

Graph Centrality를 통한 최적의 상권 입지 요건 분석

Analysis of requirements for optimal commercial location using Graph Centrality

요 약

현재 수많은 자영업자들이 창업에 뛰어들고 있다. 하지만, 여러 노력에도 수십 만 개의 가게가 폐업을 반복하며, 자영업자나 사회경제적으로나 큰 손실을 받고 있다. 따라서, 상권을 과학적으로 분석하여 가게의 흥망을 예측할 수 있는 기술이 필요하다. 해당 연구에선 상권과 관련된 여러 빅데이터와 Graph Centrality를 이용해 최적의 상권 입지 요건을 분석하고 설계한다.

1. 서 론

현재 대한민국은 대창업의 시대이다. 나이 불문할 것 없이 수많은 청년과 가정들이 부푼 꿈을 안고 새로운 가능성에 투자해 창업에 도전하고 있다. 중소벤처기업부에서 발표한 자료에 따르면 21년도 연간 창업 횟수 150만에 도달하며, 지난 해 대비 5.1% 증가한 추세를 보인다고 한다.

하지만, 그에 비해 연간 폐업자 현황은 90만에 육박하며, 60퍼에 해당한다고 한다. 이렇듯 많은 자영업은 성공을 거두지 못 하고, 그대로 사라지고 만다. 이러한 과정에서 수많은 자영업자들이 물질적 손해를 입으며, 사회경제적 문제들이 발생한다. 이러한 폐업이 반복될 경우 고용의 불안정을 야기하며, 기업 부채의 증가 등 사회적인 문제를 초래할 수 있다. 자영업에 실패하는 가장 큰 이유로는 상권 분석 실패를 들 수 있다. 입점을 준비함에 있어, 입점하는 가게의 지리적 위치는 매우 중요하게 작용한다. 심지어 입점하는 위치에 따라 해당 가게의 흥망 여부가 정해질 수도 있다.

따라서, 해당 연구에서는 상권을 빅데이터와 최대한 과학적인 기법을 이용하여 분석해 창업 전, 손쉬운 상권 분석과 창업의 위험성을 줄이고자 한다.

2. 관련 연구

2.1 Graph Centrality

해당 연구에서 가장 주축이 되는 기법이다. Graph Centrality란 그래프 혹은 사회 연결망에서 꼭짓점 혹은 노드의 상대적 중요성을 나타내는 척도이다. 이 중심성은

지수로 계산되는데 중점으로 두는 기준에 따라 다르다.

노드들 사이의 최단 경로를 가지고 계산하는 매개 중심성(Betweenness Centrality), 그래프의 노드와 다른 모든 노드들 사이의 최단 경로 길이 합의 역수를 이용해서 구하는 근접 중심성(Closeness Centrality), 한 노드에 연결된 모든 엣지의 개수로 중심성을 평가하는 연결 중심성(Degree Centrality), 주변 node의 중심성까지 고려하여 중심성을 선정하는 고유벡터 중심성(Eigenvector) 등 이 밖에도 다른 계산방법들이 다양하게 존재한다.[1]

2.2 딥러닝 데이터 분석을 통한 최적의 상권 입지 추천 기술 개발

이전에 진행됐던 유사한 연구로는 이와 같은 논문이 있다. 해당 연구에선 최적의 상권 입지를 위하여 데이터를 크게 매장 속성 정보와 매장 매출 정보로 분류하여 딥러닝 데이터 분석을 사용해 분석했다.[2] 해당 연구에선 Graph Centrality 기법을 이용하여 최적의 상권 입지를 도출한다.

