http://dx.doi.org/10.14400/JDC.2014.12.6.277

LBSNS 기반 장소 추천 시스템

정구임*, 안병익*, 김정준**, 한기준**
(주)씨온 전략마케팅팀*, 건국대학교 컴퓨터공학부**

Location Recommendation System based on LBSNS

Ku-Imm Jung*, Byung-Ik Ahn*, Jeong-Joon Kim, Ki-Joon Han****
Department of Strategy Marketing, SeeOn Inc.*
Division of Computer Science & Engineering, Konkuk University**

요 약 위치 기반 소셜 네트워크 서비스에서 사용자들은 체크인 데이터를 이용해 장소를 공유하고 커뮤니케이션을 한다. 체크인 데이터는 POI명, 카테고리, 장소 좌표/주소, 사용자 닉네임, 장소 평가 점수, 관련 글/사진/동영상 등으로 구성된다. 위치 기반 소셜 네트워크 서비스에서 이러한 체크인 데이터를 상황에 맞게 분석하면 다양한 맞춤형 서비스를 제공할 수 있다. 따라서, 본 논문에서는 사용자 체크인 데이터를 활용할 수 있는 LBSNS(Location-based Social Network Service) 기반 장소 추천 시스템을 개발하였다. LBSNS 기반 장소 추천 시스템은 체크인 데이터 중 장소 카테고리를 분석하여 가중치를 추출하고, 피어슨 상관계수를 이용해 사용자간의 유사도를 구한다. 그리고 협업적 필터링 알고리즘을 이용하여 추천 장소에 대한 선호도 점수를 구하고, 추천 대상 장소의 위치와 사용자의 현재 위치에 대해 유클라디안 알고리즘을 적용하여 거리 점수를 구한다. 마지막으로, 선호도 점수와 거리 점수에 대해 가중치 계수를 적용하여 상황에 맞게 장소를 추천한다. 또한, 본 논문에서는 실제 데이터를 이용한 실험을 통해 본 논문에서 제시한 시스템의 우수성도 입증하였다.

주제어 : 위치 기반 소셜 네트워크 서비스, 체크인 데이터 분석, 장소 추천, 사용자 유사도

Abstract In LBSNS(Location-based Social Network Service), users can share locations and communicate with others by using check-in data. The check-in data consists of POI name, category, coordinate and address of locations, nickname of users, evaluating grade of locations, related article/photo/video, and etc. If you analyze the check-in data from the location-based social network service in accordance with your situation, you can provide various customized services. Therefore, In this paper, we develop a location recommendation system based on LBSNS that can utilize the check-in data efficiently. This system analyzes the location category of the check-in data, determines the weighted value of it, and finds out the similarity between users by using the Pearson correlation coefficient. Also, it obtains the preference score of recommended locations by using the collaborated filtering algorithm and then, finds out the distance score by applying the Euclidean's algorithm to the recommended locations and the current users' locations. Finally, it recommends appropriate locations by applying the weighted value to the preference score and the distance score. In addition, this paper approved excellence of the proposed system throughout the experiment using real data.

Key Words : LBSNS, Check-in Data Analysis, Location Recommendation, User Similarity

* 본 논문은 중소기업청에서 지원하는 2013년도 산학연공동기술개발사업(No. C0114224)의 연구수행으로 인한 결과물임을 밝힙니다
Received 4 April 2014, Revised 9 May 2014
Accepted 20 June 2014
Corresponding Author: Ki-Joon Han(Konkuk University)
Email: kjhan@db.konkuk.ac.kr

ISSN: 1738-1916

© The Society of Digital Policy & Management. All rights reserved. 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

최근 사용자의 위치를 파악할 수 있는 기기가 보편화 됨에 따라 사용자의 현재 위치를 기반으로 장소를 추천해 주는 서비스들이 급증하고 있다. 그러나 이러한 서비스들은 단순히 거리 정보 제공의 성격을 띠고 있어 개인 맞춤형 서비스를 제공하는데는 아직까지 부족한 점들이 많다[1,2]. 개인 맞춤형 서비스를 위해서는 장소에 대해 방문 기록, 평가 리뷰 등의 누적된 데이터의 활용이 필요하다.

최근에는 SNS(Social Network Service) 성장으로 상호 관계가 있는 사용자들로부터의 검증된 장소 추천이 가능해져 모바일 위치 기반 장소 추천은 점점 진화하고 있다[3,4]. 또한, LBS(Location-Based Service)와 SNS가 결합된 위치 기반 소셜 네트워크 서비스(Location-Based Social Network Service, 이하 LBSNS)가 하나의 새로운 축을 형성하면서 SNS가 확장하는데 큰 역할을 담당해 오고 있다[5].

LBSNS는 유용한 지역정보, 생활정보, 할인정보 등의 서비스뿐만 아니라 체크인(Check-in) 데이터를 통해 장소에 대한 리뷰나 평가를 공유하고 체크인 한 위치의 주변 장소의 사람들과 커뮤니케이션을 할 수 있도록 해준다[6]. 이렇게 성장하고 있는 LBSNS 분야에서 최근 다양한 연구가 진행되고 있는데 위치 정보에 대한 정확도 및 정보 보안 등에 대한 기술 중심적인 연구, 행동학의 사용자 행위 관점에서의 연구, 새로운 커뮤니케이션의 수단으로 마케팅 관점에서의 연구 등이 이루어지고 있다[7].

또한, LBSNS 사용자가 증가하면서 대량으로 축적되는 데이터에 대한 활용 연구도 활발하다. 예를 들어, 개인 사용자 데이터를 기반으로 콘텐츠를 추천하는 연구, 사용자간의 공간적인 유사도와 사용자들의 체류 시간에 따라 다음 체크인 장소를 추천해주는 연구 등이 있다[8,9].

