

위치기반 소셜 네트워크에서 시간과 사용자 활동을 고려한 개인화된 POI 추천

Recommending Personalized POI Considering Time and User Activity in Location Based Social Networks

이규남*, 임종태**, 복경수**, 유재수**
충북대학교 빅데이터학과*, 충북대학교 정보통신공학과**

Kyunam Lee(rbska5645@chungbuk.ac.kr)*, Jongtae Lim(jtlim@chungbuk.ac.kr)**,
Kyoungsoo Bok(ksbok@chungbuk.ac.kr)**, Jaesoo Yoo(yjs@chungbuk.ac.kr)**

요약

위치 인식 기술의 발전 및 스마트 디바이스 사용의 활성화로 인해 위치 기반 서비스와 소셜 네트워크를 결합하여 사용자에게 정보를 공유하는 위치 기반 소셜 네트워크(LBSN: Location Based Social Network)이 활성화되고 있다. 위치 기반 소셜 네트워크에서 사용자의 체크인 기능을 이용하여 사용자가 가 흥미 있어 할 만한 장소를 추천하는 연구가 활발히 이루어지고 있다. 본 논문은 위치기반 소셜 네트워크에서 시간과 사용자 활동을 고려한 장소 추천 기법을 제안한다. 제안하는 기법은 기존 논문에서 고려하지 못한 시간에 따른 사용자의 선호도 변화와 지역의 전문가, 희귀한 장소에 대한 사용자의 관심을 고려한다. 다시 말해, 사용자의 선호도 변화를 고려하기 위해 시간에 따른 체크인 이력을 사용하고 지역의 전문가를 판별하기 위해 사용자 활동 영역을 구분한다. 그리고 사용자가 선호하는 장소에 가중치를 주기 위하여 희귀한 장소를 고려한다. 다양한 성능평가를 통해 제안하는 기법이 기존 기법에 비해 성능이 우수함을 보인다.

■ 중심어 : | 위치 기반 소셜 네트워크 | 장소 추천 | POI | 개인화 추천 | 행렬인수분해 | 랜덤워크 |

Abstract

With the development of location-aware technologies and the activation of smart phones, location based social networks(LBSN) have been activated to allow people to easily share their location. In particular, studies on recommending the location of user interests by using the user check-in function in LBSN have been actively conducted. In this paper, we propose a location recommendation scheme considering time and user activities in LBSN. The proposed scheme considers user preference changes over time, local experts, and user interest in rare places. In other words, it uses the check-in history over time and distinguishes the user activity area to identify local experts. It also considers a rare place to give a weight to the user preferred place. It is shown through various performance evaluations that the proposed scheme outperforms the existing schemes.

■ keyword : | POI(Point Of Interest) | LBSN(Location Based Social Network) | Location Recommendation | Personalized Recommendation | Matrix Factorization | Random walk With Restart |

*본 연구는 과학기술정보통신부 및 정보통신기술진흥센터의 대학ICT연구센터육성지원사업(IITP-2017-2013-0-00881), 2015년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No.2015R1D1A3A01015962), 2016년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No.2016R1A2B3007527)

접수일자 : 2017년 10월 11일

심사완료일 : 2017년 11월 02일

수정일자 : 2017년 11월 02일

교신저자 : 유재수, e-mail : yjs@chungbuk.ac.kr

1. 서론

위치 인식 기술의 발전으로 사용자의 위치를 기반으로 특정 정보를 제공하는 위치 기반 서비스가 다양한 분야에서 활용되고 있다[1-3]. 최근 무선 통신 기술의 발전과 함께 스마트 폰과 같은 모바일 디바이스의 활용이 증가됨에 따라 위치 기반 서비스와 소셜 네트워크 서비스를 결합한 위치 기반 소셜 네트워크(LBSN: Location Based Social Network) 서비스가 대두되었다[3][4]. 위치 기반 소셜 네트워크는 소셜 네트워크 관계인 사용자가 방문한 장소 이력과 리뷰 정보를 활용하여 정보를 제공하는 서비스이다. 예를 들어, Foursquare[1]와 TripAdvisor[2]는 사용자의 방문한 장소 정보를 다른 사용자들에게 제공하거나 공유하는 대표적인 위치 기반 소셜 네트워크 서비스로 Foursquare의 경우 사용자들은 체크인(check-in)을 통해 자신의 위치를 공유하고 장소에 대한 의견 또는 만족도를 부여하고 해당 정보는 소셜 네트워크의 사용자들에게 제공한다.

위치 기반 소셜 네트워크에서 사용자가 받아들일 수 있는 한계치를 넘는 정보를 받으면 서비스 이용에 혼란을 겪게 될 것이므로 사용자에게 정보를 필터링하고 제공하여 사용자에게 편의성을 제공하는 것이 필요하다. 이때, 사용자에게 유의미한 정보를 제공해야 하므로 성향을 정확히 파악하여 사용자가 선호할 만한 정보를 제공하는 것이 중요하다. 특히 장소 정보는 사용자의 방문 용이성으로 인해 선호할 만한 정보를 제공하는 것이 더욱 중요하다. 위치 기반 소셜 네트워크는 장소 이력과 리뷰 정보가 풍부하여 이를 기반으로 사용자의 성향을 파악하고 추천에 활용이 용이하다. 이로 인하여 사용자 선호도가 반영된 사용자가 흥미 있어 할만한 장소(POI: Point Of Interest)를 추천하는 기법이 활발히 연구되기 시작했다[3-8].

장소 추천을 위해서는 사용자의 개인 선호도를 잘 판별할 수 있어야 한다. 이를 위해 사용자가 방문했던 장소 이력을 활용하거나 사용자가 작성한 점수 및 리뷰를 활용하여 사용자의 개인 선호도를 판별할 수 있다. 또한 사용자의 현 위치 및 시간에 따라 각 장소 별 방문 용이성 및 선호 장소가 변화하기 때문에 사용자의 상태

정보를 활용해야 한다. 특히 현 위치에서 주로 활동하는 전문가들을 선별하여 추천에 적극 활용해야 한다. 이는 해당 지역에서 주로 활동한 사용자는 여러 장소 중 자신의 선호가 정확히 반영된 장소를 방문하므로 선호도가 정확히 반영된 장소의 선출에 보다 유의미하게 활용할 수 있기 때문이다. Tuan, Debnath[7][8]은 시간에 따른 체크인 이력을 활용하여 사용자의 선호도 변화를 고려하여 사용자가 흥미 있어 할 만한 장소를 추천한다. 그러나 다른 사용자가 방문하지 않는 희귀한 장소를 자주 방문하는 사용자에게 대한 고려가 부족했다는 단점이 있으며 지역의 전문가를 활용하지 못한다는 단점이 있다. Ying[9]는 개인화 추천을 위하여 사용자의 친구가 방문한 장소, 유사한 성향의 사용자가 방문한 장소, 추천을 받고자 하는 지역의 사용자들이 많이 방문한 장소의 3가지 요소를 고려하여 추천에 활용하였다. 그러나 3가지 요소를 완벽하게 동시에 고려하지 못하고 3가지 요소 중 2가지 요소만을 고려하여 그린 6개의 그래프 중 랜덤하게 그래프를 선택하고 확률적인 추론을 통해 추천에 활용한 단점이 존재한다. 또한 시간에 따른 사용자의 선호도 변화를 고려하지 못하였다. H. Bagci[10]는 장소에 대한 지식이 많은 사용자를 선출하기 위하여 HITS 알고리즘을 사용하여 활용하였으나 장소를 추천하지 않고 사용자의 다음 행동만을 예측하는데 그쳤다.

