



기업의 공간빅데이터 활용사례

김한국 | (주)비즈지아이에스 분석팀장

황선영 | (주)비즈지아이에스 분석팀 과장

빅데이터 시대에 기업이 원하는 것

약 5억 4,200만 년 전 대부분의 동물군이 출현하여 생물 다양성이 급격하게 증가한 시기가 있었다. 삼엽충의 시대, 고생대가 시작되는 캄브리아기 지층은 생물 진화상의 대사건이었기에 캄브리아기 대폭발이라고도 부른다.

현대인이 하루에 접하는 정보량이 19세기 말~20세기 초 사람들이 평생 접하는 정보량에 육박한다고 한다. 데이터 생산량은 2년마다 2배씩 증가하여 2011년에는 전 세계 디지털 정보량이 1.8조GB, 즉 1.8ZB에 달하는 '제타바이트 시대'로 진입했다.

빅데이터의 정의만 해도 수십 개가 쏟아져 나왔다. 빅데이터란 무엇이며 그 가치에 대한 세미나와 컨퍼런스가 붐을 이뤘고, 데이터의 처리, 저장, 관리 기술들이 쏟아져 나왔다. 미래의 어느 한 시점에서 현재를 바라본다면 이러한 데이터에 대한 관심은 갑작스러운 현상으로 보일 것이다. 마치 캄브리아기 대폭발처럼.

데이터에 기반을 둔 의사결정이 이 시대의 필요조건처럼 여겨지면서 폭발적으로 증가하고 있는 데이터들 중에서 어떤 데이터를 선택해야 하며, 어떻게 활용할 것인가에 대한 관심이 높아졌다. 정확히 말하자면 기업이 직면한 문제점을 어떤 근거에 입각해서 어떻게 해결할 것인가에 대한 관심이다. 빅데이터 시대에도 기업에서 중요하게 여기는 것은 데이터를 처리하는 기술이 아니라 여전히 데이터가 주는 시사점을 충분히 이해하는 것이다.

공간정보와 빅데이터, 그리고 활용

공간정보 분야는 오래전부터 많은 양의 데이터를 국가적 차원에서 체계적으로 축적하고 있다. 그리고 이 정보는 전자지도상에 위치를 나타낼 수 있는 위치정보를 포함하고 있다. 이는 창의적인 상상력과 실질적인 실행능력의 여부에 따라 매우 가치 있는 정보로 활용될 수 있음을 의미한다. 하지만 역설적으로 국내 공간정보산업은 그 활용에 매우 취약한 한계를 가지고 있다. 그 이유는 쌓여있는 정보를 활용하여 이슈를 해결하려는 노력보다는 새로운 공간정보를 구축하거나 기존의 것을 개선하는 것이 익숙했기 때문이다.

최근 빅데이터 이슈의 핵심은 활용이다. 근사한 언어로 표현하자면, 보다 합리적이고 효율적인 의사결정을 지원할 수 있도록 데이터를 활용하지는 것이다. 이러한 측면에서 공간정보 분야 빅데이터 사업들의 내용을 살펴보면 아쉬움이 남는다. 대부분 공공기관이 '어떻게 활용하여 무엇을 개선할 것인가'라는 핵심적인 질문보다는 데이터의 크기, 형태와 같은 형식적인 구분에 관심을 더 두고 있다.

현재 공간정보 분야는 데이터를 뛰어난 수준으로 활용하기에는 수요자와 공급자 모두 준비가 부족하다. 데이터의 뛰어난 활용과 모범적 사례는 단기간에 축적되지 않는다. 다양한 시도, 실패의 경험과 반성 등이 꾸준히 이뤄지는 시간이 필요하다. 즉, 일정 기간 '경험을 축적하는 세월'이 필요하다는 것이다. 이제부터라도 공간정보 분야의 수요자와 공급자는 데이터의 실질적인 활용을 고민하고 하나씩 시도할 필

요가 있다.

국내 공간정보 분야를 공공과 민간(비즈니스) 분야로 구분하면 환경이 조금 다르다. 다행히 민간 분야에서는 10년 넘게 다양한 시도가 이루어져 왔고, 실패와 성공의 사례가 많이 축적되어 있다. 그래서 공간정보의 활용을 얘기할 때 빠지지 않고 등장하는 것이 민간 분야 사례다. 이러한 민간 분야 사례를 참고하고 여기에서 아이디어를 얻는 것도 경험을 간접적으로 축적하는 좋은 방법일 것이다.

