

위치기반 소셜 네트워크 서비스(LBSNS)를 이용한 POI 정보 강화 방안

Enrichment of POI information based on LBSNS

조성환* · 가칠오** · 허용***
Cho, Sung-Hwan · Ga, Chil-O · Huh, Yong

Abstract

Point of interest (POI) of the city is a special place that has what importance to the user. For example, it is such landmark, restaurants, museums, hotels, and theaters. Because of its role in the social and economic life of us, these have attracted a lot of interest in location-based applications such as social networks and online map. However, while it can easily be obtained through the Web, the basic information of POI such as geographic location, another effort is required to obtain detailed information such as Wi-Fi, accepting credit cards, opening hours, romper room and the assessment and evaluation of other users. To solve these problems, a new method for correcting position error is required to link location-based social network service (LBSNS) data and POIs. This paper attempts to propose a position error correction method of POI and LBSNS data to enrich POI information from the vast information that is accumulated in LBSNS. Through this study, we can overcome the limitation of individual POI information via the information fusion method of LBSNS and POI, and we have discovered the possibility to be able to provide additional information which users need. As a result, we expect to be able to collect a variety of POI information quickly.

Keywords: POI, Location Based Social Network Service, Information Enrichment, String Similarity

1. 서 론

POI (Point of Interest) 데이터는 위치와 다양한 속성 정보를 포함하는 점(point) 타입의 공간적 객체로서 일반적으로 병원, 은행, 주유소, 편의점 등 주로 생

활편의 시설과 관련된 관심 지점 정보를 의미한다. “현 위치에서 가장 가까운 지하철역은?”, “강남역에서 근처 중국 음식점은?” 등의 질의는 웹 또는 모바일 지도 서비스를 이용하는 사용자가 가장 빈번하게 활용하는 패턴이며, POI 데이터는 이러한 서비스를 가능

* Researcher, School of Economic, Political and Policy Science, University of Texas at Dallas (first author: mission21@gmail.com)

** 한국국토정보공사 과장 Korea Land and Geospatial InformatiX Corporation (gco75@lx.or.kr)

*** 공간정보연구원 선임연구원 LX Spatial Information Research Institute (corresponding author: huhyong78@gmail.com)

하게 하는 핵심 공간 데이터이다. 현재 POI 데이터는 지도 제작, 내비게이션 업체 등 주로 민간 부문에서 구축하여 서비스를 제공하고 있으나 넓은 범위에 걸쳐 방대하게 분포하는 특성으로 인해 초기 구축뿐만 아니라 지속적인 갱신에도 많은 노동과 비용이 투입되어야 하는 문제점이 존재한다. 최근 공간정보 데이터 취득 기술이 급속도로 발전하면서 도로, 건물 등의 지형/지물에 대해서는 상당부분 자동화된 프로세스를 통하여 구축이 가능해 졌으나, POI 구축에 있어서는 여전히 수작업에 의존되고 있어 이에 대한 효과적인 구축 프로세스 개발이 시급한 실정이다.

POI 데이터는 현실 세계의 위치 또는 장소를 표현하기 때문에 그 절대적인 양과 개별 POI가 갖는 속성 정보의 풍부함이 서비스 품질을 결정짓는 중요한 요인이 된다(Braun et al., 2007). 일반적으로 POI 데이터의 구축은 현장조사에 의한 실사취득방법을 활용하며, 웹, 문헌 등 주소 정보를 포함하는 부가적인 정보를 보조적으로 활용한다(Oh, S., 2006). 따라서 이와 같은 방법은 가장 정확하게 구축할 수 있는 방법이지만 하나 전 국토에 걸쳐 분포하는 POI 데이터를 전수 조사하기란 현실적으로 불가능하다. 이와 같은 문제점으로 최근에는 사용자들에 의해 구축된 VGI (Volunteered Geographic Information) 데이터를 활용하여 POI 데이터를 구축하고자 하는 연구들이 활발하게 진행되고 있다(Ahern et al., 2007; Chen et al., 2018; Jossé, et al., 2017; Mummidi and Krumm, 2008; Zhu and Zhou, 2009).

이와 같이 풍부한 POI 데이터 구축을 위한 연구는 다소 활발하게 진행되고 있는 반면, 각 POI 데이터 속성의 풍부함을 보완하기 위한 연구는 미미한 실정이다. 현재 서비스되고 있는 POI데이터는 위치, 주소, 전화번호 등 제한적인 수준에서 제공되고 있다. 예를 들어 서울에서 일요일에 영업을 하며, 놀이방이 있는 음식점을 검색할 경우, 사용자는 직접 전화를 하여 문의를 하거나 또는 기사, 카페, 블로그 등을 통하여 재검

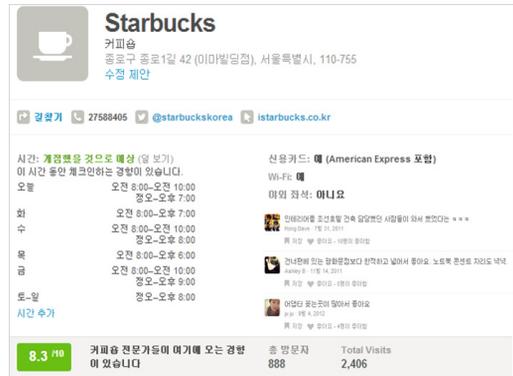


Figure 1. Attribute information from LBSNS data Source: Foursquare.com.