본 논문에서는 위치 기반 소셜 네트워크 환경에서 시간과 활동 영역을 고려한 POI 추천 기법을 제안한다. 제안하는 기법은 사용자의 활동 영역을 고려하여 지역 전문가를 활용하고 시간을 고려하여 선호도 변화에 따른 장소 추천 기법을 제안한다. 이를 통해 기존의 지역 전문가 및 시간에 따른 선호도 변화를 고려하지 못한 문제를 해결한다. 또한 제안하는 기법은 사용자가 특히 선호하는 의미 있는 장소를 고려하여 추천에 활용한다. 이를 통해 기존 논문에서 다른 사용자보다 유달리 선호하는 장소를 고려하지 못했던 문제를 해결한다.

본 논문의 구성은 다음과 같다. II장에서는 관련 기존 연구들을 분석한다. III장에서는 제안하는 기법을 기술한다. IV장에서는 제안하는 기법과 기존 기법의 성능 평가 비교를 통해 제안하는 기법의 우수성을 입증한다.

마지막 V장에서는 본 논문의 결론을 제시한다.

II. 관련 연구

위치 기반 서비스에서 사용자가 추천을 받고자 할 때 대중적인 선호도를 기반으로 추천을 받게 되면 사용자의 개인 선호도를 정확히 반영할 수 없게 된다. 그로 인해 위치 기반 서비스에서 사용자의 개인 선호도를 활용하여 개인화된 서비스를 제공하기 위한 목적으로 추천 기법에 대한 연구가 활발히 진행되어 왔다.

Tuan[7]은 개인 선호도를 고려하기 위해 사용자의 시간에 따른 사용자의 선호도 변화를 고려하였으며 이동 방향에 따라 장소 방문 가능성이 변화한다는 것을 고려하여 개인화된 장소추천을 시도하였다. 사용자의 이동 방향을 고려하기 위해 사용자가 이동하는 방향에 있는 장소에 가중치를 부여하였다. 또한 사용자의 선호도 변화를 고려하기 위해 월 별, 시즌 별, 6개월, 1년으로 다양한 기간의 가중치를 고려하여 사용자의 선호도 변화를 고려하였다. 그러나 사용자가 낮과 밤의 선호하는 장소가 다를 수 있다는 점을 간과하였다. 또한 지역의 전문가 및 희귀한 장소에 대한 선호도를 고려하지 못했다.

Debnath[8]은 시간에 따른 사용자의 선호도 변화를 세밀하게 반영하기 위해 1시간 단위로 사용자 및 타 사용자의 체크인 횟수를 이용하여 해당 시간대에 사용자가 흥미를 가질 장소를 추천한다. 이를 위해 사용자의 카테고리 이동 확률과 특정 시간의 개인 선호도를 활용하여 장소의 특정 시간의 대중적 선호도와 함께 활용하였다. 사용자의 카테고리 이동 확률은 시간 대 별로 사용자가 특정 카테고리에서 이동하여 다른 카테고리를 체크인 한 비율을 기준으로 구한다. 그러나 사용자가 유달리 선호하는 희귀한 장소에 대한 고려는 하지 못하였으며 지역의 전문가를 활용하지 못하였다. H. Bagci[10]는 장소에 대한 지식이 많은 사용자를 선출하기 위하여 HITS 알고리즘을 사용하여 활용하였으나 장소를 추천하지 않고 사용자의 다음 행동만을 예측하는데 그쳤다.

Ying[9]는 사회적 관계로 인해 방문한 요인, 사용자 개인 선호로 인해 방문한 요인, 대중적 선호로 인해 방문한 요인의 3가지 요인을 고려한 장소 추천 기법을 제안하였다. 장소 방문은 사회적 관계와 개인선호, 대중적 선호도로 영향 받는다는 가정 하에 이를 측정하여 개인화된 장소 추천에 활용하였다. 사회적 관계로 인해 방문한 요인을 고려하기 위해 사용자 체크인에 기반한 코사인 유사도와 사용자의 체크인에 기반해 구한 중심점(Base-point)을 이용하여 사용자 간 물리적 거리를 고려하였다. 다음으로 사용자 개인 선호로 인해 방문한 요인을 고려하기 위하여 해당 카테고리에 체크인한 비율로 카테고리 선호도를 구하고 사용자의 체크인에 기반해 highlight의 선호도를 고려하였다. highlight는 장소의 “카페” 카테고리에서 마신 “커피”와 같이 것을 통칭하여 의미한다. 마지막으로 대중적 선호로 인해 방문한 요인을 고려하기 위해 사용자와 사회적 관계를 맺은 사람들이 장소에 체크인한 비율과 모든 사용자가 장소에 체크인 한 비율을 이용하였다.

추천을 할 때 사용자가 아이템을 어느 정도 선호하는지 알면 추천이 손쉬울 것이다. 이를 알기 위하여 사용자와 아이템 간의 명시적인 평점이나 혹은 아이템 사용 이력과 같은 암묵적인 평점으로 아이템에 대한 사용자 별 평점 점수 행렬을 만들어 낼 수 있다. 이를 통해 사용자가 선호할 만한 아이템을 추천에 활용할 수 있다. 그러나 사용자와 관계를 정의할 수 없는 아이템의 경우 값을 정의할 수 없게 되고, 이는 사용자와 현재 관계를 정의할 수 없는 아이템은 추천을 할 수 없는 문제가 생기게 된다. 이를 해결하기 위한 방법 중의 하나로 Matrix Factorization(MF) 방법이 존재 한다 [14][16][17]. MF 기법은 명확한 평가나 관계 정보가 있을 경우 이를 바탕으로 행렬 분해를 했을 경우 잠재적인 요인을 잘 정의한다고 알려져 있다[14][16][17]. 보통 특성 행렬 U와 M으로 목표 행렬 A와의 차이 값을 기울기 강하(Gradient Descent)기법을 사용하여 특성 행렬을 학습한다[16][17]. 그러나 MF 방법은 결측치를 예측할 때 사용자와 결측치의 수가 많아질수록 추정 될 수 있는 경우의 수가 많아지기에 점점 예측하기가 힘들어지고 사용자와 관계없는 아이템의 숫자가 어느 임계

점을 넘어가는 순간 결측치로 추정되는 값의 경우의 수가 너무 많아져서 예측이 힘들어지게 될 것이다 [15][16].

