

Mobile App Recommendation using User's Spatio-Temporal Context

Younggil Kang[†] · Seyoung Hwang^{**} · Sangwon Park^{***} · Soowon Lee^{****}

ABSTRACT

With the development of smartphones, the number of applications for smartphone increases sharply. As a result, users need to try several times to find their favorite apps. In order to solve this problem, we propose a recommendation system to provide an appropriate app list based on the user's log information including time stamp, location, application list, and so on. The proposed approach learns three recommendation models including Naive-Bayesian model, SVM model, and Most-Frequent Usage model using temporal and spatial attributes. In order to figure out the best model, we compared the performance of these models with variant features, and suggest an hybrid method to improve the performance of single models.

Keywords : App Recommendation, Mobile, Personalization, Spatio-Temporal Context, Machine Learning

사용자의 시공간 컨텍스트를 이용한 모바일 앱 추천

강 영 길[†] · 황 세 영^{**} · 박 상 원^{***} · 이 수 원^{****}

요 약

스마트폰을 통해 사용자에게 제공되는 앱이 증가함에 따라 사용자들은 스마트폰에서 자신이 사용하고자 하는 앱을 매번 찾아야 하는 문제점이 커지고 있다. 이러한 앱 탐색 문제를 해결하기 위해 본 연구에서는 스마트폰에서 시간 및 장소에 따른 사용자별 앱 사용 로그를 수집하고, 이를 학습하여 사용자의 상황 정보에 따라 최적의 앱 추천 리스트를 자동으로 제공하는 방법을 제안한다. 제안 방법에서는 사용자의 앱 로그로부터 요일, 시간대, 주중주말 여부 등의 시간 속성과 주소명, POI 등의 장소 속성을 학습하여 최대사용빈도기반 예측 모델, Naive-Bayesian 예측 모델, SVM 예측 모델 등의 3가지 예측 모델을 생성한다. 최적의 예측 모델을 생성하기 위해 다양하게 조합된 학습 속성들을 학습한 예측 모델들의 추천 정확도 비교 실험을 진행하였으며, 단일 예측 모델의 성능 개선을 위한 하이브리드 추천 방법을 제안한다.

키워드 : 앱 추천, 모바일, 개인화, 시공간 컨텍스트, 기계 학습

1. 서 론

애플의 아이폰 출시로부터 시작된 스마트폰 시장은 현재까지 급격하게 성장하고 있으며, 이에 따라 스마트폰을 통해 사용자에게 제공되는 앱의 수도 기하급수적으로 증가하고 있다. 대표적인 예로, 스마트폰 보급 초기에는 안드로이드 마켓에 매달 신규 등록되는 앱의 수가 약 2000건 정도

였으나, 최근에는 약 4만 건의 앱이 마켓에 신규 등록될 정도로 앱의 생산량이 급증하였다. 그러나 안드로이드 마켓으로부터 다양한 앱을 다운받아 사용하는 사용자들은 스마트폰에 저장되어 있는 수많은 앱 중에서 자신이 사용하고자 하는 앱을 매번 찾아야 하는 어려움이 있다.

이러한 스마트폰 내에 저장된 기존 앱 탐색의 문제를 해결하기 위한 가장 기본적인 방법으로 자신이 자주 사용하는 앱들을 스마트폰 초기화면에 위젯으로 배치하는 방법이 있다. 그러나 스마트폰의 초기화면에 배치할 수 있는 앱의 숫자에는 한계가 있으며, 사용자의 상황에 따라 사용하는 앱의 종류는 다양하기 때문에 상기의 문제점을 완벽하게 해결하지는 못한다. 이러한 앱 탐색 문제를 해결하기 위해 본 연구에서는 스마트폰에서 시간 및 장소에 따른 사용자별 앱 사용 로그를 수집하고, 이를 학습하여 사용자의 상황 정보에 따라 최적의 앱 추천 리스트를 초기 화면에 자동으로 배치하는 방법을 제안한다.