색하고 다른 사용자들의 후기 등을 종합하여 판단을 내리는 과정을 거쳐야만 한다. 따라서 이와 같이 상세하고 고급의 정보를 제공할 수 있는 POI 데이터의 필요성은 모바일 및 위치기반서비스 등이 보편화될수록 지속적으로 높아질 것으로 예상할 수 있다.

본 연구에서는 단순히 POI 데이터의 양적인 구축에 비하여 상대적으로 연구가 미비한 속성 정보의 풍부함을 보완할 수 있는 방법 개발을 연구 목적으로 하였으며, 이를 위해 기 구축된 POI 데이터와 위치기반 소셜 네트워크 서비스(location-based social network service, LBSNS)를 통하여 구축된 방대한 장소의 정보 융합 방법을 제시하고자 한다. LBSNS는 위치기반 서비스와 소셜 네트워크 서비스가 결합된 서비스로 사용자들이 특정 장소를 방문하고 그 장소에 대한 위치, 영업시간, 메뉴 등 사실적인 정보와 다양한 경험 및 후기 등을 자유롭게 기술하고 다른 사용자들과 공유하는 서비스를 말한다(Figure 1). 이와 같이 LBSNS를 연계하면 위치뿐만 아니라 영업시간, 전화 번호 외에 와이파이 가능 여부, 신용카드 가능 여부, 놀이방 운영 여부, 평점 및 평가, 총 방문자 수, 이용후기 등의 다양한 사용자들로부터 축적된 고급 정보들의 취득이 가능해진다.

하지만 LBSNS에 등록된 정보는 일반적으로 사용

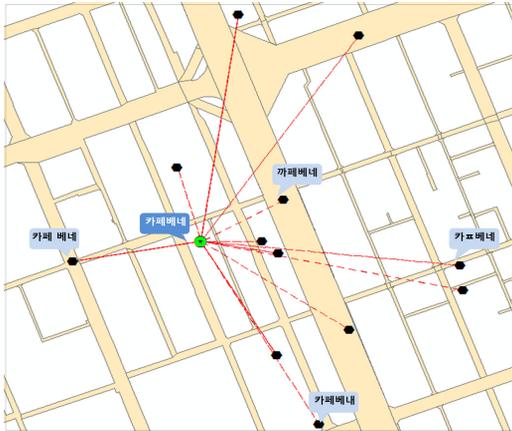


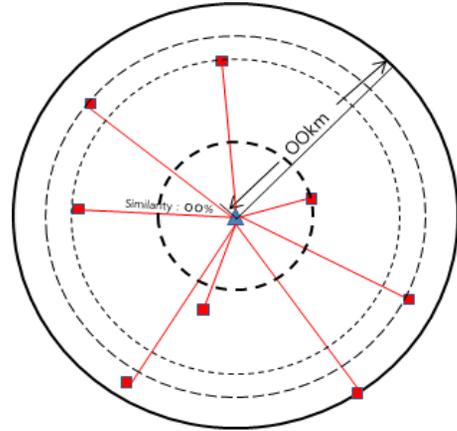
Figure 2. Positional discrepancies of a POI and its corresponding LBSNS data.

자들이 스마트폰 등의 모바일 기기를 활용하여 자유롭게 기술된 정보로 기기의 위치오차와 오타 및 비정형 텍스트를 포함하고 있다. 즉 Figure 2에서 보는 것과 같이 정밀하게 구축된 POI와 달리 LBSNS에 등록된 장소 정보는 다양한 위치오차와 명칭으로 표현되게 된다. 따라서 본 연구에서는 이러한 문제점을 해결하기 위하여 POI와 LBSNS 사이의 정합성을 탐색하기 위한 위치 및 명칭 유사도 측정방법과 그 임계값을 제시하고자 한다.

2. 연구 내용

LBSNS에 기록되어 있는 장소 정보를 활용하여 이미 데이터베이스화 되어 있는 POI 정보를 풍부하게 하기 위해서는 다음과 같은 문제점들이 존재한다.

- 1) 상호간 동일한 장소로 판단할 수 있는 임계반경을 설정하는 과정에서 반경이 증가할수록 비교 대상이 기하급수적으로 증가함에 따라 자원 및 연산 시간 증가 문제
- 2) POI 명칭과 LBSNS 데이터의 명칭 간의 틀린 맞춤법 및 정확하지 않은 외래어 표기로 인한 명칭 혼동 문제



▲ Position of POI ■ Positions of check-in venue

Figure 3. Conceptual model of proposed method.

