

# 빅데이터를 활용한 LBS시스템의 정확도 향상 기법

문창배\*, 박현석\*\*, 윤세준\*\*

\*한양사이버대학교 전기전자통신공학부

\*\*한양대학교 정보시스템학과

e-mail:changbae@hycu.ac.kr

## The LBS system with accuracy improvement function using big data

Changbae Mun\*, Hyun-Seok Park\*\*, Se-Jun Yoon\*\*

\*Dept. of Electrical, Electronic & Communication Engineering, Hanyang Cyber University

\*\*Dept. of Information System, Hanyang University

### 요약

현재 다양한 개인 통신단말기를 통한 LBS(Location-based service)가 활성화되면서 LBS의 정확도 향상 기법에 대한 중요성이 점점 더 부각되고 있다. 포털 지도정보서비스의 위치 처리 플랫폼과 연동하여 각 사용자의 위치를 활용하여 동적으로 정제된 정보를 제공하고 사용자의 주변정보를 기준으로 실질적인 서비스를 보다 정확하게 제공해 주는 것을 목적으로 한다. 이에 대한 접근 방법으로서 본 논문에서는 사용자에 대해 수집한 다양한 빅데이터를 분석하고 다양한 경로를 이용하는 사람들의 이동정보를 수집하여 가장 효율적인 경로정보를 분석하고자 한다. 특정 지역에서 두 지점에 대한 대한 경로정보를 탐색할 때, 해당 지역 거주자들이 이동하는 경로를 추출할 수 있다면 그것은 가장 정확한 정보에 해당할 수 있을 것이다. 본 연구에서는 그 방법을 모색하였고 이 시스템을 통해 사용자는 보다 향상된 수준의 위치정보를 제공받을 수 있고 이 기능은 기존의 LBS서비스와 연계하여 발전할 수 있을 것이다.

## 1. 서론

위치정보시스템 LBS(Location-based service)은 사용자의 위치정보를 파악하고 수집한 정보를 바탕으로 다양한 서비스를 제공하며 발전을 거듭하고 있다. 연평균 39.77%씩 증가하는 추세이며 2021년에는 960억 규모에 이르게 될 것으로 분석된다. 이같은 LBS시스템 기술의 발전은 우선 자동차 내비게이션 관련 다양한 애플리케이션의 서비스 영역의 확대, 참여 사업자가 증대되어 위치추위 기술과 서비스 관련 인프라 발전, 시장에 참여한 글로벌 기업들의 치열한 경쟁으로 인한 혁신적 서비스의 제공이 LBS산업 발전의 주요 근거로 분석된다[1]. 이같은 시장의 발전에 따라 지도정보서비스, 광고마케팅과 상점 정보 제공, 엔터테인먼트, 주변정보 서비스 등의 LBS에 관련한 다양한 서비스들이 출시되고 있으며 특히 모바일 커머스 산업에서 이용자에 대한 정보제공 편의를 위하여 다양한 상품정보와 할인쿠폰, 구매정보 추천등의 기능에 활용하고 있다[2]. 특히 배달관련 정보제공 서비스의 발전에 따라 지역 음식점과 사용자들의 리뷰정보가 지도정보와 함께 제공되어 새롭게 도시경관이 활성화되는 사례가 등장하고 있으며 여러 정보시스템과 융합 및 연계가 빈번한 상황이다. 이에 본 연구에서 특정 지역에 두 지점간의 경로를 보다 정확하

게 제시할 수 있는 다양한 방법을 모색하였고, 지역의 경로정보에 대해 상대적으로 정확도가 높은 정보를 갖고 있는 지역 거주민들의 데이터를 복합적 분석을 통해 활용하여 이를 LBS에 적용하는 방안을 제시하고자 한다.

