

비지도 학습 기반 클러스터링 기법을 활용한 도심 물류 배송지 최적화 연구

전형준¹, 임희석¹

¹고려대학교 컴퓨터정보통신대학원 인공지능융합학과
hjtime@korea.ac.kr, limhseok@korea.ac.kr

A Study on Optimization for Delivery Destination Clustering using Unsupervised Learning

Hyungjun Jeon¹, HeuiSeok Lim¹

¹Graduate School of Computer & Information Technology, Korea University

요 약

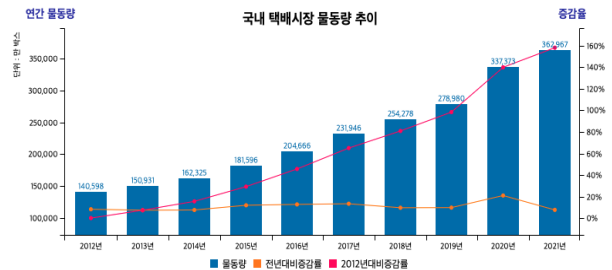
최근 이커머스 시장의 지속적인 성장으로 빠른 배송과 대용량 물류 처리를 위한 효율적 배송 시스템 마련의 필요성이 증가하고 있다. 본 연구에서는 도심 물류 거점에서의 현재 배송 물량 할당의 불균등 문제를 실무적 관점에서 정의하고, 비지도 학습 기반 클러스터링 기법을 통해 불균등 배송 할당 문제를 개선해 보고자 했다. 분석 결과 K-means++ 알고리즘 기반 클러스터링에서 최적화된 물량 할당에 대한 개선 가능성을 검증할 수 있었다. 향후 지형 정보, 교통량 등의 상세 변수를 추가하여 머신러닝 기반의 물류 배송 최적화를 위한 연구 영역을 확장할 수 있을 것으로 기대된다.

1. 서론

최근 이커머스(e-commerce) 시장의 지속적인 성장과 함께 물류에 대한 중요성이 강조되고 있다. 이와 함께 전 세계를 강타한 전례 없는 팬데믹 효과와 맞물려 온라인 비즈니스 기반 배송 시장은 폭발적 성장을 지속하고 있고, 이제는 우리의 일상 생활에서 택배가 필수적 요소로 자리 잡았다. 뿐만 아니라 주문부터 수령까지 당일 또는 수시간 내 처리를 보장하는 퀵커머스, 새벽배송 등 배송 형태의 다변화는 물론 배송 시간 단축 및 편의성 증대로 인해 온라인 구매 건수가 증가하면서 유통 물류 업계의 물동량은 지속적으로 늘어나는 추세이다. 그 예로 국내 인구 1인당 연간 택배 이용 횟수는 2000년 2.4회에서 2021년 70.3회로 20여년간 약 30배에 달하는 증가 추이를 보이고 있으며, 국내 택배 시장 전체 물동량은 2012년부터 현재까지 10년간 140,598 박스에서 362,967 박스로 약 160%에 달하는 증가율을 기록하였다. [1]

특히 COVID-19 대유행의 시작인 2020년의 물동량 급증이 뚜렷하고 이후 추세 역시 지속 증가 중인 바, 향후에도 이러한 언택트 중심의 생활 양식 변화는 가속화 될 것으로 예상된다. 이에 따라 물류 유통 전반 흐름에서 특히 소비자 배송의 중요성이 더욱 강조될

것으로 관측된다.



(그림 1) 연도별 국내 택배시장 물동량 추이

이러한 배경에서 택배 사업자와 자체 배송 체계를 갖춘 이커머스 기업들은 많은 수량의 배송 물품을 효율적으로 할당하여 배차하고, 배송에 드는 인적 리소스를 절감하며, 보다 정확한 교통량 및 지형 정보 기반으로 배송 경로를 최적화 하는 등의 대응책을 고민할 수밖에 없게 되었다. 특히 24시간 유동적인 배송 물량에 대응하기 위한 배송 기사 할당 문제는 고정비용과 직결되는 부분으로서 최근 더욱 중요성이 높아지고 있다.

본 연구에서는 대량 물류 처리를 위한 효율적 배송 시스템 마련의 필요성을 토대로, 유동적 물량에 대한 배송기사의 불균등 할당 문제에 주목하여 비지도 학

습 기반 클러스터링 기법을 적용해보고 효율성 개선 가능성에 대해 제안하고자 한다.

