

상권정보와 설문정보의 연계 분석 방법

A Linked Analysis Method between Commercial district Information and Survey Information

이원철¹ · 강만수² · 김진호^{1*}

강원대학교 대학원 컴퓨터학과¹, 서울신용보증재단 소상공인정책연구센터²

요약

소상공인은 우리나라 서민경제의 중요한 부분을 담당하고 있으나, 과다경쟁, 수익성 악화, 생활밀착형 업종 편중 등의 어려움에 직면하고 있다. 이를 위해, 정부는 소상공인을 위한 상권분석서비스를 제공하고 있다. 그러나 다양한 기관에서 제공받은 데이터는 표준화되지 않고, 제한된 데이터로 서비스의 구성에 한계가 있다. 본 논문에서는 데이터의 정합성 문제를 해결하고, 데이터 분석서비스 확장을 위해 설문정보와 상권정보를 연계하여 분석하는 방법을 제안한다. 제안하는 연계 방법은 상권정보와 설문정보를 업종과 영역을 기준으로 동일 영역 연계, 설문 정보의 개별업체 중심 연계, 설문정보의 작은 영역과 상권의 큰 영역 연계의 세 가지 방법이다. 연계된 상권정보와 설문정보는 다양한 방법으로 활용되거나 분석서비스를 확대할 수 있다. 본 논문은 기존 상권분석 서비스의 한계를 설문정보로 극복할 수 있는 방법을 제안하여 소상공인에게 필요한 상권분석 서비스를 확대할 수 있는 기반을 마련하였다.

- 중심어 : 소상공인, 설문조사시스템, 설문정보, 상권정보, 연계 분석

Abstract

In Korea, micro-enterprises are in charge of an important part of the common people's economy, but face difficulties such as excessive competition, deteriorating profitability, and concentration of life-oriented industries. In order to solve this problem, the government is providing commercial district analysis services for micro-enterprises. However, the data provided by various organizations is not standardized, and there is a limit to the composition of the service with limited data. In this paper, we propose a method of solving the data consistency problem and linking and analyzing between questionnaire information and commercial district information to expand the data analysis service. The proposed linking methods are three methods: linking the commercial area information and questionnaire information in the same area based on the type of business and area, linking the survey information centered on individual micro-enterprise, and linking a small area of questionnaire information with a large area of commercial district information. The linked commercial district information and questionnaire information can be used in various ways or expanded analysis services. This proposed a method to overcome the limitations of existing commercial district analysis services with questionnaire information and lay the foundation for expanding the commercial district analysis services necessary for micro-enterprises.

- Keyword : Micro-enterprises, Survey System, Survey information, Commercial District Information, Linked Analysis

I. 서론

규모가 작은 기업이나 생업을 위한 업종에 종사하는 자영업자를 소상공인이라 한다. 소상공인은 도·소매업, 음식점, 숙박업, 서비스업의 경우 상시근로자 5인 미만으로 운영하고 있거나, 광업, 제조업, 건설업 및 운수업의 경우 상시근로자 10인 미만의 구성된 사업자이다. 소상공인 사업체는 전체 사업체 354만 5천 개중 306만 3천개가 소상공인 사업체로 86.4%를 차지하고 있어 서민 경제에 중요한 부분을 담당하고 있다. 그러나 소강공인 간 과다경쟁으로 인한 수익성 악화, 생활밀착형 업종 편중, 소상공인 고령화, 대기업의 자영업 영역에 대한 진출로 경쟁력을 상실하여 폐업으로 이어지고 있다[1].

정부는 소상공인 자영업자를 위해 전담조직 및 소상공인기본법을 제정하였고, 상가임대료 인상률 상한선을 인하, 소상공인·자영업자 카드수수료 인하, 자영업자 고용보험지원대상 확대, 소상공인 간 협업과 온라인 진출지원, 골목상권 육성화 방안 등 다양한 정책으로 지원을 하고 있다. 특히, 준비된 창업을 유도하기 위해 창업 전 체계적 교육을 강화하고, 상권정보시스템 활용을 필수화하고 있다[2].

정부기관에서는 예비창업자 및 자영업자에게 상권정보를 제공하기 위해 노력하고 있다. 소상공인시장진흥재단은 상권정보시스템을 2006년에 구축하여 서비스하고 있고[3], 서울시는 2016년부터 독자적으로 우리마을가게 상권분석서비스를 구축하여 서울시 소상공인 및 공무원을 대상으로 서비스하고 있다[4]. 그리고 경기도는 2018년에 상권영향분석시스템을 구축하여 경기도민에게 제공하는 있다[5].

이와 같은 상권정보시스템은 다양한 기관에서 보유하고 있는 데이터를 제공받아 가공하여 상권분석서비스에 이용하고 있다. 각 기관에서 제공받은 데이터는 표준화되어 있지 않고, 데이

터 사이즈도 다르게 구성이 되어 있어, 정제 및 변환작업을 수행하여 분석 가능한 데이터로 만들고 있다. 그리고 제공되는 데이터가 제한되어 소상공인을 위한 분석 서비스 제공에도 한계가 있다.

본 논문에서는 우리마을가게 상권서비스 시스템을 중심으로 다양한 기관에서 제공받은 데이터를 분석 가능한 형태로 가공하는 과정을 설명하고, 소상공인을 위한 다양한 정책에 활용할 수 있는 데이터 분석을 위해 부족한 데이터를 설문을 통해 수집하는 절차를 설명한다. 수집된 설문정보를 상권정보와 연계하여 분석하기 위해 설문정보와 상권정보를 연계하여 방법을 제안한다.

본 논문의 구성은 다음과 같다. 2장에서는 현재 운영되고 있는 우리마을가게 상권서비스의 데이터의 구성과 분석 가능한 형태로 가공하는 방법에 대해서 설명하고, 설문과 연계를 위한 설문 데이터의 구성에 대해서 살펴본다. 3장에서는 상권정보와 설문정보를 연계하여 분석하기 위한 상권정보와 설문정보의 연계 방법과 활용방법에 대하여 설명한다. 4장에서는 연계된 상권정보와 설문정보를 분석 방법을 적용하여 분석한 결과를 보인다. 마지막으로 5장에서는 결론과 향후 연구 방향을 제시한다.