사용자와 관계를 정의할 수 없는 아이템을 추천하기 위해 그래프를 이용하여 추천에 활용하는 다양한 연구가 진행되었다[5][11]. Anastasios Noulas, J. J. Ying[6][9]는 LBSN 환경에서 POI를 추천하기 위한 목적으로 그래프 랭킹 기반 추천 기법중 하나인 Random walk With Restart(RWR)를 이용하여 추천에 사용했다. 진우정[11]은 랜덤 서퍼가 이웃 정점으로 이동하는 것에 더해 일정 확률로 어느 정점으로 이동을 하는 것으로 랜덤 워크를 페이지랭크와 RWR로 세부 기법을 분류하였다. 페이지 랭크는 랜덤 워크의 세부 기법의 하나로 그래프에서 정점들의 상대적인 중요성을 계산하는 방법이다[11]. 페이지 랭크에서 랜덤 서퍼는 임의의 정점에서 시작하여 타 정점으로 무작위로 이동가능하며 간선로 연결되어 있는 유무와 관계없이 임의의 정점으로 이동할 수 있다. 그에 반해 RWR은 랜덤 서퍼가 임의의 정점으로 이동이 불가능하며 오직 간선로 연결된 다른 정점으로 이동하거나 정해진 한 정점으로 이동할 수 있다[9-12]. 즉, 페이지 랭크는 그래프에서 글로벌 랭킹을 구하지만 RWR은 시작 정점에 맞춘 개인화된 중요성을 고려하여 계산하기 때문에 시작 정점에게 추천을 하기에 적합한 랭킹 기법이다. 또한 앞서 설명한 바와 같이 MF는 행렬에서 빈 곳이 많을수록 정확도에 영향을 받지만 RWR은 빈 곳보다는 그래프 구조에 영향을 받는 데다 POI 추천의 특성 상 결측치가 많은 장소에 대한 사용자 별 선호도 값 행렬을 이용하여 추천할 때는 MF 보다 제안하는 RWR이 적합하다. 이는 Anastasios Noulas[6]의 실험평가를 통해 입증된 바와 같다. 본 논문에서는 MF와 제안하는 RWR의 결합을 통해 사용자에게 POI를 추천한다.

사용자 활동 영역을 고려한 POI 추천기법을 제안한다. 본 논문에서는 사용자의 선호도가 정확히 반영된 장소를 추천에 활용하기 위해 자신의 선호가 정확히 반영된 장소를 방문하는 지역의 전문가를 활용한다. 또한, 사용자의 정확한 선호도를 반영하기 위해 사용자가 유달리 선호하는 희귀한 장소를 고려해야 하고 시간에 따른 사용자의 선호도 변화를 POI 추천에 활용 한다. 제안하는 기법은 시간과 활동 영역을 고려하여 MF와 제안하는 RWR을 결합하여 개인화 장소 추천을 수행한다.

[그림 1]은 본 논문에서 제안하는 기법의 전체 구조이다. 제안하는 장소 추천 기법은 학습과 장소 추천 단계로 구성된다. 학습 단계는 사용자의 과거 장소 방문 이력들을 기반으로 사용자간 유사도와 활동 영역과 전문가를 판별한다. 다른 사용자들이 방문하지 않는데 사용자가 많이 방문한 희귀한 장소에 가중치를 부여하고 Non-negative Matrix Factorization(NMF)[18]를 활용하여 카테고리 선호도를 계산하는 단계이다. 장소 추천 단계는 사용자의 현 위치/시간을 기반으로 유사 사용자와 지역의 전문가를 활용하여 RWR에 기반한 카테고리 별 장소의 중요도를 계산하고 학습에서 계산된 값을 함께 활용하여 최종적으로 사용자에게 POI를 추천하는 단계이다.

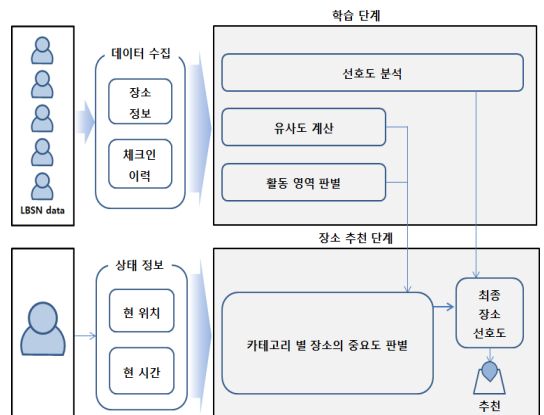


그림 1. 제안하는 POI 추천기법의 전체 구조

III. 제안하는 POI 추천 기법

1. 시스템 전체 구조

본 논문은 위치 기반 소셜 네트워크 환경에서 시간과

2. 학습 단계

2.1 사용자 유사도 계산

사용자가 방문하지 않은 장소에 대한 선호도를 구하기 위해서는 먼저 사용자의 장소 카테고리 선호도를 파악해야한다. 예를 들어, 사용자가 방문하지 않은 A와 B 장소가 있다고 하자. 이때, A장소는 클럽이고 B장소는 공원일 때 평소 사용자가 클럽보다는 공원을 선호했다면 사용자에게 B 장소를 추천해주는 것이 적절할 것이다. 다음으로 파악된 선호도를 기반으로 나와 유사한 사용자를 선별 및 활용하여 그들이 방문했던 장소를 추천할 수 있다. 예를 들어, 자신과 유사한 선호도를 보이는 사용자 k가 B 장소를 방문했다고 하자. 나와 선호도가 유사한 사용자 k는 장소 A보다는 장소 B를 선호하여 방문했기에 나 또한 장소 A보다는 장소 B를 더 선호할 것이다. 본 논문에서는 이를 파악하기 위해 학습 단계에서 선호도 분석 및 사용자 유사도를 계산한다.

