

# 자기조직화 지도와 K-평균 군집화 기법을 결합한 다단계 철도 고객세분화 방안 연구

이현주

고려대학교

lhj867@korea.ac.kr

## A Study on Multiple-Level Customer Segmentation Scheme based on SOM and K-means in Railway Business

Lee Hyun Ju

Korea Univ.

### 요약

최근 코로나19로 인해 여행 패턴이 점차 다각화되어 가고 있다. 이러한 추세를 반영하기 위해 상품의 다양성을 높이기 위한 노력을 기울이고 있으나, 이러한 개인화된 상품을 타겟마케팅 하기 위한 고객그룹에 대한 관리 역량은 아직 부족한 실정이다. 본 연구에서는 철도 이용정보를 패턴화 할 수 있는 특성변수를 생성하고 이를 활용하여 유사한 이용패턴을 보이는 고객을 마케팅 가능한 수준으로 그룹화 할 수 있는 다단계 군집화 방안을 제안한다. 이 방식은 단순한 구매활동 뿐만 아니라 이용시간 및 지리적인 특성 등 시공간적 이용패턴의 다양한 변수들을 활용하여 고객을 정교하게 분류할 수 있는 장점이 있다. 이러한 세분화로 만들어진 고객군집 데이터를 활용하여 향후 맞춤형 상품개발, 추천시스템 구축 등 실제 마케팅에 활용할 것을 기대한다.

### I. 서론

최근 전세계적 팬데믹 현상인 코로나19로 인해 여행 트렌드가 급격히 변하고 있다. 한국관광공사에서 발행한 관광 트렌드 분석 결과에 따르면 여행패턴이 점차 개별화(나홀로 또는 친밀한 관계)되고 있으며, 여행지도 유명한 관광명소보다는 개인의 성향에 따라 다양화되고 있다(한국관광공사 빅데이터를 활용한 2021년 관광 트렌드 분석, 2021). 이러한 변화는 여행의 주요한 교통수단 중 하나로 이용되는 철도에도 큰 영향을 미치고 있다. 한국형 고속철도(KTX, Korea Train Express) 도입에 따른 이동시간의 획기적 단축과 신규 노선 개통으로 지역 간 이동이 활발해지면서 철도 이용이 점차 일상화 되고 있으며, 이에 따른 이용 고객의 수도 코로나19 발생 이전까지는 꾸준히 증가하는 추세였다. 하지만, 코로나19 이후 여행자들의 안전한 여행에 대한 심리적 요인과 정책적 제한으로 철도 관광상품 판매가 중단되고 철도 이용 수요도 급감하면서 운송 서비스로 발생하는 수익 구조는 악화되고 있는 실정이다. 최근 코레일에서는 코로나 이후 여행객 회복추세에 맞춰 골프, 맛집 등과 연계한 상품의 다변화를 꾀하고 있으나, 고객의 철도 이용 특성을 분석하여 니즈에 맞는 좀 더 개인화된 마케팅으로의 전략을 변화시켜 나가야 할 필요가 있다.

따라서 본 연구에서는 유사한 특성을 그룹화하기 위해 많이 사용되는 군집화 알고리즘인 자기조직화지도와 K-평균을 적용하여 실제 철도 이용패턴이 유사한 고객의 세그먼트를 마케팅에 활용 가능한 수준에서 세분화할 수 있는 방법론을 제시하였다. 제안된 모델은 코로나19가 발생하기 이전 최대 이동 수요가 있었던 2019년 12개월 동안의 고객의 발권 이력을 추출하여 다양한 이동행태를 반영할 수 있는 특징변수를 생성하

고 이를 통해 총 20개의 세분화된 고객군을 도출하였다. 본 연구의 목표는 이러한 세분화된 고객군집 기초자료를 통해 좀 더 개인화된 맞춤형 상품개발과 회원관리시스템 구축 등에 활용 할 예정이다.

