

빅데이터를 이용한 부동산 주변 분석을 통한 입지선정 시스템 제안

송명기, 장재희, 김응모
성균관대학교 컴퓨터공학과
e-mail: ragan65@naver.com

A Proposal on Site Selection system for real estate Big Data analysis

Myeong-Ki Song, JaeHee Jang, Ung-Mo Kim
College of information & communication Engineering, Sungkyunkwan
university

요 약

본 논문에서는 실제 기업에서 활용되어오던 의사결정 과정에 사용되는 의사 결정 지원시스템(Decision Support System)에 관한 분석을 통해서 다른 방향으로 활용이 가능한지 알아보았다. 특히 부동산에 관한 의사결정에 도움이 되는 시스템을 만들어 평범한 국민들이 부동산을 구입할 때 활용 가능하도록 하고자하였다. 실제 주택 자체에만 집중하여 구매결정을 내릴 경우 주거만족도가 현저히 떨어진다. 집에서 착안하여 주택 주변을 분석하고 이를 바탕으로 주거만족도를 향상시킬 수 있도록 주변지역을 빅데이터 분석을 통해 분석하여 도출해낸 데이터를 통해서 해당 부동산이 적합한 입지조건을 가지고 있는지 확인하여 입지선정에 도움을 주는 시스템을 제안한다.

1. 서론

본 논문에서 제안하는 주변 분석을 통한 입지 결정 방식은 부동산을 거래하는 사람들이 부동산을 거래 할 때 신경을 많이 쓰는 교육에 용이한지, 교통에 편리한지, 주변에 통행량이 어느 정도인지, 집으로 오는 혹은 가게로 오는 주변 길들은 안전한지와 같은 자세히 알아보지 않고서 매물로 나온 집만 보서는 알기 어려운 사항들에 점수를 주어 매물의 상태와는 무관하게 매물이 가진 입지조건만을 고려하여 입지에 따른 추천을 해줄 수 있는 시스템이다.

부동산이라는 평범한 사람이 구매할 상품 중 가장 비싼 상품이라고 할 수 있는 상품을 고를 때 보통 사람들이 1~2번 정도 밖에는 해보지 않을 일이니, 이러한 비싼 상품을 구매하는 사람들이 얼마나 경험이 부족한 상태로 결정을 하고 있다고 봐도 무방할 것이다. 빈도수가 적은 이러한 결정을 위해서 너무 많은 공부를 하기도, 무작정 경험은 많겠지만 초면인 공인중개사에게 말기에도 어찌 보면 이상 할 수 있다. 이러한 사람들을 대상으로 부족한 경험을 조금이라도 고려해서 도움을 줄 수 있는 입지에 대한 의사결정 시스템이다.

2. 요구사항

- 주택 구매
주거환경에 대한 개념은 다양하고 포괄적이지만 광의

적으로 보면 인간의 주거생활을 영위하는 가운데 있는 모든 물리적·생리적 행동은 주거생활 감정을 포함한 정신적인 행동까지도 포함한다는 것이다. 따라서 주거환경은 주택 그 자체의 물리적 요소에서부터 주택을 둘러싼 주변지역의 물리적 환경은 물론 더 확대하여 보면 사회·문화적 환경에 이르기까지 다양하게 정의할 수 있다. 기존의 문헌 연구의 결과를 종합해보면 주거환경의 요소를 크게 물리적 요소, 사회·경제적요소, 심리적 요소로 구분하고 있다 [2].

[1]에 따르면 주거 만족도는 크게 3개 부문으로 나뉘어진 20개 항목에 따라 만족도가 나타나고 있다. 이 중 주거환경이 만족도에 기여하는 점수가 가장 크게 나타났다. 가장 중요한 세부항목은 주택의 채광, 환기 상태 점수였으며, 가장 중요하다고 여겨진 5가지 사항 중에 4가지가 주거환경에 관한사항이었다는 결과는 주택 자체보다 중요한 것이 주택 주변이라는 점을 말해 준다.

