

## 특성 유사도 기반 앱 추천

김형일\*

### 요약

정보통신의 발달로 스마트폰은 대중화를 이루었으며, 스마트폰의 대중화는 앱스토어라는 새로운 영역을 생성하였다. 스마트폰에서 사용되는 응용소프트웨어인 앱은 앱스토어를 통해 편리하게 거래될 수 있다는 장점으로 빠른 성장을 이루었다. 앱스토어에서 거래되는 앱들의 수량이 방대해짐에 따라 사용자가 원하는 앱을 정확히 추출하기란 매우 어렵다. 앱스토어에서 사용하는 일반적인 앱 추천 방식은 사용자가 입력한 질의어에 따라 앱을 추천하는 방식이다. 이러한 내용 기반 방식은 디지털 형태로 이루어진 앱을 추천할 때는 효과적인 기법이 아니다. 앱 추천의 정확성을 높이기 위해 본 논문에서는 특성 유사도 기반 앱 추천 기법을 제안한다. 본 논문에서 제안한 기법은 앱의 장르, 기능성, 다운로드 수 등을 이용하여 앱에 대한 속성을 생성한 후, 사용자의 앱 사용에 대한 성향과 비교하여 앱을 추천하는 방식을 따른다. 다양한 실험에서 본 논문에서 제안한 기법이 기본적인 앱 추출 기법보다 평균 33%의 성능 향상을 보였다.

키워드 : 추천시스템, 정보여과, 개인화, 앱

## App Recommendation Based on Characteristic Similarity

Hyung-Il Kim\*

### Abstract

The remarkable development of IT is contributed to popularization of smart phones, which in turn creates a new domain called app store. Smartphone apps have grown fast because they can be easily purchased through an app store. As the volume of apps traded in app stores is so huge that it is extremely hard for users to find the exact app they want. In general, an app store recommends an app to users based on the search words they entered. In terms of recommendation of app, this kind of content-based method is not effective. To increase accuracy in recommending app, this paper proposes a characteristic similarity-based app recommendation method. This method creates attributes on the app based on the related information such as genre, functionality and number of downloads and then compares them with the propensity to use the app. According to diverse simulations, the method proposed in this paper improved the performance of app recommendation by 33% in average, compared to the conventional method.

Keywords : Recommendation System, Information Filtering, Personalization, App

### 1. 서론

※ 교신저자(Corresponding Author): Hyung-Il Kim  
접수일:2012년 11월 06일, 수정일:2012년 11월 24일  
완료일:2012년 12월 14일

\* 나사렛대학교 멀티미디어학과

Tel: +82-41-570-1467, Fax: +82-41-570-1467

email: [hkim@kornu.ac.kr](mailto:hkim@kornu.ac.kr)

▣ 본 연구는 나사렛대학교의 2012학년도 학술비 지원에 의해 수행되었음

임베디드시스템과 무선통신의 발전으로 스마트폰이라는 새로운 영역이 발생하였다. 스마트폰은 응용소프트웨어를 이용하여 다양한 사무와 멀티미디어 정보 등을 처리할 수 있다. 스마트폰에서 동작하는 응용소프트웨어를 앱이라 하며, 이와 같은 앱들은 앱스토어라는 공간에서 거래할 수 있다.

앱은 소프트웨어로 존재하기 때문에 정보의 표현에 대한 자유도가 매우 높으며, 일반 컴퓨터용 소프트웨어보다는 적은 용량으로 존재하기

때문에 프로그램의 갱신과 전송에 매우 편리하다는 장점이 있다.

초기의 앱스토어에는 적은 수량의 앱이 존재하여, 사용자에게 앱을 추천하는 것은 쉬운 일이었다. 그러나 스마트폰의 활성화로 앱의 수량은 기하급수적으로 증가하였으며, 앱의 광범위한 증가는 사용자에게 앱 선택이라는 문제를 발생시켰다. 동일한 종류의 많은 앱들에서 자신에게 적합한 앱을 결정하기 위해, 사용자는 시간과 비용을 투자해야 한다.

일반적으로 앱스토어에서 사용자에게 앱을 추천할 때, 사용자의 질의어를 기반으로 앱을 추천한다. 이러한 방식을 내용 기반 방식이라 하며, 내용 기반 방식은 객체의 유사성 측정에 텍스트 정보를 이용한다[1][2]. 내용 기반 방식은 사용자가 요청한 질의에 대한 내용정보와 대상 객체의 내용정보를 분석하여, 특정 객체의 적합성을 판단한다. 이와 같은 내용 기반 방식은 객체에 텍스트와 같은 내용정보가 풍부한 대상에 적합하다. 그러나 앱은 디지털정보로 표현된 소프트웨어이므로, 내용 기반 방식에 의한 유사도 측정만으로 사용자에게 적합한 앱을 추출하기는 매우 어렵다.

이러한 앱의 구성적 특성을 고려하여 사용자에게 적합한 앱을 추출하기 위해서는 앱의 일반적 속성을 활용하는 것이 효과적이다. 앱스토어에 등록된 앱에는 제목, 선호도, 장르, 다운로드 정보 등이 나타난다. 이러한 앱의 일반적 속성을 앱 관별에 활용하면 사용자에게 적합한 앱을 효과적으로 추출할 수 있다.