II. 관련 연구

본 절에서는 우리마을가게 상권서비스 제공을 위해 다양한 기관에서 제공받은 데이터에 대해서 설명하고, 이 데이터를 분석 가능한 형태로 가공하는 방법에 대해서 설명한다. 그리고 상권정보에 설문정보를 연계하기 위해 필요한 설문정보의 특징에 대해서 설명한다.

2.1 상권서비스를 위한 원천 데이터 구성

우리마을가게 상권서비스는 소상공인관련 데이터를 다양한 기관에서 제공받는다. 제공기관으로는 서울시, 카드사, 이동통신사, 건강보험관리공단 등이 있다. <표 1>은 우리마을가게 상권서비스를 위해 제공받은 데이터명, 세부항목, 제공기관, 데이터 업데이트 주기, 데이터의 공간영역 사이즈 정보이다.

<표 1>에서와 같이 우리마을가게 상권서비스는 다양한 기관에서 받은 데이터를 가지고 정제 및 변환작업을 거쳐 상권데이터를 만든다. 제공받는 데이터는 시간과 공간 기준으로 특성을 구분할 수 있다. 시간 기준으로는 년, 분기, 월, 요일, 시간대가 있고, 공간 기준으로는 좌표, 그리드 50M, 집계구, 폴리곤 등이 있다. 이와 같은 원천 데이터는 데이터 크기가 다르기 때문에 데이터 분석에 적합하지 않다. 따라서 우리마을가

게 상권서비스의 시간 기준으로 월을 사용하여 데이터를 구성하고, 분기, 년 기준으로 재집계를 하였고, 공간은 블록기준으로 자료를 작성하여 행정동, 구, 시의 영역으로 재집계하였다.

2.2 데이터를 분석 가능한 형태로 가공하는 방법

본 절에서는 기관에서 제공받은 데이터를 분석 가능한 형태로 정제 및 변환하는 과정을 소개한다. 먼저 통계청의 년 매출정보를 기반으로 카드사의 월매출 실적을 이용하여 월매출액을 추정하는 방법에 대하여 설명하고, 둘째, 생활인구를 기반으로 건물단위상준인구와 길단위상준인구를 추정하는 방법을 설명한다. 마지막으로 골목상권 영역을 구축하는 방법에 대해서 설명한다.

<표 1> 우리마을가게 상권서비스 제공 데이터 구성 예

데이터명	세부항목(배분기준, 구성방법)	제공기관	주기	데이터크기
상가/업소	사업자번호, 전화번호, 우편번호, 프랜차이즈유무, 가맹점번호, 업종정보, 지번주소, 개업일, 폐업일	서울시	월	좌표계
카드매출/소비	매출액합계, 매출건수 합계, 카드사 업종별, 분류기준(성별/연령대별/시간대별/요일별 거래패턴)	BC/KB/SH 카드사	월	블록
매출정보	업체별 년 매출정보	통계청	년	좌표계
추정매출	카드사로부터 제공받은 매출데이터로 전체매출 추정	서울시· 카드사	월	블록
생활인구	서울시가 보유한 빅데이터와 KT의 통신데이터로 특정한 ‘특정 시점’에 ‘서울의 특정 지역’에 ‘존재’하는 모든 인구	서울시·KT	월	집계구
길단위 상준인구	상준인구수, 분류기준(년월, 시간대, 요일, 성별 등) 구성방법 : 서울시 생활인구를 길단위로 배분	서울신용보 증재단	요일	그리드 50m
건물단위 상준인구	건물인구수, 분류기준(년월, 시간대, 요일, 성별 등) 구성방법 : 서울시 생활인구를 건물단위로 배분	서울신용보 증재단	분기	그리드 50m
교통카드	이용수단, 승하차 일시, 승하차 정거장ID, 승하차 금액 구성방법 : 서울시 교통정책과에서 제공하는 교통카드 이용내역	스마트 카드사	월	좌표계
주거인구	행정구역별 주민등록 통계 데이터를 건물단위별 가구 수 및 성별/연령대별 추정한 인구수	서울시	년	좌표계
직장인구	국민건강보험공단 가입자 정보를 50m 그리드 단위로 집계한 성/연령별 인구수	건강보 험공단	반기	그리드 50M

2.2.1 월매출액 추정

월매출액 추정을 위해 제공받은 자료는 통계청 기업등록부(business register)의 업체별 연 매출정보와 카드 3사(신한/BC/국민)로부터 받은 가맹점과 매출관련 정보이다. <표 2>는 통계청에서 제공되는 기업등록부 정보로 연도별, 블록별, 표준업종코드 기준으로 제공되는 매출액 정보이다.

<표 3>은 카드사로부터 제공받은 가맹점과 매출관련 정보이다. 카드 3사에서 제공된 데이터는 월별 매출금액과 매출건수의 합이다. 그리고 매출 분석을 확대할 수 있도록 연령대별, 시간대별, 요일별 항목을 추가로 제공하고 있다.

제공된 데이터로 월별매출 추정을 위해 단계별로 다음과 같은 전처리 과정을 거친다.

Step 1 : 통계청 데이터 전처리 과정

- ① 표준 업종코드를 생활밀접 업종코드와 매핑
- ② 블록과 생활밀접 업종코드를 기준으로 업체 수의 합, 매출 합 계산

Step 2 : 카드사 데이터 전처리 과정

- ① 카드사별로 제공된 업종코드를 생활밀접 업종코드와 매핑
- ② 카드사별 생활밀접 업종코드를 기준으로 매출액의 합계 계산

Step 3 : 통계청의 년 매출액과 카드사별, 블록별, 생활밀접업종별, 월 매출을 기반으로 월 매출액 추정

- ① 카드사별로 보정비용을 계산
 보정금액 : (카드승인금액 합계 ÷ 매출금액) × 100
 ※ 보정금액 계산 기준 : 연도별, (블록별/행정동별/구별), 생활밀접 업종별 합계
- ② 카드사별 보정금액을 20분위로 나누어 outlier를 제거
 - 분위별로 생성된 값 중 1분위와 20분위는 outlier로 판단하여 1분위는 2분로 계산하고, 20분위는 19분로 계산
- ③ 카드사별로 보정비용을 기반으로 월별 매