객관적인 정보가 필요한 시대

-공간데이터를 활용한 시장성 평가

최근 가맹점의 피해 예방을 위해 「가맹사업법」이 개정됐다. 가맹점 수 100개 이상의 가맹본부는 가맹 계약 시 영업지역 설정을 의무화하고, 계약기간 동안 영업지역 내에는 가맹점이나 직영점을 추가적으로 출점할 수 없다. 또한 예상 매출액 자료와 산출 근거를 서면으로 제공하고, 영업 개시일로부터 1년간 예상되는 연간 매출액을 최저액과 최고액 범위로 제공해야 한다. 그동안 현업의 직관에 의존해왔던 상권 분석을 이제는 객관적인 자료를 기반으로 근거 기준을 마련해야 하는 법적 의무가 지워진 셈이다.

사실 「가맹사업법」이 개정되기 이전에도 많은 기업이 시장성을 판단하기 위해 매출을 추정해왔다. 과거 데이터 취득이 어려웠을 때 한 기업은 신규매장 입지 타당성을 조사하기 위해 2명의 직원이 3주에 걸쳐 인근 지역의 등기부등본 약 1만여 장을 발급받아 가맹 고객이 얼마나 될지 짐작하기도 했다. 외국계 담배회사는 지역별



자사 제품의 수요를 확인하고 신규 점포를 확보하기 위해 쓰레기통의 담배꽂초 개수를 세기도 했다. 형식은 다르지만 모두 데이터를 통해 시장성을 객관적으로 평가하기 위한 일련의 노력이었다.

이제는 민간기업에서도 고객 데이터가 일정 규모 이상 축적되고, 상권의 규모를 가늠할 수 있는 공간데이터를 확보할 수 있게 되면서 좀 더 타당성 있는 절차를 통해 손쉽게 시장성을 판단하고 있다.

A기업이 신규 출점지의 시장성을 평가하기 위해 시작한 첫 번째 작업은 기존 고객 분석이었다. 3년 동안 40만 명의 고객 정보를 축적한 A기업은 각 매장에 방문하는 고객의 주소 정보를 지도상의 X, Y좌표로 변환했다. 고객의 위치와 매장과의 거리 분석을 통해 매출이 발생하는 공간상의 범위를 파악했다. 또한 고객의 위치는 주거형태 및 소득 데이터와 매칭될 수 있어, 고객의 특징을 좀 더 다양하게 파악할 수 있었다.

고객 데이터와 소득, 아파트 등과 같은 인문사회 공간데이터의 결합을 통해 매출이 우수한 매장들의 배후지 여건을 파악했다. 매장 인근 빌라에 사는 중장년층의 매출비중이 높았으며, 인근에 전통시장과 소규모 은행이 있는 매장의 매출이 높았다. 이러한 분석 내용을 기초로 하여 전국을 작은 격자 단위로 세분화해서 매출 요인들을 변수화하고 회귀분석을 통해 하나의 모델을 생성했다.

기존 매장 고객들의 매출 분포와 회귀분석을 통해 추정한 매출 분포를 지도화해서 검증했다. 대부분 A사에서 가맹점을 모집하는 지역과 일치하여 A사가 판단하는 시장성에 근접하게 모델링된 것으로 판단했다. 한편 현업에서는 시장성이 좋다고 판단했지만 모델에서는 그렇지 않은 지역이 있었다. 아직은 잠재 고객이 형성되지 않은 택지개발 예정지구였다. 이러한 곳은 모델링보다는 택지개발에 대한 데이터를 지도에 표시하는 것이 의사결정에 도움이 되었다.

자동차 관련 용품을 생산하는 B기업은 가맹점과 일반 소매점에 제품을 판매하고 있다. 「가맹사업법」 개정으로 신규 가맹점에 영업지역과 예상 매출액을 제공해야만 했다. 이에 GIS를 도입한 B기업은 앞의 사례와 마찬가지로 고객 데이터를 활용하기로 했다. 하지만 고객 데이터에는 주소 정보가 제대로 입력되어 있지 않아 활용할 수 있는 데이터는 10%도 채 되지 않았다. 어떤 매장은 고객 주소를 잘 관리했지만 대부분의 매장에서는 주소 관리가 소홀했던 것이다. 블록에 한두 명의 고객만이 맵핑되기 때문에 회귀식을 도출하기 어려운 조건이었다.