POI 명칭과 LBSNS의 매칭은 각 데이터베이스에 있는 두 개의 테이블의 레코드가 얼마나 유사한가의 문제와 관련이 있다. 따라서 서로 다른 두 데이터집합의 장소 간의 공간적 유사성과 명칭 유사성을 기반으로 두 데이터 집합 간의 동일 여부 판단 방법을 제안한다. Figure 3은 본 연구에서 제안하는 개념 모델로 구축된 POI를 기준으로 LBSNS에 등록된 장소의 위치오차를 허용할 수 있는 적절한 임계반경을 설정하고, 임계반경 내에 존재하는 장소 명칭과 구축된 POI의 명칭 간 유사도를 Table 1과 같이 분석함으로써 융합 가능한 장소를 선정하게 된다. 이를 위한 처리 과정은 Figure 4와 같다.

2.1. 매칭 후보 POI 선정을 위한 반경 임계치

본 연구에서 구축된 POI는 웹지도, 내비게이션 등의 서비스에서 활용되는 POI를 말하며 이들은 일반적으로 현장 조사 등을 통하여 구축되기 때문에 비교적 정확한 위치 정보를 갖는다. 이에 반해 LBSNS 데이터는 사용자에 의해 생성된 위치로 서비스를 위하여 정밀하게 구축된 데이터가 아니기 때문에 비교적

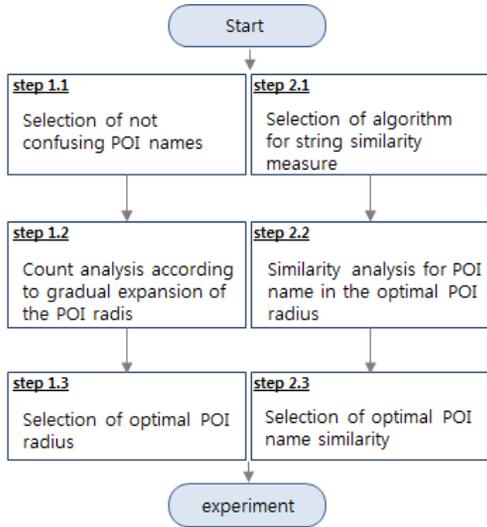


Figure 4. Workflow of this study.

Table 1. Typographical errors between corresponding POI and LBSNS data, and their string similarities

Names in POI data	Names in LBSNS data	String similarities(%)
커피숍	커피샵	83.3
회집	훗집	75
존슨즈	존슨즈	83.3
주스	쥬스	75
파이팅	화이팅	66.7
후라이팬	프라이팬	75
라이언스	라이온스	75
마포참숯불갈비	마포숯불갈비	85.7
신의주참살순대	신의주참살순대	85.7
사월에보리밥	사월에보리밥 강남점	66.7

큰 위치오차를 갖는 특성을 갖는다. 하지만, 비록 위치 오차를 갖는다 하더라도 동일한 장소에 대해서는 Figure 2와 같이 어느 정도 공간적 유사성을 나타낸다. 따라서 기 구축된 POI의 위치를 중심으로 이와 동일한 POI라고 판단할 수 있는 임계 반경이 설정되어

야 한다. 이를 위해 본 연구에서는 LBSNS 데이터의 위치 오차를 분석하기 위하여 지리적으로 고르게 분포하고 정확한 위치 정보를 갖으며 오입력이 거의 없는 전철역 POI와 전철역 LBSNS 데이터를 활용한다.

임계 반경을 결정하는 방법으로는 전체 명칭쌍 중 정확하게 일치하는 명칭쌍의 개수를 이용하여 결정한다. 전체 293개의 전철역 POI 데이터 집합과 실험에 사용한 전철역 체크인 장소 명칭은 14개(Table 2)의 차이를 보이는 것을 포함하여, 실제로 사용자들의 체크인 활동에서 전철역과 관련된 명칭의 혼동이 매우 적은 것으로 나타났다. 이와 같이 선정된 명칭쌍을 대상으로 본 연구에서는 10m 간격으로 반경을 확장해 가며 포함된 명칭쌍의 수를 분석하여 그 수가 급격히 줄어드는 거리를 임계 반경으로 설정한다.

2.2. POI 명칭 유사도 필터링

임계 반경 안에서 LBSNS 데이터들은 동일한 장소 임에도 불구하고 등록과정에서 사용자의 의도 및 실수에 의한 오타로 서로 다른 명칭으로 구축되어 있을 수 있다. 따라서 이들 간에 동일한 장소임을 판별하기 위하여 명칭 유사도로써 필터링 과정이 필요하다. 본 연구에서는 이를 위해 Levenshtein 알고리즘을 적용하였다.