2. 용어 설명 및 문제제기

‘ラスト 마일’은 유통 전체 과정 중 소비자 접점에서 이루어지는 최종 물류 단계를 뜻한다. 빠른 배송이 주요 경쟁력으로 작용하는 배송 시장에서는 최근 인구 밀집지에 배송 거점을 전진 배치하여 이 라스트 마일 배송의 절대적 배송 거리와 시간을 최소화 하기 위한 도심형 물류센터(MFC : Micro Fulfillment Center)를 도입하는 추세이다.

이 MFC 에 소속된 배송 기사는 근무시간을 수 회차의 time slot 로 나눈 ‘배송 트랙’에 할당되고, 각 배송 트랙에는 다수의 배송 물품이 배정되며 이 과정을 통틀어 ‘배송 배차’라 한다. 배송 배차 영역은 전통적으로 오프라인 프로세스 중심 도메인인 물류 업계 특성상 배송 기사에 담당 지역 기반 휴리스틱 맵핑 방식을 따르고 있다. 본 방식은 매 배송 회차별 균등한 배송 대상 물품이 발생하는 경우에는 문제가 없지만, 때로 여러 상황들로 인해 어려움이 발생하는 경우가 있다. 예컨대 일부 지역에서는 담당 배송기사의 최대 처리량을 초과하는 경우가 발생하는 반면, 한편에서는 유희 배송기사가 발생하거나 한번의 배차에 극소량의 물품이 배정되는 비효율적인 작업 분배가 다수 이루어진다. 이는 자체 플렉스 배송 시스템(일반인 배송원을 파트타임으로 모집하여 활용)을 보유하지 못하거나, 백업 자원이 없는 영세 업체 입장에서 발생 가능한 치명적인 비효율 요소로서, 배송 기사의 최대 처리량을 초과 하는 경우 주로 배송 건별로 퀵서비스 등의 외부 업체에 배송 이관을 하는 형태로 대응하여 초과 운영 비용을 발생시킨다.

3. 실험 방법

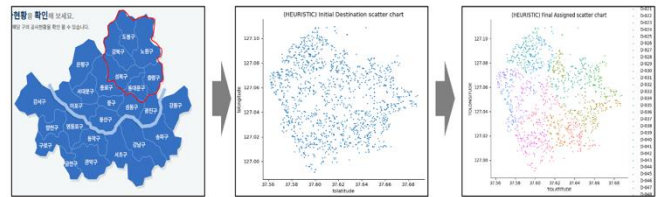
3.1 실험 데이터 정의 및 제약사항 모델링

서울 시내 인접 6 개구에서 6 시간 time slot 에서 발생한 실제 배송 주문 데이터를 활용한다. 인접 6 개구는 MFC 에서 처리할 수 있는 최대 처리 범위이며, 배송 트랙 역시 실제 현업에서 다회전 배차가 발생하는 time slot 을 참고하였다. 배송 주문 원장 데이터는 운영 데이터베이스에서 CSV 파일 형식으로 추출하여 우편번호, 텍스트로 표현된 송수하지 정보, 배송 주문별 물품 수량을 제외한 불필요 데이터는 모두 제외하고 결측치와 이상값 처리와 함께 포탈 업체 제공 좌표 변환 API 를 활용하여 송수하지 정보 기반 위경도 좌표를 획득하는 지오코딩을 수행하였다.

배송 기사 1 인당 6 시간의 배송 트랙 내 최대 배송 물품 수는 35 개로 제한하였으나, 클러스터링의 성능을 판단하여 동적으로 허용 범위를 설정하였다. 또한 모든 배송 기사는 총합 중량 350kg 급의 4 륜차량을 운행하는 것으로 가정하였다.

3.2 휴리스틱 맵핑 기반 할당 방식

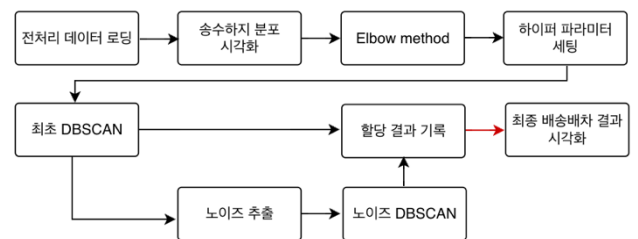
공공 데이터 제공 우편번호 리스트와 실제 MFC 소속 배송 기사 68 명을 임의대로 맵핑하고, 준비된 배송 주문 데이터와 배송 기사의 맵핑 데이터를 우편번호 기준으로 Join 하여 배송 배차를 수행한다.