[1]에 제시된 주거만족도 결정요인에 따라, 주거만족도 계산을 위해서 주거환경, 주택설계, 시공상태, 관리상태의 네가지 부문의 빅데이터 집합이 필요하다.

- 소매점 입점

가게들은 우리나라 경제에 있어서 큰 부분을 차지하고 있다. 특히나 명예퇴직을 한 경우나 혹은 구조조정에 의해서 발생한 퇴직자들이 창업에 눈을 돌리고 하나의 작은 가게를 차리는 경우가 많다. 우리나라의 경제 구조도 많이

변화하여 이러한 퇴직자들 이외에도 주부나 심지어 학생도 창업을 하고자 하는 관심을 가지고 실제로 창업을 하는 경우가 늘어나고 있다. 하지만 이중 많은 사람들이 흔히 하게 되는 생계형 창업으로 꼽히는 여관, 치킨 집 등 숙박과 음식점업의 경우 5년 후 생존율이 17.7%밖에 되지 않았다. 이들 숙박, 음식점업의 생존율은 창업 1년 뒤 55.3%였고, 3년이 지나면 그 비율은 28.9%로 줄었다. 즉 거의 절반에 가까운 수의 생계형 창업이 1년도 안 되서 망하는 추세라는 것이다. 많은 이들이 창업에 손을 대고 있어서 창업의 문턱 자체는 낮지만 이를 유지하는 일이 아주 어렵다고 할 수 있다. 이러한 창업의 가장 중요한 요소 중의 하나가 좋은 위치에 입점하는 일이다. 이러한 위치 선정은 단순히 주변에 비슷한 종류의 업종이 얼마나 있는지 얼마나 많은 사람이 주변에 사는지, 아파트단지가 근처에 있는지 만이 중요한 일이 아니라는 것이다. 더 중요한 일은 주변 사람들이 많이 지나다니는 길목에 위치하는 일이다. 특히나 그 사람들이 창업하고자 하는 업종의 잠재적 고객이 될 수 있는 대상이 많이 지나다니는지가 가장 중요한 일이 될 것이다.

3. 관련 연구

- 빅데이터

빅데이터로부터 지식 활용을 하기 위해서 5가지 단계가 필요한데, 그 단계는 그림 1과 같다[3].



그림 1 빅데이터 플랫폼 전략

이처럼 데이터 수집, 데이터 전처리, 정보 저장 관리, 정보 처리 분석 및 지식 가시화를 통해 지식을 활용하기 까지 각 단계를 지원하는 데 필요한 공통 소프트웨어를 빅데이터 처리 플랫폼이라고 한다. 해당 플랫폼을 통해 데이터가 처리 되어야만 지식으로서 활용이 가능하다[4].

- 의사결정지원시스템

의사결정지원시스템(Decision Support System)은 경영 분야에서 의사결정을 지원하기 위한 개념에서 시작된 시스템이다. 의사결정지원시스템은 경영학에서 정의하는 시스템으로, 경영활동에 있어서 의사결정자의 비구조적, 반구조적 문제를 해결하기 위한 시도로 정보시스템의 데이터와 의사결정모델 그리고 인터페이스를 통해 의사결정을 지원해주는 정보시스템이라고 정의하고 있다. 자료, 모델,

지식베이스로 구성된 시스템이 사용자 인터페이스를 통해 의사결정자와 상호작용하는 시스템이다.

- 공간적 의사결정시스템

공간데이터와 모델을 이용하여 의사결정지원시스템에 활용하는 연구는 국내외 사례가 다수 존재한다. 대표적으로 미국의 UrbanSim은 토지이용, 교통 및 환경계획수립 지원을 위해서 1996년부터 개발되었다. 해당 모형은 크게 가구와 정보, 부동산개발자와 경제활동으로 구성되어 있으며 도시계획과 관련된 각 주체들의 활동을 모형화하여 어떻게 상호작용하는지 시뮬레이션을 해주는 도구이다[5].