Levenshtein 알고리즘은 Levenshtein Distance, Damerau-Levenshtein Distance, Edit distance 등으로 알려져 있는 알고리즘으로, 두 문자열의 비슷한 정도를 측정하기 위해 널리 활용되는 알고리즘이다(Levenshtein, 1966). 이 알고리즘은 비교하고자 하는 원 문자열을 S, 대상을 T라고 할 때, 수정 연산을 통하여 이 두 문자열을 같게 만드는 과정이며 이때 수정되는 연산은 삽입(insertion), 삭제(deletion), 치환(substitution)이며, 수정 횟수를 거리(edit distance)라 한다. 예를 들어, 문자열 S가 “keyboard”이고, 문자열 T도 “keyboard”라면 거리는 0이 되며, 문자열 T가

Table 2. Different names in LBSNS data for the same POIs

Station name of POI	Station names of check-in
강남역	강남역 (Gangnam Stn.), 강남역 6번, 강남역 6번출구 IBK
서울역	서울역 (Seoul Station), 서울역 (Seoul Sta. Gyeongui Line), 서울역 공항철도 (Seoul Sta.)
암사역	암사역, 암사역 사거리
신촌역	신촌역, 신촌역 (Sinchon Sta.),
서대문역	서대문역, Seodaemun Station
영등포역	영등포역, Yeongdeunpo train station
홍대입구역	홍대입구역 (Hongik Univ. Sta.), 홍대입구역 (Hongik Univ. Strn.)
학여울역	학여울역 (Hangnyeoul Sta.), 학여울역 사거리
건대입구역	건대입구역, 2호선건대입구역
⋮	⋮

“eybaarde”인 경우를 본다면 문자열 가장 앞에 ‘k’를 삽입, ‘a’를 ‘o’로 치환, ‘e’의 삭제 과정이 필요하므로 이 경우 거리는 3이 된다.

Table 1은 동일한 POI에 대해서 기 구축된 POI 명칭과 LBSNS 데이터의 명칭 간에 Levenshtein 알고리즘에 의한 명칭 유사도 측정 예시를 나타낸다. Table 1의 예시는 비록 서로 명칭이 다르게 표현되었지만 유사도가 높게 나타남을 볼 수 있다.

이에 반하여 Table 3의 ‘마포참숯불갈비’를 중심으로 반경에 포함된 모든 LBSNS 데이터의 명칭들 간에 유사도를 보여주는 표를 보면 ‘마포숯불갈비’와 ‘마포숯불갈비’를 제외한 나머지 관심 장소들은 유사도가 매우 낮음을 알 수 있다. 즉, ‘마포숯불갈비’와 ‘마포참숯불갈비’는 POI의 ‘마포참숯불갈비’와 동일한 장소로 의도하여 등록된 관심장소로 판단할 수 있으며 나머지는 다른 장소로 판단할 수 있다.

3. 실험 및 결과

LBSNS 데이터와 POI 정보를 융합하기 위한 연관 관계를 생성하는 방법으로 첫 번째 공간적 연산 범위 설정을 위한 임계 반경을 산출하고, 다음으로 입력 문자 오류를 해결하기 위한 문자열 유사도 임계값을 계

Table 3. Candidate LBSNS data and their string similarities for one POI data

Names in POI data	Names in LBSNS data	String similarities(%)
마포참숯불갈비	마포숯불갈비	85.7
	마포숯불갈비	71.4
	마포숯불	57.1
	유가네	0
	세븐일레븐	0
	영클스	0
	하이트광장	0
	앵콜노래방	0
	맥주창고	0
	CU	0

산한다. 실험을 위한 데이터로는 서울시의 POI 264,334개와 LBSNS 데이터 404,498개 중에서 임계 반경을 산출하기 위한 실험을 위해 서울시의 지리적 범위 내의 279개 전철역 POI와 19,709개의 LBSNS 데이터를 사용하였고, 임계 문자열 유사도 결정을 위해 사용자들의 LBSNS 데이터가 가장 많은 강남역 주변의 279개의 POI와 주변 19,682개의 LBSNS를 사용하였다.

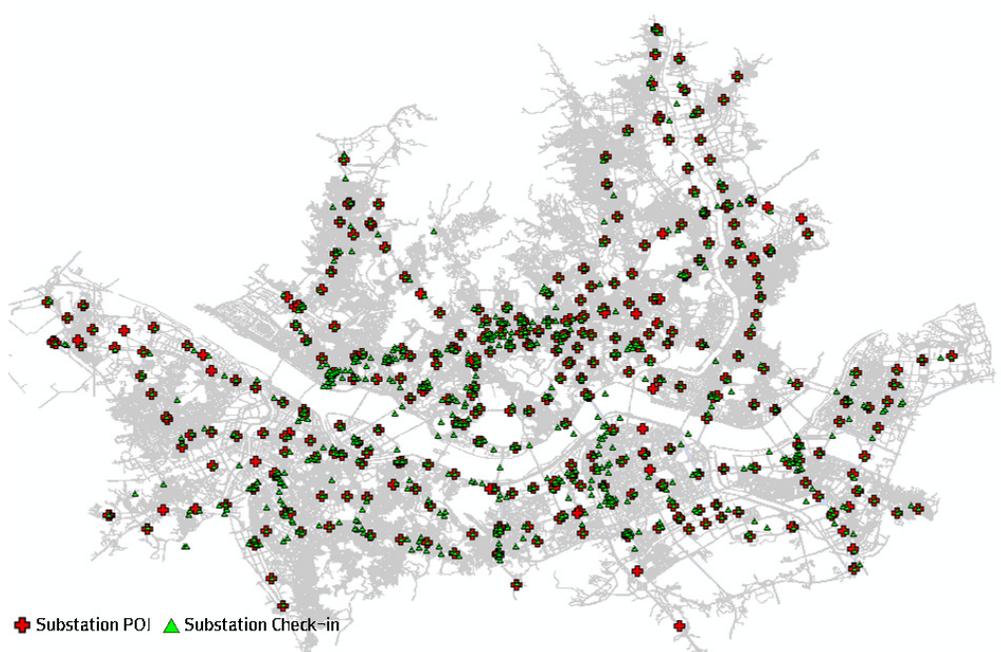


Figure 5. Map of study area to estimate radius threshold.

