는 현재 근처에 경로가 일치하는 배송원을 선택할 수 있다. 이때 배송원 추천은 TF-IDF&N-gram과 BERT를 결합한 텍스트 유사도 측정 방식으로 진행된다. 따라서 기존 택배 시스템과 달리 소비자-배송원 간의 man-to-man 방식으로 양방향 선택을 지원한다. 탑승 중인 승객을 물류 이동에 참여시킨다는 점에서 비용 최소화와 배송 기간 단축을 모두 보장할 수 있다. 더하여 운송 측면에서도 특별한 기술을 요하지 않으므로, 일자리 입지가 축소된 노동자들에게 경제 참여 기회를 제공할 수 있다는 점에서도 의의가 있다.

▶ **주제어:** 경로추천, 경로 매칭, 텍스트유사도, 단거리 매칭, 지하철배송, 플랫폼 서비스

-
- First Author: Jeongeun Song, Corresponding Author: Yoon-Ah Song
 - Jeongeun Song (stuplant@yonsei.ac.kr), Graduate School of information, Yonsei University
 - Yoon-Ah Song (dbsdk1683@yonsei.ac.kr), Graduate School of information, Yonsei University
 - Received: 2022. 08. 02, Revised: 2022. 08. 26, Accepted: 2022. 08. 26.

I. Introduction

최근 COVID-19 팬데믹의 영향으로 비대면 문화가 확산되면서 배송 서비스의 수요가 급증하였다. 특히 그중에 편의점과 편의점 간을 이동하는 편의점 택배가 최저 비용을 내세우며 서비스 시작 1년 만에 이용량 6배 증가로 큰 인기를 끌고 있다.[1] 이 서비스는 ‘반값 택배’로 지칭되며 GS25 Postbox 공식 홈페이지 배송 기한 안내에 따르면 배송 기한이 4일로 통상의 택배보다 2-3일 정도 더 소요된다. 하지만 기존 서비스 대비 반값이라는 장점 때문에 중고 거래에 주로 이용되고 있다. 반값 택배 이용 고객 5,000명을 대상으로 분석한 GS25의 자료에 따르면 이용자의 69.5%가 중고 거래를 위해 반값 택배를 이용한다고 응답하였다. 이용자의 연령은 20대와 30대가 전체의 82%로 대다수를 차지하고 있었다.[2]

2022년 3월 통계청이 작성한 ‘2021 온라인쇼핑동향조사’에 따르면 전년 동월 대비 온라인 쇼핑 총 거래액은 11.5%, 모바일 쇼핑 거래액은 17.5% 증가하였다.[3] 그에 따라 택배 물동량도 COVID-19 팬데믹 발생 이전 대비 20.9% 증가하였다.[4]

택배 물량이 증가함에 따라 기존 택배 서비스는 배송 지연이나 파손과 같은 문제가 발생하고 있으며 비용 또한 편의점 ‘반값 택배’에 비해 저렴하지 않다는 단점이 있다. 하지만 ‘반값 택배’는 기존 택배 서비스보다 2-3일 배송 기간이 더 소요된다는 단점이 있다. 따라서 본 연구에서는 전체 물량 중 근거리 배송 시스템을 개선하여 배송원 1인당 담당하는 물량의 축소를 통해 물품의 파손을 방지하고 이와 함께 비용의 최소화, 배송 기간의 단축을 해결할 수 있는 알고리즘을 제안하고자 한다.

우선 물품의 파손을 방지하기 위해서는 배송원 1인당 배송하는 품목이 적어야 할 것이다. 따라서 배송원이 다수 필요하다. 배송 인력의 수월한 확보를 위해 이동 수단으로 지하철을 선택하였다. 이는 누구나 이용할 수 있는 대중교통이어서 운전면허와 같은 조건이 필요 없기 때문이다.

지하철을 이동 수단으로 선택하였기 때문에 비용의 최소화도 함께 해결할 수 있다. 지하철로 이동하는 승객 중 배송에 참여하고자 하는 승객 본인의 경로를 입력받아 물품의 이동 경로에 이용할 예정이다. 따라서 이미 승객이 소비하는 교통비를 감면해주는 방식이나 포인트 적립 방식으로 이동 수수료를 최소화할 수 있다. 예를 들면 최근 토스 애플리케이션에서 인기를 끌고 있는 만보기 서비스와 같이 해당 장소를 방문하면 포인트나 쿠폰을 지급하는 방식으로 운영할 수 있다. 구체적으로 물품을 이동하여 해

당 장소에 큐알 코드를 찍으면 포인트가 자동 지급되는 방식이다. 토스 만보기에서 미션 장소 방문 금액은 하루 최대 140원으로 소액을 지급하고 있지만 해당 기능이 추가된 이후 2021년 8월 46만 명이었던 누적 사용자 수가 2022년 5월 기준으로 850% 이상 증가하며 400만 명을 기록했다.[5] 따라서 본 연구의 알고리즘을 이용하여 물품 이동 미션 서비스를 만보기 서비스와 같이 구축한다면 다수의 배송 인력을 마련하면서도 상대적으로 적은 비용으로 운영 가능하다.

이동 중인 승객의 정보를 입력받아 지하철역에서 역으로 물품 보관함과 같은 소형 거점을 거쳐 실시간으로 운영되는 시스템이기 때문에 배송 기간 또한 ‘반값 택배’에 비해 현저히 줄어들게 된다. ‘반값 택배’가 기존 택배 시스템보다 배송 기간이 더 소요되는 대신 저렴한 비용을 선택하였지만 본 연구에서 제시하는 시스템은 배송 기간과 비용의 최소화 두 가지 조건을 모두 만족시킬 수 있다.