본 논문에서는 사용자의 체크인 데이터를 분석하여 카테고리 가중치를 도출하고 사용자들간의 유사도를 측정하며 사용자의 특정 장소에 대한 선호도를 예측한다. 그리고 이를 기반으로 하여 최종적으로 사용자의 위치와 가까운 최적의 장소를 추천한다.

본 논문에서 연구 개발한 LBSNS 기반 장소 추천 시스템의 프로세스는 다음과 같다.

첫째, 사용자의 체크인 데이터 중 장소 카테고리를 분석하여 가중치를 추출한다.

둘째, 피어슨 상관계수를 이용해 사용자간의 유사도를 산출한다.

셋째, 유사한 사용자들이 평가한 장소를 추출하고 협업적 필터링 알고리즘을 이용하여 추천 장소에 대한 선호도 점수를 구한다.

넷째, 추천 대상 장소의 위치와 사용자의 현재 위치에 대해 유클라디안 알고리즘을 적용하여 거리 점수를 산출한다.

다섯째, 선호도 점수와 거리 점수에 대해 가중치 계수를 적용하여 상황에 맞게 장소를 추천한다.

2. 관련 연구

본 장에서는 LBSNS의 연구 분야 중에서 본 논문에서 다룰 장소 추천과 관련된 최근 연구를 분석한다.

2.1 위치 기반 장소 추천 연구

쿼드 트리를 이용한 위치 기반 협업 필터링 추천 시스템[8]은 사용자의 위치 로그를 기반으로 사용자간 공간적 유사도를 구하여 선호도가 높은 아이템을 추천하는 시스템이다. 아이템에 대한 선호도 예측 수식은 식(1)과 같다.

$$x_{k,m}=\overline{u_k}+\frac{\sum_{u_a\in C_k} s_u(u_k,u_a)(x_{a,m}-\overline{u_a})}{\sum_{u_a\in C_k} s_u(u_k,u_a)} \quad (1)$$

식 (1)에서 $x_{k,m}$은 사용자 k가 아이템 m에 대해 예측한 선호도이다. u_k는 사용자 k가 아이템에 선호도를 평가한 평균이고, u_a는 사용자 a가 아이템에 선호도를 평가한 평균이다. c_k는 사용자 후보 집합이다. $x_{a,m}$은 사용자 a가 아이템 m에 대해 예측한 선호도이고, $s_u(U_k, U_a)$는 사용자 k와 최근접 이웃 사용자 a와의 유사도이다.

실생활 정보를 이용한 사용자의 의미 있는 장소 추출 방법 연구[9]는 통화와 문자 메시지 송수신 등의 상호 작용 정보와 체류 시간 및 거리를 이용하여 의미 있는 장소

를 추출한다. 의미 있는 장소 추출 수식은 식(2)와 같다.

$$locationscore(L_a) = w_1 \times staytime(L_a) + (1-w_1) \times \sum_{i \in L_a} totalproximity(U_i) \quad (2)$$

식 (2)에서 $locationscore(L_a)$는 친밀도와 체류 시간에 기반 한 장소 점수이다. $totalproximity(U_i)$는 상호작용 정보를 통해 산출된 친밀도이고, $staytime(L_a)$는 GPS 로그를 이용한 체류 시간이다. 그리고 W_l은 친밀도와 체류 시간에 적용할 가중치 값이다.

위에서 언급한 첫 번째 연구는 사용자의 체크인 데이터를 분석하여 아이템을 추천하였지만 다양한 사용자의 취향을 고려하지 못하였다는 한계가 있다. 그리고 두번째 연구는 체류 시간의 측정이 번거로워 서비스 적용에 어려움이 있을 수 있다. 또한, 두 연구 모두 체크인 데이터를 활용하긴 했지만 사용자의 반복적인 체크인 행위 패턴에 대해 분석하여 적용하지 못했다는 단점이 있다.

2.2 협업적 필터링 알고리즘

협업적 필터링 알고리즘은 특정 아이템에 대해 유사한 성향을 보였던 사용자들은 다른 아이템에 대해서도 유사한 성향을 보일 것이라는 가정을 바탕으로 추천을 수행하는 알고리즘이다[10]. 즉, 각 사용자와 유사한 성향을 가진 사용자가 구매하였거나 좋게 평가했던 아이템 중에서 사용자가 이전에 구매하지 않은 아이템을 추천하는 것이다. 이 알고리즘은 도서나 음반, 영화 등 정량적으로 분석하기 어려운 아이템에 적용했을 때 좋은 성능을 나타내는 것으로 알려져 있다.

협업적 필터링 알고리즘을 이용한 추천 프로세스는 다음과 같다.

① 사용자 프로파일을 생성한다.

사용자가 각 아이템에 대해서 5점(1~5점) 척도로 평가한 점수를 이용하여 프로파일을 생성한다.

② 고객간 유사도를 계산한다.

고객이 각 아이템에 대해 평가한 점수를 피어슨 상관계수를 이용하여 고객간의 유사도를 구한다. 사용자 a와 u의 유사도를 구하기 위한 수식은 식(3)과 같다.

$$W_{au} = \frac{\sum_{i=1}^{m}(r_{ai} - \overline{r_a}) \bullet (r_{ui} - \overline{r_u})}{\sigma_a \bullet \sigma_u} \quad (3)$$

식 (3)에서 w_{au}는 사용자 a와 u가 공통으로 평가한 아이템을 바탕으로 계산한 두 사용자 간의 유사도를 나타낸다. m은 두 사용자가 공통으로 평가한 아이템의 수를 의미하고, r_{ai}는 예측하고자 하는 사용자 a가 i번째 아이템에 대해 수행한 평가 점수이며, r_{ui}는 사용자 u가 i번째 아이템에 대해 수행한 평가 점수이다. 그리고 r_a와 r_u는 사용자 a와 u가 수행한 모든 아이템에 대한 평가 점수의 평균이며, σ_a 와 σ_u는 사용자 a와 u가 수행한 평가 점수의 표준 편차이다.