[그림 2]는 학습 단계 중 유사도 계산 및 선호도 분석 절차를 보여준다. 먼저, 수집 모듈에서 수집된 사용자의 체크인 이력들을 시간대별로 분할한다. 시간대는 t_1 부터 t_4 까지 총 4개의 시간대로 분할하며 각 시간대의 범위는 $00시 \leq t_1 < 06시$, $06시 \leq t_2 < 12시$, $12시 \leq t_3 < 18시$, $18시 \leq t_4 < 00시$ 와 같다. 분할한 시간대 별로 사용자의 카테고리 방문 통계 값을 활용하여 사용자의 카테고리 방문 비율 행렬을 만들고 코사인 유사도를 통해 사용자간 유사도를 계산한다.

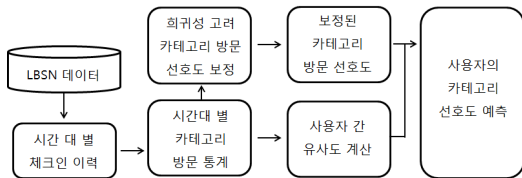


그림 2. 유사도 계산 및 선호도 분석 절차

2.2 회귀성을 고려한 카테고리 선호도 분석

다른 사용자들이 잘 방문하지 않는데 특히 사용자가 많이 방문한 장소에 대해서는 사용자가 다른 사용자에 비해 선호하는 장소라는 것을 의미하기 때문에 가중치를 반영하여 선호도를 보정해야한다. 다른 사용자들이 잘 방문하지 않는데 유달리 사용자가 선호한 장소에 대해서 가중치를 계산하기 위하여 [식 1]을 사용한다. w

는 특정 카테고리에 속하는 각 장소의 선호도이다. V_i 는 해당 장소에 사용자 i 가 방문한 횟수이다. AV 는 해당 장소에 모든 사용자가 방문한 횟수이다. CP 는 해당 카테고리의 하위 장소들이다. α 는 가중치를 얼마 만큼 반영할 것인지에 대한 보정치이다. [식 1]에서 해당 장소에 사용자가 방문한 횟수와 전체 사용자가 방문한 횟수를 고려하여 전체 사용자가 방문한 횟수가 적을 때 상대적으로 사용자가 많이 방문한 경우를 판별했다. 또한, 카테고리의 하위 장소들을 고려하여 해당 카테고리의 장소가 흔한 장소일 경우 흔하지 않은 장소보다 방문될 가능성이 많고, 이로 인해 방문 통계 값이 당연히 더 높을 수밖에 없는 현상을 미리 방지한다. 사용자 A가 놀이동산을 선호한다고 가정하자. 놀이동산은 사용자 A가 선호하는 장소이지만 음식점보다는 쉽사리 방문하기 힘든 장소이다. 그런 이유로 사용자의 카테고리 별 선호도 행렬을 구성하게 되면 사용자는 놀이동산보다는 음식점을 더 많이 방문하여 음식점을 더 선호한다는 결과가 나오게 된다. 해당 부분의 보정을 위하여 제안하는 기법에서는 카테고리의 하위 장소의 수를 활용하였다. [식 2]는 [식 1]을 이용하여 계산한 최종 사용자의 카테고리 선호도이다. CL_i 는 사용자의 i 번째 카테고리의 최종 선호도이다. CU_i 는 사용자의 i 번째 카테고리의 방문 횟수이다.

$$w_i = \log_{10} \left(\frac{V_i}{AV} * \frac{1}{CP} + 1 \right) \quad (1)$$

$$CL_i = \alpha * \left(\frac{CU_i}{\sum CU_i} \right) + (1 - \alpha) * (\sum w) \quad (2)$$

보정된 사용자의 카테고리 선호도 행렬의 결측치를 예측하기 위한 목적으로 협업필터링 기반의 NMF 방법을 수행한다. 협업 필터링을 수행하기 위해 계산된 유사도를 통해 유사한 사용자를 선출하고 해당 사용자들을 활용해 구성된 보정된 사용자의 카테고리 선호도 행렬에 NMF 방법을 활용하여 사용자가 방문하지 않은 카테고리의 선호도를 예측한다. 사용자의 카테고리 선호도 행렬은 사용자의 장소 선호도 행렬과는 다르게 결측치가 다수 존재하지 않기 때문에 NMF 방법을 활용

하여도 다수의 결측치로 인한 성능 저하가 발생하지 않는 경우가 대부분이다. 차후 산출된 사용자의 카테고리 선호도는 카테고리 별 장소의 중요성과 곱하여 사용자에게 추천한다. 이때 카테고리 별 장소의 중요성은 0~1 값이 나오므로 1이하의 값을 두 번 곱하는 경우를 없애기 위하여 카테고리의 선호도 값을 0~100 값으로 정규화를 한다.

2.3 주 활동 영역 판별

추천을 받고자 할 때 유사성만을 고려하게 되면 사용자가 위치한 지역에 몇 번 방문해보지 않은 사용자도 포함되어 추천에 활용 될 수 있다. 이러한 문제를 해결하기 위한 목적으로 사용자가 위치한 지역에서 주로 활동하는 사용자를 선별한다. 해당 지역에서 주로 활동하는 사용자는 지역에 위치한 각 장소들에 대해 지식이 많이 자신이 선호하는 선호도가 잘 반영된 장소들을 방문할 것이다. 이는 유사한 사용자가 선호도가 정확히 반영되지 않은 장소에 방문하고 해당 장소가 추천에 활용되는 오류를 감소시킬 수 있다.

사용자의 체크인 이력을 기반으로 각 지역을 분할하고 사용자들의 주 활동 영역을 판별한다. 해당 방법을 활용하면 해당 지역에 자주 방문하여 자신의 선호도가 정확히 반영된 장소를 방문하는 사용자들을 선별할 수 있다. 먼저, 사용자들의 체크인 이력을 기반으로 K-means를 수행하여 활동 영역을 군집화한다. K-means clustering은 각 개체들을 K개의 그룹으로 분할하여 군집화하는 방법으로 비계층적 군집분석중 하나로 널리 사용되는 방법 중 하나이다[19][20]. 군집화된 활동 영역을 및 각 활동 영역 별로 사용자가 체크인한 이력을 기반으로 [식 3]을 수행하여 사용자의 주 활동 영역을 판별한다. MR 은 사용자의 주 활동 영역을 의미한다. RC_i 은 사용자 A가 각 영역 별로 사용자가 체크인한 횟수를 의미한다. [그림 3]은 학습 단계에서 활동 영역을 판별 예시이다. 예시에서는 전체 사용자의 체크인 이력을 기반으로 K-means clustering을 수행한 결과 활동 영역이 5개로 군집화되었다고 가정한다. 그 중 사용자 A는 R_1 에 3번 체크인을 했고 R_2 에 0번, R_3 에 9번, R_4 에 2번, R_5 에 1번 체크인을 해서 [식 3]을 따

라 가장 많이 체크인한 영역인 R_3 를 사용자 A의 주 활동 영역으로 선출한다.