3.1. 임계 반경 결정

POI와 LBSNS 데이터 간의 동일 여부를 판단하기 위한 명칭 유사도 계산 과정은 대상 명칭쌍의 크기가 클수록 컴퓨팅 시간이 기하급수적으로 증가하기 때문에 사용자들의 행동 패턴의 분석을 통한 최적의 임계 반경 결정이 필요하다. 본 실험에서 POI와 LBSNS 데이터의 위치 오차를 분석하기 위하여 지리적으로 골고루 분포하고 있고 상대적으로 명칭 혼동이 없는 전철역 데이터를 사용하였다. 이를 위해 POI 데이터의 필터링은 ‘교통편의>>교통시설>>지하철역’ 범주에 속하는 POI 데이터를 선택하였고, LBSNS 데이터의 필터링은 LBSNS에서 제공하는 범주 중 ‘Travel & Transport’, ‘Subway’, ‘Travel & Transport’, ‘Train Station’, ‘Light Rail’ 에 속하는 데이터를 사용하였다. Figure 5는 임계 반경을 결정하기 위한 실험 지역과 데이터를 보여준다. 적색으로 표시된 십자가 모양은

정확한 좌표를 가지고 있는 전철역 POI 위치 데이터이고, 초록색 삼각형은 사용자들에 의해 생성된 LBSNS의 전철역 데이터를 나타낸다.

전철역의 해당 범주에 속하는 데이터 집합으로 279개의 전철역 POI 명칭과 19,709개의 LBSNS 데이터 명칭을 대상으로 하였을 때, POI를 중심으로 2 km 거리 이내에 있는 LBSNS와 195,815개의 명칭쌍이 생성되었다.

Figure 6는 반경이 증가할수록 연산 대상인 명칭쌍의 개수가 증가하는 것을 보여준다. 이는 기준이 되는 POI로부터 지리적 반경을 점진적으로 증가시키면서 포함되는 명칭쌍의 개수를 나타내는 것으로, 즉, 기준이 되는 POI의 명칭과 주변의 LBSNS 데이터 명칭의 유사도 연산 횟수와 동일한 값을 나타낸다. 명칭의 일치 여부를 고려하지 않은 전철역 POI로부터 주변 전철역 LBSNS의 거리별 연산 횟수 실험에서는 명칭쌍의 약 90%가 POI 위치로부터 반경 600 m(35,193회

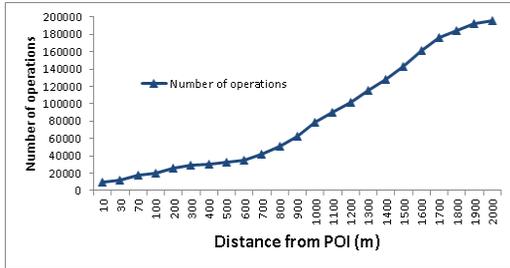


Figure 6. Number of operation by distance.

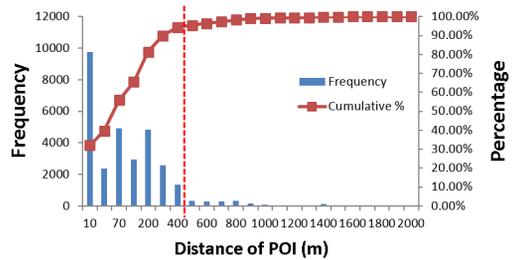


Figure 7. Distance distribution of the check-in name and POI name matches(Station name).

연산) 이내에, 약 95%가 반경 1 km(78,027회 연산), 약 98%가 1.3 km(115,196회 연산) 이내에 그리고 1.8 km 이내에 체크인 장소 99.9%가 포함되는 것으로 나타났다.

임계 반경을 구하기 위한 실험으로 POI 위치로부터의 거리 증가에 따른 정확하게 일치하는 명칭쌍의 변화량을 산출하였고, Figure 7에서 보는 바와 같이 500 m 이후로 일치하는 명칭쌍 개수의 변화가 거의 없는 것을 알 수 있었다. 따라서 본 실험을 통한 최적의 임계 반경은 전체 명칭쌍 대상 중 일치하는 명칭쌍에 대해 약 95%의 매칭 성공률을 얻을 있는 반경 500 m를 선정하였다.