③ 선호도 예측값을 산출한다.

고객간 유사도를 이용하여 각 아이템 i에 대한 사용자 선호도 예측값을 구한다. 선호도 예측값을 구하기 위한 수식은 식(4)와 같다.

$$p_{ai} = \overline{r_a} \frac{\sum_{u=1}^{n}(r_{ui} - \overline{r_u}) \bullet W_{au}}{\sum_{u=1}^{n} W_{au}} \quad (4)$$

식 (4)에서 p_{ai}는 i번째 아이템에 대한 사용자 a의 예측값을 의미한다. 그리고 r_a와 r_u는 사용자 a와 u가 평가한 모든 아이템에 대한 평가 점수의 평균을 나타내고, r_{ui}는 사용자 u가 i번째 아이템에 대해 수행한 평가 점수의 평균을 나타낸다. w_{au}는 사용자 a와 u의 유사도를 나타내며, n은 유사한 선호도를 보이는 사용자 수를 나타낸다.

④ 추천을 수행한다.

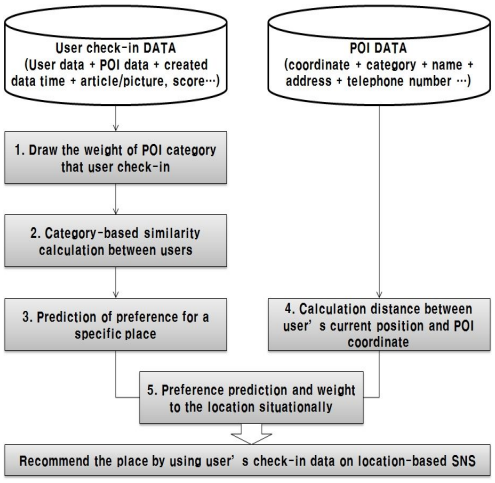
사용자에게 예측값이 높은 몇 개의 아이템을 결정해 최종 추천을 수행한다. 그리고 아이템을 추천할 때 평가 점수도 함께 피드백 받아 사용자 프로파일을 갱신한다.

3. LBSNS 기반 장소 추천 시스템

3.1 시스템 처리 과정

본 절에서는 협업적 필터링 알고리즘, 피어슨 상관계

수, 유클라디안 알고리즘 등을 기반으로 하는 LBSNS 기반 장소 추천 시스템의 주요 처리 과정에 대하여 설명한다. Fig. 1은 LBSNS 기반 장소 추천 시스템의 처리 과정을 보여준다.



[Fig. 1] Process of Location Recommendation System

[Fig. 1]에서 보는 바와 같이 LBSNS 기반 장소 추천 시스템은 사용자 체크인 데이터로부터 추출된 선호도 예측값과 POI 데이터로부터 계산된 거리를 기반으로 장소를 추천한다. 장소 추천 식은 식(5)와 같다.

$$V_u = \chi \sum_{j=1}^{m} P_{uj} + \epsilon \log \frac{1}{\sqrt{\sum_{j=1}^{m} (CL_u - PL_{uj})^2}} \cdot \chi + \epsilon = 1 \quad (5)$$

식 (5)에서 V_u는 사용자 u의 장소 추천 값이다. P_{uj}는 사용자 u의 j번째 장소에 대한 선호도 예측값이고, m은 추천 대상이 되는 장소의 개수를 의미한다. 그리고 CL_U는 사용자의 현재 위치 좌표이고, PL_{uj}는 추천 장소의 위치 좌표이다. □와 □는 상황에 따라 적용 가능한 가중치 계수이다.

3.2 체크인 데이터 분석

3.2.1 카테고리 가중치 값 도출

가) 데이터 수집

국내외 대표적인 LBSNS에서 체크인 행위시 수집되는 데이터를 분석한 결과 공통적으로 아래와 같은 데이터 구조를 형성한 것을 확인할 수 있었다. <Table 1>은 국내외 대표적인 LBSNS의 공통적인 체크인 데이터 구조를 보여준다.

〈Table 1〉 Check-in Data Structure

Location Information	User Information
POI Name	User Nickname
Category	Evaluation Score of Location
Coordinate/Address	Text/Picture/Movie

체크인 데이터는 장소 정보(POI명, 카테고리, 좌표, 주소)와 사용자 정보(사용자 닉네임, 장소 평가 점수, 남긴 글, 사진, 동영상) 등의 콘텐트로 구성되어 있다.

본 논문에서는 S서비스의 체크인 데이터를 활용하였으며 일반적으로 가장 많이 체크인하는 음식점에 대해 LBSNS 기반 추천 시스템을 적용시켜 보았다. 즉, 2012년 10월 1일부터 2012년 10월 31일까지의 S서비스에서 음식점 카테고리를 30회 이상 체크인 한 사용자 중 20명에 대한 체크인 데이터를 사용하였다.

<Table 2>는 'S서비스'의 사용자 체크인 데이터를 보여준다.

