

Skyline Query와 MapReduce 방식을 이용한 대도시에서의 창업 위치 추천 서비스

An Enterprise Location Recommendation Service in Metropolitan Region Using Skyline Query and MapReduce

이용현, 김동훈, 김응모
성균관대학교

Lee YongHyun, Kim DongEun, Kim Ummo
SungKyunKwan University

요약

본 논문은 편의점, 카페 등의 창업시 많은 후보군들 사이에서 적절한 위치를 추천하는 서비스를 만들어보고자 수행되었다. 본 연구는 Skyline Query를 이용하여 사용자가 설정한 지점으로부터의 거리에 따른 예상이익을 도출해내고, MapReduce를 사용하여 많은 후보군들을 대상으로 이를 효율적으로 처리하도록 구현하였다. 본 연구의 방법을 사용하여 창업자가 설정한 한정적 자원 및 거리 제한 조건 안에서 최적의 위치를 손쉽게 추천해줄 수 있을 것이다.

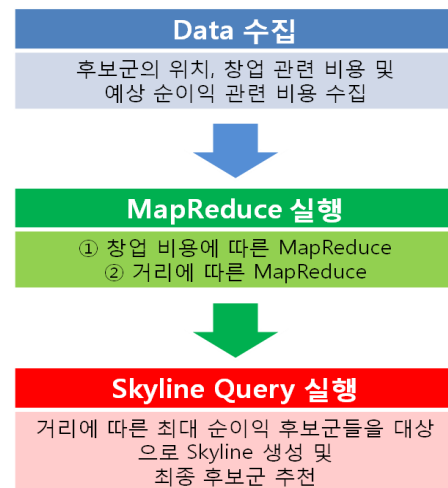
I. 서론

대도시에서 성공적인 창업을 위해서는 좋은 위치를 선점하는 것이 필수적이다. 제한된 자본으로 적절한 위치를 선택하기 위해서 수많은 후보군들을 비교 및 분석해야 한다. 이러한 후보군들을 평가하기 위해서 분석해야 할 데이터 양은 결코 적지 않으며, 충분히 빅데이터라 여길 수 있다. 이에 대한 해결책으로써 창업자의 조건에 맞는 위치를 바로 추천해줄 수 있는 서비스는 창업자의 금전적 및 시간적 자원을 상당히 아껴줄 것이며 경쟁 업체로부터 우위를 선점하게 해줄 것이다.

본 논문에서는 Skyline Query와 MapReduce를 사용하여 수많은 후보군들 사이에서 적절한 장소를 추천해주는 서비스를 제안하고자 한다. 창업에 관한 여러 조건 데이터들을 하나의 척도로 합친 후, 사용자가 설정한 기준점으로부터의 거리에 따라 Skyline Query를 그릴 것이며, 이러한 데이터들을 효율적이고 빠른 처리를 위하여 MapReduce를 사용하고자 한다.

II. 본문

본 서비스는 크게 데이터 수집, MapReduce 단계 및 Skyline Query를 적용하는 3단계로 구성된다. 그림 1은 본 서비스의 대략적인 overview를 보여준다.



▶▶ 그림 1. 본 서비스의 작동 구조

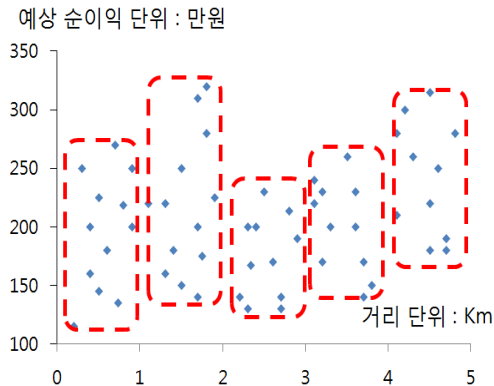
본 서비스를 위해서는 도시 내에서 후보군의 위치, 그 후보군에서의 창업 비용 및 예상 순수익에 영향을 미치는 요소들의 데이터를 필요로 한다. 창업 비용으로는 부동산 비용, 인테리어 비용 등이 있을 것이고 예상 순수익 관련 요소로는 수송비, 인건비, 임대료 등의 지출과 예상 수익과 관련된 유동 인구 및 인구 분포도 등이 있다. 해당 요소들을 각각 비교하기에는 데이터의 양은 너무나도 많고 이들을 직접 비교하는 것은 비효율적이므로 하나의 척도로 만들어서 연산을 수행한다. 해당 연산의 결과로 '창업 비용'과 '순이익'이라는 2개의 변수를 만들 수 있다.