이 시스템에서는 소비자의 물리적 위치와 시간에 따라서 보다 폭넓은 선택권을 부여하기 위해 소비자가 원하는 경로의 승객(배송원)을 추천받고 선택할 수 있는 알고리즘을 제안한다. 이때 추천 시스템은 통계적 기법인 TF-IDF와 신경망 기법인 BERT를 결합한 방법론을 기반으로, 경로에 대한 텍스트 유사도를 통해 제시한다.

이 알고리즘을 통해 소비자(이용자)에게 정확한 배송 경로와 최저 비용, 배송 기간의 단축, 개인화된 맞춤 서비스를 제공할 수 있을 것이라고 기대한다.

이 연구의 구성은 2장에서 현재 배송 서비스의 현황과 함께 선행연구를, 3장에서는 배송원 추천 시스템 알고리즘 방법론을, 4장에서는 실험 결과를 살펴보는 것으로 진행된다.

II. Research background

1. Related works

기존 택배 시스템은 근거리 이동이라도 분류를 위해 지역 허브를 거치기 때문에 배송에 하루 이상 소요되고 있다. 택배 물동량은 2012년 이후 꾸준히 증가 추세였으나 2019년과 비교하면 20.93% 급증했다.[4] 그에 따라 물량 폭주로 인한 배송 지연이나 분실 문제 등 소비자의 불편 사항도 증가했다. 한국소비자원에서 분석한 2020년 9월 소비자 빅데이터 트렌드를 살펴보면 상담 증가율 상위 품목에서 ‘택배화물운송서비스’가 55.3%로 가장 큰 증가폭을 보였다.[6]

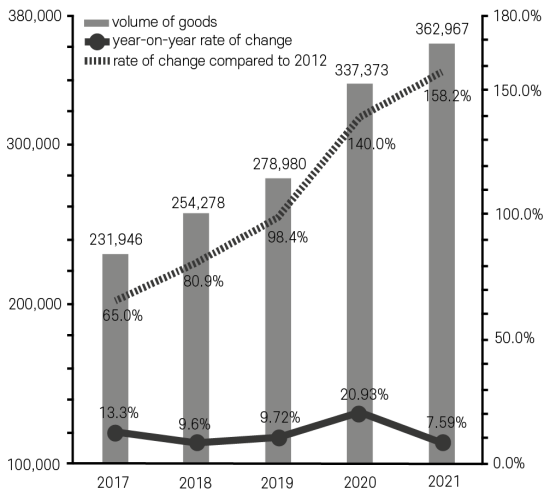


Fig. 1. Domestic market delivery volume trend

박한영 외(2021)에서는 COVID-19 이후로 물류량이 약 56.5%가 증가했으며 택배 인프라를 위한 시설 마련이 단기간에 완성되는 것이 아니므로 이에 따른 방안 모색의 시급함을 언급했다.[7]

신승진 외(2020)에서는 도시 내 물류 거점을 만드는 것의 현실적 어려움을 언급하였다. 그리고 배송 시간 단축 경쟁 심화와 함께 소형 물류 거점지의 중요성을 설파하였다.[8]

곽동재 외(2021)에서는 신속성과 배송 안전성이 온라인 식품 배송에 관한 소비자 만족도에 미치는 정도가 유의 수준 0.001에서 통계적으로 유의미하게 나타났다.[9] 추가로 물류서비스 품질의 경제성, 적시성, 안전성이 고객만족과 신뢰에 유의미한 정(+)의 관계가 있었다.[10]

본 논문의 핵심 알고리즘과 연결되는 TF-IDF, BERT는 최근 NLP연구와 함께 많은 연구가 진행되고 있다.

Rahul Shrivastava 외(2019)는 TF-IDF를 활용해 Amazon 제품 광고 간의 유사도를 통해서 제품 추천 알고리즘을 마련했다.[11]

Mithun M.sanjeev 외(2020)는 구글에서 사전 학습한 BERT를 활용하여 질의와 이메일 간의 의미 유사도 방법론을 제안하였다.[12]

Melika Golestani 외(2021)에서는 사전 훈련된 BERT로 문장의 유사도를 파악하여 여러 문장에 대한 정렬을 실험하였고 신경망 모델의 효과를 입증하였다.[13]

따라서 본 연구에서는 소형 거점 마련과 텍스트 유사도 분석을 결합하여 근거리 물류 이동에 관해 개선 방안을 제시하고자 한다.

III. Recommendation system through route matching

본 연구에서는 근거리 물류 이동의 비효율적인 경로를 단축하고 최저 비용으로 소비자의 위치와 시간에 따라 적합한 배송원(승객)을 선택할 수 있게 하는 것에 목적이 있다. 따라서 정시성을 보장하고 배송 비용의 최소화, 배송 인력의 수월한 확보를 위해 이동 수단으로 수도권 지하철을 선택하였다. 지하철은 평일 기준 대략 오전 05시 부터 밤 12시까지 운행되고 저렴한 이동 비용과 정시성을 확보할 수 있는 것이 가장 큰 장점이다. 국토교통부에서 교통카드 데이터를 기반으로 하여 분석한 대중교통 이용실태 분석에 따르면 수도권에서 하루 730만 명이 대중교통을 이용하고 있는 것으로 나타났다.[14] 이용객이 많은 만큼 본 연구에서 제안하는 배송 알고리즘 서비스의 네트워크 효과 가능성도 클 것으로 기대된다.

요금 측면에서는 편도 이용에 평균적으로 2,162원을 사용하여[14] 이동 비용 또한 저렴하기 때문에 소규모의 근거리 배송에 적합하다고 할 수 있다. 앞서 언급하였듯이 본 알고리즘을 서비스로 실현할 때에는 이동 미션을 배송원이 수행하면 포인트를 지급하는 방식으로 운영될 것이다. 토스 만보기 서비스의 사례에서 살펴본 바와 같이 소액의 포인트를 지급한다고 하더라도 이용자 증가의 가능성을 기대할 수 있다.