* This work was supported by the Industrial Strategic Technology Development Program (10035348, Development of a Cognitive Planning and Learning Model for Mobile Platforms) funded by the Ministry of Knowledge Economy(MKE, Korea), Mid-career Researcher Program through NRF grant funded by the MEST (No.2013016948) and funded by the MSIP(Ministry of Science, ICT & Future Planning), Korea in the ICT R&D Program 2013(No.13-912-03-003).

[†] 준 회 원: 숭실대학교 컴퓨터학과 박사과정

^{**} 준 회 원: 한국의국어대학교 정보통신공학과 학사과정

^{***} 종 신 회 원: 한국의국어대학교 정보통신공학과 교수

^{****} 정 회 원: 숭실대학교 컴퓨터학부 교수

논문접수: 2013년 7월 29일

수정일: 1차 2013년 8월 12일

심사완료: 2013년 8월 12일

* Corresponding Author : Soowon Lee(swlee@ssu.ac.kr)

2. 관련 연구

최근 스마트폰내의 사용자 로그를 수집하기 위한 연구가 진행되면서, 수집된 사용자 로그를 모바일 앱 추천에 활용하려는 연구들이 시작되었다[1,2]. 모바일 앱 추천 연구는 추천 대상에 따라 앱 마켓에 있는 신규 앱 추천과 스마트폰에 설치된 기존 앱 추천으로 구분될 수 있는데, 대표적인 신규 앱 추천 연구로 AppJoy[3], GetJar[4], Appazaar[5] 등이 있다. AppJoy는 스마트폰에서 사용자의 앱 사용량을 수집하여 점수화하고, 이를 활용한 아이템 기반 협력적 추천을 통해 사용자에게 신규 앱을 추천한다. GetJar는 앱 마켓에서 수집되는 데이터(view, download, reputation 등)를 활용하여 다양한 협력적 추천 모델을 통해 신규 앱을 추천하며, Appazaar는 사용자의 스마트폰에서 수집한 상황 컨텍스트를 활용한 신규 앱 추천 기술을 제안하였다.

사용자의 스마트폰내의 기존 앱 추천은 사용자의 앱 탐색 단계를 줄임으로써 앱 사용 편의성 증대를 목표로 한다. 대표적으로, 스마트폰에 설치된 앱들의 기능 명세를 활용하여 앱 기능 군집을 생성하고 이를 폴더 형태로 배치하는 앱 기능 군집 인터페이스를 제안한 연구가 있으며[6], 사용자의 상황 컨텍스트별 앱 사용 로그를 수집하고 Naive-Bayesian 학습을 통해 상황에 맞는 최적의 앱 추천 리스트를 제공하는 추천 앱을 구현하였다.[7]. 이러한 기존의 연구들에서는 사용자가 추천을 받기 위해 특정 앱을 사용하는 것을 가정하고 있다. 이에 반해 본 연구에서 제안하는 사용자의 시공간 컨텍스트를 이용한 모바일 앱 추천은 상황 컨텍스트에 따라 최적의 앱 리스트를 스마트폰 초기 화면에 제공함으로써 기존 연구에 비해 사용자에게 요구하는 탐색 과정을 줄이고, 앱 추천을 위한 상황 정보의 최적 조합을 활용함으로써 앱 추천 결과의 정확도를 높인다.

3. 연구 내용

3.1 시스템 구조도

Fig. 1은 본 논문에서 제안하는 하이브리드 앱 추천 시스템의 구조도이다. 제안된 시스템에서는 수집된 사용자의 앱 로그 데이터로부터 전처리를 통해 시간 및 장소 관련 속성을 추출하여 데이터를 생성한다. 분류기반 학습기에서는 생성된 학습 데이터를 활용하여 사용자의 가장 자주 이용하는 앱을 추천하기 위한 최대사용빈도(Maximum Frequency Usage - MFU)기반 예측 모델, Naive Bayesian 예측 모델, SVM 예측 모델의 세 가지 예측 모델을 학습한다. 앱 추천기에서는 학습된 세 종류의 예측 모델 중에 하나를 사용하여 추천하는 단일 모델 추천과 둘 이상의 모델을 결합하여 추천하는 하이브리드 추천을 통해 앱 추천 리스트를 생성하여 사용자에게 제공한다.