## 2. 시스템의 주요 구성

### 2.1 경로 분석 중심의 LBS 시스템의 사례

본 연구의 대표적 선행연구에 해당하는 연구로서 특정 경로를 이동하는 사용자의 정보를 중앙서버에서 수집하고 분석을 수행해서 동일 경로에 대해 다수가 이동한 경로를 추출하는 기법에 해당한다. 이 방법에 상점 마케팅 정보를 접목하여 특정 경로에서 사용자가 관심을 갖는 특정 지역 정보를 찾고 경로에 관련있는 상점의 특관 정보가 존재하면 그 정보를 사용자가 기선택한 방식으로 푸시, 문자, 알림전화의 방식으로 제공해준다[3]. 협업 필터링 기술에 기반한 추천 시스템 연구에서는 추천 시스템을 각 개인의 LBS시스템에 적용하여 개인화된 콘텐츠 제공 방안을 제안하였다. 이 기능은 현재 위치를 기반으로 사용자가 세팅한 거리 내에 있는 주요 장소중에서 전체 사용자에게 오픈된 콘텐츠를 선호도 단위로 분석하여 사용자에게 효용도가 높을 것으로 예상되는 콘텐츠를

제시한다[4]. LBS 기반 차량위치추적 시스템의 성능 향상 기법 연구는 위치 구역을 정하고 실시간 차량 이동객체의 위치 정보를 획득하는 방법에 대해 논하였다.[5].

### 2.2 선행연구에 대한 기능적 보완사항

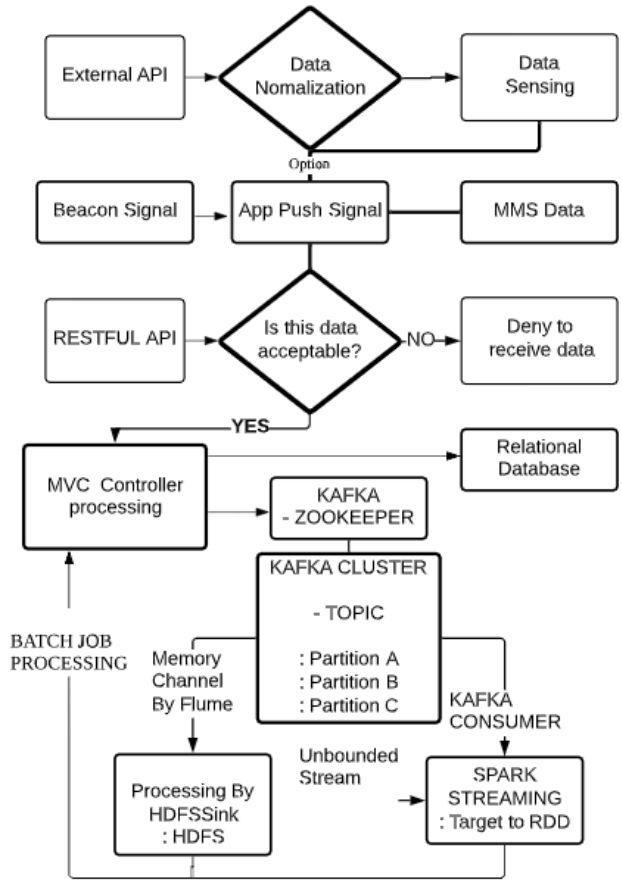
특히 [3]의 연구에서 일반 네비게이션에서 다루기 힘든 개인화된 특성을 보완하는 데 초점을 맞추었다. 즉, 기존의 상용 네비게이션은 초행길이나 야간, 우천시와 같은 기후 특수성과 사용자 특성에 따라 실제로 제시받은 경로를 이동시 정보 획득이 어려울 수 있는 상황이 있고, 초행길의 경우에는 지도를 미리 확인하고 숙지하더라도 날씨, 지도정보의 정확도 정도와 지형지물의 복잡도에 따라 이동에 소요되는 시간이 달라질 수 있다. 특히 우천시나 야간부터 새벽 시간에는 경로정보의 주요 지물에 대한 시각적 확인이 어렵고 상용 네비게이션의 주요 대상이 되는 도로가 아닌 골목길의 경우에는 상황에 따라 가로등의 점멸 등으로 인하여 정보확인이 쉽지않다. 또한 비, 눈등으로 날씨로 시야가 쾌적하지 못하거나 특정 위치의 공사 등으로 인하여 도로 통제가 있을 때 네비게이션에서는 해당 정보에 대해 즉각적인 대응이 어렵다. 이와 같은 어려움을 해결하기 위해 사용자가 초행길을 도보 이동하는 경우에 발생가능한 다양한 상황을 선가정하고 이를 지능형 정보시스템을 통해 안내하는 방법을 다룬다. 포털 지도정보 서비스의 위치 처리 플랫폼과 연동하여 각 사용자의 위치를 활용하여 동적으로 정제된 정보를 제공하고 사용자의 주변정보를 기준으로 실질적인 서비스를 보다 정확하게 제공해 주는 것을 목적으로 한다. 이에 대한 접근 방법으로서 본 논문에서는 사용자에 대해 수집한 다양한 빅데이터를 분석하고 다양한 경로를 이용하는 사람들의 이동정보를 수집하여 가장 효율적인 경로정보를 분석하고자 한다. 이를 위해 지도정보를 기반으로 한 통계적 분석시스템의 구조를 설계하고, 각 사용자의 이동 정보를 실시간으로 중앙서버에서 분석하여 특정 경로를 자주 다니는 사람이 바로 옆에서 혹은 전화로 직접 안내해주는 것과 같은 효과를 얻도록 한다. 이 서비스를 통해 사용자는 더욱 정확한 경로정보를 제공받게되므로 현재 등장한 다양한 상용 LBS서비스와 연계하여 발전할 수 있을 것이다.

