

SNS 데이터를 이용한 공공시설 매력도지수에 따른 접근성 분석기법

Accessibility Analysis Method based on Public Facility Attraction Index Using SNS Data

이지원¹⁾ · 유기윤²⁾ · 김지영³⁾
Lee, Ji Won · Yu, Ki Yun · Kim, Ji Young

Abstract

In order to expand the qualitative aspects of public facility, this study used SNS data to derive user-oriented preference factors for public facilities and then were quantified in terms of supply side and demand side. To derive preference factor, LDA, one of topic modeling, was used and attraction index was calculated for each facility. In addition we analyzed spatial accessibility to measure the degree of service experience of users by using 2SFCA model. The study area covered public libraries of Seoul, Korea. As a result of study, five topics were extracted as preference factors for the public library: Circumstance, Scale of facility, Cultural program, Parenting, Books and materials. In particular topic of circumstance and parenting were newly derived preference factors unknown in previous studies. As a result of calculating attraction index for each library, the index of Songpa Library, Jungdok Library, and Namsan Library was high. Songpa library has received good evaluation in parenting factor, and Jungdok & Namsan library in circumstance factor. The accessibility of each region seems to better in center of Seoul where public libraries are crowded, but shrinking toward the outskirts. We expect that the proposed method will contribute to user-oriented public facility evaluation and policy decision making.

Keywords : Social Network Service Data, Latent Dirichlet Allocation, Attraction Index of Public Facility, Spatial Accessibility, 2-Step Floating Catchment Area

초 록

본 연구는 공공시설의 질적인 측면의 확대를 위해 SNS (Social Network Service) 데이터를 활용하여 이용자 중심의 공공시설 선호요인을 도출하고, 이를 공급측면과 수요측면에서 정량화시키는 방법을 제안하였다. 선호요인 도출을 위해 토픽모델링 중 하나인 LDA (Latent Dirichlet Allocation)를 활용하였으며, 공급요인인 개별시설별로 매력도지수를 산출하였다. 또한 수요자 입장에서 서비스체감 정도를 측정하기 위해 공간접근성을 분석하였는데, 2SFCA (2-Step Floating Catchment Area) 방법을 활용하여 앞서 제안한 공공시설 매력도지수를 적용할 수 있도록 하였다. 실험은 서울시 공공도서관을 대상으로 진행하였다. 연구결과 공공도서관에 대한 선호요인으로 주변 환경, 시설 및 이용규모, 문화 프로그램, 육아, 장서 및 자료현황의 5가지 주제가 추출되었으며, 주변 환경이나 육아관련 주제는 새롭게 도출된 선호요인으로 선행연구와 차별성이 있다. 각 도서관별로 매력도지수를 산출한 결과 송파도서관, 정독도서관, 남산도서관의 매력도지수가 높게 나타났으며, 송파도서관은 육아요인에서 정독도서관과 남산도서관은 주변 환경요인에서 좋은 평가를 받고 있는 것으로 나타났다. 각 동별 공간접근성은 공공도서관이 많이 몰려있는 서울 중심부의 접근성이 좋은 편이며, 외곽지역으로 갈수록 줄어드는 것으로 보였다. 본 연구에 제안한 기법을 통해 이용자 중심의 공공시설 평가 및 정책의사 결정에 도움을 줄 것으로 기대한다.

핵심어 : SNS 데이터, LDA, 공공시설 매력도지수, 공간접근성, 2SFCA

Received 2018. 11. 12, Revised 2018. 12. 05, Accepted 2019. 02. 18

1) Member, Dept. of Civil and Environmental Engineering, Seoul National University (E-mail: vert802@snu.ac.kr)

2) Member, Dept. of Civil and Environmental Engineering, Seoul National University (E-mail: kiyun@snu.ac.kr)

3) Corresponding Author, Member, Institute of Construction and Environmental Engineering, Seoul National University (E-mail: soodaq@snu.ac.kr)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. 서론

1.1 연구배경 및 목적

공공시설이란 공공의 이익을 위하여 만든 공공용 시설(국토의 계획 및 이용에 관한 법률 제2조)로 국민 누구나 이용할 수 있는 공평한 접근이라는 관점에서 제공되어야 한다. 이에 입지 선정 시 기존 공업이나 상업입지 이론들의 입지의 사결정과 다르게 판단되며, 시장에서와 같은 가격경쟁 시스템이 없는 상태에서는 정부의 후생기준(government welfare criteria)에 의해 결정된다(Teitz, 1986). 하지만 과거 국내 공공시설의 입지는 정치적 기준 혹은 행정관서의 일방적 판단에 의해 결정된 부분이 많았기 때문에 공공서비스의 중복지역과 부재지역을 발생시키고, 예산의 효율성 및 형평성 문제를 유발시켰다(Park *et al.*, 2010). 특히 수도권과 같은 인구과밀지역의 경우는 이용자들의 지불의사가 존재하는 민간시설조차 입지하기가 어려워 공공시설의 경우 후순위로 밀려나게 되었다. 이는 형평성이라는 관점에서 이용자에게 고른 서비스를 제공하지 못했다는 결과를 초래하였으나, 이미 개발이 포화된 상황에서 새로운 공공시설의 추가는 쉽게 채택하기 어려운 정책적 결정이다. 이에 이러한 고정적인 입지문제를 해결을 위해 기존의 양적 차원에서의 시설 증대보다 질적인 측면에서의 확대에 초점을 맞추는 정책적 전환이 필요하다.

공공시설의 질적개선은 이용자 맞춤형으로 이루어져야 한다. 이용자들이 실질적으로 원하는 사항에 대한 개선이 이루어져야 만족도가 높아지고 방문율도 상승하게 된다. 지금까지 이러한 선호에 대한 연구는 주로 설문조사법으로 이루어졌는데, 이 경우 연구자의 주관이 들어갈 수 있다는 한계가 있다. 연구자가 집중적으로 연구하고자 하는 분야에 더 많은 수의 문항을 배분할 수 있고, 의도치 않게 원하는 방향으로 설문지의 문항이나 답을 구성할 위험이 있기 때문이다(Cho *et al.*, 2017). 이러한 한계를 극복하고자 최근에는 인터뷰나 토론회 형식의 대면조사도 실시하고 있다. 실제 서울시에서는 2017년 11월 25일 '공공 도서관 시민 대토론회'를 개최하여 공공도서관에 대한 개인의 생각과 경험을 공론화할 수 있는 자리를 마련하였다. 하지만 이 역시 참석대상 및 장소 등이 제한되어 있다는 점에서 여전히 한계가 있다.

반면 SNS (Social Network Service) 데이터를 이용하여 공공시설에 선호요인을 분석하는 것은 연구자의 주관을 최대한 배제하면서 이용자의 의견을 듣는 관점에서 적합한 방법이다. SNS 데이터는 어떤 목적과 의도를 가지고 올린 글이 아니기 때문에 연구결과를 객관화시킬 수 있다는 장점이 있다. 하지만 SNS 데이터들이 대부분 대용량이고, 이를 처리하는

방법 및 위치정보에 대한 사항이 제한적이었기 때문에 연구에 활용하기 어려웠다. 그러나 최근 하드웨어의 사양이 고급화되고 경제적 비용이 수반되지 않는 데이터 처리 오픈소스 소프트웨어가 등장하면서 이러한 대용량 SNS 데이터에 대한 분석이 가능해졌다(Kang, 2016). 이에 본 연구는 SNS 데이터를 이용하여 공공시설에 대한 선호요인을 도출하고, 이를 활용하여 분석할 수 있는 방법을 제안하고자 한다.

본 연구의 세부목적은 다음과 같다. 첫째, 공공시설의 질적 측면 개선에 초점을 맞추어 이용자 중심에서 공공시설의 선호요인을 도출하고자 한다. 공공시설 POI (Point Of Interest)를 활용하여 이용자들이 직접 올린 SNS 데이터로 시설에 대한 선호요인을 분석하였다. 둘째, 도출된 요인을 바탕으로 실제 사례에 적용하여 공공시설을 평가할 수 있는 분석기법을 제안하고자 한다. 지금까지 SNS 데이터는 처리를 통하여 데이터가 가지고 있는 함의를 도출하는 데는 많이 활용되었으나, 그 자체가 지표로 활용되는 연구는 미미하였다. SNS 데이터를 사용하여 지표를 구축할 경우 기존 지표를 통해 담아내지 못했던 사항인 공공시설의 이미지 등을 드러낼 수 있다는 장점이 있다. 이에 본 연구는 SNS 데이터를 활용하여 공급요인인 시설자체에 대한 공공시설 매력도를 측정하는 방법과 수요요인인 이용자 입장에서 어떤 체감효과를 가지는지에 대해 분석하는 방법을 제안하고자 한다. 이용자 입장에서 가지는 체감효과는 공간접근성 관점에서 비교분석할 수 있도록 하였다.