3.2 학습 데이터 및 전처리

본 연구에서는 사용자의 상황 정보 및 앱 사용 로그 정보

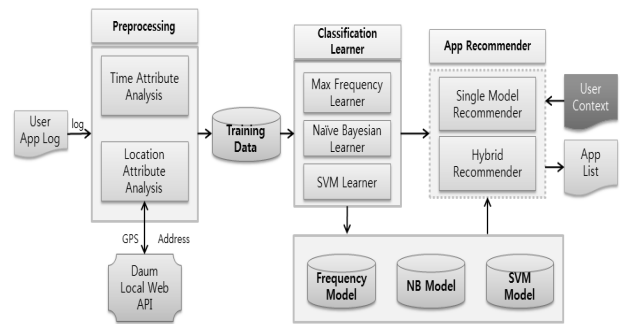


Fig. 1. Mobile app recommendation system

20130313002045		37.4017137		127.1278337		APP		S/com.kakao.talk/카카오톡
20130313002306		37.4019194		127.1277845		APP		E/com.kakao.talk/카카오톡
20130313002446		37.4020062		127.1278051		APP		S/com.sec.android.app.music/뮤직 플레이어
20130313002620		37.4017863		127.1278206		SYS		CalendarSensor.makeLogItems()
20130313002620		37.4017863		127.1278206		SYS		CallSensor.makeLogItems()
20130313002621		37.4017863		127.1278206		SYS		WebSensor.makeLogItems()
20130313002805		37.4020683		127.1277883		APP		E/com.sec.android.app.music/뮤직 플레이어

Fig. 2. App log data

를 수집하기 위해 Logical Logger를 구현하였다[2]. Fig. 2는 Logical Logger를 통해 수집된 앱 로그 데이터의 예로서, 시간, 위도, 경도, 로그타입, 시작종류, 앱ID, 앱 이름 등의 정보가 포함되어 있다.

Fig. 3은 전처리를 통해 시간 관련 속성 정보를 추출하여 생성한 학습 데이터의 예이다. 추출되는 시간 관련 속성은 앱 사용시간, 요일, 주중주말타입, 시간, 시간대 등으로 구성된다. 본 연구에서는 수집된 로그 중에서 앱 사용 시간이 최소 사용시간(10초)을 만족하는 로그에 대해서 학습을 진행한다.

본 연구에서는 수집된 로그의 장소 정보인 GPS 데이터를 학습에 활용하기 위해 Daum 지역 API를 통해 GPS 값

userID	latitude	longitude	dayOfWeek	dayType	timeOfDay	timeLine	appID
user1	37.41108	127.1291	Fri	weekday	12	afternoon	com.kakao.talk
user1	37.41027	127.1297	Fri	weekday	12	afternoon	com.facebook.katana
user1	37.4116	127.1273	Fri	weekday	12	afternoon	com.kakao.talk
user1	37.41042	127.1299	Fri	weekday	12	afternoon	com.facebook.katana
user1	37.41042	127.1299	Fri	weekday	12	afternoon	com.android.gallery3d

Fig. 3. Learning data with temporal attributes

userID	POI	address1	address2	dayOfWeek	dayType	timeOfDay	timeLine	appID
user1	office	분당구	야탑1동	Fri	weekday	12	afternoon	com.kakao.talk
user1	office	분당구	야탑1동	Fri	weekday	12	afternoon	com.facebook.katana
user1	office	분당구	야탑1동	Fri	weekday	12	afternoon	com.kakao.talk
user1	office	분당구	야탑1동	Fri	weekday	12	afternoon	com.facebook.katana
user1	office	분당구	야탑1동	Fri	weekday	12	afternoon	com.android.gallery3d

Fig. 4. Learning data with spatio-temporal attributes

을 주소 값(구, 동)으로 변환한다. 그리고 변환된 주소 값 일 낮에 주로 위치하는 곳을 직장, 그 외의 나머지를 기타로 하여 사용자의 POI를 추출하였다. 추출된 장소 속성 정보를 Fig. 3의 학습 데이터에 결합함으로써 최종적인 학습 데이터를 생성하였다(Fig. 4).