$$MR = \max(RC_i) \tag{3}$$

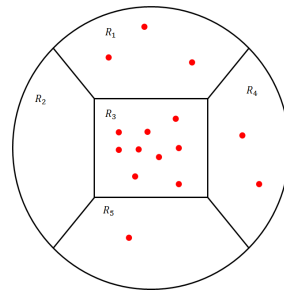


그림 3. 활동 영역 판별 예시

3. 장소 추천 단계

장소 추천 단계는 사용자가 요청할 때 사용자의 현재 위치와 시간을 기준으로 장소를 추천해주는 단계이다. [그림 4]는 장소 추천 과정을 보여준다. 먼저, 사용자의 요청이 들어오면 수집 모듈로부터 사용자의 현재 위치와 시간 값을 입력 받는다. 그리고 현재 사용자의 시간과 위치를 기반으로 사용자의 위치가 속한 영역이 주 활동 영역이면서 동시간대의 사용자 중 사용자와 유사한 사용자를 선출한다. 다음으로 지역의 대중적인 장소에 대해 지식이 많은 지역의 전문가들을 선출하기 위한 목적으로 사용자가 속한 시간대와 영역에서 각 카테고리 별로 사용자들이 많이 방문한 대중적인 장소를 선출한다. 최종 전문가는 선출된 장소에 방문한 사용자들 중에서 각 카테고리 별로 방문 횟수가 가장 높은 사용자들을 선출한다. 그 후 그래프를 활용하여 장소의 중요성을 판별하고 이를 이용하여 최종적으로 장소의 선호도를 예측한다.



그림 4. 장소 추천 과정

[그림 5]는 유사 사용자와 카테고리 별 선출된 지역

의 전문가를 활용하여 카테고리 별로 그래프를 구성한 예를 보여준다. 각 그래프에서 장소는 해당 카테고리에 해당하는 장소만을 정점(Vertex)으로 사용한다. 먼저, 처음에는 요청한 사용자를 기준으로 선출된 유사한 사용자와 지역의 전문가를 방향성 없는 간선(Edge)으로 연결한다. 다음으로 전문가를 선출할 때 활용했던 대중적인 장소를 전문가와 연결하고 유사한 사용자가 방문했던 장소와 유사한 사용자를 연결한다. 최종적으로 그래프는 추천을 요청한 사용자로부터 거리가 2 이내인 그래프로 표현된다.

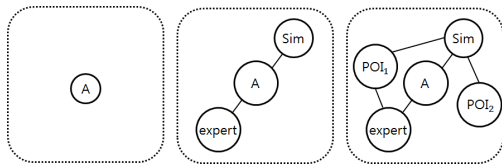


그림 5. 그래프 구성 예시

구성된 그래프는 간선에 가중치가 있는 그래프이다. 간선의 종류는 사용자를 연결하는 간선과 사용자와 장소를 연결하는 간선으로 크게 두 가지로 분할된다. [식 4]은 사용자를 연결하는 간선의 가중치를 계산하는 식이다. uu_i 는 사용자를 연결하는 간선을 의미한다. β 는 가중치이고 SC는 사용자 A와의 유사도 정도이다. PC는 해당 그래프의 카테고리를 방문한 횟수이다. [식 5]은 사용자와 장소를 연결하는 간선의 가중치를 계산하는 식이다. up_j 는 사용자와 장소를 연결하는 간선을 의미한다. AP는 사용자의 전체 체킨인 횟수이다. P_i 는 해당 장소에 사용자가 방문한 횟수이다.

$$uu_i = \beta * SC + (1 - \beta) * PC \quad (4)$$

$$up_j = \frac{P_i}{AP} \quad (5)$$

구성된 그래프를 기반으로 RWR을 활용하여 상대적인 장소의 중요성을 판별한다. 본 논문에서는 가중치가 부여된 방향성 있는 그래프를 기반으로 RWR을 활용한다. [그림 6]은 본 논문에서 사용하는 그래프 예시이다. 그래프 구성단계에서 구성된 방향성이 없는 간선에 가

중치가 있는 그래프를 기반으로 랜덤 서퍼가 정점에서 다른 정점으로 이동할 확률을 [식 9]를 활용하여 계산한다. 확률 계산은 두 가지 경우가 존재한다. 먼저, 정점과 연결된 다른 정점의 속성이 동일한 경우이다. 예를 들어, [그림 6]에서 사용자 A가 시작 정점일 때, 랜덤 서퍼는 유사한 사용자와 전문가 정점으로 이동이 가능하다. 유사한 사용자와 전문가 정점은 둘 모두 사용자 속성이므로 속성이 동일하다. 속성이 같은 경우의 [식 6]을 활용하여 계산 시 $\frac{ac_i}{\sum_{i=1}^n ac_i}$ 는 1이기 때문에 $we_{jk} = edge_{jk}$

이 된다. 다음으로, 정점과 연결된 다른 정점의 속성이 상이할 경우 [식 6]을 활용해서 방향성이 없는 그래프에서 랜덤 서퍼가 이동할 확률을 산출한다. we_{jk} 는 j번째 정점에서 k번째 정점으로 이동할 확률이다. ac는 현재 랜덤 서퍼가 있는 정점에 연결되어 있는 다른 정점 중 속성 i에 해당하는 정점의 개수이다.

$$we_{jk} = \frac{ac_i}{\sum_{i=1}^n ac_i} \times edge_{jk}, \quad \forall i \quad (6)$$

예를 들어, [그림 6]의 왼쪽 하단의 그래프에서 sim 정점에 있는 랜덤 서퍼가 이동을 수행할 때, sim 정점은 사용자 A 정점과도 연결이 되어있고 장소 정점들 (POI₁, POI₂)과도 연결이 되어있다. 이 경우 정점의 속성이 상이 하므로 동일하게 가중치를 적용하는 것은 적절하지 않다. 그러므로 [식 6]을 활용하여 속성이 상이한 경우의 확률을 계산한다. 예제에서 sim 정점으로부터 POI₁ 정점으로 이동할 확률은 $\frac{2}{3}$ 이다. 방향성이 없는 그래프에서 간선의 가중치는 0.25이므로 이를 방향성이 있는 간선의 가중치 그래프로 구성하기 위해 [식 6]을 활용하면 $\frac{2}{3} * 0.25$ 를 계산하여 약 0.17이 방향성 있는 그래프의 sim 정점으로부터 POI₁ 정점을 연결하는 간선의 가중치가 된다. 방향성 없는 그래프에서 다른 정점들에도 동일한 과정을 일괄 적용하면 최종적으로 방향성 있는 가중치 그래프의 각 정점에서 다른 정점으로

이동할 확률이 계산된다.