(그림 2) 맵핑 결과 분류값의 산점도

3.3 DBSCAN 기반 할당 방식

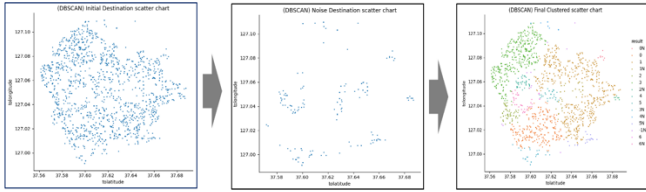
DBSCAN 은 잘 알려진 밀도 기반 클러스터링 기법 중 하나로서 밀도가 높은 데이터 포인트들을 묶어서 하나의 군집으로 구분한다. 하나의 중심 포인트부터의 범위를 유클리디안 거리 값인 Epsilon(eps)과, eps 범위 내부에 들어오는 데이터 포인트 최소 개수인 min_samples 를 하이퍼 파라미터로 지정하여, 임의의 중심 포인트로부터의 eps 범위내 데이터 포인트의 개수가 min_samples 을 충족하는지 비교하여 충족할 시 하나의 밀집 지역으로 분류한다. 이후 밀집 지역 내 데이터 포인트에 대해 더이상 eps 반경 내부로 분류된 데이터 포인트가 없을 때까지 동일 과정을 반복해 나감으로써 최종 군집을 확정한다. 전체 데이터 셋에 대하여 도출된 여러 군집 기준으로 어느 군집에도 속하지 않은 데이터 포인트는 노이즈로 분류한다. [2]



(그림 3) 최적화 DBSCAN 알고리즘의 구조

비즈니스 제약 사항을 반영한 알고리즘에서는 동일 데이터 기준으로 DBSCAN 알고리즘을 활용하여 클러스터링을 수행 하되 Elbow method 를 사용 하여 하이퍼 파라미터인 적정 Epsilon 을 도출한다. 이후 최초

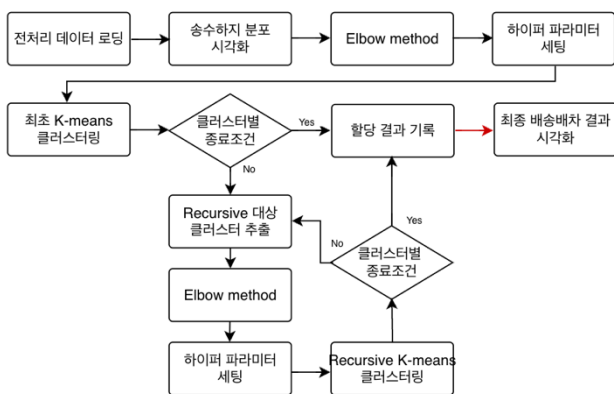
DBSCAN 으로 클러스터링을 하고 알고리즘 특성 상 발생하는 노이즈를 모아서 DBSCAN 을 1 회 더 수행 하는 최적화를 진행 한다. 또 다른 하이퍼 파라미터 인 최소 군집 수는 다회차 클러스터링을 실제로 수행 해보는 하는 Brute Force 방식으로 도출 하였으며, 25 로 설정하고 실험한다.



(그림 4) DBSCAN 클러스터링의 산점도

3.4 Recursive K-means++ 기반 할당 방식

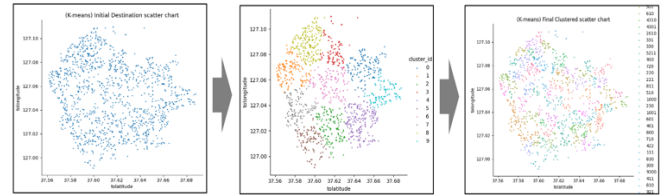
K-means 알고리즘은 임의의 데이터 포인트를 centroid 로 지정하고, centroid 와 주변 각 데이터 포인트를 탐색하면서 가장 가까운 centroid 의 군집으로 할 당해 나간다. 매 할당 마다 군집으로 지정된 데이터 포인트들을 기반으로 평균치 또는 중앙값을 기준으로 centroid 를 업데이트 하고, 더이상 centroid 가 갱신 되지 않을 때까지 반복한다. 초기 지정하는 임의의 centroid 수가 곧 군집의 수 k 가 되며, 이 k 값은 하이퍼 파라미터이다. [3] 단, K-means 알고리즘은 임의의 centroid k 개를 동시 지정 한 뒤에 알고리즘을 수행하므로 결과가 초기 centroid 에 매우 민감한 경향이 있다. K-means++ 는 이러한 약점을 보완한 기법으로 초기 centroid 를 1 개부터 k 개 까지 순차적으로 지정 해 나가되, 2 개 짝 부터의 centroid 위치를 이미 클러스터링이 끝난 군집의 centroid 로부터 가장 멀리 배치한다는 점에서 향상 된 군집화 성능을 보여준다. [4]