<표 2> 통계청 기업등록부 데이터

제공기관	항목	주요 내용
통계청 (BR정보)	시간기준	연도별
	영역기준	자치구/행정동/블록
	업종기준	표준업종코드
	제공데이터	매출액 정보

<표 3> 카드회사 제공 데이터

제공기관	항목	주요 내용
SH카드 BC카드 국민카드	영역기준	서울시 기준 6만 8천여개의 블록
	시간기준	매출 발생 년월
	업종기준	카드사 업종코드 / 표준 업종코드
	제공데이터	해당월의 매출금액의 합 및 매출건수의 합
	개인/법인 구분	개인/법인 구분 매출건수 및 매출금액
	연령대별 구분	10대, 20대, 30대, 40대, 50대, 60대이상 매출 건수 및 매출 금액
	시간대별 구분	시간대(6개의 시간대구분:0시-4시)별 매출건수 및 금액
	요일별 구분	요일별 매출건수 및 매출금액
	가맹점수	업종별/블록별 가맹점수

〈표 4〉 카드회사별 분위정보

분위수	BC		KB		SH	
	하위	상위	하위	상위	하위	상위
1	-61.6	0.75	0	0.705	0	0.762
2	0.75	2.211	0.705	1.945	0.762	2.102
3	2.212	3.71	1.945	3.193	2.102	3.383
·	·	·	·	·	·	·
·	·	·	·	·	·	·
17	25.452	28.478	19.543	21.561	18.426	20
18	28.479	33.586	21.562	25.069	20	22.277
19	33.586	51.695	25.07	39.595	22.277	29.287
20	51.706	292770	39.6	94073.93	29.288	6352055

출금액을 추정

카드사별 월 매출 금액 = (카드사별 승인 금액 합계 ÷ 보정금액) × 100

월 매출 금액 = SUM(카드사별 월 매출 금액) / COUNT(카드사별 월 매출 금액)

※ 매출금액 계산 기준 : 월별, 블록별, 생활밀집 업종별 합계

※ 블록별 보정금액이 없는 경우 블록이 포함된 행정동의 보정금액을 사용

2.2.2 건물단위 상존인구와 길단위 상존인구 추정

건물단위 상존인구와 길단위 상존인구를 추정하기 위하여 서울시와 KT가 공동으로 개발한 생활인구 데이터를 사용한다. 생활인구란 서울에 거주하거나 업무, 관광, 쇼핑 등 일시적으로 서울을 찾아 행정수요를 유발하는 인구를 포함한 개념이다. 서울시는 보유한 다양한 공공빅데이터와 KT의 통신데이터를 이용하여 서울에서 실제로 생활하는 인구의 데이터를 생성한다. 생활 인구는 연령, 시간, 남녀를 구분하여 집계구(19,1153) 단위로 제공한다. 그러나 우리마을가게 상권서비스는 블록(6.8만) 단위로 서비스되고 있어 서울시 상권서비스에 적용하기 어렵다. 이를 위해 먼저 집계구별로 되어 있는 생활 인

구를 더 작은 영역인 50m*50m 단위의 셀로 배분하고, 배분된 셀을 다시 건물단위로 또는 길 단위로 배분하여 서비스 한다. 처리과정은 다음과 같다.

Step 1 : 집계구 단위 생활인구를 그리드 단위로 배분

- ① 50m*50m으로 구성된 셀로 구성된 그리드를 구성
- ② 각 그리드에 대한 중심 좌표를 계산
- ③ 각각 집계구에 포함된 그리드의 중심좌표 개수를 계산
- ④ 각각 집계구의 인원을 그리드 중심좌표 개수(n)로 나누어 균등 배분

Step 2 : 그리드 별 생활인구를 건물단위로 배분

- ① 각 그리드에 존재하는 건물의 면적을 계산
 - ※ 2개 이상의 그리드에 접하는 건물은 속하는 그리드별로 분할
- ② 그리드별 생활인구를 그리드 내에 있는 건물의 연면적 비율에 비례하여 건물별로 인구수 배분
- ③ 2개 이상의 그리드에 접하는 건물은 건물에 속한 그리드별로 배분된 생활인구를 합하여 해당 건물의 상존인구로 계산
 - ※ 건물단위로 만들어진 건물단위 상존인

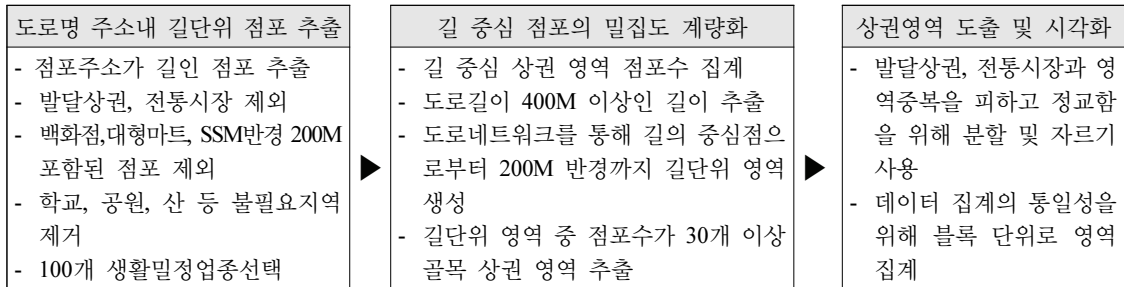
- 구를 구역, 행정동, 구 단위로 집계
- Step 3 : 그리드 별 생활인구를 길단위로 배분
- ① 각각의 길(도로형태) 영역이 포함된 그리드별 인구를 길에 배정
 - ② 각각의 길 단위로 배정된 그리드를 길 단위로 집계함
- ※ 길 단위는 여러 블록을 포함하고 있어 블록별로 분석 시 오차가 크게 발생하여 행정동 별로 분석