1.2 선행연구 고찰

공공시설에 대한 선호요인을 도출하기 위해 선행연구에서 주로 활용하는 방법은 통계데이터를 활용한 분석과 설문조사로 대표되는 여론수렴 방법이다. 통계데이터를 이용한 사례로 공공도서관의 이용자수의 영향을 미치는 요인을 분석하기 위하여 한국도서관협회에서 발간한 「2008 한국도서관 연감」을 활용한 연구가 있다. 상관회귀분석을 실시하여 좌석수, 건물연면적, 사서수 등이 도서관이용자수와 상관관계가 있음을 도출하였다(Choi, 2009). 또한 설문조사를 활용한 사례로는 한국 성인의 공공도서관 이용에 영향을 미치는 주요 요인을 분석하기 위하여 일반성인 1,000명을 대상으로 온라인패널을 통해 웹설문조사를 실시한 연구가 있다. 연구결과 우리나라 성인들의 공공도서관 방문은 인구가 많은 대도시 지역에 거주하는 독서량이 많은 개인이 일상적으로 쉽게 접근할 수 있는 중소규모의 새로 건립된 공공도서관이 있고, 이용자 개인의 요구에 부응할 수 있는 인적 자원이 충분하게 확보되어 있을 때 더 빈번히 이루어지고 있는 것으로 나타났다(Kwon and

Song, 2014). 하지만 이러한 연구들의 경우 관련 통계의 구득 여부, 연구자의 조사대상 및 설문문항 한정 등이 연구결과의 영향을 많이 미치는 한계가 있다.

공급요인인 시설의 매력도를 측정할 때 시설이 가지는 매력요소를 복합적 측면에서 산출하는 것이 필요하다. 한국국토정보공사(LX)는 2015년 공공기관 이동을 추진하면서 LX 지사의 매력도를 산출하였는데, 지사특성(건물노후도, 1인당 면적, 주차공간), 자산가치(건축가능면적, 도로접면, 어메니티), 접근성(소관청과의 거리, 고객접근성, 현장접근성) 등의 다양한 관점에서 적용하였다(Park *et al.*, 2017). 공공시설의 매력도는 시설자체가 가지는 매력요소와 외부환경으로부터 주어지는 것으로 나뉘볼 수 있다. LX의 사례에서 지사특성이나 자산가치가 전자에 해당한다면 접근성 관련 점수는 후자에 해당한다. 선행연구들의 경우 시설이 가지는 매력도 측정을 위해 다양한 변수를 활용하였으나, 데이터 구득을 위하여 통계청 등에서 제공하는 공표된 데이터만 사용하였다는 한계가 있다. 공공보건시설의 경우 의료인력수, 병상수, 상급병원여부 등이 주요 변수로 활용되었으며(Cho *et al.*, 2010; Kanuganti *et al.*, 2016; Park *et al.*, 2017), 사회복지시설은 시설종사자수, 건물규모 등이 활용되었다(Sohn *et al.*, 2007; Ahn *et al.*, 2014). 접근성 관련하여 시설이 가지는 매력도를 측정할 때는 시설자체 뿐만 아니라 외부환경으로부터 주어지는 것이 함께 고려되어 질 수 있는데, 공공도서관의 경우 접근성에 영향을 미치는 요인으로 장서수, 좌석수 외에 학교와의 거리, 도서관주변의 안전도, 지하철접근성, 공시지가, 주거지집중도 등이 적용되었다(Palmer, 1981; Won *et al.*, 2010; Kim *et al.*, 2015). 특히 Won *et al.*(2010)은 공공도서관 시설특성이 이용활성화에 미치는 영향을 분석하기 위해 대형점포시설과의 연계성, 지역중심 입지를 활용하였는데, 이때 공표된 데이터가 아닌 Arc GIS를 활용하여 새로운 지표를 구축하였다는데 다른 연구들과 차별성이 있다.

시설특성이나 장소를 평가하기 위해 비정형데이터를 활용한 연구는 최근 들어 증가하고 있는 추세이다. 비정형데이터를 지수화 시켜 공간관련 연구에 적용한 사례들은 주로 위치좌표를 이용한 SNS데이터를 활용한 경우와 텍스트마이닝(text mining) 기술을 이용하여 비정형데이터에서 추출한 정보를 적용한 경우로 나누어 볼 수 있다. Kent *et al.*(2013)은 2012년 Horsethief 계곡에서 발생한 산불기간 동안 트윗을 작성한 사용자들의 속성 및 위치좌표를 활용하여 분석을 실시하였다. Gao *et al.*(2018)은 위치좌표를 가진 트윗 데이터를 수집하여 인플루엔자 활동의 시공간적 패턴을 발견했는데, 이때 주요 키워드로 flu와 연관된 단어 fever, cough, sore throat,

runny nose 등을 활용하여 데이터를 분류하였다.

텍스트마이닝 기술은 비정형데이터인 텍스트를 벡터화 시켜 분석하는 기술로 단어빈도를 활용한 트렌드분석, 토픽모델링(topic modeling)을 활용한 주제 분석, 오피니언 마이닝(opinion mining)으로 대표되는 감성분석 등으로 나뉘볼 수 있다. 공간정보 관련 분야에서는 비정형 텍스트를 가지고 공간과 관련된 장소에 관한 평가에 대한 연구가 주로 활용되고 있다. Chae *et al.*(2017)은 감성분석 기법을 활용하여 소셜 미디어 텍스트를 통한 장소 선호도 분석 기법을 제안하였으며, Cho *et al.*(2017)은 여행 사이트 리뷰 텍스트 데이터를 활용하여 관광지 만족도 요인을 추출하고 평가하였다. 특히 Cho *et al.*(2017)의 연구는 감성분석, 주제 분석 등 다양한 텍스트마이닝 기법을 적용한 사례이다. 국내 관광지를 유형별로 구분한 후, 트립어드바이저의 리뷰 및 평점 등을 토대로 엘라스틱넷(elastic net)방법론을 통해 감정사전을 구축하고 토픽모델링 기법을 활용하여 유형별 관광지의 만족도 요인을 도출하여 최종적으로 각 유형별 관광지 만족도 요인점수를 산출하였다. 하지만 아직까지 SNS 데이터에서 추출한 데이터를 기반으로 공간적 방법론에 적용한 사례는 많지 않다.

수요요인인 이용자들의 체감효과 측정하기 위한 공간접근성 분석관련 연구는 이미 다방면에서 선행된 주제이다. 공간적 접근성을 측정하는 방법론은 현재까지 컨테이너 방식, 최소거리 방식, 총(평균) 비용 방식, 중력 포텐셜 방식 등이 제안되었고, 이들 중 중력 포텐셜 방식이 가장 진보된 것으로 받아들여지고 있다(Lee, 2012). 중력 포텐셜 방식은 인구와 시설 간의 잠재력으로 정의될 수 있는데, 거리와 시간에 따라 그 잠재력의 크기가 달라질 수 있다. 최근 연구에서는 중력 포텐셜 방식 기반의 발전된 형태로 Luo(2004)가 제안한 2SFCA(2-Step Floating Catchment Area) 방법이 주목받고 있는데, 해석이 직관적일 뿐만 아니라 하나의 공급지가 둘 이상의 수요지를 관할하더라도 공급량이 분산되지 않도록 개선하였다(Cho *et al.*, 2010). 특히 2SFCA 방법론은 현재까지 개발된 방법론 중 시설의 공급에 대한 세부적인 자료의 구축이 가능하다면 가장 타당한 것으로 판단되는데(Lee, 2014), 실제 구체적인 공급요인을 넣은 사례는 많이 없다. 본 연구는 SNS 데이터를 이용하여 구축한 시설의 매력도를 공급요인으로 적용한 공간접근성을 측정하기 위하여 2SFCA 방법을 적용하였다. 2SFCA 방법은 주로 의료서비스 등 공공시설 관련 접근성 측정연구에 주로 활용되고 있다(Kanuganti *et al.*, 2016). Cho *et al.*(2010)는 이 방법론을 통해 농촌지역 공공 보건서비스에 대한 공간적 접근성을 분석하였는데 공급요인을 각 보건소별 의료인력의 전문성으로 적용하였으며, 여주군 보건소 관계자와

의 인터뷰를 참고로 의료인력의 가중치를 공중보건과의사와 보건진료원, 간호사를 1:0.3:0.1로 설정하여 분석하였다. Ahn *et al.*(2014)은 구미시 노인복지시설을 중심으로 공간적 접근성의 변화 특성을 연구하였는데, 이 연구에서는 시설의 공급요인을 노인복지시설 종사자수로 한정하였다. 최근 2SFCA 방법과 관련한 연구는 모형자체를 개선한 수정모형 관련한 연구가 증가하고 있는 추세이다. Wan *et al.*(2012)는 의료서비스 수요는 근처 의료기관의 이용가능성에 의해 영향을 받는 것으로 가정하여 선택적 이용가중치의 단계를 추가한 3SFCA 방법을 제안하였으며, Park *et al.*(2017)은 더 나아가 임계거리 차별화, 연령별 응급의료 가중치, 관내 응급의료기관 선택가중치 등을 적용한 M3SFCA (Modified 3-Step Floating Catchment Area Model)을 제안하였다. 하지만 아직까지 공급에 대한 구체적인 수정모형을 제안한 연구는 많이 없다.