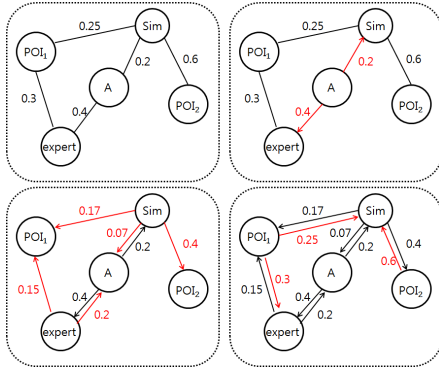


그림 6. 방향성 있는 가중치 그래프 구성 예시

각 카테고리 별 방향성 있는 가중치 그래프를 기반으로 RWR를 수행하면 각 카테고리 별로 장소의 상대적인 중요성이 계산된다. 제안하는 기법에서는 계산된 장소의 상대적인 중요도 값과 학습 단계에서 계산했던 사용자의 카테고리 선호도를 결합하여 최종 장소 추천을 수행한다. 만약 사용자가 선호했던 카테고리라면 선호도가 높으므로 해당 카테고리에 해당하는 장소들의 중요도 값이 증가할 것이고, 반대로 선호하지 않았던 카테고리라면 선호도가 낮으므로 해당 카테고리에 해당하는 장소들의 중요도 값이 감소할 것이다. 그러나 선호하지 않는 카테고리에 속하는 장소들 중에서도 중요도가 높은 장소라면 최종적으로 계산된 장소의 선호도가 높으므로 추천 목록에 포함될 수 있다. 반대로 선호하는 카테고리 중에서도 상대적으로 중요도가 떨어진다면 추천 목록에 포함 되지 않을 수 있다.

IV. 성능 평가

본 논문의 성능평가 환경은 Windows 7 Enterprise K 64bit i5 8792MB RAM에서 수행되었으며 R version 3.3.3 64 bit로 구현되었다. 제안하는 기법의 성능평가를 위해 Weeplace 사용자 약 8,000명의 약 1,000,000건의 사용자들의 체크인 데이터를 활용하였다. [표 2]는 성능평가 데이터 요약으로 New York의 2009/04/26 ~

2011/06/28의 기간 동안 4,789명의 686,910건의 체크인 수와 San Francisco의 2003/12/05~2011/06/28의 기간 동안 3,189명의 307,550건의 체크인 데이터를 이용하여 평가를 수행하였다. 카테고리는 Arts & Entertainment, Education, Universities, Food, Outdoors, Home / Work / Other, Nightlife, Parks, Shops, Travel의 총 10개로 구성하였다.

표 2. 성능평가 데이터 요약

도시	New York	San Francisco
기간	2009/04/26~2011/06/28	2003/12/05~2011/06/28
총 사용자 수	4,789	3,189
총 카테고리 수	10	10
총 체크인 수	686,910	307,550

성능평가를 위해 데이터를 2010/10/28 10:00pm을 기준으로 훈련 데이터와 테스트 데이터로 분할하였다. 분할된 테스트 데이터 기간 동안 사용자가 체크인 한 장소를 정답 셋으로 간주하였다. 그 후 New York의 훈련 데이터로 기법을 훈련시키고 San Francisco의 테스트 데이터를 기반으로 테스트를 진행하였다. 사용자의 활동 영역을 판별하기 위해 K-means clustering을 수행하였다. K-means clustering에서 정점 k의 개수를 정하기 위해 k의 개수를 1~15로 변화시키면서 분할된 군집의 표준편차를 통해 적절한 k의 개수를 추론하였다.

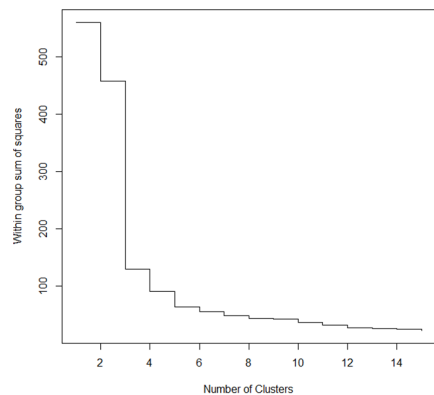


그림 7. San Francisco의 k의 변화에 따른 집단의 표준편차

[그림 7]은 San Francisco의 k의 개수의 변화에 따른 집단의 표준편차를 보여준다. 그림을 보면 군집을 5개까지 분할할 때는 표준편차의 변화의 폭이 크다는 것을 알 수 있지만 6개부터는 분할을 지속해도 표준편차의 변화의 폭이 작아 더 이상 분할을 하는 것이 크게 의미없음을 알 수 있다. 본 논문에서는 k의 개수를 5로 설정하였다.

본 논문에서 사용자가 설정하는 모수는 회귀성 고려가중치 α 와 유사한 사용자와 지역 전문가 가중치 β 와 제안하는 RWR에서 시작 정점으로 돌아가는 확률 c 이다. 세 개의 모수의 값을 구하기 위해 α 와 β 와 c 는 0~1까지 0.1간격으로 값을 변화시키면서 대입하여 정확도를 계산했다. 그 결과 α 는 0.6, β 는 0.3, c 는 0.9를 대입하였을 때가 가장 정확도가 높았다. 시작 정점으로 돌아가는 확률이 높은 점이 특이한데 이는 시작 정점에서 최대 먼 정점까지의 거리가 최대 2인 작은 그래프이기 에 발생하는 것 이라 사료된다.

성능평가를 위해 4가지 방법[6][8][20]과 비교를 하였다. PLTSRS[8]는 시간의 변화에 따른 사용자의 선호도를 고려한 기법이다. CF(collaborative filtering)[6][20]는 사용자와 성향이 유사하다고 판단되는 사용자의 장소 선호도를 기반으로 사용자가 방문하지 않은 장소의 선호도를 예측하는 방법이다. MF는 사용자 별 장소의 선호도 행렬을 이용하여 사용자가 방문하지 않은 장소의 선호도 값을 예측하는 기법이다[6]. WRWR(Weighted RWR)은 친구와 친구가 방문한 장소를 정점으로 설정하고 사용자 i 와 사용자 j 를 연결하는 간선의 가중치는 사용자 i 정점에 연결된 간선의 개수를 역수 취해서 설정하고 사용자 i 에서 장소 j 를 연결하는 간선의 가중치는 전체 사용자가 장소 j 에 방문한 횟수를 기준으로 사용자 i 가 장소 j 에 방문한 비율로 설정하고 장소 j 에서 사용자 i 의 가중치는 사용자 i 가 장소 j 에 방문한 비율로 설정하는 방법이다[6]. 본 논문에서는 New York의 데이터를 기반으로 San Francisco에서 테스트를 하는 방법을 수행하는데 해당 방법은 거리가 멀어 친구의 수가 적어 결과가 의미 없을 정도의 적은 값이 나와서 사용자의 친구를 사용자와 유사한 사용자로 대체하고 사용자 i 에서 사용자 j 를 연결하는 간선의