2.2.3 골목상권의 영역 설정 방법

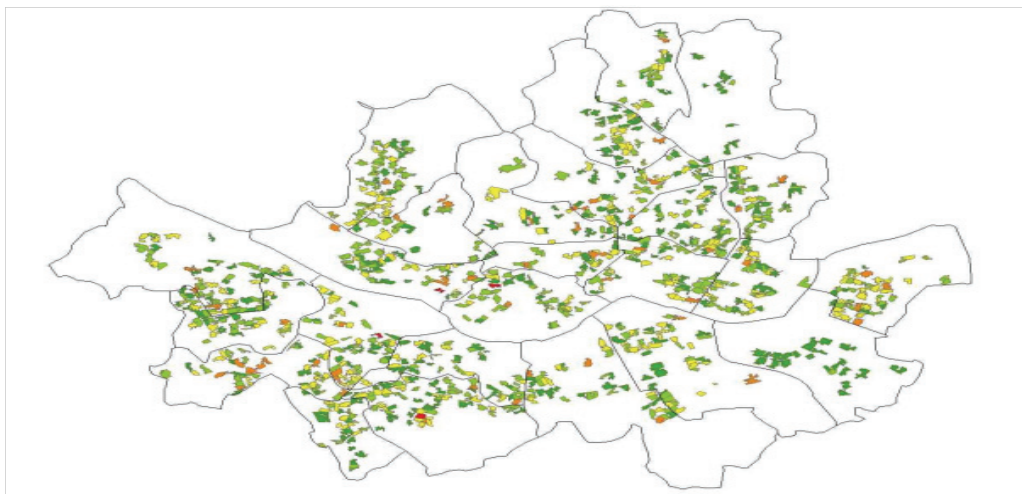
서울시는 골목상권, 발달상권, 전통상권으로 구분하여 관리하고 있다. 그 중 골목상권은 대로변이 아닌 거주지 안의 좁은 도로를 따라 형성

되는 상업세력 중 점포밀집도가 높은 상권으로 총 1,010개 골목상권을 정하였다. 골목상권은 문헌조사, 전문가 인터뷰, 관련 근거조사를 거쳐 다음과 같은 개념적인 정의를 통해 구성하였다. 첫째, 도보로 이동할 수 있는 거리 둘째, 주거지역이 밀집된 곳에 형성된 상권 셋째, 대형유통시설을 상권영역에 포함된 점포 제외 넷째, 대로를 제외한 길 단위 구성 다섯째, 30개 이상 점포수를 포함한 곳이 골목상권지역이다. 이를 구축하기 위해 그림 1과 같은 구축 과정을 거친다.

그림 2는 이와 같은 구축과정을 거쳐 만들어진 골목상권을 서울시 지도에 영역으로 표시한 것이다.



〈그림 1〉 골목상권 구축과정



〈그림 2〉 골목상권영역 표시

2.3 상권정보와 연계하기 위한 설문의 특징

설문조사란 표본 집단으로부터 정보를 수집한 결과를 모집단으로 일반화하려는 방법이다. 따라서 모집단을 대표할 수 있는 표본집단을 구성하는 방법이 중요하고, 상권정보와 연계 분석을 위해 표본집단의 업종과 위치정보가 필요하다. 이를 위해, 표본집단의 위치 및 업종 정보를 미리 파악하거나, 설문을 통해 수집한다.

이러한 방법으로 수집된 설문데이터는 정량적인 특성과 정성적인 특성을 모두 가지고 있다. 정량적인 데이터는 상권데이터와 같이 측량을 하거나 분석이 가능한 수량적인 데이터이고, 정성적인 데이터는 사용자의 경험 및 사고를 보여주는 기술적인 데이터로 사람이 어떻게 생각하고, 어떻게 느끼지를 표현할 수 있다. 즉 자연 언어에 의한 서술로 표현되는 범주형 데이터이다. 데이터 분석에 있어 정량 데이터는 객관적인 결론을 도출하고, 정성데이터는 주관적인 결론을 도출한다.

데이터 분석에 있어 정량데이터의 분석과 정성데이터의 분석은 서로를 보완할 수 있다. 정량데이터는 일반적인 결과를 증명할 수 있는 수치데이터이고, 정성데이터의 정량데이터의 의미를 설명할 수 있다. 따라서 상권분석 서비스를 제공하기 위해 정량데이터의 수집과 정성데이터 수집이 가능한 설문이 필요하다.

Ⅲ. 상권정보와 설문정보 연계 방법과 활용 방법

본 절에서 상권서비스 제공을 위해 기관에서 제공받은 데이터의 한계에 대해서 설명하고, 상권정보와 연계를 위해 필요한 설문을 구성하는 절차에 대해서 설명한다. 그리고 설문정보와 상권정보를 연계 하는 방법에 대하여 제안하고, 제안된 연계방법으로 분석에 활용하는 방법을 설명한다.

3.1 기관에서 제공된 데이터의 한계

우리마을가게 상권분석서비스는 다양한 기관에서 제공받은 데이터를 통합 및 가공하여 상권 분석 서비스를 제공하기 때문에 다음과 같은 데이터의 한계가 있다.

첫째, 다양한 기관에서 제공받은 데이터를 통합 가공하여 만든 상권서비스는 제공받은 데이터의 정합성 검증이 어렵다. 제공받은 데이터가 개별 기관의 제한된 데이터로 데이터 간 교차 검증이 어렵다.

둘째, 데이터 부족에 따른 상권 데이터 분석의 한계이다. 데이터 분석을 위해 여러 범주의 다양한 데이터들이 필요하지만 현실적으로 기관에서 제공되는 데이터는 제한적이다. 예를 들어, 창업에 관련되어 분석에 필요한 정량데이터로 창업자금, 직원고용형태, 직원 수 등이 필요하나 이러한 데이터를 제공받을 수 있는 기관이 없다.