B기업 담당자에게 현재의 조건을 충분히 설명하고 다른 대안 모델들을 함께 고민했다. 결국 현업과의 논의를 통해 잠재 소비자가 해당 점포를 선택할 확률을 계

〈그림 1〉 공간데이터를 활용한 매출추정 사례



산하는 중력모형을 적용하기로 결정했다. 고객의 위치 정보가 풍부하지 않은 조건에서, 배후지 인구 규모와 경쟁점의 수, 그리고 거리 조건만으로 시장성을 평가할 수 있는 방법이다.

여기에 B기업의 매출을 구성하는 조건을 반영했다. 제한적인 조건이기는 하지만 기존 고객의 위치를 활용하기로 했다. 주거지에 위치한 고객과 직장지역에 위치한 고객 구성비율을 계산하고 잠재 고객으로 삼았다. 또한 출점 담당자들의 현장 경험을 토대로 입지유형을 구분했다. 입지유형에 따라서 매출이 발생할 것으로 예상되는 영역이 조정되는 것이다. 자동차 관련 용품을 판매하는 매장인만큼 인근지역에 잠재 고객이 드물더라도, 통행량이 많고 매장으로 유입될 수

있는 조건이 갖춰진 입지는 매출 영역이 좀 더 확장되는 식이다.

기존 매장의 실제 매출액과 추정 매출액을 검증하고 현업의 의견을 반영하여 모델을 수정하는 과정을 거쳐서 B사에 맞는 영업지역과 매출 추정모형을 만들었다. 기업의 내부 데이터(고객 데이터)와 세대 및 인구 데이터, 직장인구 데이터, 주택유형 데이터 등 외부 데이터의 결합, 기업의 현실에 맞는 모델의 선택 그리고 현업의 의견이 조화롭게 적용된 사례였다.

루이지 살바네쉬(Luigi Salvaneschi)는 매출 추정의 이론적 방법들을 설명하면서 “이는 마치 진공 속에서 축구공의 속도와 방향을 계산하는 것에 대하여 허리케인 동안 축구공의 속도와 방향을



계산하는 것과 같다”고 했다. 매출추정에 국한된 것만은 아니다. 그 어떤 훌륭한 데이터와 모델이 있더라도 기업마다 처해 있는 특수한 환경과 현업에서 경험한 다양한 요인들을 반영하지 못한다면, 데이터와 모델만으로 중요한 시사점을 얻기는 어려울 것이다.

공간데이터를 활용한 지역 마케팅

기업이 원하는 것은 현명하게 고객에게 접근하는 것이다. 입지도 고객에게 다가가기 위한 마케팅의 일환이다. 그러나 물리적인 입지만으로는 고객이 매장에 발을 들여놓게 하기는 어렵다.

기업은 고객이 누구이며 이들이 어떻게 행동하고 어디에서 찾을 수 있을 것인지 고민한다. 지리적으로 광역적인 스케일이나 넓은 의미에서의 고객보다는 개별적이고 세분화된 그룹으로 마케팅 활동을 펼쳐 타깃 확률을 높이고 싶어 한다. 불특정 다수를 대상으로 하는 매스마케팅(mass-marketing) 대신 고객과 지역에 대한 데이터를 활용해서 지역마다 차별화된 마케팅 의사결정의 수립을 원한다.

의류기업 C는 점점 치열해지고 있는 패션산업 환경에서 매출 하락을 우려하고 있었다. 각 가맹점주에게 마케팅 지원을 하고 있었던 C기업은 각 매장에 적합한 마케팅 방법을 적용해서 매출을 신장시킬 수 있는 방안을 고민했다.

먼저 C기업 고객의 소비 패턴을 살펴볼 필요가 있었다. 어느 한 매장에서 상품을 구매한 고객들을 지도에 뿌려봤다. 매장과 거리 상관없이 재방문율은 낮았고, 고객 분포는 인구 분포와 유사했다. 고객이 해당 지역에 비해 특정한 속성을 가진 그룹은 아니라는 의미였다. 그러나 매출 분포는 매장 인근지역 중에서도 상대적으로 소득이 낮은 곳에서 오히려 매출 비중이 높았다.

이번에는 구매한 상품을 맵핑했다. 상대적으로 소득이 낮은 곳에서는 고가 제품인 점퍼를 많이 구매했고 소득이 높은 곳에서는 저가 제품인 티셔츠를 많이 구매했다. 일반적인 소매점의 구매 패턴과는 다른 결과였다. 이러한 분석 결과를 기반으로 소득 수준이 높지만 저가 제품을 주로 구매하던 지역의 고객 재방문율을 높이고 좀 더 고가의 상품을 구매하도록 유도하기로 했다. 또 기존 고객 외에도 해당 지역의 잠재 고객을 신규 고객화하기로 했다.