3.3 앱 사용 예측 모델 학습

본 논문에서는 상황 정보에 대한 앱 사용 예측을 위해 최대사용빈도(MFU)기반 예측모델, Naive-Bayesian(NB) 예측 모델, Support Vector Machine(SVM) 예측 모델의 3가지 예측 모델을 생성한다. 최대사용빈도기반 모델은 사용자의 앱 로그에서 앱의 출현빈도를 기준으로 상위 N개의 앱을 추천한다.

Naïve Bayesian 예측 모델은 클래스가 알려지지 않은 미지의 샘플 X에 대해 이 샘플이 특정 클래스 C에 속할 확률 P(C|X)를 결정하여 확률 값이 가장 큰 클래스로 예측하는 방법이다. 이 때 P(C|X)는 Bayes 이론에 의해 P(X|C)*P(C)로 계산이 가능하다. P(C)의 값은 학습 예제를 통해서 구할 수 있으며 P(X|C)는 클래스 변수를 제외한 모든 변수가 서로 독립이라는 가정 하에 Equation (1)과 같이 학습 예제로부터 구할 수 있는 각 변수별 확률 값의 곱으로 계산할 수 있다.

$$P(X | C) = \prod_{k=1}^n P(x_k | C) \tag{1}$$

SVM(Support Vector Machine) 예측 모델은 두 클래스의 경계면으로 마진(margin)을 가장 최대화하는 초평면을 찾는 모델로서 1개의 학습 표본{(x1,y1)...(xl,yl)}이 주어졌을 때 결정 함수인 Equation (2)를 학습한다. SVM은 기본적으로 이진 분류이며, 입력 x가 들어왔을 때 클래스 해당 여부를 커널함수 K(x · xi)을 통해 결정한다. 선택 가능한 커널함수는 선형(linear), 다항식(polynomial), RBF(radial basis function), 시그모이드(sigmoid) 등이 있으며, 다수의 실험을 통해 분류 정확도가 높은 커널 함수를 선택한다.

$$f(\mathbf{x}) = \text{sgn} \left[\sum_{i=1}^l y_i \alpha_i K(\mathbf{x} \cdot \mathbf{x}_i) + b \right] \tag{2}$$

3.4 하이브리드 앱 추천 기술

본 연구에서는 단일 예측 모델들의 추천 정확도 성능 개선을 위해 개개의 예측 모델에 가중치를 적용하여 결합하는 병렬 하이브리드 앱 추천 모델을 제안한다. Equation (3)은 가중치를 적용한 병렬 하이브리드 추천 스코어 계산식으로 rec_k(u, c, a)는 사용자 u의 상황 컨텍스트 c에서 앱 a에 대한 추천 스코어이며 전 단계의 앱 사용 예측 모델에서 예측한 앱 사용 예측 확률 값이 된다. β_k는 각 예측 모델 별 가중치로서 총합은 1이 된다. 이와 같이 가중치를 적용하여

예측된 추천 스코어를 결합한 rec_{weighted}(u, c, a)가 높은 상위 N개를 추천한다.

$$rec_{weighted}(u, c, a) = \sum_{k=1}^n \beta_k \times rec_k(u, c, a) \tag{3}$$

4. 실험 및 평가

4.1 단일 예측 모델 추천 정확도 평가 실험

사용자의 상황 정보에 최적화된 앱 추천의 결과는 사용자가 스마트폰 화면을 활성화한 시점에서 확인할 수 있어야 한다. 이에 따라 본 연구에서는 안드로이드 홈 화면의 위젯 영역에 추천된 앱들을 노출시키는 것을 가정한다. Fig. 5는 안드로이드 스마트폰의 초기 화면의 예로서, 본 연구에서 제안하는 앱 추천 영역을 보여준다.