본 연구는 다음과 같은 측면에서 선행연구들과 차별성을 지닌다. 첫째, SNS 데이터를 활용하여 공공시설에 대한 실질적인 이용자의 선호를 분석할 수 있다. 공공시설 선호에 대한 선행연구들의 경우 설문조사 등 정형적인 여론수렴방법을 채택하기 때문에 연구자에 의도에 의한 문항 설계, 조사대상 제한 등으로 연구결과에 한계가 있을 수 있다.

둘째, 이용자 선호에 바탕을 둔 공공시설 매력도를 지수화시키는 방법론을 제안하였다. 정형데이터를 활용하여 지수화시키는 선행연구는 이미 많이 있으며, 비정형데이터의 경우도 최근 SNS 데이터를 활용하여 장소 등을 평가하고 있는 연구가 증가하고 있는 추세이나 아직 미비한 실정이다. 본 연구는 SNS 데이터를 활용하여 이용자들의 선호요인을 도출한 후 이를 고려하여 시설별 매력도지수를 구축하는 방법을 체계화시켰다.

셋째, 공간접근성 분석 시 공공시설 매력도를 반영할 수 있는 방법론을 제안하여 시설 선호도에 따른 이용자 체감효과를 분석할 수 있도록 하였다. 최근 많이 연구되어지고 있는 2SFCA 방법의 경우 거리조각 문제나 수요의 다양성을 반영한 수정모형 등이 등장하고 있으나 공급요인을 구체화시켜 적용한 사례는 거의 없다. 본 연구는 SNS 데이터를 이용한 공급시설의 매력도를 측정된 결과값을 공간접근성 방법론에 적용하여 접근성 수치를 정교화시켰다는데 의의가 있다.

2. 공공시설 매력도지수에 따른 접근성 분석방법

본 연구는 이용자 관점에서 공공시설 선호요인을 도출하고 이를 적용하여 공급요인인 공공시설의 복합적인 매력도와 수요요인인 이용자의 체감효과를 측정하는 방법을 제안하였다.

SNS 데이터인 블로그글을 수집하여 공공시설 선호요인을 도출화하고, 이를 지수화시켜 공공시설 매력도를 산출하였다. 또한 앞에서 도출된 공공시설 매력도지수를 고려한 공간접근성을 측정하여 이용자입장에서 체감효과가 지역별로 어떻게 나타나는지 분석하는 방법을 제안하였다. 전체적인 연구방법론은 Fig. 1과 같다.

2.1 공공시설 선호요인 도출

빅데이터로 분류되는 SNS 데이터는 자료의 양이 방대하기 때문에 원하는 자료를 구득하기에 쉽지 않다. 특히 블로그와 같이 대용량 데이터의 경우 필터링이 매우 중요하다. 공공시설에 대한 이용자의 의견을 듣기 위하여 블로그를 검색하면 실제로 방문과 관련된 글보다는 광고성 위주의 글이 많이 검색된다. 이러한 문제점을 개선하고자 본 연구에서는 문서 분류 알고리즘인 SVM (Support Vector Machine) 기법을 적용하였다. SVM은 통계적 학습이론을 기반으로 개발되어 분류문제를 해결하는데 좋은 성능을 보이고 있으며, 대부분의 학습 알고리즘이 경험적 위험(empirical risk)을 최소화하는 데 초점을 맞췄다면 SVM 기법은 구조적 위험(structural risk)의 최소화에 기반을 둔 기법으로 분류 문제에 있어 일반화가 용이하기 때문에 다양한 분야에서 응용되고 있다(Lee *et al.*, 2017). 블로그 API (Application Programming Interface)를 통해 해당 공공시설에 대한 글을 수집한 후, 전체 문서 중 일부를 라벨링하여 이용관련 글이면 1, 그렇지 않은 경우는 -1을 부여하여 학습시키고 전체 데이터에 적용하여 최종적으로 공공시설 이용관련 문서만 분류하였다.

분류된 문서는 자연어처리 과정인 NLP (Natural Language Process)을 통해 불용어 제거, 형태소 분석 등을 실시하여 단어단위의 명사만 추출될 수 있도록 하였다. 추출된 명사들을 대상으로 토픽모델링을 실시하여 최종적으로 이용자입장의 공공시설 선호요인을 도출하였다. 토픽모델링은 방대한 양의 텍스트를 분석하여 주제를 발견하고 해당 주제들을 토픽들로 분류하는 기법으로 텍스트마이닝의 주요 분야이다.

본 연구에서는 토픽모델링 중 사전확률을 고려하기 때문에 분류성공률이 높고 해석이 다른 기법에 비해 용이하다는 장점을 지닌 LDA (Latent Dirichlet Allocation)을 적용하여 분석하였다(Bae *et al.*, 2013). LDA는 토픽별 단어의 분포 및 문서별 토픽의 분포를 추정하는 확률모형이다. Eq. (1)은 이러한 문서 생성과정을 나타내고 있다(Blei, 2012). LDA 분석을 통해 전체 문서에 대한 토픽을 도출한 후, 토픽별 비중 및 각 토픽별 단어의 확률분포를 구할 수 있다.

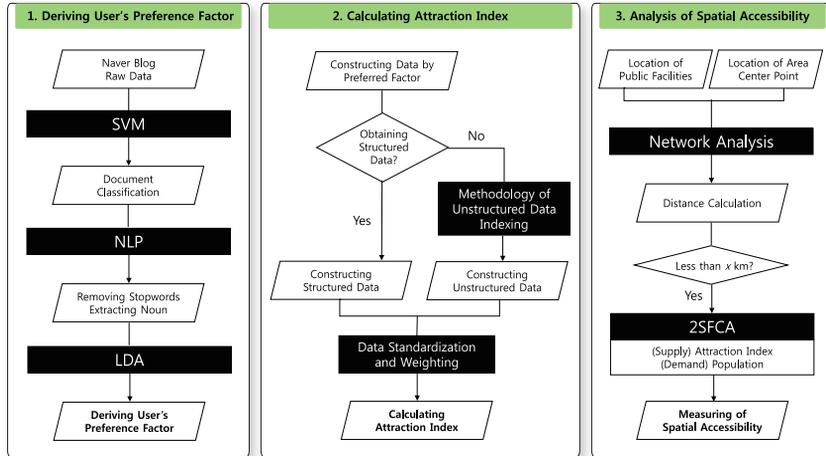


Fig. 1. Study flow chart

$$p(\beta_{1:K}, \theta_{1:D}, z_{1:D}, w_{1:D}) = \prod_{i=1}^K p(\beta_i) \prod_{d=1}^D p(\theta_d) \left(\prod_{n=1}^N p(z_{d,n} | \theta_d) p(w_{d,n} | \beta_{1:k}, z_{d,n}) \right) \quad (1)$$

where β_k is a distribution over the vocabulary; $\theta_{d,k}$ is the topic proportion for topic k in document d ; $Z_{d,n}$ is the topic assignment for the n th word in document d ; $W_{d,n}$ is the n th word in document d , which is an element from the fixed vocabulary.

2.2 공공시설 매력도지수 산출

지수는 구체적인 숫자 자체의 크기보다는 시간의 흐름에 따라 수량이나 가격 등 해당 수치가 어떻게 변화되었는지를 쉽게 파악할 수 있도록 만든 것이다(Statistics Korea, 2015). 이에 단순히 어떤 자료에 대한 통계적 수치만을 표기한 지표와는 구별된다. 공공시설 매력도지수는 시설의 한 면만이 아닌 복합적인 측면에서 매력도를 측정하기 위해 앞에서 도출한 선호요인별로 지표를 구축한 후 이를 하나의 수치로 합쳐 복합지수로 산출해야 한다. 이때 선호요인별로 지표 구축 시 요인을 나타낼 수 있는 통계자료 등 구체적인 정형데이터가 있을 경우에는 그 자료를 지표로 구축하되, 관련된 정형데이터가 없는 경우 비정형데이터를 활용하여 지표를 구축한다.

여기서 활용한 비정형데이터는 SNS 데이터로 이용자들이 개별시설에 대해 올린 SNS 데이터를 수집하여 2.1절에서 도출된 선호요인에 대한 단어 노출도를 수치화 한 것이다. 예를 들어 공공시설 선호요인으로 “육아”라는 주제가 도출되었고, 주제를 이루고 있는 단어가 “수유실, 영유아검진, 유모차”등

이라면 “육산구청”이라는 특정시설에 대한 SNS 데이터를 수집하여 관련 단어가 육산구청에 관한 블로그 글에 얼마나 노출되어있는지를 파악하는 방법이다. LDA 분석은 전체 글에 대한 주제들과 주제를 이루는 단어 및 단어가중치를 결과값으로 도출한다. 앞서 제시한 사례에서는 “육아”라는 주제 외에 “접근성”이나 “시설노후화” 등이 주제로 도출될 수 있고 육아라는 주제를 구성하는 단어 및 단어가중치가 함께 제시된다. 이때 단어 가중치는 그 주제를 구성하는 단어의 영향력을 의미하는 것으로 “수유실”이 0.05, “영유아검진”이 0.03, “유모차”가 0.01로 나타났다면, “수유실”이 육아라는 주제를 가장 잘 나타내는 단어라고 해석할 수 있다.