수도권 배송의 경우에는 일반 택배 시스템보다 이동 시간도 단축이 가능하다. 기존 택배 시스템은 같은 지역 내에서 배송이더라도 분류를 위해 분류 센터를 거친다. 따라서 기본적으로 배송에 하루 이상 소요된다. 하지만 본 연구에서 제시하는 알고리즘 시스템으로 배송한다면 분류를 거치지 않고 원하는 이동 경로를 입력하는 즉시 플랫폼을 통해 'man-to-man'으로 매칭되어 더 빠른 속도를 보장 받을 수 있다. 또한 역내 물품보관소나 소매점을 소형 거점으로 활용하므로 물류센터가 따로 필요 없기 때문에 물류센터 유지 비용과 같은 부가적인 비용이 소요되지 않는다. 이 또한 서비스 비용 최소화에 기여한다.

국토교통부의 대중교통 이용실태 분석 자료에 따르면 수도권 출근시간 평균 이동시간은 같은 지역 내에서 서울 47분, 경기도 1시간 36분, 인천광역시 50분이었다. 지역 간 이동은 서울-경기도 최대 1시간 27분, 서울-인천광역시 최대 1시간 33분, 경기도-인천광역시 최대 1시간 52분 정도였다.[14] 혼잡 시간에 이와 같은 기동성을 보인다면 혼잡도가 낮은 시간에는 더 높은 기동성으로 더욱 빠르게 이동할 수 있을 것이라 예상된다. 수도권 출근 시간 지역

사이 평균 이동 시간을 정리해보면 그림 2와 같다. 지역 사이의 이동은 출발점이 어딘가에 따라서 2-3분 정도 차이가 있다.

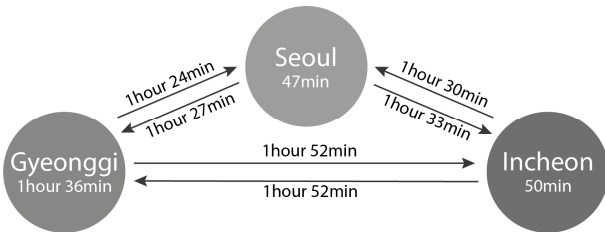


Fig. 2. The average travel time between regions in the Seoul metropolitan area during rush hour

물품 이동 인력은 수도권 지하철에 탑승하는 승객들을 대상으로 한다. 앞에서 언급한 바에 따르면 수도권에서 하루에 이동하는 인구는 약 730만 명이다. 이들 중 물류 이동에 참여하고자 하는 승객이 본인의 이동 경로를 입력하면 이동 경로 상에서 픽업할 수 있는 물품을 가진 사용자의 위치 그리고 사용자의 입장에서 현재 근처에 경로가 일치하는 배송원(승객)을 선택할 수 있다. 기존에 지하철로 이동 중인 승객의 경로를 이용하여 물류 이동에 참여시킨다는 점에서 비용의 최소화 그리고 신속성 두 가지를 모두 해결할 수 있다. 또한 기존 배송 시스템에서는 사용자에게 배송원에 대한 선택권을 부여하지 않았는데 이 시스템에서는 사용자에게 직접 배송원을 선택할 수 있도록 하여 사용자의 선택권을 증가시킬 수 있다. 사용자가 신뢰하는 배송원을 스스로 선택할 수 있기 때문에 배송 서비스에 대한 만족도가 높아질 것이라 기대한다.

1. Subway Station Dataset Construction Process

1-1. Subway Route API

본 연구에서 제안하고자 하는 추천 시스템은 소비자가 원하는 배송 경로 데이터를 기반으로, 가장 적합한 배송원을 추천하는 것이다. 따라서 해당 시스템을 실험하기 위해서 임의로 경로 데이터를 구축하였다.

관련 데이터는 공공데이터 포털(www.data.go.kr)의 서울특별시_대중교통환승경로 조회 서비스 API를 호출하여 데이터셋을 마련하였다. 그중에서 지하철 이동 경로 API를 활용하였다. 해당 API는 출발역과 도착역의 '위도'와 '경도' 좌표를 입력하면 이에 상응하는 경로를 출력해주는 시스템이다. '위도'와 '경도' 좌표는 국가 철도 공단을 통해 마련할 수 있었다. 이때 국가 철도 공단을 통해 수집한 수도권 지하철역의 '위도'와 '경도' 좌표를 토대로 '출발역'과 '도착역'세트를 랜덤으로 구성하였고 API를 통해 최단

경로를 추출하였다. 그러나 해당 API는 중간 경로를 생략한 채, '출발역-환승역-도착역' 정보에 관해서만 데이터 추출이 가능하였다. 이는 정확한 경로 데이터를 확인할 수 없어, 사용자와 배송원 간의 경로 유사도 측정에 있어 어려움이 있었다. 따라서 보다 정확한 경로 추천을 위해서는 출발역과 도착역 사이에 생략된 역 데이터를 채워 넣어야 했다. 또한 수도권 노선 수와 경로가 복잡하여 이를 구조적으로 해결할 수 있는 방안을 모색할 필요가 있었다.