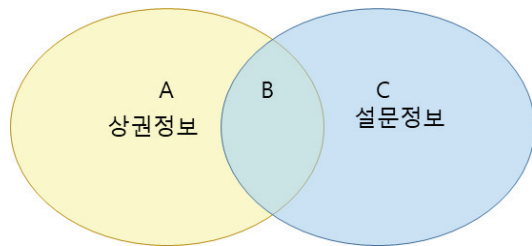
셋째, 다양한 기관에서 수집되는 데이터는 점차 작은 영역의 집계성 데이터에서 넓은 영역의 집계 데이터로 제공된다. 이는 개인정보 보호정책의 강화에 따른 것으로, 작은 영역의 분석 서비스가 점차 어려워지고 있다. 예를 들어, 과거에 제공받은 데이터는 구역별 집계 정보로 구성되어 있어 행정동별, 구별로 집계하여 서비스가 되었으나, 기관에서 점차 행정동별 집계 형태로 데이터를 제공하여 구역별 서비스가 불가능하게 된다.

3.2 상권데이터와 연계를 위해 필요한 설문데이터

설문문항 설계 시 설문정보와 상권정보와 연계를 위해 설문대상자의 업종과 업소의 위치정보가 반드시 필요하다. 이는 상권 서비스가 업종별, 영역별(위치정보)로 구성되어 있기 때문이다. 이를 위해, 100개로 구분된 생활밀접업종

을 수집 및 관리하고, 설문조사 시 설문 대상자의 업소의 위치를 자동으로 수집하거나, 위치 조회 기능을 사용하여 수집해야 한다.

상권정보와 연계분석을 위한 설문정보를 구성하기 위해 설문문항 설계 시 상권분석이 선행되어야 하고, 분석 후 추가적으로 필요한 데이터에 대해서 설문을 실시한다. 예를 들어, 상권정보 분석 시 월 매출액 추정에 대한 상권정보 검증이 필요한 경우 설문문항 설계 시 월매출액을 문항에 추가하여 데이터를 수집한다.



<그림 3> 상권정보와 설문정보의 영역

그림 3은 상권정보와 설문정보의 영역을 그림으로 표현한 것이다. 설문정보는 공통의 영역(B)를 통해 상권정보를 검증하고 설문영역(C)를 통해 상권 분석서비스를 확대할 수 있다.

3.3 상권정보와 설문정보를 연계하는 방법과 분석 활용 방법

상권 정보와 설문정보를 연계하기 위해 업종과 영역(위치좌표)을 기준으로 데이터를 연계한다. 이를 위해, 첫째, 영역별로 집계된 상권 정보와 영역별로 집계한 설문 정보를 동일한 영역을

기준으로 1:1 형태로 매핑된 데이터를 구성하여 분석하는 방법이 있다. 둘째, 설문조사 된 개별 업체를 중심으로 연계하여 분석하는 방법이다. 설문조사 된 개별업체의 위치좌표를 중심으로 반지름이 100m~1000m 사이에 있는 업소의 상권데이터를 집계하거나, 구역, 행정동, 구와 비교하여 분석하는 방법이다. 셋째, 설문의 작은 영역과 설문의 작은 영역이 포함된 상권의 넓은 영역을 연계하여 상권의 작은 영역에 대한 집계값을 추정하는 방법이다.

3.3.1 동일영역 연계 분석

상권정보와 설문정보를 연계하여 분석하기 위해, 각각의 영역을 동일한 크기의 동일한 영역으로 집계하고 업종도 동일한 업종에 대하여 분석하는 방법이다. 이를 위해, 설문정보도 영역별(블록, 동, 구, 시)로 집계자료를 만들어 상권정보와 동일한 영역에 대해 분석 가능하도록 구성하는 방법이다. 예를 들어, 서초구에서 편의점 30곳에 대하여 대표자의 성별, 임차료 월세, 인건비, 종업원수, 창업자금에 대하여 설문을 하였을 경우, 설문정보를 블록기준으로 집계하고, 상권정보는 서초구 상권데이터의 점포수, 프랜차이즈수, 개업수, 폐업수, 주거인구, 직장인구를 블록 영역기준으로 2차원 테이블을 구성하여 분석하는 방법이다.

위의 <표 5>는 상권정보와 설문정보를 블록별로 매핑한 예이다. 블록별로 설문정보를 집계하고, 블록별로 집계된 상권정보와 연계하여 분석하기 위해 동일한 블록을 매핑하여 2차원 테이블 형태로 데이터를 구성한다. 이와 같이, 우

<표 5> 설문정보와 상권정보 동일 영역 매핑

블록 코드	설문정보					상권정보					
	대표자 성별	임차료월세 (만원)	인건비 (만원)	종업원 외부종사자	창업자금 (만원)	점포수	프랜차이즈	개업수	폐업수	주거인구	직장인구
156426	1	80	0	0	5000	480	38	22	24	227	337
16215	1	50	0	0	20000	378	27	10	1	670	636

리마을가게 상권분석서비스에서 제공하는 상권 정보와 설문정보를 영역별, 업종별로 매핑 하면 상권정보와 설문정보를 연계한 분석이 가능해 지고, 동일 항목 설문 시 데이터 검증이 가능해 진다.

3.3.2 동일영역 연계를 통한 분석 활용 방법

상권정보와 설문정보를 동일영역, 동일업종으로 매핑을 하면, 첫째, 상권정보와 설문정보 간의 정합성 검증이 가능하다. 둘째, 이를 기반으로 회귀식을 적용하여 미래 정보에 대하여 추정이 가능하다. 셋째, 상권정보와 설문정보를 조합하여 분석하고자 하는 파생정보를 생성이 가능해 진다.

3.3.2.1 상권정보와 설문정보 간의 정합성 검증

상권정보는 과거의 데이터로 각 기관에서 수집된 데이터를 상권서비스에 적합한 형태로 변환하여 서비스 한다. 이를 위해, 상권정보는 시간 기준으로 월별, 분기, 또는 년 단위 과거형 데이터의 집계로 구성이 된다. 적재되는 상권정보는 시계열 데이터 형태로 지속해서 쌓이기 때문에 현재의 설문정보와 상권정보가 일치하는 시점이 발생하게 된다. 이 시점에서 상권정보와 설문정보를 서로 검증할 수 있게 된다. 소상공인을 대표하는 설문 대상자들에 대한 설문결과를 모집단의 결과를 일정한 오차 비율로 추정이 가능하기 때문이다.