### II. 본론

#### 2-1 연구 방법

본 논문에서 제안한 다단계 고객 세분화 방법론의 특징은 자기조직화지도를 통해 만들어진 10\*10의 2차원 cluster별 각 노드의 가중치 벡터값을 K-평균 군집화를 통해 재그룹화하는 과정을 총 3단계에 걸쳐 수행함으로써 특이그룹, 우수/일반그룹, 우수/일반 그룹을 상세화하는 세부 분류 체계를 만드는 것이다. 이를 위해 첫번째로 고객의 발권정보를 수집하여 고객별 이용 특성정보로 변환하고 두번째로 모델에 입력할 변수들의 전처리와 정규화를 수행한다. 세번째로 3단계 군집화를 수행하고 네번째로 마케팅 전략 수립을 위한 군집별 고객특징을 분석하였다. 세번째와 네번째 단계에서 군집별 고객 특징이 잘 나타나도록 변수를 다양화하여 반복적인 군집화 과정을 수행하고 최종적으로 33개의 변수를 선정하였다. 군집화 기반 다단계 고객 세분화 방법론의 절차는 다음 그림1과 같다.

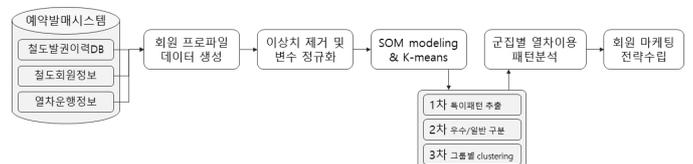


그림 1. 철도이용패턴 분석 프로세스

## 2-2 데이터 수집

2019년 1월부터 2019년 12월까지 한국철도공사의 고속·일반 열차 발권 실적 내역 약 9.4천만 건의 실제 데이터를 활용하여, 회원관리번호를 키값으로 데이터 처리 및 고객 특성정보를 결합하여 개인별 프로파일 데이터를 생성한 후 분석을 진행하였다. 분석에 활용한 변수는 표 1과 같다.

표 1. 회원 기준 데이터 내 활용변수

분류	변수명
회원특성	회원번호, 연령, 성별, 회원등급, 주소
발권정보	연간 발매 수, 연간 발매 금액, 승차권 종류, 캐린더 정보(공휴일, 주말, 주중, 평일), 열차종(KTX, 무궁화, 새마을 등), 노선(경부선, 호남선, 강릉선 등), 객실정보(특실, 일반실), 정기권, 이용시간대(출근시간, 오전시간, 오후시간, 퇴근시간, 심야시간) 등
생성변수	건당 발권 매 수, 발권 대비 반환율

## 2-3 군집분석

철도회원 군집화 분석에 코호넨에 의해 개발된 SOM기법과 K-means 군집화 알고리즘을 적용하였다. 경쟁학습을 통해 주변의 유사한 노드들이 존재한다는 특징을 활용해 다시 한번 K-means 군집화로 묶어주는 과정을 통해 실제 마케팅에 접목 가능한 수준의 결과를 도출하였다.

1차 SOM 모델링 과정에서 이용금액의 과대편차로 인해 대군집 1개와 소군집 2개로 분류되는 특징을 해소하기 위하여 대군집과 소군집을 별도의 군집으로 업데이트하여 2차 군집화를 진행하였다. 2차 군집 또한 이용 금액의 가중치가 크게 작용하여 다른 특성별 사용패턴으로 구분되지 못하여 우수그룹과 일반그룹을 별도로 업데이트하여 3차 군집화를 진행하였다. 최종적으로 최적의 군집 수 결정 방법(엘보우기법)을 통해 우수그룹 9개, 일반그룹 10개로 세분화하고 군집별 2020년도 사용패턴을 확인해봄으로써 결과를 검증하였다.