Fig. 5. Android home ui

Table 1. App log information

user	period	num of log
user1	2013.01.22~2013.03.15	3,078
user2	2013.01.25~2013.03.31	8,271
user3	2013.01.27~2013.03.06	7,009
user4	2013.02.05~2013.03.26	33,495
user5	2013.01.25~2013.03.30	29,850

본 연구에서는 Fig. 5의 추천 영역에 사용자의 상황에 맞는 최적의 앱을 노출시키는 것을 추천 상황 정보를 포함한 앱 사용 로그를 선호도 학습 및 앱 추천 실험에 사용하였다. 실험에서 앱 추천은 1회에 4개의 앱을 추천하고, 추천 리스트에 속한 4개의 앱 중에 사용자가 선택한 앱이 있으면 추천이 성공한 것으로 가정하였다, 이에 따라 추천 정확도 평가 척도로 전체 추천 회수에 대한 추천 성공 회수의 비율을 나타낸 Success Rate를 사용하였다. Table 1은 5명의 실제 사용자에게 의해 수집된 앱 로그 데이터의 정보로서, 사용자들의 로그 수집 기간과 수집된 로그 데이터의 건수를 보여준다.

Table 2. Attribute list for training

case	Learning Feature
case0	weekDay, timeLine
case1	weekDay, timeLine, POI
case2	weekDay, timeLine, POI, timeOfDay
case3	weekDay, timeLine, POI, timeOfDay, dayOfWeek
case4	weekDay, timeLine, POI, timeOfDay, dayOfWeek, address2
case5	weekDay, timeLine, POI, timeOfDay, dayOfWeek, address1, address2
case6	weekDay, timeLine, POI, timeOfDay, dayOfWeek, address1, address2, recentAppID

본 연구에서 학습에 사용한 시간 속성은 dayofWeek, weekDay, timeOfDay, timeLine이며, 장소 속성은 address1 (ex., 관악구등), address2(ex. 상도동 등), POI(ex. home, office, unknown)이다. 추가적으로 앱 사용 순서의 연관성을 반영하기 위해 최근 사용한 앱 정보(recentAppID)도 학습에 활용하였다. 그리고 학습 속성들 간의 최적 조합을 찾기 위해 Table 2의 학습 속성 리스트를 활용하여 학습한 앱 선호도 모델들 간의 추천 정확도 비교 실험을 진행하였다. 비교 실험에서는 전체 앱 로그 데이터 80%를 학습 데이터로 20%를 평가 데이터로 사용하였다. Table 3은 사용자별 학습 데이터와 평가 데이터의 분포를 보여준다.

Table 3. Experiment Data Distribution

user	num of training data	num of test data
user1	2,458	615
user2	6,616	1,655
user3	5,607	1,402
user4	26,796	6,699
user5	23,880	5,970

Fig. 6은 Table 2의 학습 속성 리스트에 따라 학습한 Naive-Bayesian 모델들의 추천 정확도를 비교한 결과이다. 여기에서 대다수의 사용자가 다양한 속성 값을 갖는 address1, 2가 추가되는 case 3, case 4에서 추천 정확도가 급격히 낮아지는 것을 확인할 수 있었다. 반면에, case 6의 경우에는 모든 사용자의 추천 정확도가 큰 폭으로 상승하는 것을 나타냈으며, 보다 정확한 앱 추천을 위해서 앱의 연속적인 사용 특성을 반영할 필요가 있는 것으로 분석되었다.

Fig. 7은 Table 2의 학습 속성 리스트에 따라 학습한 SVM 모델들의 추천 정확도를 비교한 결과이다. case 3에서 추천 정확도가 급격히 낮아졌던 Fig. 6의 결과와는 달리 대다수 사용자의 추천 정확도가 서서히 증가하는 추세를 보이고 있다. 특히 user4는 학습 대상 속성 수가 증가함에 따라 추천 정확도가 선형으로 증가하는 특징을 보였다. 이는 학습 데이터의 양도 적으면서, 특정 앱을 주로 이용하는 다른