예를 들어, 2019년 1월 A 구역에서 설문을 통한 편의점 매출액 평균이 27,000,000원으로 신뢰수준 95%로 표본오차가 ±5%일 때, 2019년 1월 A구역의 상권에 저장된 편의점 매출액 평균이 26,340,000원 이라면 설문결과와 상권에 저장된 데이터와 비교할 수 있다. 설문을 기준으로 표본오차가 ±5%일 때, 실제 매출액 평균은 25,650,00~28,350,00 범위에 있다. 실제 상권에 저장된 데이터는 26,340,000으로 표본오차 내에

데이터가 존재하는 것을 검증할 수 있다.

3.3.2.2 과거 상권 정보와 현재 설문정보를 기반으로 미래 상권정보를 예측

시계열로 저장된 상권 데이터는 회귀식을 기반으로 미래를 예측할 수 있다. 상권데이터는 월, 분기, 년도 별 기준으로 시계열 데이터형태로 저장되고 있어 회귀식을 사용할 수 있다. 아래 표는 저장된 상권정보와 설문정보이다.

〈표 6〉 매출액 상권정보

구분	년월	매출액
상권정보	2014. 1	24,100
상권정보	2015. 1	24,200
상권정보	2016. 1	24,300
상권정보	2017. 1	24,400
설문정보	2018. 1	24,500

〈표 6〉은 4년 동안 저장된 월별 매출액과 2018년 1월의 설문 정보를 기반으로 2019년 1월의 매출액을 예측할 수 있다. 위 데이터는 선형형태로 표현이 가능하며 최소제곱법에 의한 선형 회귀식을 사용하면 오차값을 최소화하는 최소제곱회귀직선(least squares regression line)을 찾을 수 있다. 수식은 다음과 같다.

$$\hat{y} = ax + b \text{로 표현}$$

$$\bullet a = \frac{SS_{xy}}{SS_{xx}} \text{ 및 } b = \bar{y} - a\bar{x}$$

$$\bullet SS_{xy} = \sum xy - \frac{\sum x \sum y}{n} = 244945000 - \frac{1224720000}{5} = 1000$$

$$\bullet SS_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} = 101606400 - \frac{20321290}{5} = 10$$

$$\bullet a = \frac{1000}{10} = 100$$

$$\bullet b = 24300 - 100 * 2016 = -177300$$

$$\hat{y} = 100x - 177300$$

위 식을 기반으로 2019년을 계산하면 $100 * 2019 - 177,300 = 24,600$ 된다. 따라서 2019년 1월의 매출액은 24600으로 예측가능하다.

3.3.2.3 동일영역 연계 분석을 통한 파생 정보 생성

상권정보와 설문정보를 연계하여 새로운 정보를 파생할 수 있다. 이를 위해, 상권정보와 설문정보를 같은 영역으로 집계 값을 작성한다. 1:1로 매핑이 되는 정보를 편의점에 대한 월 평균 순수익을 알고 싶은 경우 아래와 같은 수식으로 계산할 수 있다.

$$\text{순수입} = \text{월매출액} - \text{임대료} - \text{상품구입비} - \text{인건비} - \text{그외 지출금액}$$

업종별 순 수입을 알기 위해 상권데이터에 없는 정보를 설문조사를 통해 확보하는 방법으로 <표 7>은 월평균 순수익을 계산하기 위한 설문데이터의 사용 예이다.

3.3.3 개별업체 중심 연계 분석

개별 업체를 중심으로 데이터 분석을 위해 개별업체가 포함된 시점 및 영역을 기반으로 데이

터를 분석하는 방법이다. 예를 들어, 서울신용보증보험의 수혜업체를 대상으로 설문조사를 한 경우 수혜업체의 상권 환경을 분석하기 위해 사용하는 방법이다.

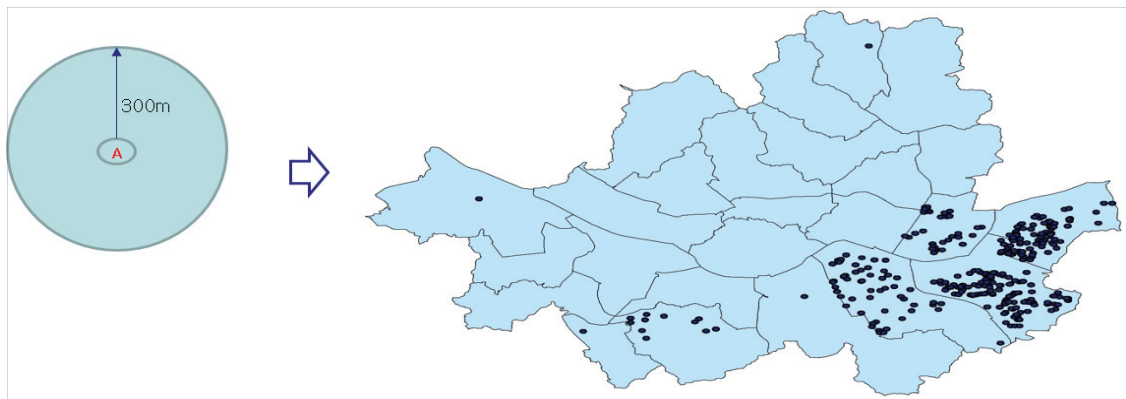
그림 4는 신용보증수혜업체 설문대상자 영업점 위치를 지도에 표시한 것이다. 설문완료 후 각각의 점포에 대하여 300m이내에 있는 블록들의 점포수, 프랜차이즈, 개업수, 폐업수, 주거인구, 직장인수 정보를 조회할 수 있다.

3.3.3.1 개별업체 중심 연계 분석 활용 방법

개별업체 중심 분석을 통해 생성된 상권정보와 연계된 설문정보를 통해 설문 조사된 특정업체의 상권 환경 분석이 가능하고 이를 기반으로 개별업체에 대한 컨설팅이 가능하다. 개별업체를 중심으로 상권데이터 비교 범위를 구역, 행정동, 구로 확장할 수 있고, 업종도 선택하여 비교할 수 있다.