(그림 5) 최적화 Recursive K-means++ 알고리즘의 구조

동일 데이터를 활용하여 K-means++ 알고리즘을 활용하되 Elbow method 를 사용하여 하이퍼 파라미터인 적정 k 를 도출 한다. 이 k 값은 K-means 알고리즘의 시작점이 되는 임의의 데이터 포인트의 개수를 의미 하며 이는 K-means 알고리즘의 성능에 큰 영향을 미

치는 값이다. 좌표간 거리측정 방식으로는 유클리디 안을 채택하였다. K-means++ 알고리즘의 최적화는 최 초 클러스터링 수행 후 1 차 클러스터링 결과 군집 내부의 데이터 포인트의 개수를 기준으로 Recursive 클러스터링 대상 하위 군집을 추출한다. 이후 과정은 재귀 함수 형태로 구현하여 모든 하위 군집들이 종료 조건에 들어가게끔 최적화 한다. 하위 군집들에 대하여 Recursive 클러스터링 대상 여부를 판단하는 데에 는 비즈니스 제약 사항인 기사 1 인당 최대 배송 물 품 수량인 35 가 사용 된다.



(그림 6) Recursive K-means++ 클러스터링의 산점도

4. 실험 결과

4.1 결과 기술 통계 및 클러스터링 평가 지표

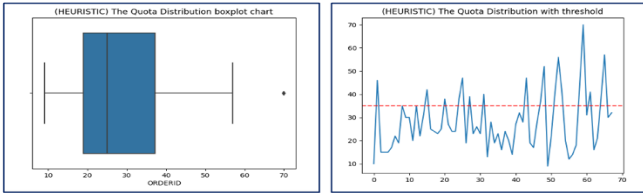
각 실험 결과에서 분류 된 하나의 군집은 배송 기사 1 인이 처리하는 배송 물품의 범위를 의미한다. 이는 최적화 로직에서 비즈니스 제약 사항을 미리 반영하여 최종 군집이 포함하는 데이터 포인트 수를 어느정도 적정 수준으로 유도해 둔 결과이다.

성능 평가를 위해서는 각 실험에 대한 기술 통계를 측정하고, 실루엣 계수를 별도 계산하여 하나의 비교 지표로 활용한다. 기술 통계는 최종 군집 수, 배송 기사에 할당 된 평균 수량, 분산을 사용하며, 실루엣 계수는 최종 군집들의 평균 계수값을 구하여 비교 지표로 사용한다. 각 실험 케이스 마다 전체 배송 기사 수 대비 초과된 배송 물품을 할당 받은 기사의 비율과 초과 할당 되어 처리 불가능한 배송 물 품의 수량을 측정한다.

4.2 휴리스틱 맵핑 할당 결과

별도의 알고리즘을 활용 하지 않고 매핑 기반의 할 당 규칙을 따랐으므로 보유 배송 기사의 수 만큼 군 집이 1:1 로 매칭 되었다. 이는 현실에서 발생하고 있 는 비효율 요소가 모델링 된 실험 환경에 그대로 재 현 된 결과이다.

배송 기사당 평균 할당 수량은 28.3 개로 실제 환경 에서도 적정치에 어느정도 근접 한다고 볼 수 있으나, 분산이 높으므로 불균등 할당이 주로 이루어 지고 있 다는 점을 확인 할 수 있다. 또한 총 배송 기사의 26% 가량이 초과 물량에 할당 되는 결과와 함께 매 배송 트랙마다 발생한 183 개의 미처리 수량이 관측된다.



(그림 7) 맵핑 분류의 Box plot 과 할당 수량 분포

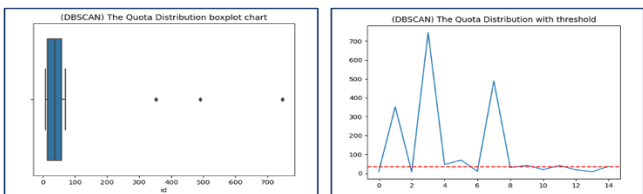
4.3 DBSCAN 클러스터링 할당 결과

인풋 데이터의 분포 형태가 전반적으로 비슷한 밀도를 가진 데이터 셋이므로 노이즈가 상당수 발생하였고, 세밀한 클러스터링이 이루어지지 않았음을 확인하였다.