우선 수도권 전 지역에 운행 중인 호선 24개에 대한 모든 경우의 수를 목록으로 만들었다. 한 개 노선의 끝과 끝 부분에서 임의로 한쪽 종착역을 시작 지점으로 삼고 그 반대편을 종착점으로 삼았다. 그리고 시작점에서 종착점까지를 순방향으로, 종착점에서 시작점까지를 역방향(re)으로 정의했다. 하지만 순환 고리를 가지고 있는 2호선과 6호선, 여러 갈래 노선을 가지고 있는 1호선과 4호선의 경우는 추가 데이터 처리가 필요하였다. 1호선과 4호선 같이 갈래 노선이 있는 경우에는 모든 갈래의 종착점을 제 n의 호선으로 취급하였다. 예를 들어 갈래가 두 개인 노선이 있다면, 종착점을 1과 2로 나누어 시작점에서 종착점1까지, 시작점에서 종착점2까지를 기록하는 방식으로 데이터를 구축하였다. 역방향의 경우는 종착점1에서 시작점, 종착점2에서 시작점으로 구축하였다. 마지막으로 2호선과 6호선과 같은 순환선의 경우는 더 복잡한 경우의 수를 가지고 있었다. 시작점과 종착점을 이어 노선 데이터를 구축하되, 순환 고리를 고려할 필요가 있었다. 기본적으로 시작점과 종착점의 모든 경우의 수를 마련하였으며, 추가로 갈래 노선을 시작/종착점으로 설정하여 순환 노선을 또다시 구성하였다. 순환선 구조인 2호선을 예로 들면, 기본 시작점과 종착점인 까치산과 신철동을 기준으로 총 4개의 제 n호선을 마련하였다. 그리고 갈래 노선인 신도림을 기준으로 신도림→대림, 신도림→문래 방향으로 또다시 호선을 구축하였다. 덧붙여 또 다른 갈래 노선인 성수역을 기준으로 성수역→독섬역, 성수역→건대입구역 방향으로 다시 호선을 마련하였다. 수집한 노선별 지하철역 목록은 부록에 첨부하였다.

2. Theoretical Approach of Methodology

2-1. Statistical-based method and Neural network-based method

텍스트 유사도 측정 방식으로는 크게 통계 기반 방식과 신경망 기반 방식이 있다. 통계 기반 방식의 경우는 단어 또는 단어의 집합인 구를 단위로 유사도를 측정한다. 이에 반해 신경망 기반 방식의 경우는 단어와 구를 넘어 문맥을

고려하여 유사도 측정을 진행한다. 이 때문에 신경망 기반 방식이 최근 NLP연구에서 기존 한계를 극복할 수 있는 해결책으로 제시되어 왔다. 그럼에도 불구하고, 신경망 기반 방식이 100% 실험 결과를 해석할 수 없다는 점에서 통계 기반 방식을 도외시킬 수 없다. 이 같은 이유로 본 연구에서는 통계 기반 방식과 신경망 기반 방식을 모두 활용하고자 한다.

해당 추천 시스템에서는 일반적인 텍스트 유사도와 다르게 노선 경로 방향을 인지하는 것이 필수적이다. 즉 역방향으로 이동하는 배송원을 유사도가 비슷하다는 이유로 추천할 수 없다. 따라서 통계적 기반 방식인 TF-IDF&N-gram으로 순서를 고려한 유사도를 1차적으로 진행하고, 신경망 기반의 방식인 BERT를 통해 2차적으로 유사도를 측정하여 배송원을 추천하고자 한다.

2-2. TF-IDF

TF-IDF(Term Frequency - Inverse Document Frequency)는 텍스트 마이닝과 분석에서 활용하는 대표적인 통계적 기법이다. 이 중에서 TF는 단어 빈도를 의미하며, 문서 내에서 해당 단어가 얼마나 빈번하게 사용되는가를 통계적으로 산출한다. 그 값이 클수록 문서 내 중요도가 높다고 판단할 수 있다. 하지만 TF만을 기법으로 활용할 경우, 문서 내에서 큰 의미를 갖지 않을 수 있는 조사, 어미의 중요도를 높게 산출할 가능성이 있다. 이를 위해서 문서 빈도의 역수를 의미하는 IDF를 활용한다. IDF는 단어 하나가 문서 전체에서 얼마나 공통적으로 등장하는지를 나타내는 값이다.[15]

TF-IDF는 TF와 IDF를 곱한 값으로 그 수식은 다음과 같다.

$$tf(t, d) = 0.5 + \frac{0.5 * f(t, d)}{\max f(w, d) : w \in d}$$

$$idf(t, D) = \log \frac{|D|}{|d \in D : t \in d|}$$

$$tfidf = tf(t, d) * idf(t, D)$$

2-3. N-gram

N-gram은 문장 내에서 연속하는 n개의 단어들을 하나의 단어들로 묶어서 한 문치로 만드는 기법이다. 만약 '용산역-서울역-시청역'이라는 지하철 노선이 있을 때, N-gram을 적용하지 않는다면 위 경로를 각각의 토큰으로써 '용산역', '서울역', '시청역'으로 인식한다. 하지만

N-gram(이때 n=3)적용 시, 하나의 토큰으로써 '용산역 서울역 시청역'으로 인식한다.

2-4 BERT

BERT(Bidirectional Encoder Representations from Transformers)는 Transformer의 Encode부분을 stack하여 만든 모델로, Transformer와 달리 양방향으로 텍스트를 파악하기 때문에 NLP연구에서 주목하고 있는 모델이다. BERT는 그림 3.과 같이 Pre-training을 통해 언어를 학습하고 Fine-tuning에서 태스크에 맞게 활용 가능하다. 그 태스크로는 번역, 분류, Q&A, 유사도 측정 등이 있다.[16][17]

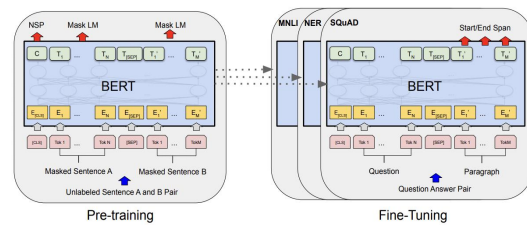


Fig. 3. BERT[16][17]

3. Recommendation System Experiment

본 연구에서는 서비스 이용객이 배송 요청을 했을 때를 가정하여, 그 경로에 대한 유사도를 바탕으로 배송원을 추천하는 실험을 진행한다. 전체 개요는 다음 그림 4.와 같다.