프로모션 타깃으로 선정한 지역의 건물 데이터에서 주소를 추출해 쿠폰이 동봉된 DM(Direct Mail)을 발송했다. 프로모션 기간이 종료되면 반응률을 체크해서 해당지역의 고객화 가능성을 다시 판단하게 된다. 지속적으로 신규 고객 창출

〈그림 2〉 티셔츠 구매 고객 분포(좌)와 점퍼 구매 고객 분포(우)



에 자원을 투입할 것인지 기존 고객을 충성 고객화하는 것에 중점을 둘 것인지 판단하는 피드백 과정을 거치는 것이다.

그러나 아직 고객이 없는 신규 매장은 어떻게 지역 마케팅 활동을 펼칠 수 있을까? 영국 체셔(Cheshire) 지방에 주유소를 오픈했다. 판촉활동을 위해 자동차로 약 15분 거리 내에 거주하는 고객들에게 약 5만여 장의 전단을 배포했다. 그러나 많은 제작비가 들어간 전단은 제대로 정의되지 않은 고객들에게 전달됐고 결국 직접적인 이윤으로 돌아오지 않았다.

따라서 최대 이윤이 발생할 수 있는 최적의 전단지 배포 구역을 찾는 것이 중요하다. 이를 위해 새로운 주유소의 고객이 될만한 사람이 누구인지, 또 그러한 사람의 비중이 높은 곳은 어디인지 찾아야 했다. 하지만 아직 해당 주유소에 방문한 고객이 없는 만큼 고객에 대한 정보를 알 수 있는 방법이 없었다. 이때 사용한 데이터가 매년 2만 5천 명의 패널조사를 통해 만들어진 Target Group Index였다. 여기에는 주유소 방문과 관련된 항목이 있기 때문에 주유소에서의 소비가 많은 지역 구분이 가능했던 것이다. 해당 주유소는 가구당 주유비 지출이 높은 지역의 5만 가구에

집중적인 판촉활동을 벌였다. 차량으로 15분 이내의 모든 가구에 전단을 배포하는 일부 비용으로 더 많은 고객을 유치할 수 있었다.

우리나라에서도 세부 공간단위에서 지역을 파악할 수 있는 데이터들이 공개되면서 지역 마케팅이 가능한 조건이 만들어지고 있다. 인구, 가구, 주택, 산업 등 네 가지 대분류에 성/연령별 인구, 종교별 인구, 점유형태별 가구, 유형별 주택, 산업분류별 사업체 수와 종사자 등 열두 가지 중분류로 구분된 통계청의 집계구 데이터가 대표적이다. 기존 지역통계의 기본단위인 행정동보다 약 20배 이상 정교하며 가장 공신력 있는 데이터다. 활용하는 데 별도의 비용이 들어가지 않아 GIS를 도입한 기업들이 가장 손쉽게 활용할 수 있는 데이터다.

카드사에서도 매일, 매시간 발생하는 카드 사용 기록을 분석해 새로운 데이터를 생성하고 있다. 단순히 카드 사용 금액과 건수를 집계하는 것이 아니라 고객들이 결제한 상점의 속성을 통해 그 사람의 현재 상태를 추정하는 것이다. 육아용품 전문점의 결제 빈도가 높은 사람은 아이가 있을 것이고, 자동차 정비소에서 카드로 결제한 사람은 최소한 1대 이상의 자동차를 보유하



고 있을 것이라는 추론이 가능하다. 이러한 데이터들은 앞서 설명한 주유소 사례와 같이 고객 데이터가 없는 조건에서의 지역 마케팅을 가능하게 한다. 데이터들은 개인정보 유출의 위험이 없도록 합법적인 형태로 공간단위별로 합역해서 제공하고 있다.