본 연구는 선호요인별로 구성된 단어별로 단어가중치에 전체 문서에 대한 단어의 언급횟수를 곱한 후 요인별로 다 더해서 최종지표를 산출한다. 이때 단어의 언급횟수는 전체 단어를 고려한 값으로 표준화 시킨다. 이는 대형 공공시설의 경우 SNS 노출도가 상대적으로 높아 같은 단어라도 절대값만 가지고 보면 언급횟수가 더 많은 것으로 보이기 때문이다. 하지만 노출도가 낮은 시설의 경우도 전체 단어 중 해당 단어의 언급횟수가 높으면 상대적으로 선호도가 높은 것으로 간주될 수 있다. 예를 들어 “육산구청”에 관한 SNS데이터가 “금천구청”에 비해 월등히 많다면, 금천구청의 “육아”관련 사항이 “접근성”이나 “시설노후화”에 비해 특화되어 있음에도 불구하고 낮은 점수를 산출할 수 있기 때문이다.

다음은 구축된 지표를 하나의 지수로 합치는 과정이다. 지표의 단위가 다르고 지표별로 중요도가 상이하기 때문에 통계적인 방법론을 사용하여 복합지수를 개발해야 한다. 먼저 선호요인별로 지표를 구축한 후 데이터를 표준화시킨다. 원

자료를 표준화하는데 가장 많이 이용되는 방법인 선형 변형 (linear scale transformation) 중 점수범위화 방법을 활용하였다. 원자료와 최소값 간의 차이를 주어진 기준의 범위(최대~최소)로 나누어 줌으로써 표준화 점수를 산출한다(Lee, 2003).

또한 공공시설 매력도지수 같이 여러 지표를 종합해서 하나의 지수로 산출해야 하는 경우 각 지표들의 중요도가 서로 다를 수 있기 때문에 지표들의 상대적 중요도를 반영하기 위해 가중치를 부여해야 한다. 본 연구에서는 도시재생사업단에서 적용한 요인분석(factor analysis)을 이용한 가중치 부여방식을 사용하였다(KURC, 2010). 이 연구는 개별도시의 쇠퇴정도를 건물의 노후도, 인구이동 정도 등 복합적인 측면에서 측정할 '복합쇠퇴지수'를 개발하였다. 요인분석은 수집된 많은 변수들을 유사한 특성을 지닌 변수들끼리 묶어 적은 수의 요인으로 축소시키는 방법으로 설명변수와 종속변수의 구분이 없다(Lee et al., 2012). 본 연구에서는 적용한 요인분석을 통한 가중치 산정방법은 요인분석에서 고유치(eigenvalue)가 1보다 큰 요인들을 추출하고 각 요인들의 공통분산을 토대로 하여 각 지표의 가중치를 산정하였다. 즉, 각 변수의 공통변량(communality)을 추출된 요인들의 고유치의 합계로 나눈 값이다. 요인의 추출 방법으로는 주성분 분석방법(principal component)을 사용하였으며, 요인회전의

방법으로는 varimax를 이용하였다. 요인분석을 통해 얻은 가중치 결과 값을 표준화시킨 지표에 곱한 후 전체를 합산하여 최종 공공시설 매력도 지수를 도출하였다. 공공시설 매력도 지수를 산출하는 전체 방법론은 Fig. 2와 같다.

2.3 공공시설 매력도를 고려한 공간접근성 분석

본 연구에서는 공공시설 매력도지수를 공급요인으로 고려하여 공간접근성을 분석하기 위해 2SFCA 모형을 활용하였다. 2SFCA는 두 단계로 나뉘서 계산되는데, 첫 번째는 공급시설을 중심으로 수요를 측정하고 두 번째로는 수요지를 중심으로 이용할 수 있는 공급시설에 대한 접근성을 계산한다. Eq. (2)은 이러한 단계를 수식으로 나타내고 있다(Luo, 2004).

R_j 는 공급시설 j 의 공급대비 인구비율인 PPR (Provider to Population Ratio)을 나타내며, 이때 R_j 가 공급시설 j 의 공급능력을 의미하는 지표로 여기에 앞 절에서 구한 매력도지수를 적용할 수 있다. 공급시설 j 가 서비스를 제공할 수 있는 최대 거리를 d_0 라고 본다면 P_k 가 이 거리 내에 있는 수요량을 의미한다. 두 번째 단계는 수요지를 중심으로 이용할 수 있는 공급시설의 PPR을 합산하는 과정이다. A_j 가 최종결과 값으로 각 수요지에서 이용할 수 있는 공급시설 j 에 대한 접근성을 의미하며, 수요지가 도달할 수 있는 거리 내에 있는 공급시설 j 의 PPR을 더한 값이다. 2SFCA는 결과 값이 직관적일 뿐만 아

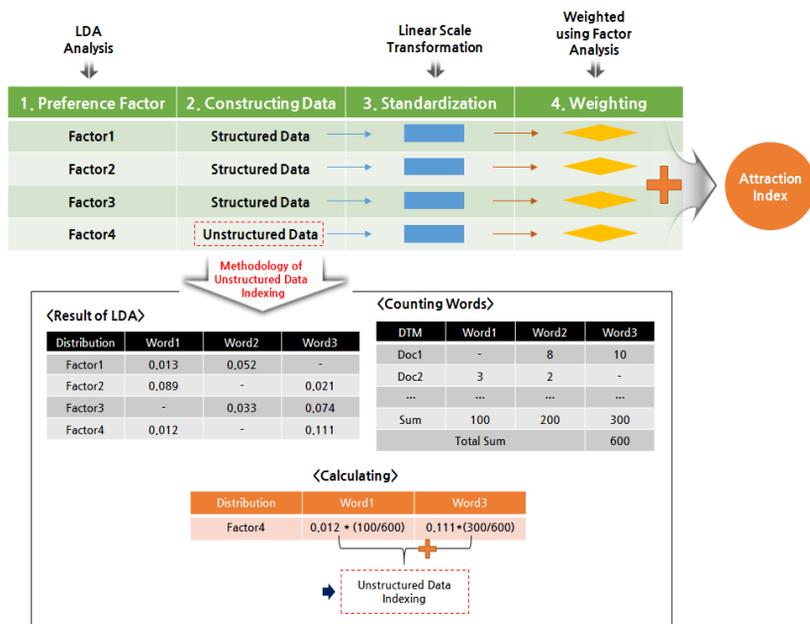


Fig. 2. Methodology of calculating attraction index

나라 하나의 공급지가 둘 이상의 수요지를 관할하더라도 공급량이 분산되지 않게 산출할 수 있다는 장점이 있다(Yang *et al.*, 2006; Cho *et al.*, 2010).

$$\begin{aligned} \text{<Step 1>} \quad R_j &= \frac{S_j}{\sum_{k \in d_{jk} \leq d_0} P_k} \\ \text{<Step 2>} \quad A_i &= \sum_{i \in d_{ik} \leq d_0} R_j \end{aligned} \quad (2)$$

where R_j is the provider-to-population ratio at service supply location j ; S_j the measure of service supply volume at location j ; P_k the population count at location k that lies within the service area catchment of j ; d_{jk} is the cost in distance between locations k and j , d_0 is a user-specified maximum threshold cost; A_k is the final 2SFCA accessibility score assigned to the population at location k .

또한 수요량 측정은 행정구역이나 집계구의 인구중심점이 많이 활용되며, 거리 설정 시 직선거리는 현실성과 정확성이 떨어져 네트워크거리를 사용한다(McGrail, 2012; Langford *et al.*, 2016; Park *et al.*, 2017). 본 연구에서는 공급시설의 위치와 행정구역 상 중심점의 거리를 네트워크 분석을 통해 산출한 후 공급시설 서비스가 최대 도달할 수 있는 거리 내에 있는 값을 2SFCA 모형에 적용하여 접근성을 계산하였다. 이때 공급지표 S_j 는 공공시설 매력도지수를 활용하였으며, 수요량 P_k 는 인구를 사용하였다.

3. 실험 및 결과

3.1 실험대상 및 데이터 수집

본 연구에서 제안한 공공시설 매력도 및 접근성 분석기법을 통해 실제 네이버블로그 데이터를 대상으로 공공도서관에 대한 분석을 수행하였다. 서울시 공공도서관 147개 중 국가도서관 통계에서 발행하는 공공도서관 통계결과표 참고하여 방문자수 상위 15개 도서관에 대해 실험을 진행하였다. 이는 공공시설 특성상 SNS 노출도에 대해 시설별로 격차가 크기 때문에 방문자수가 상대적으로 많은 공공도서관이 관련 데이터도 많을 것으로 예상하여 선정하였다. 수집기간은 최근 1년으로 2017년 6월 1일부터 2018년 5월 31일로 한정하였다. 네이버블로그 오픈API를 이용하여 총 13,948건을 수집하였으며 실험 대상 및 각 블로그 수집건수는 Table 1과 같다.