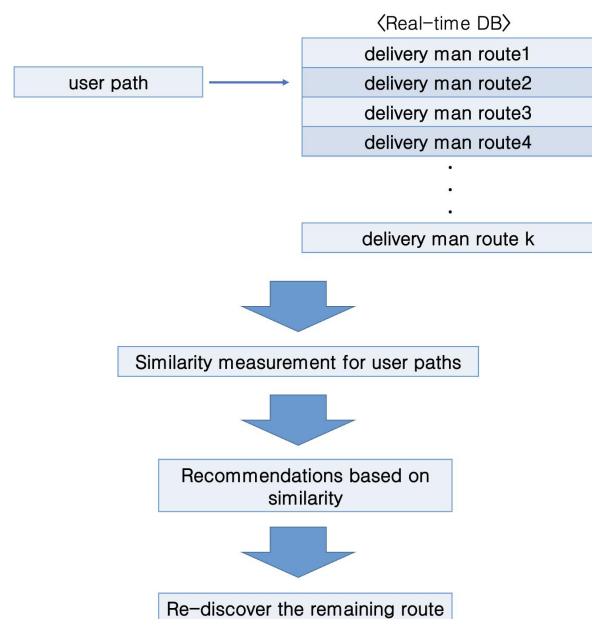


Fig. 4. System Summary

3-1. Data Building

제안하는 데이터 구축 방법은 다음과 같다.

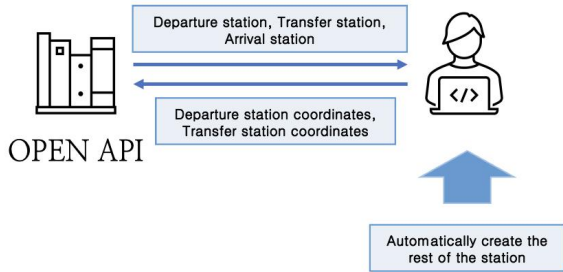


Fig. 5. data building process

본 연구에서는 ‘출발역-도착역’ 정보에 대하여 그림 5.와 같이 OPEN API를 통해 추천 경로를 얻고, 그중에서 가장 짧은 최단 경로 하나만을 택하였다. 그리고 앞서 구축한 수도권 지하철 노선 목록에 따라 자동으로 생략된 역을 메꾸어 주었다.

해당 추천 시스템 실험에서 사용한 배송원 데이터는 1만 8000천 건이다. 이는 2022년 5월 시간대별로 평균 승하차 인원이 가장 많은 시간대와 역이 08-09시, 가산디지털단지역(18,711명)으로 조사되었기 때문이다.

3-2. Similarity Measurement

유사도 측정은 그림 6. 도식과 같이 진행된다.

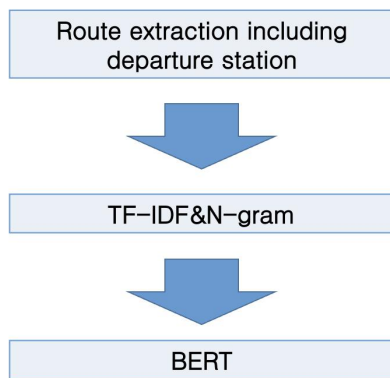


Fig. 6. Similarity Measurement Process

유사도 실험에서 소비자(이용자)가 원하는 배송 경로는 ‘가산디지털단지역’을 출발역으로 하였다. 이는 위에서 언급한 바와 같이 출퇴근 승/하차 인원이 가장 많은 역으로 ‘가산디지털단지역’이 조사되었기 때문이다. 도착역의 경우는 랜덤으로 선정하였고 ‘논현역’으로 하였다.

추가로 내선순환 이용 시의 배송원 추천을 알아보고자, 일일 승하차 인원이 가장 많은 2호선 ‘잠실역’을 대상으로

도 실험을 진행하였다. 이 경우는 앞선 실험과 반대로 ‘잠실역’을 도착역으로 선정하였다. 출발역의 경우는 랜덤으로 선정하였고 ‘사당역’으로 하였다.

유사도 측정에 앞서, 소비자(이용자)의 출발역인 ‘가산디지털단지역’과 ‘사당역’을 포함하고 있는 경로만을 추렸다. 이는 물품보관소에 있는 소비자(이용자)의 물품을 픽업하기 위한 과정이라 볼 수 있다. 결과적으로 1만 8천 건에 대하여, 각각 2143건, 1892건이 추려졌다.

1차 유사도 측정으로는 TF-IDF&N-gram을 활용하였다. 이는 경로의 방향성을 고려하기 위함이다. N-gram은 4로 설정하였다. 이때 추천 시스템의 작동 시간을 단축하고자 2143건, 1892건 중 각각 유사도가 높은 100명의 배송원 경로만을 출력하였다.

2차 유사도 측정으로는 1차 유사도 측정에서 출력된 100명의 경로에 대하여 진행하였다. 이때 한국어에 일반적으로 활용할 수 있도록 훈련된 BERT를 활용하여 유사도 순위와 유사도를 출력하였다. 해당 모델은 2018년 구글에서 공개한 다국어 BERT로, 104개국의 언어를 학습한 모델이다.