2개의 변수를 생성 후 모든 후보군들을 대상으로 창업 비용에 따른 MapReduce를 실행한다. 생성된 2개의 변수

중 예상 창업 비용을 기준으로 후보군들을 그룹화할 수 있고[1], 사용자가 가지고 있는 자본금을 초과하는 창업 비용이 요구되는 후보군들은 쉽게 제외된다.

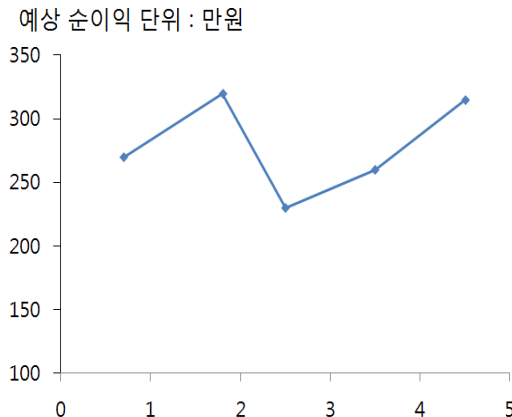
창업 비용을 초과하는 후보군들을 제외한 다음에는 사용자가 선택한 지점으로부터의 거리를 X축, 위에서 계산한 예상 순이익을 Y축으로 놓는 Skyline을 그린다. 사용자가 선택한 지점은 도심, 번화가, 사용자의 집 근처 등 사용자가 임의로 선택할 수 있다. 이러한 그래프를 그렸을 때 각 후보군들의 거리에 따른 예상 순이익을 알 수 있다. 거리에 따른 MapReduce를 사용하기 위하여 후보군들을 거리 단위로 그룹화한다. 표 1은 반경 5km내에서 1km의 단위로 후보지점들을 그룹화한 그래프를 보여준다.

표 1. 각 후보지점들을 거리에 따라 그룹화



후보군들을 거리 단위로 그룹화를 하고, 그 그룹에서 최상위의 후보군을 선택하면 선택된 거리 범위 내에서 가장 높은 순이익을 낼 수 있는 후보군이 선택된다. 각 거리에 따라 이런 후보군들을 모두 이으면 각 거리 구간에서 최대 순이익을 내는 지점들을 연결한 Skyline이 생성된다.[2] 해당 Skyline에서 사용자가 원하는 거리를 설정한 후, 가장 높은 지점을 선택하면 사용자의 기본 자금을 바탕으로 원하는 거리 내에서 가장 높은 순이익을 창출할 수 있는 지점을 선택하게 되는 것이다.

표 2. 그룹화된 요소 중 최고 순이익을 이은 Skyline



각각의 그룹화된 부분에서 최고 위치의 후보점은 해당 거리에서의 최고 순이익을 내리라고 예상되는 후보 지점이다. 창업자의 목표는 최대의 이익을 내는 것이 목적이므로 최대 순이익 창출 예상 지점을 제외한 나머지 지점들은 고려할 필요가 없다. 그러므로 표1의 각 그룹에서 Y값에 최고 값을 갖는 지점들끼리 연결하면 표2와 같은 Skyline을 생성하고 되고, 창업자는 원하는 거리 내에서 최고의 Skyline을 선택하면 된다.

III. 결론

본 논문에서는 대도시에서 수많은 창업 위치 후보군들 중에서 사용자에게 적절한 위치를 Mapreduce와 Skyline Query를 어떻게 사용하여 추천해줄 것인지를 알아보았다. 또한 약간의 변형을 추가한다면 대도시 뿐만 아니라 더 넓은 지역을 대상으로도 위치 후보군을 평가할 수 있을 것으로 기대된다. 본 연구의 한계점은 정형화된 데이터들만을 대상으로 구현하였다는 점이다. 추가적인 연구에서는 데이터 마이닝이나 비정형 데이터를 읽어옴으로써 더욱 데이터를 모으기 더 편하고, 효과적인 서비스가 되어야 할 것이다.

■ 사 사 표 기 ■

본 연구는 미래부가 지원한 2013년 정보통신·방송(ICT) 연구개발사업의 연구결과로 수행되었음. 이 논문은 2013년도 정부(교육부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임(NRF-2013R1A1A2008578)

■ 참고 문헌 ■

- [1] Jeffrey Dean, Sanjay Ghemawat, "MapReduce : Simplified Data Processing on Large Clusters", 6th Symposium on Operating System Design and Implementation, Google Inc, 2004, pp. 2-4
- [2] Dimitris. Papadias, Yufei Tao, Greg Fu, Bernhard Seeger, "An optimal and Progressive Algorithm for Skyline Query", SIGMOD 2003, pp. 2-3