3.2 서울시 공공도서관 이용자 선호요인 도출결과

서울시 공공도서관 관련 네이버블로그 수집 건 중 SVM을 통하여 실제 도서관 이용에 대한 후기와 그렇지 않은 문서로 분류하였다. 총 13,948건의 문서 중 3,428건이 적합한 문서로 선정되었는데, 이는 공공도서관명으로 블로그데이터를 수집할 경우 근처 아파트분양, 맛집 광고 등 다양한 형태의 광고성 데이터가 포함되어 있기 때문이다. 실험에서 사용한 프로그래밍 언어는 Python 3.6.0이며, 단어 단위분리, 불용어 제거, 품사 태깅 등 자연어 처리를 위해 KoNLPy (Korean NLP in Python)패키지를 활용하였다.

Table 1. List of experiments

Rank	Name	Visitors	Num of naverblog
1	Songpa public library	1,831,215	419
2	Jungnanglib Seoul library	1,787,600	300
3	Yangcheon public library	1,761,949	239
4	Mapo lifelong learning center public library	1,688,509	605
5	Jungdok public library	1,622,421	2,690
6	Sungdong public library	1,547,923	1,172
7	Nowon public library	1,537,522	187
8	Gangbuk culture information library	1,525,642	364
9	Nowon lifelong learning center	1,451,290	185
10	Gwangjin district public library	1,449,453	155
11	Seochogu banpo library	1,306,337	179
12	Namsan public library	1,233,637	1,652
13	Gocheok public library	1,218,497	366
14	Jongno public library	1,153,409	268
15	Seoul metropolitan library	1,148,307	5,167

LDA 모델링을 위해 먼저 전처리과정을 통해 추출된 단어 중 명사(NNG)만 추출하여 사전으로 만든 후 gensim 패키지의 doc2bow 메소드를 사용하여 term id와 term frequency로 구성된 문서-단어 매트릭스를 구성하였다. 파라미터값을 바꾸면서 실험하였으며 토픽수 5개, 단어수 5개로 지정하여 1,000번 반복수행한 결과 주변환경, 시설 및 이용규모, 문화프로그램, 육아, 장서 및 자료현황의 5가지 주제로 추출되었다. 이는 기존 선행연구에서 도서관 방문 선호요인으로 선정하지 못하던 주변환경이나 육아 등에 대한 토픽을 비정형데이터를 통해 찾아냈다는데 의의가 있다. LDA를 적용하여 도출된 공공도서관 방문 선호요인과 요인들을 대표하는 단어들은 Table 2와 같다.

“주변 환경”이라는 선호요인에 대해서는 사진, 건물, 사람, 카페, 벚꽃이라는 단어가 추출되었다. 이는 공공도서관의 방문목적이 단순히 책을 빌리는 목적이 아니라 사진을 찍거나 근처 카페를 방문하는 등 외부 환경적 요소에 대해 좌우된다는 것을 알 수 있다. 특히 벚꽃이 추출된 것을 통해 벚꽃명소로 이름이 난 도서관은 그 시즌에 방문자수가 증가함을 의미한다. “시설 및 이용규모” 요인에서 도출된 주제단어는 도서관, 이용, 인문학, 열람실, 시설이다. 이용자들은 도서관에 대한 전체적 이용뿐만 아니라 열람실 활용여부나 시설 자체의 편리성에 관심을 가지고 있다. 인문학의 경우 최근 인문학에 대한 관심이 증가하면서 관련 장서나 시설의 여부가 방문의 선택을 고려하는 요소 중 하나로 꼽히고 있다. “문화프로그램” 요인에 대한 단어는 문화, 시민, 전시, 공간, 프로그램이 추출되었다. 이는 도서관의 이용목적이 책에서 문화적 요소로 바뀌고 있음을 시사한다. 복합적인 문화공간으로서의 역할이 확대된 것이다. “육아” 선호요인에 대한 주제단어는 아이, 시간, 생각, 사람, 엄마로 나타났다. 이는 공공도서관을 이용하는 주요 계층 중 하나로 아이와 육아를 주로 담당하는 엄마 등으

로 해석할 수 있다. 향후 공공도서관 관련 정책 수립 시 이용자에 대한 정확한 분석을 통한 맞춤형 지원을 제공한다면 효과를 극대화 할 수 있을 것이다. 마지막 “장서 및 자료현황” 요인에 대한 단어는 도서관, 도서, 자료실, 자료, 정보이다. 이는 도서관의 전통적 목적을 나타내는 것으로 아직까지 공공도서관을 방문하는 주요이유가 독서나 자료 등을 찾기 위해서라는 것을 알 수 있다.

3.3 서울시 공공도서관 매력도지수 산출결과

앞 절에서 도출한 공공도서관 방문 선호요인 5가지에 대한 지표를 구축하여 전체 매력도지수를 산출하였다. 공공도서관의 매력도를 시설자체가 가지는 매력요소와 외부환경으로부터 주어지는 것으로 나뉠 수 있는데, “시설 및 이용규모”, “문화프로그램”, “장서 및 자료현황”이 시설 자체가 가지는 매력요인으로 꼽을 수 있고 “주변 환경”은 외부환경으로부터 주어지는 요소라고 볼 수 있다. “육아”에 대한 요인은 복합적인 관점을 가지고 있는데, 공공도서관 내에 육아관련 책이나 편의시설이 있는 것을 보면 시설자체에 대한 매력요소이나 근처에 육아기 인구가 많아서 엄마들 주도의 모임이나 입소문 등으로 방문자수가 증가한다는 점을 고려한다면 외부적인 면도 간과할 수 없다.

이에 본 연구에서는 시설 자체가 가지는 매력요소에 대한 점은 국가도서관 통계에서 발행하는 공공도서관 통계결과표에 구축된 자료를 활용하였고, 외부환경으로부터 주어지는 요소에 대한 지표는 SNS 데이터를 활용하여 구축하였다. 이는 외부환경으로부터 주어지는 매력요소가 기존 지표로는 측정이 쉽지 않기 때문이다. “육아”관련 요인은 본 연구가 SNS 데이터를 통해 매력도를 산정한다는 연구목적에 초점을 맞추고 있기 때문에 시설 자체보다는 외부환경적 요소가 더 영향력이 큰 것으로 설정하여 분석하였다.

Table 2. Result of LDA analysis

Topics	Extraction of words
Circumstance	$0.012 \times \text{"photo"} + 0.007 \times \text{"building"} + 0.007 \times \text{"person"} + 0.006 \times \text{"cafe"} + 0.006 \times \text{"cherry blossom"}$
Scale of facility	$0.115 \times \text{"library"} + 0.018 \times \text{"overall use"} + 0.015 \times \text{"humanities"} + 0.013 \times \text{"reading room"} + 0.011 \times \text{"facility"}$
Cultural program	$0.012 \times \text{"culture"} + 0.011 \times \text{"citizen"} + 0.010 \times \text{"exhibition"} + 0.010 \times \text{"space"} + 0.009 \times \text{"program"}$
Parenting	$0.016 \times \text{"children"} + 0.014 \times \text{"time"} + 0.014 \times \text{"thinking"} + 0.012 \times \text{"person"} + 0.008 \times \text{"mom"}$
Books and materials	$0.019 \times \text{"library"} + 0.016 \times \text{"reading"} + 0.014 \times \text{"reference room"} + 0.013 \times \text{"material"} + 0.012 \times \text{"information"}$

Table 3. Constructing indicators by public library preferred factor

Library name	Circumstance indicator (using SNS)	Number of seats	Number of cultural program courses	Parenting indicator (using SNS)	Number of reference room users
Songpa	1.569	1,176	100	4.948	1,293,101
Jungnanglib	0.326	600	47	2.545	1,449,119
Yangcheon	0.924	1,612	45	3.965	1,312,345
Mapo Lifelong	0.869	885	285	4.056	805,628
Jungdok	2.990	1,080	85	4.121	958,385
Sungdong	0.643	558	124	4.131	1,332,501
Nowon	0.903	685	24	2.936	619,902
Gangbuk	0.795	524	64	2.740	1,189,979
Nowon Lifelong	0.755	853	327	2.830	937,546
Gwangjin	0.934	949	117	2.648	1,286,599
Seochogu Banpo	0.764	450	66	4.578	914,436
Namsan	1.674	1,303	70	3.978	825,933
Gocheok	0.821	665	49	3.359	926,555
Jongno	1.020	506	30	3.446	532,553
Seoul	1.358	390	1	3.815	1,148,307

통계결과표에 의해 지표를 선정할 때는 선행연구에서 사용한 지표를 고려하여 선정하였다. Choi(2009)는 공공도서관의 이용자수의 영향을 미치는 요인을 분석하기 위하여 다음의 12개 변수를 사용하였다. 인적측면은 사서수와 도서관 전체직원 수, 장서측면은 장서수와 연간 증가 책 수, 예산측면은 연간 자료구입비 예산과 연간 도서관 전체 예산, 행사 측면은 문화 프로그램 실시횟수와 도서관 및 독서관련 프로그램 실시횟수, 그리고 기타 측면은 좌석 수, 도서관연면적, 개관일수, 개관경과년도 등이다. 본 연구에서는 선행연구에서 선정한 지표를 최우선순위에 두었으며, 비슷한 위계에 있는 지표가 존재하는 경우는 연구목적에 맞게 선택하였다.