## 2-4 실험 결과

표 2. 철도이용패턴 기반 고객 세분화 결과

Group	Clustering according to train use pattern	
VIP (9)	Travel agency	High-class client
	Weekend couple	Weekend trip
	Night owl(gold)	Woopies
	Commuter by Mugunghwa	short commuter
	Mainly using Gangneung line	
Normal (10)	Near metropolis city	Prefer KTX
	Prefer Saemaeul	Night owl(silver)
	Commutation ticket user	Non-peak
	Group consisting of 4 passengers	
	Train traveler using Gangneung line	
	High-class client in normal group	
Holiday(1)	Train service user who only use in nat'l holiday	

클러스터링 결과, 우수 집단 내 9개, 일반 집단 내 10개의 세그먼트를 생성하였다. 각 군집을 연령, 선호 열차종, 주 이용노선/요일/시간대, 정기권 사용 여부 등을 기준으로 특성을 도출하였다. 고객 세분화 결과는 다음의 표2와 같다.

본 고객세분화 방법론을 적용한 결과 이용량의 차이가 크더라도 유사한 이용패턴을 보이는 군집을 발견할 수 있었고, 기존에는 주중/주말 이용객으로 단순하게 활용하였던 방식에서 주말에만 주로 열차를 이용하는 여행객, 금요일 저녁과 월요일 아침에 열차를 이용하는 주말부부패턴 등으로 이용패턴을 상세화할 수 있었다.

## III. 결론

본 논문에서는 철도회원의 이용패턴을 이해하고 목적에 맞는 서비스를 제공하기 위하여 발권정보를 활용한 K-means와 SOM 기반 다단계 군집분석 방법론을 제안하였다. 이러한 다단계 군집분석 방법론은 특성변수가 다른 집단에 비해 크게 나타나 군집별 특징이 한 집단에 의해 잘 나타나지 않아 집단을 소규모로 줄여나가야 하는 경우에 유용하게 사용될 수 있을 것이다. 본 방법론을 철도 고객 세분화에 적용한 결과 총 19개 집단으로 분류할 수 있었으며, 이를 통한 적정규모의 고객 그룹 관리와 타겟 마케팅 정책에 활용이 가능할 것으로 기대된다.

본 연구의 한계점은 특성변수를 고객의 인구통계학적 변수를 제외하고 주로 발권량과 금액을 활용하여 고객을 분류하기 때문에 상품개발 등 마케팅 영역에서는 좋은 분류이나 고객 등급 산정, 고객 이탈 모형 등 고객관계관리에 활용은 제한적이다. 또한 철도 이용정보만을 대상으로 하고 있어 고객의 라이프스타일 특성을 반영하지 못한다는 한계점도 있다.

후속 연구로는 철도 이용정보 외 고객의 신용정보, 취미, 선호도 등 외부 데이터를 결합하여 철도 고객을 좀 더 입체적으로 분석하고, 이러한 외부 변수들을 활용하여 마케팅 정책의 민감도를 높일 수 있도록 그룹을 재분류하는 것이 필요하다. 또한 본 연구결과의 그룹별 코로나 전후 열차이용 패턴 변화를 분석하여 이용목적별 이탈 방지 정책에 대한 추가적인 연구도 진행할 예정이다.

## 참고 문헌

- [1] Kohonen, Teuvo. "Self-organizing maps," Springer Science & Business Media, Vol. 30, 2012.
- [2] Min-Suk Chang, Hyung Joong Kim. "A Customer Segmentation Scheme Base on Big Data in Bank" Journal of Digital Contents Society 19(1), pp. 85-91, January 2018.
- [3] Yong Jun Cho, Jun Hoh. "A Application Stratey of Customer Scoring Model : Focusing on Cosmetics part of Department Store" Journal of the Korean Data Analysis Society, Vol. 8, No 1, pp. 335-348, Febrary 2006.
- [4] Young Sung Cho, Song Chul Moon, Keun Ho Ryu. "Clustering Analysis by Customer Feature based on SOM for Predicting Purchase Pattern in Recommendation System" Journal of the Korea Society of Computer and Information, Vol. 19, No. 2, pp. 193-200, February. 2014.