〈Table 2〉 User's Check-in Data of 'S-service'

LOC_NM	SCATE	ADDR	LAT	LNG	SCORE	…
STARBUCKS City Theater	Au	Yeoksam-dong 816	37.5	127	5	…
Caffebene Gwangjin	Au	Hwayang-dong 121	37.5	127	5	…
Paris Croissant	Ar	Seocho2-dong	37.5	127	0	…
Homestead Coffee Gangnam	Au	Teheran-ro	37.5	127	0	…
Mr.Donuts Gangnam	Au	Yeoksam-dong 112	37.5	127	0	…
…						

<Table 2>에서 LOC_NM은 POI명, SCATE는 카테고리, ADDR은 주소, LAT와 LNG는 좌표, SCORE는 장소 평가 점수이다. S서비스에서 음식점 카테고리는 <Table 3>과 같다.

〈Table 3〉 Restaurant Categories of 'S service'

SCATE	EXP	SCATE	EXP	SCATE	EXP
Aa	Hanjeongsik	Ai	Chicken	Aq	Shabu-Shabu
Ab	Pork Belly	Aj	Beef Intestines	Ar	Bakery
Ac	Ribs of Pork	Ak	Pettitoes	As	Hamburg
Ad	Bulgogi	Al	Steak House	At	Pizza
Ae	Marbled Beef	Am	See Food	Au	Coffee Shop
Af	Smoked Duck	An	Japanese Food	Av	Fusion Restaurant
Ag	Chicken Roast	Ao	Chinese Food	Aw	Italy Restaurant
Ah	Samgyetang	Ap	Curry	Ax	French Restaurant

나) 가중치 값 계산

20명 사용자들이 장소 체크인 횟수에 따른 카테고리 내역은 <Table 4>와 같다.

〈Table 4〉 Categories According to the Number of Check-in

$User_i$	1	2	3	4	5	…	…	…	m
User1	Ab	Aa	Ab	Ag	Ag	Ac	Ac	Af	Ab
User2	Ab	Aa	Ab	Af	Aa	Ab	Ab	Aa	Ab
…	Ag	Ag	Aa	Ab	Ac	Ag	Af	Ag	Ab
User20	Ab	Aa	Ag	Ab	Ag	Ab	Af	Ab	Af

사용자 u의 k 카테고리 가중치를 구하는 수식은 식(6)과 같다.

$$W_{uk} = \sum_{i=1}^{m} uf_{uk} \quad (6)$$

식 (6)에서 W_{uk}는 사용자 u의 체크인 데이터에 나타난 카테고리 k의 가중치를 나타낸다. m은 사용자 u가 체크인한 음식점 카테고리 개수를 나타내고, uf_{uk}는 사용자 u의 체크인 데이터에 나타난 카테고리 k의 출현 횟수를 의미한다.

<Table 5>는 <Table 4>를 기반으로 추출된 사용자들의 카테고리에 대한 가중치 값을 보여준다.

〈Table 5〉 Weight Values for User's Categories

User1		User2		User3		User4	
SCATE	CNT	SCATE	CNT	SCATE	CNT	SCATE	CNT
Au	73	Au	35	Au	25	Au	104
Ah	46	Ap	11	Aa	12	Av	41
Av	24	Af	9	At	10	Ap	33
Aq	16	Ah	7	Aq	9	Af	23
Aa	13	Av	6	Ap	5	Aq	21

3.2.2 사용자간 유사도 측정

본 논문에서는 피어슨 상관계수를 이용하여 사용자들 간의 카테고리 유사도를 측정한다. 유사도 값은 -1 < N < 1 이며 1에 가까울수록 비슷한 카테고리에 체크인을 한 사람이다. 사용자간의 유사도 측정 수식은 식(7)과 같다.

$$W_{uv} = \frac{\sum_{x=1}^{m} (S_{ui} - \overline{S}_u) \times (S_{vi} - \overline{S}_v)}{\sigma_u \times \sigma_v} \quad (7)$$

식 (7)에서 W_{uv}는 사용자 u와 사용자 v의 유사도이다. m은 장소의 개수를 나타낸다. S_{ui}와 S_{vi}는 사용자 u와 사용자 v가 각각 i번째 장소에 대해 체크인한 평가 점수이며, S_u와 S_v은 사용자 u와 v의 i번째 장소에 대해 체크인한 평가 점수의 평균이다. 그리고 σ_u와 σ_v은 사용자 u와 v가 체크인 장소들의 평가 점수의 표준 편차를 나타낸다.

<Table 6>은 사용자간 유사도 측정값을 보여준다.

〈Table 6〉 Similarity Values between Users

$User_i$	$User_j$	Similarity
User1	User2	0.281718085
	User3	-0.265580337
	User4	-0.917970806
User2	User3	0.576104636
	User4	-0.036749697
User3	User4	0.405469572

사용자간의 유사도를 이용하여 m개의 장소에 대해 특정 사용자가 어느 정도의 선호도를 가질 것인지 예측한다. 그리고 선호도 예측치가 높은 장소 중 사용자의 현 위치와 거리를 계산하여 최종 추천하게 된다.

3.2.3 특정 장소에 대한 선호도 예측

장소에 대한 선호도 예측값은 사용자들간의 카테고리 유사도를 이용하여 구한다. 사용자 u의 j번째 장소에 대한 선호도 예측값을 구하는 수식은 식(8)과 같다.

$$P_{uj} = \overline{S_u} + \frac{\sum_{j=1}^{m} (S_{vj} - \overline{S}_v) \times W_{uv}}{\sum_{j=1}^{m} W_{uv}} \quad (8)$$

식 (8)에서 P_{uj}는 사용자 u의 j번째 장소에 대한 선호

도 예측값을 나타낸다. m은 추천 대상이 되는 장소의 개수를 의미하며, S_u와 S_v는 사용자 u와 v의 j번째 장소에 대해 체크인한 평가 점수의 평균이다. 그리고 S_{vj}는 j번째 장소에 대한 사용자 v의 평가 점수를 나타내며, W_{uv}은 사용자 u와 v의 유사도를 나타낸다.