<표 2> DBSCAN 클러스터링에 따른 지표

Metric1	The number of Clusters	15
Metric2	Average quota by drivers	128.3333
Metric3	Variance quota by drivers	49106.2381
Metric4	Number of over-allocated drivers	53.33%
Metric5	Number of exceeded quota	1543
Metric6	Mean of Silhoutte coefficients	-0.1136

노이즈를 별도 처리하는 최적화를 거쳤음에도 불구하고 클러스터링 자체의 성능이 낮게 측정되어 재귀 DBSCAN 을 적용해도 하위 군집 내에서 상세 군집화가 더이상 진행 되지 않음을 확인 할 수 있다. 그 결과로 총 할당 배송 기사 수가 현저히 적게 나타났으며 이에 따라 실루엣 계수가 음수로 집계되었고 전체적으로 클러스터링의 설명력이 부족하다는 것을 알 수 있다.



(그림 8) DBSCAN 의 Box plot 과 할당 수량 분포

4.4 Recursive K-means++ 클러스터링 할당 결과

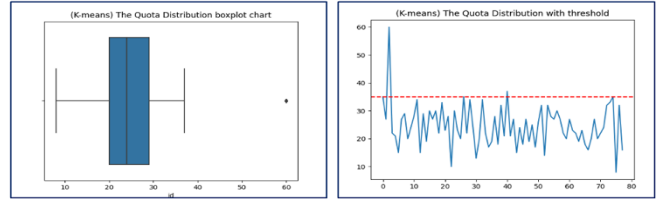
K-means++ 알고리즘을 기초로 하여 재귀 형태의 최적화 과정을 거치며 균등한 배송 배차와 함께 운영에 필요한 적정 배송 기사 수가 도출 되었다.

<표 3> Recursive K-means++ 클러스터링에 따른 지표

Metric1	The number of Clusters	78
Metric2	Average quota by drivers	24.6795
Metric3	Variance quota by drivers	57.4414
Metric4	Number of over-allocated drivers	2.56%
Metric5	Number of exceeded quota	27
Metric6	Mean of Silhoutte coefficients	0.3797

군집 수는 상대적으로 늘어났지만, 초과 물량을 할

당 받은 배송 기사 수 전체의 3% 미만으로 크게 개선되었으며, 분산 역시 휴리스틱 맵핑에 비해 낮아진 것을 확인 할 수 있다. 배송 트랙별 미처리 수량도 27 건으로 대폭 감소되었으며 실루엣 계수는 DBSCAN 대비 양수 전환하여 합리적 클러스터링이 수행 되었음을 알 수 있다.



(그림 9) K-means++의 Box plot 과 할당 수량 분포

5. 결론

본 연구에서는 효율적인 ‘ラスト 마일’ 배송 배차를 위한 첫번째 단계인 배송지 그룹핑 단계부터 비지도 학습 알고리즘을 재귀적 형태로 최적화하여 다회차 배차 시 발생할 수 있는 불균등 배송 할당 문제를 해결하기 위한 방법을 제시해 보았다.

분석 결과 K-means++ 알고리즘 기반 클러스터링에서 두드러진 개선 가능성을 검증 할 수 있었다. 또한 DBSCAN 알고리즘의 경우, 향후 데이터 차원을 높여 최적화를 수행해 나간다는 관점으로 볼 때 eps 값을 지정 하기 어려워 지므로 적합하지 못한 알고리즘이라는 통찰도 얻을 수 있었다.

현재 배송 할당 효율화 문제에 대한 머신러닝 적용 연구는 활발히 이루어지고 있지 않은 바, 추후 지형 정보, 교통량 정보 등의 상세한 변수를 정의 하여 최적화 과정을 추가 해 나간다면 유연한 배송 할당 대응은 물론 배송 경로 최적화 문제로 확장 할 수 있을 것으로 기대된다.

참고문헌

- [1] 이승원, "ラスト마일은 어떻게 다른가?", Industrial Engineering Magazine 27(3), 30-35, 2020
- [2] K. Mahesh Kumar, A. Rama Mohan Reddy "A fast DBSCAN clustering algorithm by accelerating neighbor searching using Groups method", Elsevier, Pattern Recognition 58, 39-48, 2016
- [3] Shi Na, Guan yong, "Research on K-means Clustering Algorithm : An Improved k-means Clustering Algorithm", 3rd International Symposium on Intelligent Information Technology and Security Informatics, 63-67, 2010
- [4] An Gong, Yun Gao, Xingmin Ma, WenJuan Gong, Huayu Li, Zhen Gao "An Optimized Artificial Bee Colony Algorithm for Clustering", International Journal of Control and Automation Vol.9 No.4, 107-116, 2016