3.2 임계 문자열 유사도 설정

LBSNS 데이터는 모바일 단말기를 이용하여 불특정 다수의 사용자들의 참여를 통하여 구축된다. 따라서 앞에서 언급한 바와 같이 동일한 점포나 위치라 할지라도 상이한 명칭이 부여된 별도의 객체일 수 있다. 이는 단순히 모바일 단말기를 사용하는 과정에서 발생한 사용자의 오타 및 외래어 또는 차용어를 사용하는 과정에서 발생하는 표기의 혼돈이 복합적으로 발생한 결과이다. Figure 8은 임계 문자열 유사도를 결정하기 위한 실험 지역과 데이터를 보여주고, Table 4는 이러한 단어와 제안된 방법에 의하여 측정된 명칭 유사도를 보여준다. Figure 8에서 적색으로 표시된 십자가 모양은 정확한 좌표를 가지고 있는 강남역 주변

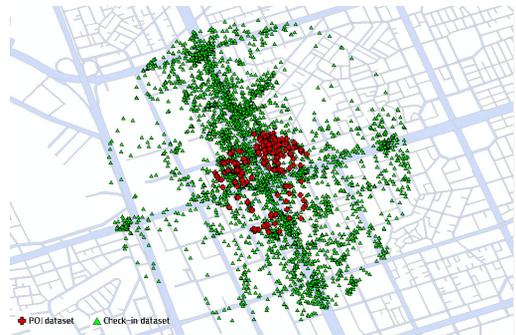


Figure 8. Map of study area for setting of threshold string similarity.

POI 위치 데이터이고, 초록색 삼각형은 사용자들에 의해 생성된 강남역 주변 LBSNS 데이터를 나타낸다. 임계 문자열 유사도 결정을 위하여 사용자들의 체크인 수가 가장 많은 강남역 주변의 279개의 POI와 주변 19,682개의 LBSNS 데이터를 사용하였다.

Table 4는 명칭 유사도의 임계값을 결정하기 위하여 Figure 8 지역에서 선정한 7개 POI와 각각 측정된 유사도를 기준으로 가장 높은 유사도를 가지는 LBSNS의 명칭들을 보여준다. Levenshtein 알고리즘으로 측정된 유사도는 POI의 명칭 중 어느 정도가 LBSNS의 명칭에 유사하게 포함되어있는가를 기준으로 유사도가 측정된다. 따라서 ‘던킨도너츠’의 경우 ‘반포점’ 또는 ‘DUNKIN’ DONUTS’이라는 문자열이 LBSNS의 명칭에 포함되어 있더라도 ‘던킨도너츠’라는 문자열이 LBSNS의 명칭에 포함되어 있으면 유사

Table 4. String similarities of confusing foreign words in LBSNS data

Names in POI data	Names in LBSNS data	String similarities(%)
가쓰라	가쓰라	100
	가츠라	67
	가스라	67
빍스	빍스	100
	VIPS	0
사보텐	사보텐	100
	샤보텐	66.7
스무드킹	스무드킹	100
	스무디킹	75
쥬스떡볶이	쥬스떡볶이	100
	쥬스떡볶이	80
	쥬스떡볶이	60
김리떡볶이	김리떡볶이	100
	김리떡볶이	80
	김리떡볶이	60
COFFEE	Coffee	100
	Coffeenie	67
	Coffine	71
	Coffeacoffee	50

도가 100으로 측정되기 때문에 더욱 안정적인 결과를 얻을 수 있다. 하지만, 이러한 특성은 ‘가쓰라’, ‘사보텐’ 또는 ‘파리바게뜨’와 같이 표기상의 차이가 작은 경우에는 완만한 유사도 변화를 보여주지만, 그렇지 않은 명칭 사이에서는 유사도가 급격하게 변하는 결과를 반환하게 된다. 이러한 경우에는 정규분포를 가정하는 일반적인 임계값 추정 기법은 적합하지 않다.

따라서 본 연구에서는 비모수통계학 분야에서 임계값 결정에 사용되고 있는 Box-Whisker plot 기법을 적용하고자 한다. 이 기법은 어떤 분포를 가지고 있는 자료의 극단값, 즉 임계값을 결정하는 방법으로 자료 분포의 제1사분위와 제3사분위인 Q1, Q3 를 양 끝으로 하는 상자를 가정한다. 이 때 Q1, Q3의 차이를 R이라고 정의한다면 Q1과 Q3로부터 좌우로 ±1.5R 지점

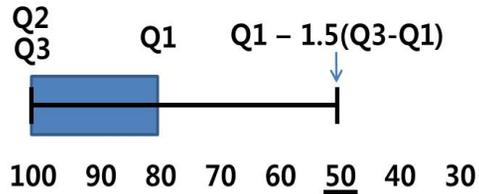


Figure 9. Estimation of string similarity threshold with Box-Whisker plot method.

을 극단값(outlier), 즉 임계값으로 가정하게 된다. Table 5의 자료에 LBSNS의 명칭의 빈도를 고려하여 이 기법에 적용하면 Figure 9와 같이 Q1과 Q3이 각각 100(%)와 80(%)이므로 50(%)가 임계값으로 결정된다.