- 데이터 마이닝

기존에는 데이터베이스 안에서 무의미하게 집합되어 존재하는 거대한 데이터 더미 안에서 의미 있는 정보를 추출 하여 데이터 모델을 찾아내어 사용하고자 하는 활동을 충칭한다. 데이터에 고급 통계 분석과 모델링 기법을 적용하여 유용한 패턴과 관계를 찾아내는 과정을 통해서 정보를 발견하는 것이다. 이 데이터 마이닝에서 가장 중요한 요소는 신뢰도로, 신뢰할 수 있는 자료더미의 확보가 필수적이다.

4. 시스템 제안

입지 결정을 하기 위해서는 입지결정 목표를 설정하는 수요설정이 기본이 되어 주어야한다. 본 시스템에서 중요하다고 생각하는 요건에 가중치를 주어서 수요설정을 바탕으로 GIS의 정보와 취합하여 현황분석을 할 것이다. 여기서 수요설정을 위해 사용자입력을 필요로 한다. 학군이나 도심지와외의 교통 등 중요하게 생각하는 수요를 정한다. 그 후 후보지를 사용자로부터 입력을 받아 가장 사용자의 수요에 맞는 후보지를 제시하고, 그 외의 대안도 그림 2와 같이 함께 제시한다.

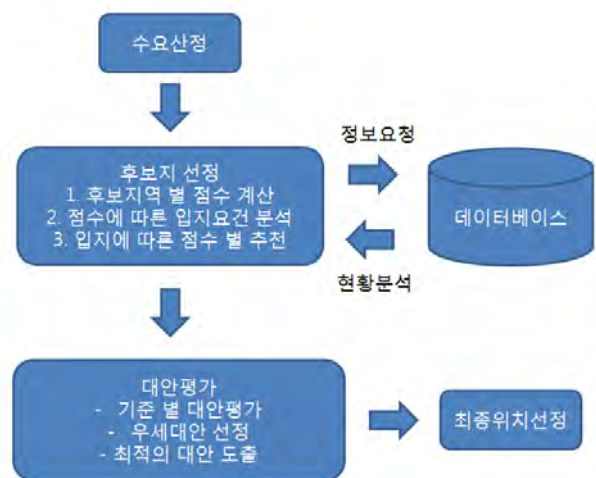


그림 2 제안시스템

표 1 제안시스템 상세사항

수요산정	수요설정을 위해 중요도를 결정한다. 그리고 살 생각이 있는 부동산 정보를 입력하여 사용한다.
후보지 선정	1단계 입지중요도에 따른 가중치 계산 2단계 입지요건 분석 3단계 입지에 따른 점수 제시
대안평가	후보지 선정 단계에서 생성된 대안들을 주거기준 별 대안평가를 통해 5가지 우수 대안에 대한 점수 산정
최종위치 선정	대안평가에서 제시된 최적대안 도출을 사용하여 사용자에게 최종적인 추천 장소를 제시한다.

이렇게 주택 선택을 예측하고 주거환경, 주택설계, 관리상태, 시공상태의 4가지 부문에서 가장 중요하게 생각하는 항목에 가중치를 주고 이에 따라 [1]에서 조사한 만족도 기여 점수와 중요도에 따라 만족도 기여 점수를 새롭게 산출하여 주거 만족도를 계산한다.

만족도 모형은 [1]의 연구에서 사용한 접근방법을 보완하여 사용하고자 한다. 모형설정에 관한 구체적인 내용은 생략하고 필수적인 몇 계산식과 간단한 설명으로 대체한다.