<Table 7>은 사용자 u의 장소에 대한 선호도 예측값을 보여준다.

〈Table 7〉 Preference Predictive Values for User's Specific Locations

User		LOC_NM	PREDICTION
User u	U2 Recommendation	Angel In-us Coffee	4.884515425
		Hollys Coffee 1	4.8
		Caffebene 1	4.940859042
		Caffebene 2	4.856343617
		Thinking Coffee	4.771828192
	U3 Recommendation	Dunkin Donuts	4.879674101
		Caffebene 3	4.853116067
		Caffebene 4	4.826558034
		Coffee Bean 1	4.773441966
		Ice Bean	4.720325899
	U4 Recommendation	Albero	4.763281168
		Hollys Coffee 2	4.671484087
		Coffee Bean 2	4.855078248
		Pascucci	4.579687006
		Caffebene 5	4.487889926

3.3 사용자와 추천 장소간의 거리 계산

본 논문에서는 선호도 예측값이 높은 순으로 추천 장소를 정렬하고 사용자 u의 현재 위치와 거리를 계산하여 가까운 장소를 최종 추천한다.

사용자의 현재 위치와 추천 장소와의 거리 계산은 유클라디안 알고리즘을 이용한 식(10)을 사용한다.

$$L_{uj}(Distance\ Between\ CL_u \text{ and } PL_{uj}) = \sqrt{\sum_{j=1}^{m}(CL_u - PL_{uj})^2} \quad (9)$$

식 (9)에서 CL_u는 사용자의 현재 위치의 좌표이고, PL_{uj}는 추천 장소의 좌표이다. <Table 8>은 사용자의 현재 위치와 추천 장소와의 거리 계산값을 보여준다.

〈Table 8〉 Distance Values between User's Current Location and Recommended Location

User's Coordinate	LOC_NM	X	Y	Distance Calculation
37.49817 127.02869	Angel In-us Coffee	37.496516	127.030704	0.00260613
	Hollys Coffee 1	37.499792	127.032366	0.004017942
	Caffebene 1	37.499432	127.035939	0.007358033
	Caffebene 2	37.504982	127.025081	0.00770897
	Thinking Coffee	37.505088	127.023075	0.008909935
	Dunkin Donuts	37.498549	127.02829	0.000551036
	Caffebene 3	37.499517	127.026283	0.002758271
	Caffebene 4	37.495959	127.025338	0.004015523
	Coffee Bean 1	37.50293	127.024725	0.006195065
	Ice Bean	37.506095	127.004467	0.025486454
	Albero	37.482375	127.044346	0.022239433
	Hollys Coffee 2	37.530144	126.964934	0.071324329
	Coffee Bean 2	37.479861	126.952084	0.078763562
	Pascucci	37.513394	126.943332	0.086705008
	Caffebene 5	37.510633	127.021025	0.014631425

사용자의 현재 위치와 추천 장소의 거리 계산값은 소수점으로 나오므로 역수로 하여 값이 큰 장소를 사용자 u의 현재 위치와 가까운 것으로 간주한다.

사용자 u의 추천 장소값 V_u를 보다 간편하게 계산하기 위하여 거리 계산값을 선호도 예측값과 동일한 자리수로 변환한다. 이 때 자리 수 변환은 로그 함수를 사용한다.

<Table 9>는 로그 함수를 사용한 거리 계산값 변환 내역을 보여준다.

〈Table 9〉 Conversion of Distance Values

LOC_NM	Distance Calculation Value	Reciprocal	Log Value
Angel In-us Coffee	0.00260613	383.710757	2.584003975
Hollys Coffee 1	0.004017942	248.883616	2.395996309
Caffebene 1	0.007358033	135.905893	2.133238287
Caffebene 2	0.00770897	129.7190085	2.113003621
Thinking Coffee	0.008909935	112.2342592	2.050125444
Dunkin Donuts	0.000551036	1814.762488	3.258819794
Caffebene 3	0.002758271	362.5459789	2.559363093
Caffebene 4	0.004015523	249.0335627	2.396257882
Coffee Bean 1	0.006195065	161.418818	2.207954163
Ice Bean	0.025486454	39.23652884	1.593690579
Albero	0.022239433	44.96517602	1.652876298
Hollys Coffee 2	0.071324329	14.02045545	1.146762122
Coffee Bean 2	0.078763562	12.69622619	1.103674651
Pascucci	0.086705008	11.53335924	1.061955819
Caffebene 5	0.014631425	68.34604341	1.834713378

3.4 가중치 적용한 장소 추천

추천 장소가 사용자의 현재 위치에서 거리가 너무 멀리 떨어져 있을 경우에는 아무리 선호도 예측값이 높다고 해도 사용자에게 의미가 없을 수 있다. 이런 경우 각 요소에 가중치를 적용하여 장소를 추천할 수 있다.

가중치 없이 추천 장소를 추출했을 경우 <Table 10>과 같은 결과가 나왔다.

〈Table 10〉 Score of Recommended Location without Weight

LOC_NM	PREDICTION	Distance Score	Score of Recommendation (no Weight)	
Dunkin Donuts	4.879674101	3.258819794	8.138493895	1
Angel In-us Coffee	4.884515425	2.584003975	7.4685194	2
Caffebene 3	4.853116067	2.559363093	7.41247916	3
Caffebene 4	4.826558034	2.396257882	7.222815915	4
Hollys Coffee 1	4.8	2.395996309	7.195996309	5
Caffebene 1	4.940859042	2.133238287	7.07409733	6
Coffee Bean 1	4.773441966	2.207954163	6.981396129	7
Caffebene 2	4.856343617	2.113003621	6.969347238	8
Thinking Coffee	4.771828192	2.050125444	6.821953636	9
Albero	4.763281168	1.652876298	6.416157466	10
Caffebene 5	4.487889926	1.834713378	6.322603304	11
Ice Bean	4.720325899	1.593690579	6.314016478	12
Coffee Bean 2	4.855078248	1.103674651	5.958752899	13
Hollys Coffee 2	4.671484087	1.146762122	5.818246209	14
Pascucci	4.579687006	1.061955819	5.641642826	15

거리 점수 보다 선호도 점수에 높은 가중치를 주었을 땐 <Table 11>과 같은 결과가 나왔다.