시설 및 이용규모의 경우 주제관련 단어가 열람실이나 시설 등으로 도출되었기 때문에 선행연구에서 도서관 내 시설 규모를 정량적으로 드러낼 수 있는 지표인 “총 좌석수”를 선택하였다. 문화 프로그램 요인 역시 선행연구에서 사용한 “문화 프로그램 실시횟수”를 사용하였다. 하지만 장서 및 자료현황에 관한 요인은 자료나 자료실에 대한 단어가 주제관련 단어로 선정되었기 때문에 선행연구에서 사용한 “장서수”보다는 “자료실 이용자수”를 관련 지표로 선정하였다.

SNS 데이터를 이용하여 지표를 구축하는 방법론은 SNS 상에서 해당 단어를 통해 나타난 이미지 노출도를 지표화하는 작업이다. 따라서 선호요인별 구축된 주제단어들에 대한 언급비율을 가중치와 곱한 후 합산하는 방식으로 지표를 구축하

였다. 예를 들어 송파도서관의 주변 환경에 대한 점수는 송파도서관에 대한 SNS 데이터를 수집한 후, 사진, 건물, 사람, 카페, 붓꽃에 대한 단어 언급횟수를 카운팅하였다. 이는 블로그 상에서 해당 도서관이 관련 주제에 얼마나 노출되어 있는지 보여준다. 이때 전체 단어에 대한 언급비율을 적용하는데, 이는 각 도서관별 SNS데이터의 양을 고려한 것이다. 또한 각 단어별 가중치로 앞 절에서 분석한 LDA 결과값에 대한 단어가중치를 사용하였는데, 주변 환경이라는 주제에 대해서 사진이라는 단어가 붓꽃이라는 단어에 비해 영향력이 높기 때문이다. 송파도서관에 사진이라는 단어는 10번 밖에 언급되지 않았는데, 붓꽃이라는 단어가 100번 언급되었다면 주변 환경 점수는 크지 않을 수 있다. 이는 LDA 분석을 통해 붓꽃이라는 단어가 좀더 특정 도서관에 국한되어 나타난 지표라는 것을 알 수 있다.

방문자수 상위 15개 서울시 공공도서관의 선호 요인별 지표구축 결과는 Table 3과 같다. SNS 데이터를 이용한 주변 환경 지표의 경우 정독도서관, 남산도서관 등에서 높게 나왔으며, 육아 관련 지표는 송파도서관에서 높게 나타났다. 좌석수 관련 지표는 양천도서관, 남산도서관, 송파도서관 등에서 높게 나타났으며, 문화프로그램 수는 노원평생학습관, 마포평생학습관 등에서 높게 나타났다. 자료실 이용자수에 관한 지표는 중랑구립도서관, 성동도서관, 광진정보도서관 등에서 높게 나타났다.

Table 4. Result of correlation analysis

Name of indicator		Circumstance indicator	Number of seats	Number of cultural program courses	Parenting indicator	Number of reference room users
Circumstance indicator	Pearson correlation	1	.397	-.122	.437	.034
	Sig.(2-tailed)		.143	.666	.103	.905
	N	15	15	15	15	15
Number of seats	Pearson correlation	.397	1	.156	.283	-.308
	Sig.(2-tailed)	.143		.580	.308	.264
	N	15	15	15	15	15
Number of cultural program courses	Pearson correlation	-.122	.156	1	-.028	-.407
	Sig.(2-tailed)	.666	.580		.922	.133
	N	15	15	15	15	15
Parenting indicator	Pearson correlation	.437	.283	-.028	1	.122
	Sig.(2-tailed)	.103	.308	.922		.665
	N	15	15	15	15	15
Number of reference room users	Pearson correlation	.034	-.308	-.407	.122	1
	Sig.(2-tailed)	.905	.264	.133	.665	
	N	15	15	15	15	15

Table 5. Result of factor analysis

Component	Initial eigenvalues			Extraction sums of squared loadings			Rotation sums of squared loadings		
	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %
1	1.762	35.248	35.248	1.762	35.248	35.248	1.743	34.859	34.859
2	1.546	30.913	66.161	1.546	30.913	66.161	1.565	31.302	66.161
3	.727	14.530	80.691						
4	.489	9.787	90.479						
5	.476	9.521	100.000						

Table 6. Weighting process and results for each factor

Name of indicator	Factor1	Factor2	Communality	Weight #1	Weight #2
Circumstance index	0.833	-0.091	0.702170	29.91777	0.212212
Number of seats	0.763	-0.142	0.602333	25.66395	0.182039
Number of cultural program courses	0.678	0.466	0.676840	28.83852	0.204556
Parenting index	-0.088	0.775	0.608369	25.92113	0.183863
Number of reference room users	0.001	-0.848	0.719105	30.63933	0.217330
Sum	2.187	0.16			1

Note: Factor1 and Factor2 are result of rotated factor pattern.

5가지 선호요인에 대한 지표들을 합산하여 하나의 매력도 지수를 만들기 위해서는 지표들 표준화시키는 것과 가중치를 부여하는 것이 필요하다. 지표 표준화를 위해서는 선형방법 중 점수범위화 방법을 사용하였으며, 표준화 결과 최대치와 최소치가 1과 0의 값으로 나타난다.

요인분석을 이용한 가중치 부여를 위해서 먼저 구축된 지표들을 대상으로 상관분석과 요인분석을 실시하였다. Table 4는 상관분석, Table 5는 요인분석 결과를 나타낸다. 상관분

석 결과 상관성이 높은 변수는 거의 없는 것으로 나타났다. 요인분석 결과 고유치가 1보다 큰 요인은 2개로 추출되었는데, 성분1은 34.9%의 설명력을 가지고 성분2는 31.3% 설명력을 가져 두 개의 요인을 통해 전체 데이터의 66.2%를 설명할 수 있다.

Table 6은 각 요인별 가중치도출 과정 및 결과 값을 보여준다. 요인분석에 의한 가중치는 각 지표들 중에서 요인으로 설명되지 않은 unique factor를 제거한 후 각 요소들이 전체 지

Table 7. Calculating public library attraction index

Library name	Circumstance index	Scale of facility index	Cultural program index	Parenting index	Books and materials index	Total attraction index
Songpa	0.099	0.182	0.132	0.056	0.146	0.614
Jungnanglib	0.000	0.000	0.035	0.026	0.000	0.061
Yangcheon	0.048	0.108	0.205	0.025	0.093	0.478
Mapo Lifelong	0.043	0.114	0.083	0.160	0.070	0.470
Jungdok	0.212	0.119	0.116	0.047	0.091	0.585
Sungdong	0.025	0.120	0.028	0.069	0.179	0.422
Nowon	0.046	0.030	0.049	0.013	0.096	0.234
Gangbuk	0.037	0.015	0.022	0.036	0.156	0.266
Nowon Lifelong	0.034	0.022	0.078	0.184	0.021	0.338
Gwangjin	0.048	0.008	0.094	0.065	0.190	0.405
Seochogu Banpo	0.035	0.154	0.010	0.037	0.101	0.337
Namsan	0.107	0.109	0.153	0.039	0.065	0.472
Gocheok	0.039	0.062	0.046	0.027	0.185	0.359
Jongno	0.055	0.068	0.019	0.016	0.217	0.377
Seoul	0.082	0.096	0.000	0.000	0.180	0.359

표를 설명하는 분산 중 개별 지표를 설명하는 분산의 정도를 나타낸 것으로, 가중치가 가장 높은 지표는 전체를 대표하는 지표로 판단할 수 있다(KURC, 2010). 요인분석을 통한 가중치 도출을 위해 먼저 변수별 요인부하의 제공함을 통해 공통 분산을 구한 후 선택된 요인에 의해 설명되는 분산의 합을 구해서 이 값으로 공통분산을 나눠주어 산출하였다. 최종적으로 각 선호요인별 가중치의 합이 1이 될 수 있도록 표준화시켰다. Table 6에서 Weight #2로 명기된 값이 최종적으로 부여한 가중치인데, 0.18에서 0.21 사이의 값을 가지며 자료실 이용자 수나 주변 환경 관련 지표가 상대적으로 가중치가 높게 나타났다. 이는 이들 지표가 다른 지표에 비해 도서관의 매력도를 더 반영한다고 해석할 수 있다.