가중치는 사용자 i 에서 사용자 j 가 유사한 정도 설정하고 수행하였다. 정확도와 재현율, F-measure는 각 [식 9, 10, 11]을 이용하였다. 정확도는 제안하는 기법이 추천한 장소 중 정답 셋과 일치되는 비율을 뜻한다. 즉, 제안하는 기법이 정답이라 예측한 것 중 정답인 것의 비율을 뜻한다. 재현율은 정답 집합 중 제안하는 기법이 추천한 장소와 일치되는 비율을 뜻한다. 즉, 실제 정답인 것 중에서 제안하는 기법이 정답이라 예측한 것의 비율을 뜻한다. F-measure는 정확도와 재현율의 조화 평균으로 두 값의 중요성을 동일하게 두고 평균을 계산한 결과이다.

$$precision = \frac{recommended\ POI \cap ground\ truth}{Number\ of\ Recommended\ POIs} \quad (9)$$

$$recall = \frac{recommended\ POI \cap ground\ truth}{Number\ of\ Ground\ truth} \quad (10)$$

$$F-measure = 2 \times \left(\frac{precision \times recall}{precision + recall} \right) \quad (11)$$

[그림 8]은 추천하는 개수를 1, 3, 5로 변화시키면서 기법 별 정확도를 측정된 결과이다. 1개 추천 할 때 제안하는 기법을 제외하고 성능이 가장 좋은 기법과 제안하는 기법을 비교하면 약 172%정도 성능이 좋아진 것으로 관측되었다. 마찬가지로 3개, 5개 추천할 때 약 131%, 121% 성능이 향상되는 것을 볼 수 있었다. 기존 논문에서 고려하지 못한 사용자의 시간에 따른 선호도 변화와 주 활동 영역을 고려함으로써 사용자가 현재 위치한 장소의 정보가 많은 사용자들을 선별적으로 선택 활용하여 정보가 없는 사용자들을 활용하게 되는 경우를 없앨 수 있었으며 회귀한 장소의 선호도를 보정한 결과 다른 사용자에 비해 사용자가 유달리 선호하는 장소에 대한 고려를 할 수 있었고 이를 통해 사용자의 개인 선호도를 보다 정확히 반영할 수 있었다. 본 논문의 제안하는 기법의 정확도가 1개 추천할 때에서 3개 추천할 때 많이 낮아지는 이유는 테스트 기간 동안 사용자의 체크인 수가 적기 때문에 발생하는 현상이다. 특히 제안하는 기법은 시간대를 분할하여 추천을 해주고 있

어서 각 시간 대 별로 사용자의 체크인 수가 적절히 있어야 하지만 테스트 기간 동안 사용자가 각 시간대에 체크인 한 횟수가 1인 경우가 많아 1개를 맞춘 후에는 무엇을 추천해도 무조건 틀리는 경우가 발생하여 정확도가 많이 감소하게 되었다. 위와 같은 고려 사항으로 인해 기존 기법대비 성능 향상을 나타내게 되었다.

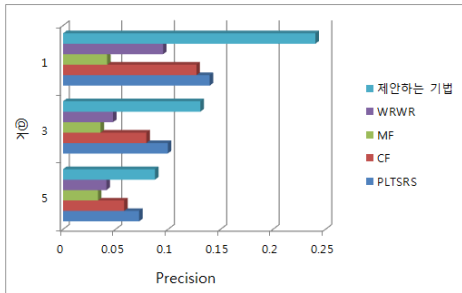


그림 8. 기법 별 @k precision

[그림 9]는 추천하는 개수를 1, 3, 5로 변화시키면서 기법 별 재현율을 측정된 결과이다. 제안하는 기법을 제외하고 성능이 가장 좋은 기법과 제안하는 기법의 차이를 살펴본 결과 약 212%, 204%, 179%만큼의 성능 향상을 보였다.

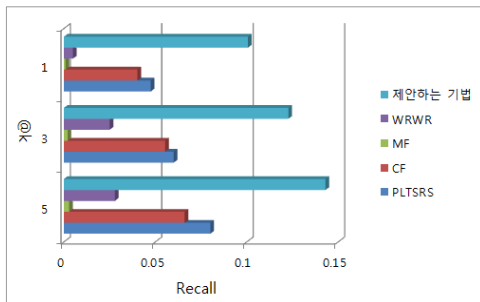


그림 9. 기법 별 @k recall

[그림 10]은 추천하는 개수를 1, 3, 5로 변화시키면서 기법 별 F-measure를 측정된 결과이다. 마찬가지로 안하는 기법을 제외하고 성능이 가장 좋은 기법과 제안하는 기법의 차이를 살펴본 결과 약 179%, 169%, 143%만큼의 성능 향상을 보였다.

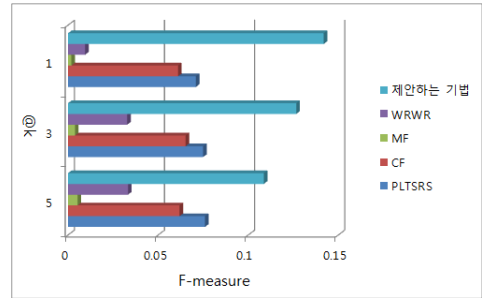


그림 10. 기법 별 @k F-measure

V. 결론

본 논문은 사용자의 체크인 데이터를 기반으로 개인화된 POI 추천 기법을 제안하였다. 제안 기법은 Foursquare나 Tripadvisor 등과 같은 위치 기반 소셜 네트워크에서 여행 장소를 추천하거나 기존에 방문하지 않은 새로운 장소를 추천하는데 활용된다. 사용자의 주활동 영역, 시간, 장소의 희귀성, 전문가 정보를 활용하여 장소를 추천하기 때문에 추천 정확도를 향상시킨다. 또한, RWR을 기반으로 사용자의 카테고리 별로 중요한 장소들을 선출하기 때문에 새로운 장소를 추천할 수 있다. 실험 평가를 수행한 결과 제안 기법이 기존 기법 중 가장 성능이 뛰어나게 나온 PLTSRS에 비해서 약 172%, 131%, 121% 만큼의 정확도 향상을 보였으며 약 212%, 204%, 179%만큼의 재현율이 향상되었다. 제안하는 기법은 체크인 수가 감소할 경우 일부 성능이 저하되는 문제점이 있다. 또한, 사용자 활동 시간대를 4구간으로 분할하여 24구간으로 분할한 경우보다 시간에 따른 선호도 변화를 민감하게 산출하지 못하는 문제가 발생한다. 향후 연구로 체크인 수가 적은 환경에서도 정확도를 향상시키고 시간대를 다양하게 적용하기 위한 연구를 수행할 예정이다. 또한 제안하는 기법을 실제 추천시스템에 적용하여 프로타입 시스템을 개발할 예정이다.