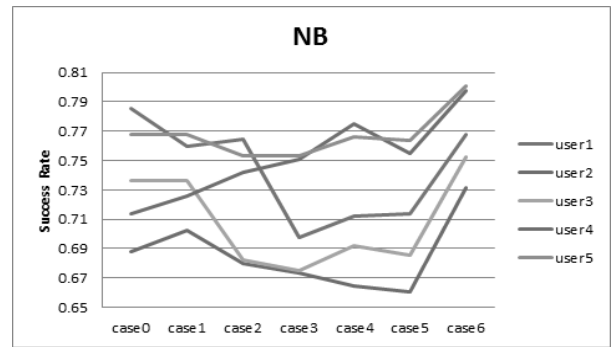


Fig. 6. Success rate for naive-bayesian

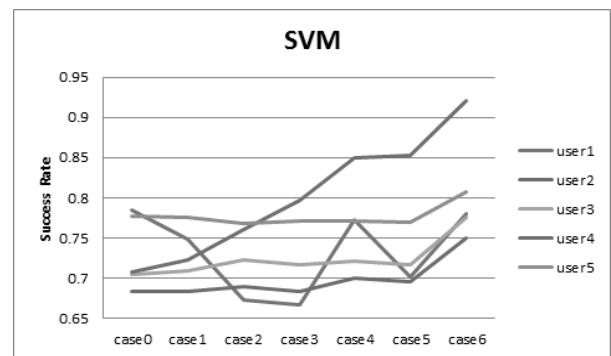


Fig. 7. Success rate for svm

사용자들에 비해 user4는 많은 양의 다양한 앱을 사용한 학습 데이터를 갖고 있기 때문에 보다 정확한 학습이 이루어진 것으로 보인다.

Fig. 8은 모든 속성을 학습할 경우(case6)의 예측 모델별 추천정확도를 비교한 것이다. user1의 경우 분류기반 예측 모델들에 의한 추천 정확도보다 최대사용빈도기반(MFU) 추천에 의한 추천 정확도가 높게 나왔다. 반면에, user2, 3, 4, 5는 SVM, NB, 최대 빈도 기반 예측 모델의 순으로 추천 정확도가 높게 나타났다. 특히, user4의 SVM 모델은 약 92%의 가장 높은 추천 정확도를 보였다.

4.2 하이브리드 추천 정확도 변화 비교 실험

Fig. 9는 user5의 앱 사용 모델별 가중치를 변화시켜가며 측정된 하이브리드 추천 정확도 변화 그래프이다. 하이브리

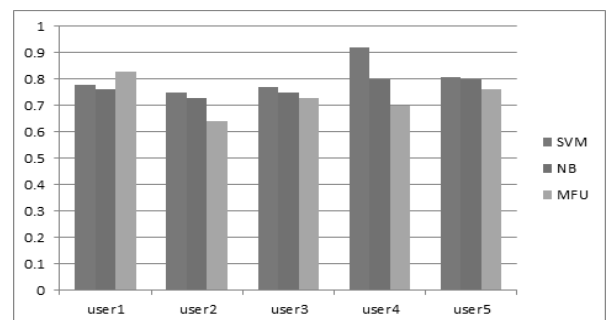


Fig. 8. Success rate for each models (case6)

드 추천 모델은 단일 모델 추천에 비해 약간의 성능 개선이 있었으나 그 차이는 크지 않았다. 하이브리드 추천에 사용된 예측 모델들은 동일한 속성으로 학습하여 모델별로 예측 범위의 차별성이 적기 때문에 모델 결합에 의한 성능 개선의 여지가 적었던 것으로 분석된다. 이에 따라 향후 연구로 예측 범위의 차별성이 큰 예측 모델들을 활용한 하이브리드 추천 연구를 진행할 필요가 있다.

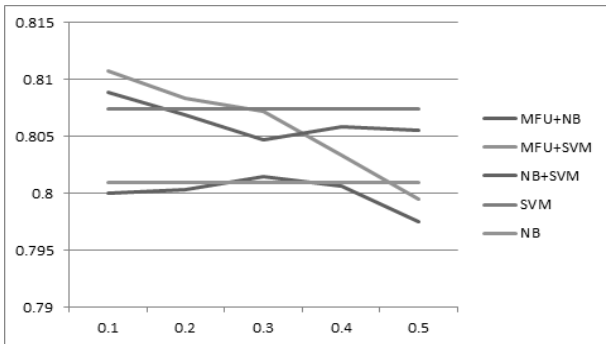


Fig. 9. Success rate for hybrid recommendation - user5

5. 결론 및 향후 연구

본 연구에서는 실제로 스마트폰에서 수집된 사용자의 상황 정보별 앱 로그를 활용하여 사용자의 상황에 맞는 최적의 앱 추천을 제공하는 시공간 사용자 앱 로그를 이용한 모바일 앱 추천 기법을 제시하였다. 제안된 기법을 활용하여 사용자의 앱 사용 예측에 필요한 시공간 컨텍스트를 정의하고 이를 학습한 분류 기반 앱 예측 모델을 활용한 앱 추천 시스템을 구현하였다. 특히, 시공간 컨텍스트 이외에도 최근 사용한 앱과 같은 컨텍스트도 앱 사용 예측에 중요함을 실험을 통해 확인하였다.