최종적으로 표준화시킨 값에 가중치를 부여한 다음 5가지 선호요인에 대한 값을 합산하면 전체 도서관에 대한 매력도 지수를 산출할 수 있다. Table 7은 방문자수 상위 15개 공공 도서관에 대한 매력도지수를 보여준다. 결과 값을 보면 송파도서관, 정독도서관, 남산도서관 등에서 비교적 매력도가 높은 것으로 평가되었다. 이는 SNS 데이터에 대한 노출도와 최종 매력도지수간의 상관관계가 존재하며, 시설 자체의 매력 요소뿐만 아니라 외부 환경으로부터 주어지는 것의 비중이 점점 커지고 있음을 나타낸다. 실제로 SNS 상에서 좋은 평가를 받은 시설들은 앞으로 방문하고자 하는 예비 이용자들에게 좋은 이미지를 심어주기 때문에 방문결정 여부에 중요한

변수로 작용할 수 있다.

매력도지수가 높게 나온 공공도서관을 선호요인별 지표와 함께 분석해보면 다음과 같다. 먼저 송파도서관은 전체적으로 매력도가 높지만 특히 육아부문에서 점수가 높게 나타난다. 이는 송파구가 육아기 인구가 상대적으로 많은 지역이며, 송파도서관에 육아 관련 시설이나 장소 등을 보유하여 SNS 상에서 엄마들을 중심으로 아이와 함께 방문하고 싶은 좋은 도서관으로 평가되는 것으로 해석할 수 있다. 다음으로 정독도서관과 남산도서관의 경우는 주변 환경 관련 지표에 대한 점수가 높게 나타났다. 이는 이 두 도서관이 다른 곳에 비해 환경적 요인에 의해 도서관을 찾는 사람이 많다는 걸 의미한다. 실제로 정독도서관은 벚꽃이 아름다운 도서관으로 유명하기 때문에 이 시즌이 되면 사진명소로 꼽혀 방문객이 증가하고, 남산도서관은 주위의 서울N타워 등과 연계하여 장소성을 가지고 있는 곳으로 부각된다. 이러한 점이 해당 도서관이 SNS 상에서 이용자들에게 의한 노출도가 큰 이유이다. 이용자들은 독서를 위한 목적보다도 아름다운 도서관을 배경으로 하는 멋진 사진을 찍거나, 테이트코스의 하나로 도서관을 방문하여 그 후기를 SNS에 올린다. 이는 앞으로 공공도서관의 역할이 단순히 책을 읽고 자료를 찾는 곳에서 벗어나 교육, 문화, 환경적 요소 등 다각화된 측면에서 고려되어야 함을 시사한다.

3.4 서울시 공공도서관 매력도지수에 따른 공간접근성 분석결과

이용자 입장에서 느끼는 체감효과를 정량화시키기 위하여 앞에서 산출한 공공도서관 매력도지수를 2SFCA모형에 적용하여 공간접근성을 분석하였다. 공간 분석단위는 읍면동으로 설정하였으며, 읍면동 중심점에서 공공도서관까지 도로 네트워크 거리를 사용하였다. 임계거리는 10km를 설정하였는데 이는 일반적으로 차로 20분 정도를 감안한 것이다. 실제 공공도서관 건립 시 도시지역에서는 도보로 20분 이내(2km)로 접근이 가능해야 한다는 규정이 있으나 대중교통을 이용할 경우 달라질 수 있다는 예외조항을 적용한 것이다(MCST, 2013).

ArcGIS 10.3에 네트워크 분석방법(Network analysis) 중 하나인 OD cost matrix를 통해 읍면동 중심점부터 공공도서관 거리까지 산출한 다음 10km 이내에 있는 것만 선정하여 분석하였다. 이때 수요량 P_k 는 읍면동 인구수를 적용하였으며, 공급량 S_j 는 앞 절에서 산출한 공공도서관 매력도지수를 적용하였다.

분석결과는 Fig. 3과 같다. 2SFCA의 최종 결과 값으로 공공시설 i에 대한 접근성지수인 A_i 를 natural break를 이용하여 7개의 구간으로 나누어 지도화 시켰다. A_i 는 각 읍면동에서 이용할 수 있는 공공도서관에 대한 접근성을 의미한다. 방문자수 상위 15위내에 공공도서관이 많이 몰려있는 서울 중심부의 접근성이 좋은 편이며, 외곽지역으로 갈수록 접근성이 줄어든다. 특히 은평구 진관동, 금천구 시흥2동 및 삼성동의 경우 A_i 가 0으로 나타나는데, 이는 이 지역을 중심으로 10km 이내에 존재하는 공공도서관이 없음을 의미한다. 따라서 이들 지역에 대한 공공도서관 접근성 개선 마련 정책이 시급한 것으로 보인다. 접근성이 좋은 지역 중 노량진1동이 중심부에서 떨어진 지역임에도 불구하고 높은 결과 값을 나타내는데 이는 도로네트워크 분석 결과 이 지역의 교통 입지적 여건이 좋아서 총 8개 도서관에 10km 이내로 접근이 가능한 것으로 나타났다.

공공도서관의 양적인 면보다 질적 측면이 좋은 경우 공간접근성이 개선된 사례도 분석결과 확인되었다. 노원구와 송파구의 경우 절대적 도서관의 개수는 노원구가 높으나, 매력도지수가 높은 송파도서관이 위치한 송파구의 접근성이 전체적으로 더 높게 나타났다. 이는 공공시설의 질적 확대가 중요하다는 본 연구의 가설을 입증하는 사례로 시설의 물리적 개수만 늘릴 것이 아니라 이용자들이 원하는 서비스를 제공하는 공

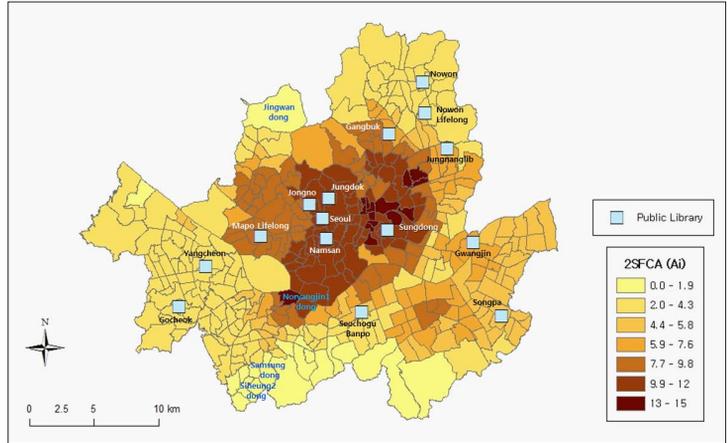


Fig. 3. Result of spatial accessibility to public libraries

공시설을 늘리는 것이 실제 체감효과를 높일 수 있는 좋은 정책 수단으로 활용될 것이다.

4. 결론

국민 중심의 복지를 실현하기 위해 필요한 것은 복지 수요자가 원하는 서비스를 제공하는 것이다. 본 연구는 수요자 입장에서 올린 SNS 데이터를 바탕으로 공공시설 선호요인을 분석하고, 이를 활용한 공공시설 매력도지수를 산출하는 방법을 제안하였다.

서울시 공공도서관 방문자수 상위 15개 도서관을 대상으로 지난 1년간 네이버 블로그 글을 수집하여 분석한 결과, 도서관 선호요인으로 주변환경, 시설 및 이용규모, 문화프로그램, 육아, 장소 및 자료현황 등의 주제가 도출되었다. 이는 기존 연구들에서 데이터의 한계로 인해 주로 내부시설이나 인적자원, 장소수 등으로 한정하여 선호요인을 도출하였던 점과 차별화된다. 공공시설 매력도지수는 이러한 다양한 선호요인을 반영한 지표이다. 각 선호요인별로 정형데이터 혹은 비정형데이터를 이용하여 지표를 구축한 다음 전체 지표에 대한 표준화 및 가중치 부여과정을 거쳐서 매력도지수를 산출하였다. 본 연구에서는 주변환경 및 육아 관련 사항은 비정형 데이터를 활용하여 지표를 구축하였으며, 나머지 요인은 관련 통계자료를 이용해 구축하였다. 산출결과 송파도서관, 정독도서관, 남산도서관 등에서 비교적 매력도가 높게 평가되었다. 송파도서관은 전체적인 요인이 모두 높게 평가되었으며, 특히 육아관련 점수가 높은 것으로 나타났다. 정독도서관 및 남산도서관은 주변환경에 대한 점수가 높게 나타났는데 이는 사람들이 공공도서관을 방문하는 이유가 도서관 내부 시설

보다는 외부환경 및 편의시설 등과 긴밀하게 연관되어 있음을 보여준다.

최종적으로 SNS 데이터를 이용하여 측정된 공공시설 매력도지수를 적용한 공간접근성을 분석하였다. 공간접근성은 수요자의 실제 서비스 체감정도를 측정할 수 있는 지표 중 하나로, 특히 공공시설의 경우 국민의 보편적인 복지라는 차원에서 매우 중요하다. 본 연구에서는 공급요인에 대한 공간접근성을 구할 수 있는 2SFCA 모형을 사용하였으며, 공공시설 매력도지수를 공급요인으로 적용하였다. 분석결과 방문자수 상위 15개 도서관이 몰려있는 서울 중심부에서 접근성이 높은 것으로 나타났으며, 외곽지역으로 갈수록 접근성이 줄어들었다. 또한 공공시설의 질적 개선이 양적증가 보다 중요하다는 점이 확인되었는데, 송파구의 경우 도서관의 개수가 노원구에 비해 적으나 송파도서관의 매력도지수가 높아 전체적인 접근성은 더 높은 것으로 나타났다. 이를 통해 공공시설의 양적인 확대보다는 시설별로 수요자가 원하는 서비스에 대한 질적인 측면을 개선하는 것이 바람직할 것으로 보인다.