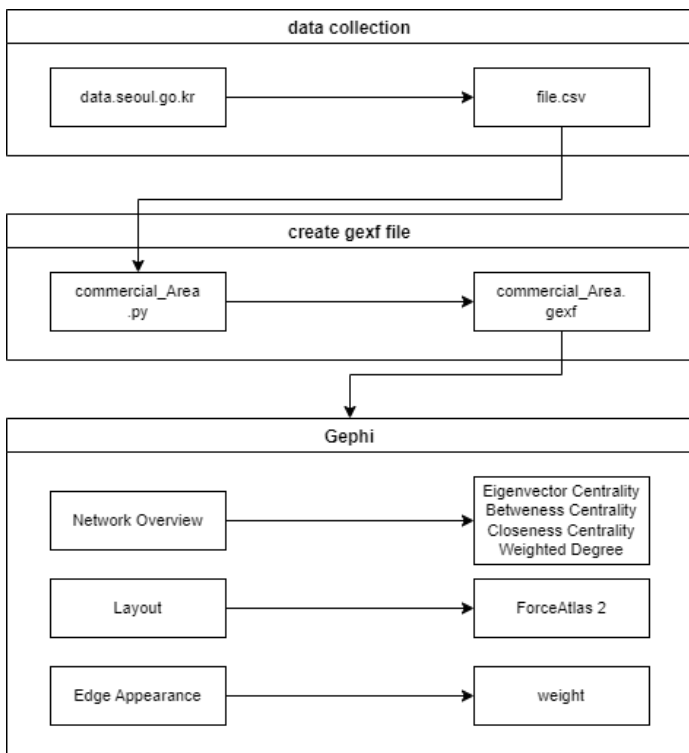
3. 연구 목적 및 연구 방법

3.1 연구 목적

제시하고자 하는 결론은 이와 같다. 첫 번째, Graph Centrality를 이용해 최적의 상권 입지에 유리한 기준을 분석한다. 두 번째, Graph의 특성과 Gephi를 이용한 편리한 시각화와 데이터 정렬로 해당 지역의 업종이 지니는 특성을 확인할 수 있게 해, 창업에 있어 유리한 상권을 확보할 수 있도록 한다.

3.2 연구 방법

다양한 상권, 업종을 분석하는 것이 목적이기에 서울특별시: 우리마을가게 공식홈페이지에서 제시하는 1091가지 상권과 100가지 업종을 분석 데이터로 삼는다. 이후, 기존 사회 분석 시스템에서 상권 분석에 사용되는 여러 데이터를 수집한다. 이때 사용되는 데이터로는 해당 행정동의 거주인구, 생활인구, 직장인구, 아파트 단지 수, 가게 매출, 해당 인구 월 평균 소득, 개업률, 폐업률이 있다. 각 행정동과 기준별 데이터를 노드 무리로 선정한 이후, 이를 엣지를 이용해 연결한다.



[그림 1: 프로그램 모듈]

본 연구에서 모듈의 양상을 이런 형태를 띤다. 서울 열린데이터광장에서 데이터를 수집한다. 이후, 수집된 데이터를 시각화 프로그램에 적합한 형태로 정제 및 가공하는 과정을 거친다. 그렇게 시각화되어 생성된 그래프는 4가지 중심성(Eigenvector, Betweenness, Closeness, Weighted)에 넣어 도출된 결과를 대조하며 확인한다. 이후, 적절한 중심성을 선정하여 원하는 결과를 도출한다.

3.3 데이터 추출 및 정제

상권 분석을 위한 데이터 제공으로는 서울 열린 데이터광장을 사용한다. 해당 사이트는 서울특별시에서 직접 운영하는 공식홈페이지로 연구, 관리, 서비스 제공 등 서울시 시정활동 과정에서 수집된 다양한 데이터가 존재한다. 해당 년도에서 2021년도 1분기 데이터를 위주로 행정동의 거주인구, 생활인구, 직장인구, 아파트 단지 수, 가게 매출, 해당 인구 월 평균 소득, 개업률, 폐업률 데이터를 수집한다. 이후 이러한 데이터를 Gephi에서 구동하기 적합한 데이터로 정제하는 과정을 거친다.

A	B	C	D	E	F	G	H	I
기준_년_코	기준_분기	상권_구분	상권_구분	상권_코드	상권_코드	서비스_업	서비스_업	분기당_매
2021	1	A	골목상권	2110001	이북5도청	CS100001	한식음식점	1.25E+09
2021	1	A	골목상권	2110001	이북5도청	CS100005	제과점	1.21E+08
2021	1	A	골목상권	2110001	이북5도청	CS100008	분식전문점	1.96E+08

[그림 2: 정제 전 가게 매출 데이터의 일부]

가게 매출을 담은 파일의 일부분이다. 년도와 분기, 상권명, 업종명, 그리고 분기당 매출 데이터를 추출해 csv로 저장돼 있다.

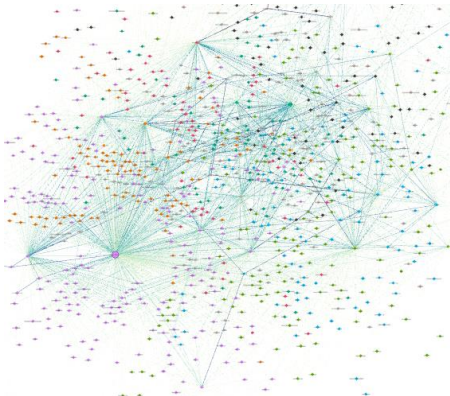
3.4 그래프 생성 및 시각화

이러한 데이터를 시각화하고 보다 정확하게 분석하기 위해 가장 적합한 자료구조 ‘그래프’를 채택했다.