IV. Experiment result

해당 실험이 비지도 학습이기 때문에 추천 시스템 평가는 BERT모델을 통한 유사도로 검증하였다. 추가로 ‘출발역 이전 불일치 역 수’, ‘소비자(이용자)와의 경로 일치 역 수’, ‘마지막 경로 일치 이후 불일치 역 수’를 분석하였다. 아래는 추천 배송원과 그 유사도를 출력한 결과이다.

Table 1. Similarity measurement result 1

	mismatch before departure ststion	match path with consumer	mismatch since last matching path	similarity
Gasan Digital Complex-> Konkuk university entrance	0	15	5	0.9852
Gasan Digital Complex-> Ogeum	0	15	13	0.9539
GwangmyeongInter section-> Shinbanpo	2	15	1	0.9372

Cheonwang->Geumho	3	15	5	0.9130
Seoksu->Daechi	3	15	6	0.9069
Doksan->Sangcheon	1	15	20	0.8946
Doksan->Ganeung	1	15	21	0.8762
GwangmyeongIntersection->Yangsu	2	15	17	0.8727
Onsu->Sangbong	4	15	7	0.8721
Seoksu->Guliy	3	15	10	0.8635

표 1은 출퇴근 승/하차 인원이 가장 많은 역인 '가산디지털단지역'을 출발역으로, 도착역을 '논현역'으로 설정한 실험의 결과를 제시한다.

Table 2. Similarity measurement result 2

	mismatch before departure ststion	match path with consumer	mismatch since last matching path	similarity
Dongjak->Gangdong Office	2	11	2	0.9258
Sillim->Songpa	4	11	2	0.8999
Seoul National university->Hagye	2	11	16	0.8329
Sindaebang->Seolleung	5	7	0	0.8188
Daelim->Ewha university	7	11	18	0.8026
Sindaebang Samgeori->Ahyeon	10	11	17	0.7969
Sillim->Sinbanghwa	4	11	31	0.7925
Bongcheon->Yeongdeungpo Office	3	11	23	0.7911
Geumjeong->Guui	10	11	3	0.7652
Seryu->Seokchon	17	11	1	0.7629

표 2에서는 내선순환이자 일일 승하차 인원이 가장 많은 역인 '잠실역'을 도착역으로, 출발역을 '사당역'으로 설정한 실험의 결과를 제시한다.

유사도 측정을 통한 결과 해석은 다음과 같다.

표 1을 참고할 때 해당 알고리즘은 '가산디지털단지역->논현역' 배송을 희망하는 소비자에게 15개 역이 동일한 배송원을 추천한다. 이 경우 출발역이 동일한 상태에서 가산디지털단지역~건대입구역까지 총 15개 역에 대해 동일한 경로를 보이며, 이후 5개 역에서 경로 비유사를 보인다. 결과적으로 유사도는 0.9852로 나타났다.

다음으로 표 2를 보면 '사당->잠실' 배송을 희망하는 소비자의 경우는 동작~강동구청역을 이동하는 배송원을 추천받는다. 이때 배송원은 소비자보다 2개 역을 앞서며 11개의 동일한 경로를 가지고 있고, 이후 2개 역에 대해 비유사를 보인다. 이때 유사도는 0.9258로 나타났다.

두 실험에서 알 수 있듯이 유사도는 소비자의 요청 경로의 출발역과 가까울수록, 배송원의 전체 이동 경로 수가 짧을수록, 배송원과 소비자의 일치 경로가 많을수록 높게 측정되었다. 이는 방향성을 고려하면서도 소비자와 출발역이 더 가깝고 전체 경로가 길지 않은 배송원을 선별해, 빠르게 배송 물품을 픽업하여 배송할 수 있기 위함이다. 다시 말해, 소비자와 배송원 사이의 출발역, 도착역, 전체 경로를 고려하여 추천하므로 단순히 통계적인 역의 일치만으로 추천이 이루어지지 않는다는 점에서 의의를 지닌다.

마지막으로 소비자가 요청한 전체 경로를 지나는 배송원이 없을 경우에는 첫 번째 매칭 이후에 남은 경로에 대해서 배송원을 다시 추천받는다.

V. Conclusions

본 연구는 COVID-19 팬데믹 이후 비대면 문화가 확산되면서 폭발적으로 증가하고 있는 배송 서비스 수요에 맞춰 새로운 알고리즘을 제시하여 증가하는 물량에 따라 발생하는 문제점 해결과 서비스 개선을 위해 출발하였다. 해당 알고리즘은 배송원 1인당 배송하는 물량이 많아 발생하고 있는 서비스 지연 문제나 파손 문제 그 외에도 비용 문제, 기존에 고려되지 않았던 사용자의 선택권 부족 문제를 해결하는 것을 목적으로 하고 있다.

그 목적에 따라 본 연구에서는 문제 해결 방안으로 근거리 지하철 배송 시스템을 제안하였다. 전체 배송 물량 중 근거리 이동은 지하철 배송 시스템으로 물량을 분산시켜 충분한 인력을 확보하고 'man-to-man' 매칭을 통해 비용의 최소화와 1인당 물류의 축소, 배송 기간 단축을 보장하는 시스템이다.

본 시스템의 알고리즘은 TF-IDF, N-gram, BERT를 활용하여 순서를 고려한 유사도를 측정하여 유사도가 높은

상위 10명의 지하철 배송원(승객)을 배송 이용자에게 추천한다. 이후 이용자(소비자)가 배송원을 선택하면 선택된 배송원이 출발역의 물품보관소 또는 소형 거점에서 물품을 픽업하여 소비자가 원하는 도착역의 소형 거점으로 물품을 전달한다. 소형 거점을 통하여 배송원과 소비자 모두 물품에 대한 접근성이 높아지기 때문에 배송을 더욱 신속하게 할 수 있다. 본 연구의 실험에서는 특정 상황에서 유동 인구가 많은 두 역을 대상으로 두 번의 실험을 진행했다. 이때 유사도는 배송 물류의 출발지와 배송원(승객)의 출발지가 가까울수록, 배송원의 전체 경로가 길지 않을수록, 배송원의 전체 경로와 배송 물류의 전체 경로가 유사할수록 높게 나타났다.