〈Table 11〉 Score of Recommended Location with High Preference Score Weight

LOC_NM	PREDICTION	Distance Score	Score of Recommendation (Preference score : Distance score = 0.8 : 0.2)	
Dunkin Donuts	4.879674101	3.258819794	4.55550324	1
Angel In-us Coffee	4.884515425	2.584003975	4.424413135	2
Caffebene 3	4.853116067	2.559363093	4.394365472	3
Caffebene 4	4.826558034	2.396257882	4.340498003	5
Hollys Coffee 1	4.8	2.395996309	4.319199262	6
Caffebene 1	4.940859042	2.133238287	4.379334891	4
Coffee Bean 1	4.773441966	2.207954163	4.260344406	8
Caffebene 2	4.856343617	2.113003621	4.307675618	7
Thinking Coffee	4.771828192	2.050125444	4.227487642	9
Albero	4.763281168	1.652876298	4.141200194	10
Caffebene 5	4.487889926	1.834713378	3.957254616	14
Ice Bean	4.720325899	1.593690579	4.094998835	12
Coffee Bean 2	4.855078248	1.103674651	4.104797529	11
Hollys Coffee 2	4.671484087	1.146762122	3.966539694	13
Pascucci	4.579687006	1.061955819	3.876140769	15

선호도 점수 보다 거리 점수 점수에 높은 가중치를 주었을 땐 <Table 12>와 같은 결과가 나왔다.

〈Table 12〉 Score of Recommended Location with High Distance Score Weight

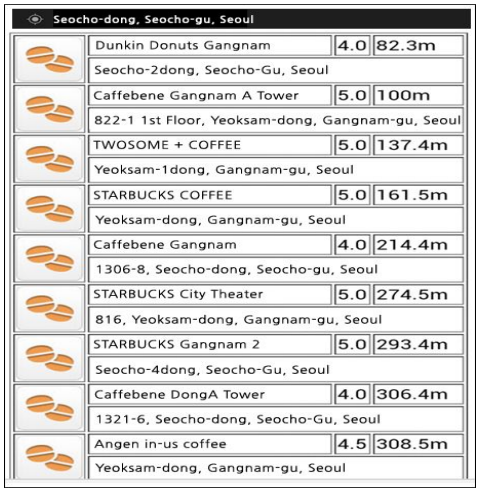
LOC_NM	PREDICTION	Distance score	Score of Recommendation (Preference score : Distance score = 0.2 : 0.8)	
Dunkin Donuts	4.879674101	3.258819794	3.582990655	1
Angel In-us Coffee	4.884515425	2.584003975	3.044106265	2
Caffebene 3	4.853116067	2.550363093	3.018113688	3
Caffebene 4	4.826558034	2.396257882	2.882317912	4
Hollys Coffee 1	4.8	2.395996309	2.876797047	5
Caffebene 1	4.940859042	2.133238287	2.694762438	7
Coffee Bean 1	4.773441966	2.207954163	2.721051724	6
Caffebene 2	4.856343617	2.113003621	2.66167162	8
Thinking Coffee	4.771828192	2.050125444	2.594465994	9
Albero	4.763281168	1.652876298	2.274957272	11
Caffebene 5	4.487889926	1.834713378	2.365348688	10
Ice Bean	4.720325899	1.593690579	2.219017643	12
Coffee Bean 2	4.855078248	1.103674651	1.85395537	13
Hollys Coffee 2	4.671484087	1.146762122	3.966539694	13
Pascucci	4.579687006	1.061955819	3.876140769	15

선호도 점수 및 거리 점수에 대한 가중치는 카테고리의 종류에 따라 적용하면 보다 의미 있는 추천이 될 수 있다. 예를 들어 카페 카테고리인 경우 가까운 곳으로 방문하는 경우가 많기 때문에 거리 점수에 더 높은 가중치를 줄 수 있고, 스테이크 또는 스파게티 등의 고급 레스토랑인 경우 거리가 멀어도 검증되어 있는 곳을 가고 싶어 하기 때문에 선호도 점수에 더 높은 가중치를 두어 장소를 추천하면 만족도 높은 결과를 얻을 수 있다.

4. 구현 및 실험

4.1 시스템 구현

본 논문에서는 S서비스의 체크인 데이터를 활용하여 스마트폰 모바일 웹 환경으로 구현하였다. [Fig. 2]는 본 논문에서 구현한 시스템의 모바일 웹 실행 화면을 보여준다.



[Fig. 2] Mobile Web Execution Screen of the Location Recommendation System

4.1.1 사용자의 체크인 데이터 분석

본 논문에서는 사용자의 체크인 데이터 분석은 다음과 같은 단계를 거쳐 수행된다.

(1) 사용자 u의 체크인 데이터를 가지고 지난 1개월 내 30회 이상 체크인을 했었는지 확인하고 음식점 카테고리에서 가장 많이 체크인한 카테고리 순으로 데이터를 추출한다. <Table 13>은 사용자 u의 체크인 데이터를 분석한 수행 결과를 보여준다.