이렇게 결정된 임계 반경과 유사도를 평가하기 위하여 Table 5과 같이 4개의 POI를 대상으로 제안된 방법을 적용하였다. 임계 반경 500m와 임계 유사도 50%를 적용한 결과 정확도는 80% (8/10)였다. 2개의 잘못된 탐색은 모두 스타벅스라는 명칭을 가지고 있는 POI에서 발생하였다. 이는 ‘스타’라는 문구가 주변 건물이나 점포의 명칭에 많이 사용됨으로써 문자열 유사도가 50%로 측정되었기 때문이다. 비록 앞의 과정을 통하여 50%라는 유사도가 얻어졌지만 만약 POI나 LBSNS의 명칭에 일반적으로 사용되는 명칭이 존재할 경우 이 유사도로는 오류가 발생할 수 있다. 이러한 문제는 Table 4에서 ‘떡볶이’나 ‘coffee’라는 문구는 유사 업종의 점포명에 일반적으로 포함되어 있어 50%라는 유사도 임계값을 적용할 경우 유사 업종의 다른 점포의 명칭이 정합될 수 있는 점에서 역시 확인할 수 있다. 이를 해결하기 위해서는 ‘떡볶이’나 ‘coffee’와 같이 고유명사의 성격이 아닌 일반명사의 성격을 가지는 문구를 제외하고 명칭의 문자열 유사도를 측정하는 것이 필요하다.

Table 5. POIs to assess the proposed thresholds.

POI name	Top 5 similar LBSNS names (No. of LBSNS data)	String similarity (%)
스무드킹	스무드킹 (1)	100
	스무디킹 (SMOOTHIE KING)(3)	75
	나무그늘 (2)	25
	나무그늘 Coffee (1)	25
	스시코 (1)	25
스타벅스	스타벅스 (1)	100
	스타타워 (1)	50
	스타호프 (1)	50
	스윙타임바 (1)	40
Mass	Nass cafe & bar (1)	75
	Compass Media (1)	47
	HANS (1)	40
	Mad for Garlic (1)	40
	MARSIL (1)	33
파파이스	파파이스 (Popeyes) (1)	100
	카페루이스 (1)	40
	파워라이드 (1)	40
	강남파이낸스센터 (1)	38
	아카스페이스 (1)	33

4. 결론

본 연구는 LBSNS에 누적되어 있는 방대한 정보로부터 POI의 정보를 더욱 풍부하게하기 위한 방법으로 LBSNS 정보를 POI를 통해 제공하기 위한 연관 관계를 생성하는 방법을 제안하였다. 사용자의 소셜 미디어 사용 패턴에 따른 체크인 장소와 POI의 위치 오차를 해결하기 위한 방법으로 첫 번째 명칭쌍 생성을 위한 연산 범위 설정을 위해 임계 반경을 산출하고, 다음으로 입력 문자 오류를 위한 문자열 유사도 임계값을 계산하였다. 그 결과 임계 거리 결정을 통하여 명칭쌍의 개수를 현저하게 줄일 수 있어 명칭 유사도 연산에 필요한 컴퓨터 자원 사용 및 연산 시간을 낮출 수 있었다. 또한, POI 명칭과 LBSNS 장소 명칭의 동일 여부

판단을 위한 임계 문자 유사도를 제시하였다.

본 연구는 POI와 LBSNS의 정보 융합을 가능하게 하고, 조사에 의해 생성될 수밖에 없는 개별 POI 정보량의 한계를 극복하여 와이파이 및 신용카드 가능 여부, 야외 좌석 여부, 놀이방 운영 여부, 사용자들의 의견 및 평가 등과 같은 추가적인 정보를 구축할 수 있는 가능성을 보여주었다. 이러한 정보는 특정 장소의 정보를 얻고자하는 사용자들에게 풍부하고 빠른 정보 수집의 기회를 제공해 줄 수 있을 것으로 판단된다. 향후 제안하는 방법을 활용하여 POI 위치 오차 해결을 위한 더 많은 연구가 필요하다.

감사의 글

본 연구는 국토교통부 도시건축개발사업 연구비지원 (13도시건축A02)에 의해 수행되었습니다.

참고문헌

References

- Ahern, S., Naaman, M., Nair, R., and Yang, J.H.I. 2007. World Explorer: Visualizing aggregate data from unstructured text in geo-referenced collections, *In Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries*, New York, USA, pp. 1-10.
- Braun, M., Scherp, A., and Staab, S. 2007. Collaborative creation of semantic points of interest as linked data on the mobile phone, *In Extended Semantic Web Conference*. Springer
- Brown, P. F., Desouza, P. V., Mercer, R. L., Pietra, V. J. D., and Lai, J. C. 1992. Class-based n-gram models of natural language. *Computational linguistics*, Vol. 18, No. 4, pp. 467-479.
- Chen, S., Wang, Z., Liang, J., & Yuan, X. 2018. Uncertainty-aware visual analytics for ex-