비정형 공간데이터의 사내 지식화

출점 담당자는 경험을 통해 관리지역에 대한 해박한 지식을 쌓게 된다. 업력이 오래될수록 자신만의 노하우로 발전하게 된다. 관리지역의 지형, 교통, 주거형태, 상업시설 등 물리적 공간뿐만 아니라 지역 주민의 연령 구성, 인구 동태, 유동인구, 소비 패턴, 경쟁점 점주들의 영업력 등을 자신만의 데이터베이스로 만들어간다. 하지만 해당 지역에 대한 경험적 지식이 많은 출점 담당자는 첫 번째 스카우트 대상이 된다. 이들이 이직을 하게 되면 기업 입장에서는 해당 지역에 대한 공간 빅데이터 웨어하우스가 사라지게 되는 셈이다. 이런 문제점들을 경험한 기업은 현장 업무의 노하우를 사내 지식 역량으로 축적하기 시작했다. 발로 뛰면서 보고 듣고 경험한 지역에 대한 현황을 공간데이터화하기로 한 것이다.

현장 업무는 늘 바쁘게 돌아가기 때문에 일련의 형식을 갖춘 보고서보다는 간단한 메모와 사진을 전송하게 했다. 외부 현장답사에서 얻은 정보를 스마트폰으로 간략히 작성하고 전송하면 좌표와 함께 기업 내부 서버에 저장된다. 현장에서 전송한 정보들은 개인 PC와 스마트폰을 통해 지도상에서 확인이 가능하고 다른 사람들과 공유도 가능하다. 기업 내부에서는 현장의 정보를 실시간으로 받아서 의사결정을 내릴 수 있을 뿐만 아니라 지역 정보의 축적을 통해 현업 담당자의 변동에도 안정적으로 대응할 수 있다.

기업 스스로 자사에 필요한 공간정보를 생산하고 축적하는 체계를 갖추기 시작했다. 스스로가 공간정보의 생산자이자 소비자가 되었다. 철저하게 문제해결에 목적을 둔 정보자산의 창출이었다.

데이터 자체보다 활용

그동안 공간데이터 분야에서의 큰손은 국가였다. 지표면과 관련된 모든 데이터를 생성하고 관리하는 곳이기 때문이다. 그런 국가가 브이월드(Vworld)라는 이름으로 지도를 무상으로 서비스하기 시작했다. 민간 입장에서는 핵심적인 데이터가

아닌데도 불구하고 큰 비용을 지불해야 했던 지도가 무상으로 제공되면서 GIS 도입을 적극적으로 검토할 수 있게 되었다. 이에 따라 공간데이터에 대한 수요도 많아졌다. 전통적인 GIS에서는 지형, 도로 등 물리적 환경을 나타내는 데이터가 가장 중요했지만, 기업에 있어 가장 중요한 데이터는 고객이다. 기업은 고객을 잘 파악할 수 있는 데이터를 필요로 한다.

이제 민간기업들이 자사가 경험한 고객의 특성을 기반으로 고객과 상권을 더 잘 이해할 수 있도록 하는 공간빅데이터를 만들어 상품화하고 있다. 공간데이터의 생산이 국가에서부터 민간기업으로 점차 확산되면서 데이터는 좀 더 다양하고 복잡한 형태를 띠게 되었다.

나무는 저마다의 특성이 있어 옮겨 심을 때는 살던 지대의 높이나 방향, 깊이를 맞춰줘야 한다고 한다. 데이터도 마찬가지로 기업 환경에 맞는 형태로 옮겨 심어야 한다. 특히 다른 데이터와의 결합을 통해서 의미를 갖게 되는 공간데이터는 각 기업의 의사결정 목적에 따라 다른 형태로 가공해서 내부에 안착시켜야 한다. 이를 위해서는 데이터를 갖고 있는 곳과 활용하는 곳을 중계하는 사람이 필요하다. 그러한 사람은 뛰어난 빅데이터 처리기술을 갖고 있기보다는, 해당 비즈니스를 잘 이해하여 고객의 언어로 해석해줄 수 있는 분석가일 것이다. 데이터는 그것이 주는 의미를 이해할 수 있을 때 비로소 유용성을 가지기 때문이다.

국내 민간기업이 공간데이터를 활용하기 시작한 지 10여 년이 훌쩍 지났다. 그럼에도 여전히 대부분의 기업들은 공간데이터가 마치 새로운 형식의 데이터인 것처럼 낯설고 어려워한다.

빅데이터에 대한 관심이 급격히 증가하고 있는 지금, 공간데이터에 대한 민간기업의 수요에 현실적으로 대응하고 길안내를 해서 공간정보 산업의 발판을 마련해야 할 시기다. 🍷

참고문헌

루이지 살바네쉬, 2008. 상업용부동산 입지: 이론과 실제. 서울 : 부연사.