본 연구는 서울시 전체 도서관 중 방문자수 상위 15개 도서관만을 대상으로 실험을 한정하였으며, 공간접근성 산출 시 2SFCA 모형의 정교화를 위한 거리조각의 문제 등을 해결하지 못했다는 점에서 한계가 있다. 하지만 이는 추후 연구대상 범위를 넓히고, 최근 연구되어지고 있는 2SFCA의 다양한 수정모형을 적용함으로써 해결할 수 있다. 추후에는 공공시설 매력도 측정을 위해 SNS 데이터 중 텍스트데이터 뿐만 아니라 최근 급격히 늘어나고 있는 이미지나 동영상 자료 등을 통해서도 의견을 수집하는 연구도 필요할 것으로 보인다.

감사의 글

본 연구는 “국토교통부 국토공간정보연구사업 국토공간정보의 빅데이터 관리, 분석 및 서비스 플랫폼 기술개발(18NSIP-B081023-05)과제”의 연구비 지원에 의해 연구되었음.

References

- Ahn, J.S., Kim, L.B., and Park, M.R. (2014), An analysis of variation of spatial accessibility pattern based on 2SFCA : A case study of welfare facilities for the aged in Gumi city, *Journal of the Korean Association of Geographic Information Studies*, Vol. 17, No. 4, pp. 112-128. (in Korean with English abstract)
- Bae, J.H., Son, J.E., and Song, M. (2013), Analysis of twitter for 2012 South Korea presidential election by text mining techniques, *Journal of Intelligence and Information Systems*, Vol. 19, No. 3, pp. 141-156. (in Korean with English abstract)
- Blei, D. (2012), Probabilistic topic models, *Communications of the ACM*, Vol. 55, No. 4, pp. 77-84.
- Chae, I.Y., Lee, Y.M., Yu, K.Y., and Kim, J.Y. (2017), A method for analysis of preferences of places using social media text, *Journal of the Korean Society for Geospatial Information Science*, Vol. 25, No. 4, pp. 55-64. (in Korean with English abstract)
- Cho, D.H., Shin, J.Y., Kim, K.Y., and Lee, G.H. (2010), An analysis of spatial accessibility to public healthcare services in rural areas, *Journal of The Korean Association of Regional Geographers*, Vol. 16, No. 2, pp. 134-153. (in Korean with English abstract)
- Cho, S.H., Kim, B.S., Park, M.S., Lee, G.C., and Kang, P.S. (2017), Extraction of satisfaction factors and evaluation of tourist attractions based on travel sites review comments, *Journal of the Korean Institute of Industrial Engineers*, Vol. 43, No. 1, pp. 62-71. (in Korean with English abstract)
- Choi, H.K. (2009), An analysis of impact factors on the number of users of public libraries, *Journal of the Korean Society for Information Management*, Vol. 26, No. 4, pp. 129-145. (in Korean with English abstract)
- Gao, Y., Wang, S., Padmanabhan, A., Yin, J., and Cao, G. (2018), Mapping spatiotemporal patterns of events using social media: a case study of influenza trends, *International Journal of Geographical Information Science*, Vol. 32, No. 3, pp. 425-449.
- Joseph, A. and Phillips, D. (1984), *Accessibility and Utilization*, Haper and Row, New York, N.Y.
- Kang, A.T. (2016), *A Study on Regional Characteristics on the Stress Sentiment and Topics Extracted from Tweet Data*, Ph.D. dissertation, Ewha University, Seoul, Korea, 183p.
- Kanuganti, S., Sarkar, A.K., and Singh, A.P. (2016), Evaluation of access to health care in rural areas using enhanced two-step floating catchment area method, *Journal of Transport Geography*, Vol. 56, pp. 45-52.
- Kent, J.D., Henry, T., and Capello, J. (2013), Spatial patterns and demographic indicators of effective social media content during the Horseshoe canyon fire of 2012, *Cartography and Geographic Information Science*, Vol. 40, No. 2, pp. 78-89.

- Kim, H.J., Lee, J.G., and Yeo, K.H. (2015), Regional disparities and determinants of spatial accessibility of public libraries in Seoul, *Seoul Studies*, Vol. 16, No. 4, pp. 109-127. (in Korean with English abstract)
- KURC. (2010), *Development of Urban Deprivation and Potential Diagnostic Index*, Research report, Korea Urban Regeneration Center, Daejeon, pp. 73-82.
- Kwon, N.H. and Song, K.J. (2014), A national study explaining the public library use among Korean adults – Examining the influence of individual characteristics, local library inputs and local government investments, *Journal of the Korean Biblia Society for Library And Information Science*, Vol. 25, No. 4, pp. 291-312. (in Korean with English abstract)
- Langford, M., Fry, R., and Higgs, G. (2016), Multi-modal two-step floating catchment area analysis of primary health care accessibility, *Heath & Place*, Vol. 38, pp. 70-81.
- Lee, H.Y. (2003), *Geographic Information Systems*, Bobmunsa, Seoul.
- Lee, H.Y. and Roh, S.C. (2012), *Advanced Statistical Analysis*, Moonwoosa, Gyeonggi.
- Lee, J.M. (2014), *Regional Disparity in the Accessibility of Public Healthcare Facilities, and the Determinants of the Accessibility*, Ph.D. dissertation, Dong Shin University, Jeollanam-do, Korea, 141p.
- Lee, S.I. (2012), The nature and subject matter of spatial interaction studies: a consilient approach to human geography, *Journal of the Association of Korean Geographers*, Vol. 1, No. 1, pp. 137-151. (in Korean with English abstract)
- Lee, T.W. and Hong, T.H. (2017), Integrated sentiment classification model using SVM+MTL: Application to the customer review of online store, *Entrue Journal of Information Technology*, Vol. 16, No. 1, pp. 81-95. (in Korean with English abstract)
- Luo, W. (2004), Using GIS-based floating catchment method to assess areas with shortage of physicians, *Heath & Place*, Vol. 10, pp. 1-11.
- McGrail, M. and Humphreys, J. (2009), Measuring spatial accessibility to primary care in rural areas: Improving the effectiveness of the two-step floating catchment area method, *Applied Geography*, Vol. 29, No. 4, pp. 533-541.
- MCST. (2013), *A Manual on Construction and Management for Public Libraries*, Research report, Ministry of Culture, Sports and Tourism, Sejong, pp. 28-34.
- Palmer, E.S. (1981), The effect of distance on public library use : A literature survey, *Library Research*, Vol. 3, No. 3, pp. 315-334.
- Park, H.Y., Chung, I.H., and Kim, C.J. (2010), A study on the optimal location decision of public service facilities: focused on Paju city, *The Korea Spatial Planning Review*, Vol. 66, pp. 149-168. (in Korean with English abstract)
- Park, J.H., Lee, S.G., and Byeon, J.H. (2017), *A Study on the Optimal Location Selection Criteria of LX Branch*, Research report, Korea Land and Geospatial Informatix Corporation Publication, Jeonju, pp. 9-60.
- Park, J.H., Woo, H.J., and Kim, Y.H. (2017), Measures of spatial accessibility to emergence medical services with a modified three-step floating catchment area model : A case study of the Chungnam province, *Journal of The Korean Association of Regional Geographers*, Vol. 23, No. 2, pp. 388-402. (in Korean with English abstract)
- Sohn, J.Y. and Oh, S.K. (2007), Measuring accessibility of day care centers for the elderly in Seoul using GIS spatial analysis techniques, *Journal of The Korean Association of Regional Geographers*, Vol. 13, No. 5, pp. 576-594. (in Korean with English abstract)
- Statistics Korea. (2015), Understanding statistics index, *Statistics Korea Homepage*, <http://kostat.go.kr/understand/info/> (last date accessed: 14 December 2018).
- Teitz, M.B. (1968), Towards a theory of urban public facility location, *Papers in Regional Science*, Vol. 21, No. 1, pp. 35-51.
- Wan, N., Zou, B., and Sternber, T. (2012), A three-step floating catchment area method for analyzing spatial access to health services, *International Journal of Geographical Information Science*, Vol. 26, No. 6, pp. 1073-1089.
- Won, J.J. and Ahn, K.H. (2010), The effect of locational and facility characteristics on public library use : Focused on public Libraries in Seoul, *Journal of the Architectural Institute of Korea Planning & Design*, Vol. 26, No. 2, pp. 79-86. (in Korean with English abstract)
- Yang, D., Goerge, R., and Mullner, R. (2006), Comparing GIS-based methods of measuring spatial accessibility to health services, *Journal of Medical Systems*, Vol. 30, No. 1, pp. 23-32.