〈Table 13〉 Analysis Result of User's Check-in Data

SCATE	SCATE_NM	CNT
Ak	Coffee Shop	65
Aa	Hanjeongsik	23
Av	Italy Restaurant	16
Ah	Pizza	13
Af	Hamburg	13

(2) 사용자 u와 유사한 카테고리를 갖는 유사한 사용자의 체크인 데이터를 분석한다. <Table 14>는 유사한 사용자 4명의 체크인 데이터를 분석한 수행 결과를 보여준다.

〈Table 14〉 Analysis Result of Similar User's Check-in Data

U1		U2	
SCATE	CNT	SCATE	CNT
Ak	73	Ak	35
Ah	46	Ap	11
Av	24	Af	9
Aq	16	Ah	7
Aa	13	Av	6
U3		U4	
SCATE	CNT	SCATE	CNT
Ak	25	Ak	104
Aa	12	Av	41
At	10	Ap	33
Aq	9	Af	23
Ap	9	Aq	21

(3) 유사한 사용자의 체크인 장소에 대한 위치 좌표와 평가 점수를 추출한다. 그리고 추출된 결과를 기반으로 추천 장소를 선정한다.

<Table 15>는 유사한 사용자의 체크인 장소에 대한 위치 좌표와 평가 점수를 추출한 수행 결과를 보여준다.

〈Table 15〉 Location Coordinate and Evaluation Score for Similar User's Check-in Location

LOC_NM	SCATE	ADDR	LAT	LNG	SCORE
Dunkin Donuts Gangnam	Ak	...	37.49855	127.0283	4
STARBUCKS City Theater	Ak	...	37.49817	127.0287	5
Caffebene Gangnam A Tower	Ak	...	37.49862	127.0285	5
TWOSOME+ COFFEE	Ak	...	37.49917	127.0274	5
STARBUCKS COFFEE	Ak	...	37.4993	127.027	5
.....					

(4) 사용자 u의 현재 위치와 추천 장소의 위치에 대하여 거리를 계산하여 가까운 장소를 추출한다.

<Table 16>은 사용자의 현재 위치와 추천 장소와의 거리를 계산한 수행 결과를 보여준다.

〈Table 16〉 Distance Result between User's Current Location and Recommended Location

current location coordinate of user U latitude: 37.497963, longitude : 127.027721						
LOC_NM	SCATE	ADDR	LAT	LNG	SCORE	DIST
STARBUCKS City Theater	Ak	...	37.49817	127.0287	5	88.54044
Caffebene Gangnam A Tower	Ak	...	37.49862	127.0285	5	99.98726
TWOSOME+ COFFEE	Ak	...	37.49917	127.0274	5	137.3635
STARBUCKS COFFEE	Ak	...	37.4993	127.027	5	161.5144
STARBUCKS Gangnam 2	Ak	...	37.50005	127.0257	5	293.3949
.....						

4.1.2 가중치를 부여한 최종 장소 추천

본 논문에서는 선호도 점수와 거리 점수 가중치를 부여하여 최종 장소 추천 과정은 다음과 같은 단계를 거쳐 수행된다.

(1) 유사한 사용자의 추천 장소 50개 중에 사용자 현재 위치와 가까운 기준으로 정렬한다.

<Table 17>은 추천 장소를 사용자 현재 위치 기준으로 정렬한 수행 결과를 보여준다.

〈Table 17〉 Sort Result of Recommended Locations by using User's Current Location

LOC_NM	SCATE	ADDR	LAT	LNG	SCORE	DIST
Dunkin Donuts Gangnam	Ak	...	37.49855	127.0283	4	82.26193
STARBUCKS City Theater	Ak	...	37.49817	127.0287	5	88.54044
Caffebene Gangnam A Tower	Ak	...	37.49862	127.0285	5	99.98726
STARBUCKS COFFEE	Ak	...	37.4993	127.027	5	161.5144
.....						

(2) 사용자의 현재 위치와 가까운 추천 장소 중에 평가 점수가 높은 순으로 정렬한다.

<Table 18>은 추천 장소를 평가 점수 기준으로 정렬

한 수행 결과를 보여준다.

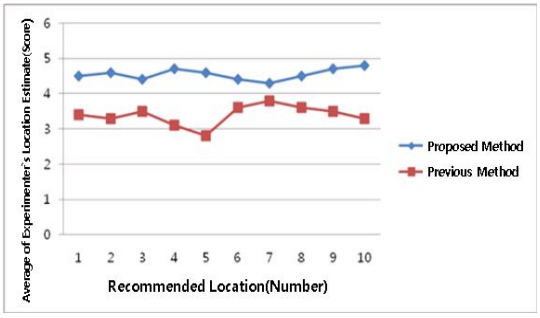
〈Table 18〉 Sort Result of Recommended Locations by using Evaluation Score

LOC_NM	SCATE	ADDR	LAT	LNG	SCORE	DIST
STARBUCKS City Theater	Ak	...	37.49817	127.0287	5	88.54044
Dunkin Donuts	Ak	...	37.50495	127.0244	5	830.4709
Can More Noryangjin	Ak	...	37.51383	126.9416	5	7801.598
Joo-Coffee Yongsan	Ak	...	37.5289	126.9658	5	6451.505
TOM N TOMS Sinnonhyun	Ak	...	37.50435	127.0237	5	795.1785
.....						

4.2 만족도 비교

본 논문에서 제시한 방법을 이용하여 장소 추천을 할 경우 사용자뿐만 아니라 유사한 사용자들의 체크인 데이터를 분석하여 장소를 추천받기 때문에 LBSNS 기반 개인 맞춤형 장소 추천이 가능해지게 된다.

본 논문에서 제안한 방법과 기존의 공간적 유사도를 기반으로 추천 장소를 추출한 방법을 비교하기 위해 실험자 20명을 대상으로 10개 추천 장소에 대해 만족도 평가를 진행하였다. [Fig. 3]은 10개의 추천 장소에 대한 만족도 평가 수치를 보여준다.