배송 기한 측면에서는 기존 택배 시스템과 달리 물품을 분류하기 위해 분류 센터를 거치지 않기 때문에 더 빠른 배송이 가능하다. 따라서 비용 측면에서도 물류센터 유지 비용이 소모되지 않아 비용 절감 효과가 있다. 또한 지하철을 이용하고 있는 승객들을 물품 이동 인력으로 참여시킨다는 점에서 이동 수수료를 최소화하고 수도권 배송 네트워크를 더 광범위하게 확장할 가능성이 있다. 그리고 기존 배송 서비스에서 이용자(소비자)가 선택할 수 없었던 배송원 영역을 이용자가 직접 선택할 수 있게 하여 이용자의 서비스 신뢰도와 만족도를 상승시킬 수 있다.

이 시스템을 통해 배송원 1인당 물량의 과부하로 인한 물품의 파손, 분실, 배송 지연 등과 같은 문제점을 해결할 수 있을 것으로 기대한다. 그뿐만 아니라 이 시스템이 확장되어 플랫폼 노동과 같은 부업, 시스템 관리 인력 등의 간접적 일자리 창출, 대중교통을 통한 물류 이동으로 탄소 배출 감소 등에 기여할 수 있기를 바란다. 물론 서비스 시작 단계에서는 사회 문제를 해결할 만큼의 효과를 기대할 순 없겠으나, 앞서 예시를 든 만보기 포인트 지급 서비스와 같이 사용자 증가로 인한 네트워크 효과가 발생한다면 충분히 사회 문제 해결 방안으로 고려될 것이라 판단한다.

그러나 해당 연구에서는 최단 경로 마련에 있어 API에 의존하고 있다. 향후 연구에서는 강화학습을 통해 직접 최단 경로를 마련하여 API의 의존도를 낮춰 API가 야기할 수 있는 문제에 유연히 대처할 수 있는 알고리즘 연구가 필요하다. 또한 시스템 운영과 이동 비용에 관해 큐알 코드를 통한 포인트 지급 형식이라 앞서 요약하였지만 본 연구에서는 경로 매칭 알고리즘 원리에 집중하였기 때문에 상세한 언급을 하지 못하였다. 따라서 구체적인 서비스 운영에 관한 사항은 후속 연구를 통해 더욱 상세히 기술할 예정이다.

ACKNOWLEDGEMENT

This work was supported by BK21 of Yonsei Univ.

REFERENCES

- [1] Son Jeongbin, "Half-price parcel service usage increased 6 times in a year". Newsis, May.06.2021. https://newsis.com/view/?id=NI_SX20210506_0001431338&cID=13001&pID=13000# quoted in GS25, Apr., 2021.[assed: July., 28, 2022]
- [2] Park Min-joo, "who will use convenience store delivery? 'A lot of used deals in the 20s'", Seoul Economy, Mar..05.2020. <https://m.sedaily.com/NewsView/1Z02X9XWWG> quoted in GS25, Mar., 2020.[assed: Aug., 01, 2022]
- [3] Statistics Korea, Online Shopping Trends Survey, 2021, <https://kosis.kr/search/search.do?query=%EC%98%A8%EB%9D%BC%EC%9D%B8%EC%87%BC%ED%95%91>
- [4] National Logistics Integrated Information Center, Household Logistics Statistics by Year, 2022, <https://www.nlic.go.kr/nlic/parcelServiceLogistics.action> [assed: June. Feb, 2022]
- [5] Hee-Kyung Jung, "Toss pedometer users surpass 4 million, 'Providing both fun and health benefits'", https://m.businesspost.co.kr/BP?command=mobile_view&num=279736 [assed: May. 03, 2022]
- [6] Korea Consumer Agency, September 2020 Consumer Big Data Trend Analysis, 2020, <https://www.kca.go.kr/home/main.do>
- [7] Han-Young Park, So-Hyung Kim, Seung-Joo Jeong, Sang-Beom Seo, "Changes in Domestic Delivery Service Usage Behavior due to the Spread of COVID-19", Transportation Research, 51-66.[5], Jun 2021 (16page)
- [8] Seung-Jin Shin, Hong-Seung Noh, Han-Young Park, "Measures to Securing Living-Friendly Urban Logistics Facilities", Transportation Technology and Policy, 17(1), 37-45.[6], Feb 2020, (9pages)
- [9] Dong-Jae Gwak, Kyung-Doo Nam, "Effects of Marketing Mix Factors of Online Grocery Delivery Platforms on Consumer Satisfaction", Korea International Trade Research Institute, Trade Research, 17, 1, 663-678, Feb .2021, DOI: 10.16980 /jitic.17.1.220 2102.663
- [10] Yang Liu and Kyungsook Jeong. "A study on the relationship between logistics service quality and customer loyalty of a cross-border e-commerce platform through satisfaction and trust mediating effects", International Commerce, 37(1), 107-132, Mar 2022, DOI: 10.18104/kaic.2022.37.1.107
- [11] R. Shrivastava and D. S. Sisodia, "Product Recommendations Using Textual Similarity Based Learning Models," 2019