한편, 단일 예측 모델의 추천 정확도 성능 개선을 위해 가중치를 이용한 병렬 하이브리드 시스템에 대한 추천 정확도 실험을 진행하였으나, 동일한 속성으로 학습된 예측 모델들 간의 낮은 차별성으로 인해 큰 성능 향상은 이루지 못하였다. 향후에는 앱 사용 예측에 요구되는 학습 속성을 구분하여 개별 속성에 특화된 앱 사용 예측 모델을 생성하고, 이를 결합한 하이브리드 추천 시스템에 대한 연구를 진행할 필요가 있다.

참고 문헌

[1] H. Falaki, R. Mahajan, S. Kandula, D. Lymberopoulos, R. Govindan, and D. Estrin. "Diversity in smartphone usage". In Proceedings of Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services, pp.179-194, 2010.

[2] D. Kim, J. Shin, S. Park, "Decision tree based application recommendation system". In Proceedings of Korea Computer

Conference 2012, pp.140-142, 2012.

[3] B. Yan and G. Chen, "AppJoy: personalized mobile application discovery", In Proceedings of the 9th International Conference on Mobile Systems, Applications, and Services, pp.113-126, 2011.

[4] K. Shi and K. Ali, "GetJar mobile application recommendations with very sparse datasets", In Proceedings of the 18th ACM SIGKDD Conference, pp.204-212, 2012.

[5] P. Yin, P. Luo, W. Lee, M. Wang, "App recommendation: a contest between satisfaction and temptation", In Proceedings of the 6th ACM International Conference on Web Search and Data Mining, pp.395-404, 2013.

[6] D. Lavid, B. Lulu, T. Kuflik, "Functionality-based clustering using short textual description: helping users to find apps installed on their mobile device", In Proceedings of the 2013 International Conference on Intelligent User Interfaces, pp.297-306, 2013.

[7] C. Shin, J. Hong, and A. K. Dey, "Understanding and prediction of mobile application usage for smart phones", In Proceedings of the 2012 ACM Conference on Ubiquitous Computing, pp.173-182, 2012.



강 영 길

e-mail : dudrf34@gmail.com
 2005년 숭실대학교 컴퓨터학부(학사)
 2007년 숭실대학교 컴퓨터학과(석사)
 2007년~현 재 숭실대학교 컴퓨터학과 박사과정
 관심분야 : Data Science, Personalization, Text Mining



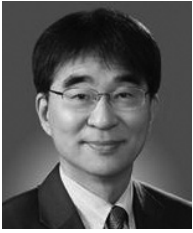
황 세 영

e-mail : hwangs88@naver.com
 2007년~현 재 한국외국어대학교 정보통신공학과 학사과정
 관심분야 : Database, Mobile Application, Context Awareness



박 상 원

e-mail : swpark@hufs.ac.kr
 1994년 서울대학교 컴퓨터공학과(학사)
 1997년 서울대학교 컴퓨터공학과(석사)
 2002년 서울대학교 컴퓨터공학과(박사)
 2002년~2003년 세종사이버대학교 디지털 콘텐츠학과 전임강사
 2003년~현 재 한국외국어대학교 정보통신공학과 교수
 관심분야 : Flash Memory, Embedded Database, Mobile Computing



이 수 원

e-mail : swlee@ssu.ac.kr

1982년 서울대학교 계산통계학과(학사)

1984년 한국과학기술원 전산학과(석사)

1994년 University of Southern California
전산학과(박사)

1995년~현재 숭실대학교 컴퓨터학부
교수

관심분야: Data Science, Machine Learning, Artificial Intelligence 등