- ploring human behaviors from heterogeneous spatial temporal data. *Journal of Visual Languages & Computing*.
- Foursquare. <https://ko.foursquare.com> (last date accessed: 5 September 2018)
- Hall, P. A. and Dowling, G. R. 1980. Approximate string matching. *ACM Computing Surveys (CSUR)*, Vol. 12, No. 4, pp. 381-402.
- Heo, Y. and Cha, H. 2012. Evaluating similarity metrics for accurate POI recognition. *Korea Institute of Communication and Information Sciences*, pp.175-176. (in Korean with English abstract)
- Jin, A., Lee, J., and Lee, J. 2013. Measuring method of string similarity for POI data retrieval, *Journal of KISS: Computing Practices and Letters*, Vol.19, No.4, pp.177-185. (in Korean with English abstract).
- Jonietz, David, and Alexander Zipf. "Defining fitness-for-use for crowdsourced points of interest (POI)." *ISPRS International Journal of Geo-Information* 5.9 (2016): 149.
- Jossé, G., Schmid, K. A., Züfle, A., Skoumas, G., Schubert, M., Renz, M., & Nascimento, M. A. 2017. Knowledge extraction from crowdsourced data for the enrichment of road networks. *Geoinformatica*, 21(4):763-795.
- Kang, B., Lee, J., and Choi, K. 1999. Phonetic similarity measure for the korean transliterations of foreign words, *The Korea Institute of Information Scientists and Engineers*, Vol. 26, No. 10, pp. 1237-1246. (in Korean with English abstract)
- Lcvenshtcin, V. 1966. Binary codes capable of correcting deletions, insertions, and reversals. *In Soviet Physics-Doklady*, Vol. 10, No. 8.
- Lee, W. 1998. *Approximate matching algorithms for ODP trader*, MD. Dissertation, Pohang Technical University, Pohang, Korea. (in Korean with English abstract)
- Lee, H., Choi J., and Park Y. 2012. Semantic point of interest detection from large-scale GPS data of mobile users, *The Korea Institute of Information Scientists and Engineers*, Vol. 39, No. 3, pp. 175-184. (in Korean with English abstract)
- Lin, J., Xiang, G., Hong, J. I., and Sadeh, N. 2010. Modeling people's place naming preferences in location sharing. *In Proceedings of the 12th ACM international conference on Ubiquitous computing*, ACM, pp. 75-84
- Lian, D. and Xie, X. 2011. Learning location naming from user check-in histories. *In Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, ACM, pp. 112-121
- Novack, Tessio, Robin Peters, and Alexander Zipf. "Graph-Based Matching of Points-of-Interest from Collaborative Geo-Datasets." *ISPRS International Journal of Geo-Information* 7.3 (2018): 117.
- Mummidi, L. N. and Krumm, J. 2008. Discovering points of interest from user's map annotations, *GeoJournal*, Vol. 72, No. 3, pp. 215-227
- Oh S. 2006. Current status and planning of POI construction, planning and policy, *Korea Research Institute for Human Settlements*, Vol. 291, pp. 152-157
- Touya, Guillaume, et al. "Assessing crowdsourced POI quality: Combining methods based on

reference data, history, and spatial relations.”
ISPRS International Journal of Geo-Information 6.3 (2017): 80.

pp.730-734.

Zhu, X. and Zhou, C. 2009. POI inquiries and data update based on LBS, *IEEE International Symposium on Information Engineering and Electronic Commerce*, Washington D.C., USA,

2018년 10월 05일 원고접수(Received)

2018년 10월 10일 1차심사(1st Reviewed)

2018년 10월 29일 2차심사(2st Reviewed)

2018년 11월 22일 게재확정(Accepted)

초 록

도시의 POI는 사용자들에게 어떤 중요성을 가지고 있는 특별한 장소이다. 예를 들어 식당, 박물관, 호텔, 극장 그리고 랜드마크 등이다. 이들은 그 역할 때문에 우리의 사회·경제적 생활 속에서 온라인 지도나 소셜 네트워크 등의 위치기반 어플리케이션에서 많은 관심을 받고 있다. 하지만, 지리적 위치 등의 POI에 대한 기본 정보는 웹을 통해서 쉽게 얻을 수 있는 반면, 와이파이 가능 여부, 신용카드 가능 여부, 야외 좌석 여부, 놀이방 운영 여부, 개점 시간, 다른 사용자들의 평가 및 평점 등의 세부 정보를 얻기 위해서는 또 다른 노력이 필요하다. 이러한 문제를 해결하기 위해서는 LBSNS 데이터와 POI 위치의 동일 여부를 판단하여 연결해주는 작업이 필요하다. 이 논문은 LBSNS의 누적되어 있는 방대한 정보로부터 POI의 정보를 더욱 풍부하게 만들기 위한 방법으로 LBSNS 데이터와 POI의 위치 오차 해결 방법을 제안하여 두 데이터 집합의 융합 정보를 생성하고자 한다. 본 연구의 POI와 LBSNS의 정보 융합 방법을 통하여 개별 POI 정보의 한계성을 극복하고, 사용자들이 필요로 하는 부가 정보를 제공할 수 있는 가능성을 발견하였다. 이를 통해 POI에 대한 풍부하고 빠른 정보 수집이 가능할 것으로 판단된다.

주요어 : 관심지역, 위치기반 소셜 네트워크 서비스, 정보 풍부화, 위치오차, 문자열 유사도