<표 7> 파생정보 생성을 위한 상권정보와 설문정보 매핑

2019년 1월 상권정보			2019년 1월 설문정보			월평균 순수입
영역	매출액	임대료	상품 구입비	인건비	그외 지출금액	
압구정동	25,312	3,413	13,240	3,214	3,145	2,300
강남구	26,245	3,304	12,840	2,812	3,015	4,274



<그림 4> 신용보증수혜업체 설문대상자 영업점 위치

〈표 8〉 점포 위치 기반의 상권정보와 설문정보

점포명	상권정보					설문정보				
	점포수	프랜 차이즈	개업수	폐업수	주거 인구	직장 인구	임차료 월세 (만원)	인건비 (만원)	종업원 외부종사 자수	창업자금 (만원)
포즈**	100	150	1	5000	648	74	32	25	311	178
조유**	100	0	0	3000	827	64	26	27	425	141
현대**	400	1000	5	10000	401	49	19	13	471	148
O2**	80	0	0	3000	512	66	15	13	400	180
삼진**	100	1000	4	10000	544	63	23	18	407	91
진일**	300	400	1	10000	62	8	1	1	469	90
써밋**	100	0	0	3000	428	36	15	12	452	290

또한 설문 조사된 개별 업체를 업종별로 설문 정보를 집계 하여 상권정보과 매핑하여 분석이 가능하다. 특정 목적으로 설문된 개별업체는 업체들만의 특징을 가지고 있고, 이 특징을 가진 업체들이 상권에 어떤 영향을 미치고 받고 있는지 검토할 수 있다.

3.3.4 설문정보의 작은 영역과 이 영역이 포함된 상권정보의 넓은 영역을 연계하여 상권정보의 작은 영역에 대한 값을 추정하는 방법

제공되는 정보가 넓은 영역의 집계 정보라면 작은 영역에 대해서는 분석하는 것은 불가능하다. 그러나 설문대상자 선정 시 표본의 전체를 대표할 수 있는 샘플 표본을 통해 설문조사한 것을 가정한다면, 표본의 적은 량의 설문이라도 전체를 대표할 수 있다. 즉, 작은 영역의 설문을 통해 작은 영역의 상권정보 값을 추정할 수 있다.

예를 들어, 서초동은 구역 가, 나,의 블록으로 구성이 되어 있고, 서초동의 2018년 2월 평균 매출액은 상권정보로 200만원이라고 가정하고, 설문결과로 가 구역은 180만원 나 구역은 225만원이라는 설문 결과가 나왔다고 가정한다. 그러면, 가 구역의 평균 매출액은 가구역과 나 구역의 평균을 구하고 평균에 비해 차이 비율을 계산하여 상권정보의 매출액과 곱하여 계산한다. 수식

은 아래와 같다.

$$\text{가 구역 추정 평균 매출액} = (180 - (180 + 225)/2) / ((180 + 225)/2 - 180) * 200 + 180$$

IV. 연계된 상권정보와 설문정보의 분석 방법 적용

상권정보와 설문정보를 분석하기 위해 데이터를 1:1 형태로 매핑을 구성하면 기존의 상권정보 간 분석, 설문정보 간의 분석뿐만 아니라 상권정보와 설문정보를 연계한 분석 적용이 가능해진다.

본 절에서는 상권정보와 설문정보의 연계분석 방법으로 T-test, 상관분석, 회귀분석을 기준으로 설명한다.

4.1 T-test분석

T-test분석은 차이 분석으로 데이터간의 평균이 유의미한 차이가 있는지를 분석하는 방법이다. <표 9>는 강남구 블록별 설문정보와 블록별 상권정보를 매핑하여 구성한 데이터이다.

성별에 따른 창업자금이 유의미한 차이가 있는지 분석하기 위해 먼저, 집단 통계량을 계산한다. <표 10>은 집단통계의 결과이다.

〈표 9〉 강남구 블록별 설문정보와 블록별 상권정보 매핑 예제

블록 코드	설문정보					상권정보					
	대표자 성별	임차료월세 (만원)	인건비 (만원)	종업원 외부종사자	창업자금 (만원)	점포수	프랜차이즈	개업수	폐업수	주거 인구	직장 인구
156426	1	80	0	0	5000	480	38	22	24	227	337
16215	1	50	0	0	20000	378	27	10	1	670	636
18058	2	70	0	0	5000	1309	107	50	54	207	270
188153	2	100	0	0	4000	492	30	15	14	370	272
24444	1	20	0	0	1000	1695	100	58	56	84	353
24484	2	200	700	3	10000	1426	125	48	39	108	744
24561	1	300	1200	6	10000	939	116	35	33	195	1745
24578	2	400	1300	4	20000	1431	212	52	50	163	497
25445	1	100	100	1	18000	504	49	9	11	181	527
266233	1	100	100	1	5000	194	17	2	6	240	227
267111	1	50	200	2	5000	592	35	13	6	191	622

〈표 10〉 성별기준의 창업자금 집단통계량

창업자금	성별	N	평균	표준편차	평균의 표준오차
	남(1)	17	9411.76	6671.383	1618.048
	여(2)	10	6500.00	5126.185	1621.042

〈표 11〉 성별기준의 창업자금 T-test 분석 결과

구분	levene의 등분산 검정	평균의 동일성에 대한 T검정								
		F	유의 확률	t	자유도	유의확률 (양측)	평균차이	표준오차 차이	차이의 95%신뢰구간	
								하한	상한	
창업 자금	등분산가정	1.152	.293	1.186	25	0.247	2911.765	2454.896	-2144.187	7967.717
	등분산을 가정하지 않음			1.271	23.016	0.216	2911.765	2290.384	-1826.072	7649.601

집단통계량을 작성하여 데이터의 특성을 파악한 후 levene의 등분산 검정과 평균동일성에 대한 T검정을 진행한다. <표 11>은 성별 기준의 창업자금 T-test 분석 결과 값이다.

결과에서 유의확률이 0.05보다 높기 때문에 성별에 따른 창업자금은 유의미한 차이가 없다.