참고 문헌

- [1] <https://ko.foursquare.com/>
- [2] <https://www.tripadvisor.com/>
- [3] 김석현, 김지욱, 김현정, 박동규, “위치 기반 서비스를 이용한 스마트폰 관광 정보 시스템,” 멀티미디어학회논문지, 제15권, 제5호, pp.677-691, 2012.
- [4] 김서경, B. Ahmed, 강재우, “LBSN (Location-based Social Network) 의 위치정보 데이터를 활용한 사용자 이동경로 추천 시스템에 관한 서베이 연구,” 한국정보과학회 학술발표, pp.280-282, 2015.
- [5] V. Martínez, F. Berzal, and J. C. C. Talavera, “A Survey of Link Prediction in Complex Networks,” ACM Computing Surveys, Vol.49, No.4, pp.1-33, 2017.
- [6] A. Noulas, S. Scellato, N. Lathia, and C. Mascolo, “A Random Walk around the City: New Venue Recommendation in Location-Based Social Networks,” Proc. International Conference on Privacy, Security, Risk and Trust and International Conference on Social Computing, pp.144-153, 2012.
- [7] C. Tuan, C. Hung, and Z. Wu, “Collaborative location recommendations with dynamic time periods,” Pervasive and Mobile Computing, Vol.35, pp.1-14, 2017.
- [8] M. Debnath, P. K. Tripathi, and R. Elmasri, “Preference-Aware Successive POI Recommendation with Spatial and Temporal Influence,” Proc. International Conference on Social Informatics, pp.347-360, 2016.
- [9] J. J. Ying, W. Kuo, V. S. Tseng, and E. H. Lu, “Mining User Check-In Behavior with a Random Walk for Urban Point-of-Interest Recommendations,” ACM Transactions on Intelligent Systems and Technology, Vol.5, No.3, pp.1-26, 2014.
- [10] H. Bagci and P. Karagoz, “Random walk based context-aware activity recommendation for location based social networks,” Proc. International Conference on Data Science and Advanced Analytics, pp.1-9, 2015.
- [11] 진우정, 정진홍, 강유, “랜덤 워크를 활용한 그래프 랭킹 기반 추천 시스템,” 정보과학회지, 제34권, 제6호, pp.30-35, 2016.
- [12] L. Lu and T. Zhou, “Link Prediction in Complex Networks: A Survey,” CoRR abs/1010.0725, 2010.
- [13] L. Laszlo, “Random walks on graphs: A survey,” Combinatorics, Paul Erdős is Eighty, Vol.2, pp.1-46, 1993.
- [14] Y. Koren, R. M. Bell, and C. Volinsky, “Matrix Factorization Techniques for Recommender System,” IEEE Computer, Vol.42, No.8, pp.30-37, 2009.
- [15] B. Vandereycken, “Low-rank matrix completion by Riemannian optimization,” SIAM Journal on Optimization, Vol.23, No.2, pp.1214-1236, 2013.
- [16] E. J. Candès and B. Recht, “Exact Matrix Completion via Convex Optimization,” Foundations of Computational Mathematics, Vol.9, No.6, pp.717-772, 2009.
- [17] 김상화, 오병화, 김문중, 양지훈, “협력적 필터링과 콘텐츠 정보를 결합한 영화 추천 알고리즘,” 정보과학회논문지: 소프트웨어 및 응용, 제39권, 제4호, pp.261-268, 2012.
- [18] D. D. Lee and H. S. Seung, “Algorithms for Non-negative Matrix Factorization,” Proc. Annual Conference on Neural Information Processing Systems, pp.556-562, 2000.
- [19] W. S. Bae and S. W. Roh, “A Study on K - Means Clustering,” Communications for Statistical Applications and Methods, Vol.12,

No.2, pp.497-508, 2005.

[20] G. Adomavicius and A. Tuzhilin, "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions," IEEE Transactions on Knowledge and Data Engineering, Vol.17, No.6, pp.734-749, 2005.

저 자 소 개

이 규 남(Kyunam Lee)

준회원



- 2016년 2월 : 충북대학교 심리학 과(문학사)
- 2016년 2월 : 충북대학교 정보통 신공학과(이학사)
- 2016년 3월 ~ 현재 : 충북대학 교 빅데이터협동과정 석사과정

<관심분야> : 빅데이터, 추천 시스템, 머신러닝, 자연어 처리, 그래프처리 등

임 중 태(Jongtae Lim)

정회원



- 2009년 2월 : 충북대학교 정보통 신공학과(공학사)
- 2011년 2월 : 충북대학교 정보통 신공학과(공학석사)
- 2015년 8월 : 충북대학교 정보통 신공학과(공학박사)

• 2015년 9월 ~ 현재 : 충북대학교 정보통신공학과 박사후연구원 (Post.doc)

<관심분야> : 시공간 데이터베이스 시스템, 이동 객체 질의 처리, 위치기반 서비스, P2P 네트워크, 빅데이터, 그래프 처리, 머신러닝 등

복 경 수(Kyoungsoo Bok)

중심회원



- 1998년 2월 : 충북대학교 수학과 (이학사)
- 2000년 2월 : 충북대학교 정보통신공학과(공학석사)
- 2005년 8월 : 충북대학교 정보통신공학과(공학박사)

• 2005년 3월 ~ 2008년 2월 : 한국과학기술원 정보전자연구소 Postdoc

• 2008년 3월 ~ 2011년 2월 : 가인정보기술 연구소 연구원

• 2011년 3월 ~ 현재 : 충북대학교 전자정보대학 정보통신공학부 초빙교수

<관심분야> : 데이터베이스 시스템, 이동 객체 데이터베이스, 이동 P2P 네트워크, 소셜 네트워크 서비스, 빅데이터 등

유 재 수(Jaesoo Yoo)

중심회원



- 1989년 2월 : 전북대학교 컴퓨터공학과(공학사)
- 1991년 2월 : 한국과학기술원 전산학과(공학석사)
- 1995년 2월 : 한국과학기술원 전산학과(공학박사)

• 1995년 2월 ~ 1996년 8월 : 목포대학교 전산통계학과 전임강사

• 1996년 8월 ~ 현재 : 충북대학교 전자정보대학 정교수

<관심분야> : 데이터베이스 시스템, 멀티미디어 데이터베이스, 센서 네트워크, 바이오 인포메틱스, 빅데이터 등