## 3. 시스템의 설계와 검증

### 3.1 시스템의 주요 기능 설계

분산처리시스템은 다수의 시스템들을 접근투명성과 서능 투명성을 기반으로 전산망으로 연결 구성하여 서로 다른 공간에 분산된 시스템 자원을 병렬처리 시켜서 통합한 시스템을 뜻한다. 이는 분산을 통해 시스템 자원을 공유하고 속도, 용량

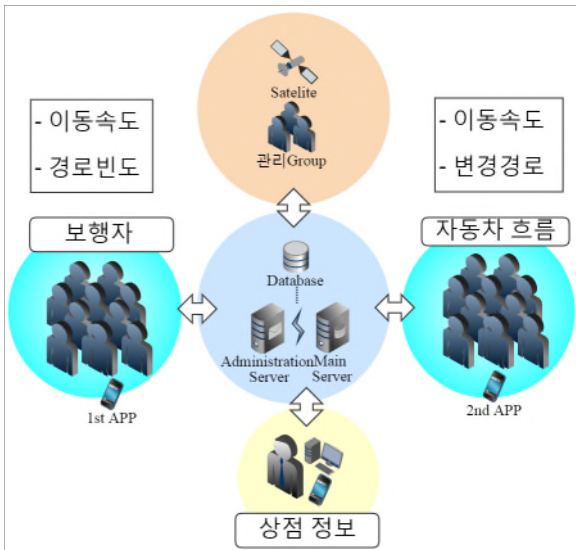
등의 효율을 높이는 것에 목적이 있다. 전체 시스템의 목표 처리량이 증가하더라도 시스템 개별 처리량 분배에 효율성을 가진다. 데이터 분산저장 측면에서 널리 사용되는 Hadoop기술은 구글 파일 시스템을 대체할 수 있는 하둡 분산 파일 시스템(HDFS: Hadoop Distributed File System)과 맵리듀스를 구현한 것에 해당한다.