(1) 세부항목의 만족도

$$IS_{ij} = \frac{\sum_{k=1}^5 f_{ijk} \cdot K}{\sum_{k=1}^5 f_{ijk}}$$

IS_{ijk} : i 부문의 j 세부항목의 만족도
 f_{ijk} : i 부문의 j 세부항목의 k 번째 만족수준의 빈도수
 k : k 번째 만족수준의 점수 (1~5)

세부항목의 만족도를 나타내는 위의 식은 먼저 5단계의 척도에 따라 나누고 각각의 빈도수(응답자수)에 따라 단계별 점수를 곱해서 나온 결과의 합을 전체 빈도수로 나눈 값이다.

(2) 부문별 만족도

$$WSF_i = \sum_{j=1}^m WS_{ij}$$

$$WS_{ij} = IS_{ij} \cdot IW_{ij}$$

WSF_i : i 부문의 만족도
 WS_{ij} : i 부문의 j 부문 만족수준 기여도
 IW_{ij} : i 부문의 j 세부항목 가중치

WSF_i 는 i 부문의 만족도를 계산하는 식이다. 이는 각 부

문별 세부항목 만족도에 가중치를 곱해서 나온 결과를 합산한 것이다.

(3) 종합만족도

$$TSF = \sum_{i=1}^N TS_i$$

$$TS_i = WSF_i \cdot TW_i$$

TSF = 종합만족도
 TS_i = i 부문의 만족도 기여도
 TW_i = 부문가중치
 WSF_i = i 부문의 만족도

부문을 구성하는 세부항목의 중요도에 따라 그 부문의 전체 만족도를 계산한다. 여기서는 가중치를 사용자에게 입력 받도록 하여 종합 만족도 (TSF)를 계산한다.

5. 예측모형 도출 및 실험결과

어떤 주택을 선택할지에 대한 예측모형을 바탕으로 주택을 추천하기 위해서 예측모형을 만들고 분석하였다. 다항로지분석의 목적은 로지계수를 추정, 이를 바탕으로 특정 범주에 속할 확률(P) 혹은 선택할 확률을 계산하는 것이다. 이를 적용, 예측모형을 도출한 결과는 다음과 같다.

$$P_1 = \frac{\frac{P_1}{P_3}}{\frac{P_1}{P_3} + \frac{P_2}{P_3} + 1} = \frac{A}{A + B + 1}$$

여기서 P_1 = 주택1을 선택할 확률
 $A = \exp(b_{10} + b_{11}X_{11} + \dots + b_{1k}X_{1k})$
 $B = \exp(b_{20} + b_{21}X_{21} + \dots + b_{2k}X_{2k})$

이를 일반화시켜 예측확률 계산 공식으로 나타낸 확률 예측모형 식은 아래와 같다.

$$P_c = \frac{\exp(L_1)}{\exp(L_1) + \exp(L_2) + 1}$$

= 주택 1을 선택할 확률
 = 주택 2를 선택할 확률 [6]

앞서 도출된 예측모형 식을 토대로 수요자의 연령과 성별, 교육수준에 따른 주택 선택에 대한 설명 및 예측분석을 하였으며, 사용 변수는 아래와 같다.

종속변수 : 주택(1=다세대 주택, 2=아파트, 3=연립주택)
 독립변수 : 성별(1=남, 2=여), 연령, 교육수준(기간(년))
 표본크기 : 1453

예측모형 식을 통한 결과 분석에 앞서 확률선택모형

에 기초한 다항로직 통계분석을 이용한 변수의 유의미한 검증 결과 아래 [표 2]과 같은 모수 추정 값을 얻었다.

표 2 모수 추정 값

주택	추정 값	표준오차	Wald	자유도	유의확률	Exp(B)
(아파트)						
절편	.729	.492	2.195	1	.138	
age	_.032	.005	39.058	1	.000	.968
educ	_.022	.028	.608	1	.435	.979
[sex=1]	.298	.146	4.153	1	.042	1.347
[sex=2]	0b	.	.	0	.	.
(연립주택)						
절편	_.366	.399	.841	1	.359	
age	_.003	.004	.650	1	.420	.997
educ	.033	.022	2.209	1	.137	1.033
[sex=1]	_.478	.121	15.696	1	.000	.620
[sex=2]	0b	.	.	0	.	.