4.2 상관분석

상관분석은 연속 변수로 측정된 두 변수간의 선형관계를 분석하는 방법으로 운영점포수와 프랜차이즈 수를 분석한다.

운영점포수와 프랜차이즈수는 0.868로 아주 높은 선형관계가 있으며, 운영점포수가 많은 곳은 프랜차이즈 수가 많다고 추정할 수 있다.

〈표 12〉 운영점포수와 프랜차이즈수 상관분석 결과

구분		운영점포수	프랜차이즈수
운영점포수	Person 상관	1	.868
	유의확률(양측)		.000
	N	27	27
프랜차이즈수	Person 상관	.868	1
	유의확률(양측)	.000	
	N	27	27

상관관계가 0.01 수준에서 유의합니다.

〈표 13〉 직장인구와 창업자금 회귀 분석 결과

계수^a

모형	비표준화 계수		표준화 계수 베타	t	유의확률
	B	표준화 오류			
1 (상수)	6258.133	1673.878		3.739	.001
직장인구	4.092	2.393	.324	1.710	.100

a. 종속변수: 창업자금
(만원)

4.3 회귀분석

회귀분석이란 독립변수가 종속변수에게 얼마나 영향을 미치는지를 알아보고자 하는 방법으로 단순 선형회귀분석을 적용하여 직장인구 수가 창업자금에 어떤 영향을 미치는지 알아본다. 결과적으로 유의확률이 0.01로 유의미하며 직장인구에 따른 창업자금은 아래의 식으로 계산할 수 있다.

$$\text{직장인구수} = \text{창업자금} * 4.092 + 6258.133$$

V. 결론

소상공인을 위한 상권분석서비스는 다양한 기관에서 제공받은 데이터를 기반으로 제공되고 있으나, 데이터가 표준화되지 않고, 제한된 데이터로 서비스의 구성에 한계가 있다. 본 논문에서는 상권분석 서비스를 확대를 위해 설문정보를 연계하는 방법과 활용, 분석하는 방법 대하여 제안하였다.

이를 위해, 본 논문에서는 우리마을가게 상권 서비스 시스템을 중심으로 다양한 기관에서 제공받은 데이터를 분석 가능한 형태로 가공하는 기존의 전처리 과정을 설명하였다. 둘째, 상권정보와 설문정보를 연계하기 위한 설문정보 구성 과정 및 설문정보 특징에 대하여 설명하였다. 셋째, 상권정보와 설문정보를 업종과 영역을 기준으로 동일 영역 연계, 설문 정보의 개별업체 중심 연계, 설문정보의 작은 영역과 상권의 큰 영역 연계의 세 가지 방법을 제안한다. 그리고 연계된 상권정보와 설문정보를 다양한 방법으로 분석 및 활용하는 방법을 설명한다. 마지막으로, 상권정보와 설문정보를 연계하여 다양한 분석 방법을 적용하여 분석한 결과를 보였다.

본 논문은 기존 상권분석 서비스의 한계를 설문정보로 극복할 수 있는 방법을 제안하여 소상공인에게 필요한 서비스를 확대할 수 있는 기반을 마련하였다. 향후 연구 과제로는 소상공인 설문조사 모집단을 대표할 수 있는 패널구성 방법과 점진적으로 쌓이는 시계열 상권정보와 설문

정보를 이용하여 미래를 예측할 수 있는 ARIMA (Autoregressive Integrated Moving Average) 분석과 RNN(Recurrent Neural Networks)의 LSTM(Long Short-Term Memory Models) 대한 연구가 필요하다.

참고 문헌

- [1] 강만수, 윤재웅, 정경훈, 위평량, “상권시스템으로 본 서울 영세자영업 통계 및 젠트리피케이션 양상 분석과 시사점”, 서울신용보증재단 소상공인 정책연구센터, 연구보고서, 2019.12.
- [2] 대한민국 정책브리핑 소상공인·자영업자 지원대책, <https://www.korea.kr/special/policyCurationView.do?newsId=148856781&pWise=main&pWiseMain=L5>
- [3] Micro Enterprise And Market Service, <https://sg.sbiz.or.kr/> .
- [4] Seoul Metro City, <https://golmok.seoul.go.kr/main.do>.
- [5] Gyeonggi Market Revitalization Agency, <https://sbiz.gmr.or.kr/main.do> .

저자 소개



이원철(Won-Cheol Lee)

- 2000 강원대학교 행정학과 (학사)
 - 2002 강원대학교 컴퓨터과학과 (석사)
 - 2007 강원대학교 컴퓨터과학과 (박사수료)
 - 2019~현재 한림대학교 소프트웨어중심대학사업단 산학협력중점교수
 - 관심분야 : 상권데이터 분석, 데이터베이스 설계 및 튜닝, 접근제어 설계, 업무시스템 개발, 머신러닝 등
- E-mail: woncheol@kangwon.ac.kr



강만수(Man-Su Kang)

- 1997 강원대학교 경영학과 (학사)
 - 2000 강원대학교 경영학과 (석사)
 - 2011 강원대학교 경영학과 (박사)
 - 2013~2016 신용보증재단중앙회 연구원
 - 2016~현재 서울신용보증재단 소상공인 정책연구센터 데이터인프라팀장
 - 관심분야 : 금융 빅데이터, 상권분석서비스, 소상공인 컨설팅시스템, 다차원 정보수집 시스템 개발 등
- E-mail: jazz@kangwon.ac.kr



김진호(Jinho Kim)

- 1982 경북대학교 전자공학과 (학사)
 - 1985 KAIST 전산학과 (석사)
 - 1990 KAIST 전산학과 (박사)
 - 1990~현재 강원대학교 컴퓨터과학과 교수
 - 2020~현재 강원대학교 빅데이터·융합보안사업단 단장
 - 관심분야 : 대용량 빅데이터 저장 및 처리, 하둡/맵리듀스 분산/병렬처리 기술, 빅데이터 분석 기법, 데이터 마이닝, 클라우드 컴퓨팅, 데이터 웨어하우스, OLAP 다차원 분석, 데이터베이스 시스템 개발
- E-mail: jhkim@kangwon.ac.kr