[그림 1] 분산처리 프로세스

이 HDFS을 통해 Hadoop이 구성된 시스템 네트워크에 통합된 복수의 시스템들이 각 노드를 통해 데이터에 접근 가능하며 분산처리 기술 Map Reduce를 사용하여 클러스터링된 시스템 환경에서의 효율적인 병렬 처리를 지원한다. 세부 노드 단위의 스케줄링을 데이터 프로세싱 아키텍처에 도입하여 스트리밍 데이터를 보다 안정적으로 처리하도록 하였다. 본 연구에서는 선행연구를 참조하여 그림1과 같이 이동 데이터를 수집 및 분석하고 사용자가 제공한 정보들에서 개인정보로 사용될 수 있는 주요 정보를 제외하여 전체 수집용 데이터 스트림을 구성한다.[3] 시스템의 스키마는 정보처리 시스템에서는 수집 정보의 분석처리기능과 스마트폰 애플리케이션을 통한 사용자 안내기능으로 구성된다. 특정 경로의 출발지와 목적지의 주요 갈림길과 빠른 경로에 대해 잘 알고 있는 지역주민에 해당하는 사람들이 경로 정보 제공자로 역할을

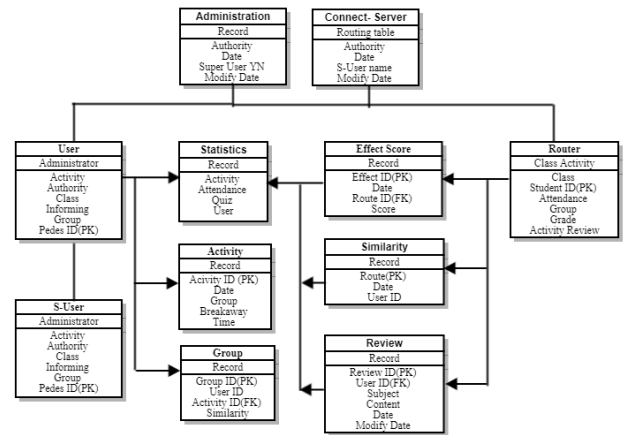
하도록 하기위해 보상시스템을 적용하여 이 활동을 활성화시키고 더욱 정확한 경로정보를 상비할 수 있게된다. 전체 시스템 아키텍처는 이동정보를 제공하는 사용자로부터 경로정보를 수집 및 분석하기 위해 데이터의 분산처리 목적의 SPARK, 세트의 분산전달 목적의 KAFKA, Hadoop에서 옵션기능의 작업수행 목적의 FLUME모듈을 통합 적용하였다. FLUME모듈은 대량의 로그 데이터를 다수의 수집서버 소스에서 이관저장하기 위한 용도로 사용되며 확장성과 신뢰성이 높은 특징이 있으며 실시간 스트리밍 서버에서 대용량 데이터를 효율적으로 처리 및 적재하기 위한 플랫폼을 제시한다. Flume모듈은 오픈소스로서 시스템 내 서버 로그를 비롯하여 정형 혹은 비정형 데이터를 전체를 대상으로 정보 수집을 시행하며 다양한 포맷과 다양한 네트워킹 프로토콜을 통해 이종의 시스템과의 호환성 높은 싱크를 설정할 수 있으며 프로퍼티만 수정하여 다중 스토리지에 단일 Channel과의 커넥션을 연결한다.이 환경에서는 복합 싱크에서 단일Source로 정보 연동을 제공하도록 하였으며 복합 데이터 입출력을 동시에 처리할 수 있는 Spring Cloud Data Flow를 통해 효율을 높인다. 특히 컨트롤러에서 API의 증계정보를 수집하고 Spark Stream과 Flume기반으로 데이터를 전달한다. 또한 Flume모듈은 추가적인 배치 처리를 위하여 HDFS에 텍스트 로그를 저장하는 방식으로 시스템 구조를 설계하였다.