첫 결과는 기준범주인 다세대 주택에 비해 아파트를 선택하는 가능성에 미치는 3개의 독립변수의 효과를 검증한 것이다. 연령은 부가적으로 유의미하고 성별은 정적으로 유의미하게 나타났으며, 교육수준은 부가적인 영향을 준다. 그러나 이는 통계적으로 유의미하지는 않다. 두 번째 결과 역시 검증결과와 부호와 값의 차이일 뿐 동일한 결과이다.

이에 따른 통계분석의 유의미한 결과 값을 토대로 예측 모형 식을 이용하여 각 사례의 연령, 교육수준, 성별정보를 바탕으로 어떤 주택을 선택할 것인지 예측한 결과 값은 아래 [표 3]와 같다.

표 3 주택 선택 예측모형 실험결과

주택	예측 값			
	다세대 주택	아파트	연립주택	결과(%)
다세대 주택	547	5	109	82.9
아파트	225	2	51	0.4
연립주택	385	5	124	24.3
종합(%)	79.7	0.8	19.6	46.4

실제로 다세대 주택을 선택한 661명 중에서 548명은 다세대 주택을 선택 할 것으로 예측되었다. 전체적인 예측적중률은 46.5%로 나타났다. 적중률을 향상시키기 위해서는 주택선택에 관련이 높은 독립변수를 추가하면 적중률이 높아질 것으로 기대된다.

6. 결론

국민이 구매할 수 있는 품목 중에서 가장 비싸다고 할 수 있는 품목이 바로 주택이다. 그러나 주거만족도에 있어

서 주택 자체만의 물리적 요소만이 아니라 주택을 둘러싼 주변지역의 물리적 환경, 더 확대하여 사회·문화적 환경까지 다양하게 정의할 수 있다. 주거환경의 차이는 주거만족도에 큰 영향을 미치므로 주택 자체만이 아니라 주변 분석에 의한 입지결정 시스템을 통해서 주택 구매에 도움을 주고자 하였다. 주택만족도에 대한 빅데이터 분석을 통해서 어떤 조건이 중요한지 분석하여 입지결정을 보완할 수 있도록 하였다.

향후 주택구매에 관련한 분석만이 아니라 소매점 입점에 관한 입지결정에 도움을 줄 수 있도록 부동산의 상권에 대한 경제적 가치에 대한 연구가 필요하다. 미국 SBA(Small Business Administration)의 시장분석을 사용하여 산업정보, 타깃정보, 시장크기 등의 정보를 분석하여 각각의 score에 극성값을 주어 total score를 구하는 방법을 이용할 수 있다. 이런 방법을 통해 해당 부동산의 실시간 상권의 정보를 분석하여 사용자에게 정보를 전달 할 수 있게 시스템을 개선할 계획이다.

감사의 글

본 연구는 미래부가 지원한 2013년 정보통신·방송(ICT) 연구개발사업의 연구 결과로 수행되었음(1391105003). 이 논문은 2013년도 정부(교육부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임 (NRF-2013R1A1A2008578).

참고문헌

- [1] 권창기, "주거만족도 결정요인에 관한 연구" 한국지방자치연구, 2005
- [2] 이주택, "우리나라 대도시 주거단지의 인구밀도에 관한 조사연구", 대한건축학회, 1993
- [3] 신신애, "빅데이터 기술분류 및 현황", 한국정보화 진흥원, 2013
- [4] Kambatla, "Trends in big data analytics", Journal of Parallel and Distributed Computing, 2014.
- [5] 백영기, "소매중심지 입지를 위한 GIS 기반의 공간적 의사결정지원시스템", 대한지리학회, 2001
- [6] 오대식, "주택 의사결정지원시스템 구축을 위한 예측모형에 관한 연구" 대한건축학회, 2014