가장 기준이 되는 상권 행정동을 우선 노드를 선정한 이후, 상권 분석에 필요한 데이터 무리들을 노드로 형성해준다. 편의점 업종 분석을 예시로 할 경우, 창덕궁 상권의 편의점 데이터를 노드로 형성한다. 그 이후, 해당 창덕궁 편의점 상권과 일치되는 거주인구, 생활인구, 직장인구, 아파트 단지 수, 해당 인구 월 평균 소득, 개업률, 폐업률 노드와 엣지로 연결해준다. 이때, 가중치는 해당 가게 매출을 기준으로 각기 다르게 부여해준다.

그래프를 쉽게 파악하기 위해 데이터 시각화 도구 ‘Gephi’ 프로그램을 사용했다. ‘Gephi’는 대규모 데이터를 시각화하고 분석하기 위한 오픈 소프트웨어이다[2]. 노드, 엣지, 가중치를 설정했다면 Gephi를 이용해 그래프 중심성 수치를 계산하고 쉽게 파악할 수 있다.

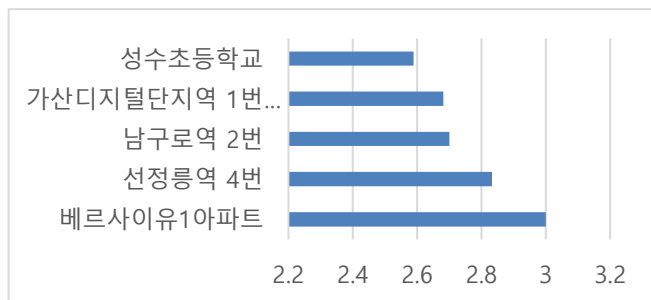
3.5 그래프 분석



[그림 3: 시각화한 그래프]

가게 매출, 거주 인구, 직장 인구, 평균 월 소득 데이터만을 사용해 이들을 노드로 형성한 이후 임시 가중치를 입력 Gephi로 형성한 그래프이다. 만들어진 그래프의 총 노드 개수는 1,135개, 엣지의 개수는 3,225개이다. 서로 집단을 이루는 모듈끼리 색 조정을 하였으며, 앞으로 남은 추가 상권 분석 기준을 도입한다면 최종 제작된 그래프 또한 이와 같은 형상을 띤다.

3.6 중심성 분석



[그래프 1 상권 가중연결 중심성 (Weighted Degree Centrality) 상위 5명]

단순히 매출을 0~1로 가중치화 한 그래프에서 가중 연결 중심성은 이와 같은 결과를 보였다. 베르사이유1아파트부터 성수초등학교 순까지 상위 5의 가중 연결 중심성을 보인 것을 확인할 수 있다.

4. 결론 및 향후연구

본 연구에선 Graph Centrality를 통한 상권 분석을 제안하였다. 아직 모든 상권 분석 요소를 입력하지 않았기 때문에 온전한 그래프를 띄지 못 했지만, 노드와 엣지를

형성하여 시각화함으로써 이들의 관계를 손쉽게 파악할 수 있었다.

다만, 아직 가중치의 공식의 신뢰성이 많이 떨어지기 때문에 향후 적절한 가중치를 선정하여 해당 연구의 신뢰성을 더 높일 예정이다. 또한, 제작한 그래프의 형태에서 가중연결 중심성을 제외한 나머지 중심성을 채택하여 적용할 경우 해당 그래프의 특징 상 동일한 중심성이 여러 번 반복되어 도출되는 것도 확인할 수 있었다. 따라서, 추가 기준 데이터를 입력하고, 가중치에 대한 공식을 다시 검토하여 신뢰성을 높이고 나머지 중심성의 수치도 확인해볼 예정이다.

참고문헌

- [1] 조태수, 한치근, 이상훈, “그래프 중심성들을 이용한 그래프 유사도 측정”, 한국컴퓨터정보학회논문지, 2018.
- [2] 박형빈 (Hyeong-bin Park), 김소희 (So-hee Kim), 남지수 (Ji-su Nam), 조운빈 (Yoon-bin Cho), 전희국 (Hee-gook Jun), 임동혁 (Dong-hyuk Im), 2022. 딥러닝 데이터 분석을 통한 최적의 상권 입지 추천 기술 개발. 한국정보처리학회 학술대회논문집, 29(1): 602-605
- [3] <https://gephi.org/>