[그림 2] 시스템 내 전체 장비의 구성

시스템의 전체 구조는 [그림 2]와 같이 시스템 사용자에게 해당하는 정보 제공자와 수집의 주체, 관리자등이 각 장치의 주체로서 구조를 이룬다. 정보제공자인 사용자가 정보 수집을 위해 경로를 이동하기 전에 목적지 정보를 해당 모바일 어플리케이션에서 입력하고, 시작 버튼을 클릭한 이후에 이동을 시작한다. 중앙서버에서는 사용자의 위치를 2초 단위의 GPS정보를 서버로 전송하게 세팅되어있다.사용자의 이동정보들이

지속적으로 서버에 축적되며 각 이동정보에 대한 옵션 데이터들까지 누적되게 된다. 이를 통해 [그림3]과 같이 구성된 분석정보가 사용자에게 최상의 경로를 제공하는 것을 이 서비스의 목표로 한다. 정보제공자의 움직임 특성에 따라 혹은 날씨나 도로환경에 따라 같은 지역 주민이라도 여러 경로들이 존재할 수 있으며, 사용자들은 최신 정보를 기준으로 추천받은 경로를 이용하여 판단하게 된다.



[그림 3] 주요 데이터 구성도

[표 1] 분석 데이터와 대상

분석 대상 정보	분석 방법	분석 주기	대상자
유사 경로 정보	협업 필터링	10분	관리자
경로 옵션정보	SQL	사용시	사용자
전체 사용자의 경험통계	통계	일작업	사용자
삭제할 예외정보	멀티 레이블	1시간	관리자

이와 같이 구성된 시스템에서 정확한 경로정보 안내를 위하여 3가지의 대표적인 기능 요소를 갖는다.

- 기능 1 : 사용자들의 공통적 이동 경로를 지름길로 유추
- 기능 2 : 사용자의 이동경로에서 예외정보를 추출
- 기능 3 : 일시적인 예외정보를 검증하고 사용자에게 제시

첫 번째로 기능1은 특정 지역에 거주하는 사용자들이 다니는 경로를 분석하여 경로내에 공통적으로 이동하는 세부경로를 분석하였다. 이는 지역 주민들이 다니는 경로로 경험적으로 입증된 경로임을 알 수 있다. 그리고 이 경로는 버스, 지하철 등의 대중교통과도 접점이 높고 확률적으로 최단 지름길일 확률이 높다. 이 경로를 분석하기 위해 일단위로 지역 주민들의 이동경로 전체를 위치 단위로 수집하여 이동경로의 분포

의 점점 수치를 분석하였다. 이를 통하여 사용자들의 경험으로 수집된 최적경로를 구성하였다.

두 번째로 기능2는 사용자의 이동경로에서 예외정보를 추출한다. 기능 1에서 사용자들의 기능을 분석하는 점에 있어서 정기적으로 발생하는 예외상황을 추출하면 더욱 정보를 정확하게 가공할 수 있을 것이다. 예를 들어 2주단위로 미용실에 가거나 3일단위로 마트에 방문하는 등의 특정 좌표로의 이동이 경로에 노이즈를 주게될 수 있다. 이 문제점을 해소하기 위해 사람들에게 공통적으로 발생하는 이같은 이동정보를 취합하고 분석하는 과정을 거친다. 이를 통해 보다 완결성 높은 정보를 구성할 수 있게 되었다. 이같은 정보분석을 위해 멀티레이블 분석기법을 도입하였다. 다중이진분류인 경우, 아래 수식과 같이 binary classification에는 binary cross entropy loss(BCELoss)를 활용하며 이 기법은 교차 엔트로피 손실 함수에 포함된다. n개의 각 노드에 시그모이드 함수를 적용해서 한 종류의 모델로 binary classification을 진행하게 된다.