- International Conference on Computer Communication and Informatics (ICCCI), pp. 1-7, Jan 2019 doi: 10.1109/ICCCI.2019.8821893.
- [12] M. M. Sanjeev, B. Ramalingam and S. Kumar T.K., "Realtime Semantic Similarity Analysis of Bulk Outlook Emails Using BERT," 2020 International Conference on Advances in Computing, Communication & Materials (ICACCM), pp. 89-94, Aug 2020, DOI: 10.1109/ICACCM50413.2020.9212979.
- [13] M. G. Pour, S. Z. Razavi and H. Faili, "A New Sentence Ordering Method using BERT Pretrained Model," IEEE, 2020 11th International Conference on Information and Knowledge Technology (IKT), pp. 132-138, Dec 2020, DOI: 10.1109/IKT516189.340
- [14] Ministry of Land, Infrastructure and Transport, Analysis of Public Transportation Usage Based on Transportation Card Data, 2020, <http://www.molit.go.kr/portal.do>
- [15] W.-S. Choi and S. B. Kim, "N-gram Feature Selection for Text Classification Based on Symmetrical Conditional Probability and TF-IDF," Journal of Korean Institute of Industrial Engineers, vol. 41, no. 4. Korean Institute of Industrial Engineers, pp. 381-388, Aug-2015, DOI: 10.7232/jkiie.2015.41.4.381.
- [16] Devlin, J., Chang, M. W., Lee, K., and Toutanova, K., "Bert: Pre-training of deep bidirectional transformers for language understanding", Minneapolis, Minnesota. Association for Computational Linguistics, In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pp. 4171-4186, Jun 2019, DOI: 10.18653/v1/N19-1423
- [17] H. Choi, J. Kim, S. Joe and Y. Gwon, "Evaluation of BERT and ALBERT Sentence Embedding Performance on Downstream NLP Tasks," IEEE, 2020 25th International conference on pattern recognition (ICPR), pp. 5482-5487, Jan 2021, DOI: 10.1109/ICPR48806.2021.9412102.

Authors



Jeongeun Song received her bachelor's degree in Korean Language and Literature from Myongji University in 2010. She is currently pursuing the M.S. degree in Graduate School of information, Yonsei University.

Jeongeun Song is interested in deep learning, NLP.



Yoon-Ah Song received her bachelor's degree in Economics and Finance from Sangmyung University in 2022. She is currently pursuing the M.S. degree in Graduate School of information, Yonsei University.

Yoon-Ah Song is interested in reinforcement learning, deep learning.

부록

<p>[1호선] #1호선: 인천역→소요산역 #1호선_re: 소요산역→인천역 #1호선_s: 인천역→신창역 #1호선_s_re: 신창역→인천역 #1호선_k: 인천역→광명역 #1호선_k_re: 광명역→인천역 #1호선_d: 인천역→서동탄역 #1호선_d_re: 서동탄역→인천역</p>	<p>[2호선] #2호선: 까치산역→설동역 #2호선_re: 설동역→까치산역 #2호선_2: 까치산역→신설동역 #2호선_2_re: 신설동역→까치산역 #2호선_순환: 신도림역→대림역 #2호선_순환_re: 대림역→신도림역 #2호선_순환2: 성수역→독섬역 #2호선_순환2_re: 독섬역→성수역</p>
<p>[3호선] #3호선: 대화역→오금역 #3호선_re: 오금역→대화역</p>	<p>[4호선] #4호선: 오이도역→진접역 #4호선_re: 진접역→오이도역</p>
<p>[5호선] #5호선: 방화역→하남검단산역 #5호선_re: 하남검단산역→방화역 #5호선_m: 방화역→마천역 #5호선_m_re: 마천역→방화역</p>	<p>[6호선] #6호선: 신내역→구산역 #6호선_re: 구산역→신내역</p>
<p>[7호선] #7호선: 석남→장암 #7호선_re : 장암→석남</p>	<p>[8호선] #8호선: 암사→모란 #8호선_re: 모란→암사</p>
<p>[9호선] #9호선: 개화→중앙보훈병원 #9호선_re: 중앙보훈병원→개화</p>	<p>[수인분당] #수인분당: 인천→청량리 #수인분당_re: 청량리→인천</p>
<p>[공항철도] #공항철도: 인천공항2터미널→서울역 #공항철도_re: 서울역→인천공항2터미널</p>	<p>[자기부상] #자기부상: 인천공항1터미널→용유역 #자기부상_re: 용유역→인천공항1터미널</p>
<p>[경의중앙] #경의중앙선1: 임진강역→지평역 #경의중앙선1_re: 지평역→임진강역 #경의중앙선2: 임진강역→서울역 #경의중앙선2_re: 서울역→임진강역</p>	<p>[에버라인] #에버라인: 기흥역→전대에버랜드역 #에버라인_re: 전대에버랜드역→기흥역</p>
<p>[경춘] #경춘1: 청량리역→춘천역 #경춘1_re: 춘천역→청량리역 #경춘2: 광운대역→춘천역 #경춘2_re: 춘천역→광운대역</p>	<p>[신분당] #신분당선: 광교역→강남역 #신분당선_re: 강남역→광교역</p>
<p>[의정부] #의정부: 발곡역→탑석역 #의정부_re: 탑석역→발곡역</p>	<p>[경강] #경강: 판교역→여주역 #경강_re: 여주역→판교역</p>
<p>[우이신설] #우이신설: 신설동역→북한산우이역 #우이신설_re: 북한산우이역→신설동역</p>	<p>[서해] #서해: 소사역→원시역 #서해_re: 원시역→소사역</p>
<p>[김포골드] #김포골드: 양촌역→김포공항역 #김포골드_re: 김포공항역→양촌역</p>	<p>[인천] #인천1: 계양역→송도달빛축제공원역 #인천_1re: 송도달빛축제공원역→계양역 #인천2: 검단오류역→운연역 #인천_2re: 운연역→검단오류역</p>