[Fig. 3] Comparison Proposed Method with Previous Method

[Fig. 3]을 보면 만족도 평가 결과 본 논문에서 제시한 방법이 기존 방법보다 만족도가 40% 향상된 것을 알 수 있다. 따라서 본 논문에서 제시한 방법이 기존 방법보다 사용자 맞춤형 추천 장소를 추출하는데 더욱 효율적이라 판단된다.

5. 결론

본 논문에서는 최근 성장하고 있는 LBSNS 분야에서 사용자의 체크인 데이터를 분석하는 LBSNS 기반 장소 추천 시스템을 연구 개발하였다. 본 논문에서 개발한 LBSNS 기반 장소 추천 시스템은 장소 카테고리 추출, 사용자간의 유사도, 협업적 필터링 알고리즘, 피어슨 상관계수, 유클라디안 알고리즘, 가중치 계수 등을 기반으로 사용자 맞춤형 장소를 추천할 수 있다.

그리고 본 논문에서는 실제 사용자의 체크인 데이터를 이용한 실험을 통해 제안 방법이 기존 연구 방법 보다 우수하다는 것을 입증하였다. 그러나 본 논문에서 사용한 협업적 필터링 알고리즘은 많은 사용자의 체크인 데이터가 있어야 한다는 단점이 있다. 또한 새로운 장소를 추천하고자 할 때 그 장소가 사용자로부터 평가를 받았어야만 추천할 수 있다. 따라서, 추후 협업적 필터링 알고리즘 이외에 내용 기반 필터링 알고리즘 등을 활용한 장소 추천 연구가 필요할 것이다.

ACKNOWLEDGMENTS

This work (Grants No. C0114224) was supported by Business for Cooperative R&D between Industry, Academy, and Research Institute funded Korea Small and Medium Business Administration in 2013.

REFERENCES

[1] H. Y. Kim and J. H. Oh, The current state and social meaning of domestic and foreign SNS. Hongik University, 2012.

[2] J. M. Ryu, C. P. Hong, K. B. Kang, D. H. Kang, D. Y. Yang and J. W. Jwa, Development of mobile context awareness restaurant recommendation services. The Korea Contents Association, Vol. 7, No. 5, pp. 138-145, 2007.

[3] B. I. Ahn, Location-based services of the mobile service. KT Economic Research, Trend Report, 2011.
[4] eMarketer, Social Network Ad Revenues Worldwide. Market Trend Report, 2012.
[5] J. Y. Oh, Comparison analysis of Korean, American and Japanese's SNS service. National Information Society Agency, IT Policy Research Series, Vol. 11, 2009.
[6] S. Y. Kim, Pronoun of LBS the change of Foursquare. KT Economic Research, Trend Report, 2012.
[7] S. B. Cha, An analysis of structural equation models on university students' Social Network Service participation and learning outcomes. Konkuk University, 2011.
[8] H. C. Shin, A location-based collaborative filtering recommender using quadtree. Yonsei University, 2011.
[9] S. H. Lee, B. K. Kim, T. B. Yoon and J. H. Lee, The method for extraction of meaningful places based on behavior information of user, Korean Institute of Intelligent Systems, Vol. 20, No. 4, pp. 503-508, 2010.
[10] S. J. Bae, Study of personalized recommendation algorithms for new books on internet bookstores. Chungnam University, 2003.

정 구 임(Jung, Ku-Imm)

- 2005년 2월 : 백석대학교 경상학부 (경영학사)
- 2013년 2월 : 건국대학교 정보통신학과 (공학석사)
- 2010년 7월 ~ 현재 : (주)씨온 전략마케팅 팀장
- 관심분야 : LBS, LBSNS 플랫폼, 빅데이터, 위치 기반 모바일 플랫폼
- E-Mail : jki0721@seeon.kr

안 병 익(Ahn, Byung-Ik)

- 1993년 2월 : 동국대학교 컴퓨터공학과(공학석사)
- 2007년 2월 : 연세대학교 컴퓨터과학과 (공학박사)
- 2002년 10월 : 스탠포드대학교 경영대학원 SEIT수료
- 2010년 5월 ~ 현재 : (주)씨온 대표이사
- 2010년 9월 ~ 현재 : 건국대학교 정보통신대학원 겸임교수
- 관심분야 : LBS, SNS, 모바일 컴퓨팅, 소셜 컴퓨팅
- E-Mail : biahn99@seeon.kr

김 정 준(Kim, Jeong-Joon)

- 2003년 2월 : 건국대학교 컴퓨터 공학과(공학사)
- 2005년 2월: 건국대학교 컴퓨터공학과(공학석사)
- 2010년 8월 : 건국대학교 컴퓨터공학과(공학박사)
- 2013년 3월 ~ 현재 : 건국대학교 정보통신대학 컴퓨터공학부 조교수
- 관심분야 : 공간/시공간 데이터베이스, 공간 빅데이터, 시멘틱 웹, GIS, LBS, USN
- E-Mail : jjkim9@db.konkuk.ac.kr

한 기 준(Han, Ki-Joon)

- 1979년 2월 : 서울대학교 수학교육학과(이학사)
- 1981년 2월 : KAIST 전산학과(공학석사)
- 1985년 2월 : KAIST 전산학과(공학박사)
- 1985년 3월 ~ 현재 : 건국대학교 정보통신대학 컴퓨터공학부 교수
- 관심분야 : 공간 데이터베이스, GIS, LBS, 텔레메틱스, 정보시스템 감리
- E-Mail : kjhan@db.konkuk.ac.kr