$$L = \sum_{i=1}^n BCELoss(y^i, \hat{y}_i) \quad (1)$$

$$\text{조건 } BCELoss(y^i, \hat{y}_i) = -(y_i \times \log \hat{y}_i + (1 - y_i) \times \log(1 - \hat{y}_i))$$

binary classification를 넘어간 경우에는, 다른 종류의 레이블에 대해서 각각 분류를 수행하기 위해 n개의 softmax를 적용해야하며 이 경우의 최종적인 손실 함수는 다음과 같다.

$$L = \sum_{i=1}^n CrossEntropy(y^i, \hat{y}_i) \quad (2)$$

$$\text{조건 } y^i \in \mathbb{R}^3, y_i \in \{0, 1\}^3 \text{ and } |y_i| = 1.$$

$$\text{교차엔트로피 } :(y^i, \hat{y}_i) = -y_i \times \log \hat{y}_i = -\sum_{j=1}^3 y_{i,j} \times \log \hat{y}_{i,j}$$

세 번째로 기능3은 기능2보다는 빈도가 낮은 일시적으로 발생하는 예외정보를 검증하고 사용자에게 제시하고자 한다. 예를 들어 특정 시간에 발생한 도로공사나 눈길에 특정 지점에 이동이 막혀있는 경우가 예가 될 수 있다. 이러한 경우에 사용자들의 이동 속도가 느려지거나 특정 지점에서 돌아가는 등의 좌표를 분석하여 이를 즉시 인근 사용자들에게 제공하도록 한다. 이 점은 인근 정보를 이동하는 사용자들의 이동이 기존의 경로와 차이가 발생했을 경우에 이 사용자들을 대상으로 협업필터링을 실시한다. 특정 상황에서는 분석케이스에 대해 사용자 데이터가 부족한 경우가 있을 수 있으며 충분치 않은 이력의 숫자는 제공정보의 정확도를 하락시킬 가능성이 크므로 이런 유형에서는 평가 인원의 최소수를 설정하여 10 이상일때만 사용자간의 유사도를 계산하기로 기준을 설정한다. 이를 통해 실험에서 발생가능한 예외적 오차를 줄이고자 하였다.



[그림 4] 시뮬레이션 실험의 결과

이와 같이 구현한 시스템을 통해 이동성 실험을 시행하여 아래와 같은 결과를 얻었다. 초록색 실선이 특정 지도정보서비스의 이동 경로이고, 화살표로 표시한 것이 본 시스템의 정보이다. 적색의 지점에 가상의 공사현장과 회피지점을 배치하여 이 시스템의 효과성을 입증하였다.

#### 4. 결론

본 연구에서는 LBS 관련 정확도 향상 기법 관련하여 선행기술과 기술 동향, 정확도 향상에 대한 주요 기술적 해결과제에 대해 살펴보고, 앞으로 지속적인 성장이 기대되는 LBS 시장을 위한 기능적 발전방향성을 제공하고자 한다. 후속연구를 통하여 본 연구의 장점을 더욱 발전시켜서 보다 정확도 높은 LBS시스템을 구현하는 데 도움이 되길 바란다.

#### 참고문헌

- [1] 한국인터넷 진흥원, “국내·외 LBS 산업 동향 보고서”, 2018년.
- [2] Tiwari, Shivendra, et al. "A survey on LBS: system architecture, trends and broad research areas." International Workshop on Databases in Networked Information Systems. Springer, Heidelberg, 2011.
- [3] 문창배, & 박현석. "빅데이터에 기반한 지역 상점 관련 정보제공 서비스." 한국콘텐츠학회논문지, 제 20권2호, pp. 561-571, 2월 2020년.
- [4] 권형준, & 홍광석. "협업 필터링 기반 추천 시스템을 이용한 LBS의 개인화". 인터넷정보학회논문지, 제 11권 6호, pp. 1-1, 2020년..
- [5] 유홍진, 오준환, & 채진석. "LBS 기반 차량위치추적 시스템의 성능 향상 기법". 한국정보과학회 학술발표논문집, 36호, 291-294, 2009년