본 논문에서 앱의 일반적 속성을 활용하여 사용자에게 적합한 앱을 추천하는 특성 유사도 기반 앱 추천 기법을 제안한다. 본 논문에서 제안한 앱 추천 기법은 앱에 대한 사용자 성향과 앱의 일반적 속성을 활용하여, 사용자의 요구에 적합한 앱을 추천하는 장점이 있다.

본 논문의 2장에서는 관련 연구에 대해 기술하고, 3장에서는 본 논문에서 제안한 앱 추천 기법인 특성 유사도에 대해 설명한다. 4장에서 제안한 기법에 대한 실험결과를 분석하고, 5장에서 결론과 향후 연구에 대해 기술한다.

## 2. 관련 연구

스마트폰을 중심으로 사용되는 앱은 소용량으로 존재하여 저장 및 전송에 효율적이고, 앱의 완성과 동시에 앱스토어에 빠르게 등록할 수 있기 때문에 배포와 갱신에 편리하다는 장점도 있다. 또한 앱은 전 세계의 스마트폰 사용자를 대상으로 한다는 장점이 있으며, 판매 및 결제에 대한 중요 작업은 앱스토어 업체가 담당하기 때문에 개발사 입장에서는 편리하게 시장을 접근할 수 있다는 장점도 있다. 이러한 앱의 장점들로 인해 IT기업들뿐만 아니라 소자본 창업자들까지도 앱 개발에 집중하고 있는 것이 현재의 실정이다.

이러한 상황으로 인해 앱스토어에 등록되는 앱들은 기하급수적으로 증가하고 있으며, 완성도가 낮은 앱들까지도 앱스토어에 등록되고 있다. 그러므로 방대한 수량의 앱에서 사용자에게 적합한 앱을 사용자 스스로가 선택한다는 것은 매우 어려운 작업이다. 앱스토어에서 사용하는 대중화된 앱 추출 기법인 내용 기반 기법으로 사용자에게 적합한 앱을 추출하기도 쉬운 일은 아니다.

다양한 객체 추출 방법에서 많이 활용되는 기법으로 내용 기반 기법, 인구통계학적 기법, 협동적 여과 기법이 있다.

내용 기반 기법에서 객체 추출 시 사용되는 대표적인 방식은 단어빈도 방식이다[3][4]. 단어빈도 방식은 질의어의 내용정보와 대상 객체의 내용정보를 비교하여 상호 유사성을 측정하는 방식이다. 이러한 단어빈도 방식은 뉴스나 서적 등을 추천할 때 많이 활용하는 방식이다. 이러한 내용 기반 기법은 텍스트 정보가 많이 포함된 객체에 적합하다.

인구통계학적 기법은 사용자의 직업, 성별, 나이 등과 같은 특수한 신상정보를 주로 이용한다[5][6]. 객체 추출에 이러한 인구통계학적 정보를 활용할 경우에는 근접 사용자의 객체 선호도를 계산하여, 특정 사용자의 객체 선호도를 예측하는 방식을 따른다. 이러한 인구통계학적 기법은 객체 집단을 쉽게 구별할 수 있기 때문에 대량의 객체를 포함한 영역에서 효과적인 기법이다[7][8]. 인구통계학적 기법을 적용하기 위해서는 사용자의 신상정보가 필요하다.

협동적 여과 기법은 사용자들의 객체 평가정보를 이용하여 객체를 추출한다[9][10]. 협동적 여과 기법은 특정 객체 집단에 나타난 사용자들

의 객체 평가정보를 이용하여 임의의 사용자들과 거리를 측정하고, 이러한 사용자 거리를 기반으로 특정 객체의 사용자 선호도를 예측한다. 협동적 여과 기법은 객체의 텍스트 정보를 이용하지 않지만, 객체의 평가정보라는 사용자의 명시적 반응을 요구한다[11][12].

본 논문에서 제안한 특성 유사도 기반 앱 추천 기법은 앱의 일반적 속성과 사용자의 행위정보를 고려하여 앱 속성을 생성한다. 특성 유사도 기법을 활용하면 사용자에게 적합한 앱을 효과적으로 제공할 수 있다.

### 3. 특성 유사도

일반적인 카테고리에 의한 군집화는 방대한 수량의 앱에서는 효과적이기 어렵다. 사용자의 성향에 맞추어 앱을 추천하기 위해서는 앱의 일반적 속성을 고려한 특성 군집화가 효과적이다.

특성 군집화를 수행하기 위해 사용할 수 있는 앱 속성으로 기능, 비용, 장르, 완성도, 다운로드 등이 있다. 앱은 기능적 측면에서 다기능과 소기능으로 나눌 수 있으며, 사용자는 자신의 성향에 맞게 기능적 측면을 고려하여 앱을 선택한다. 앱의 비용적 측면에서 유료와 무료로 나뉘며, 사용자는 이러한 비용적 측면을 고려하여 앱을 선택한다. 앱에 대한 완성도는 일반적으로 5단계로 구분되며, 사용자들은 앱 사용에 대한 경험을 통해 앱에 대한 완성도를 표현한다. 다운로드에 대한 정보는 앱의 인기도를 반영하며, 다운로드 수량을 통해 앱 인기도를 등급화할 수 있다. 본 연구에서는 이와 같은 앱의 다양한 일반적 속성을 객체 특성화 군집에 활용하였다.

군집화에 적용하는 대표적인 기법은 k-means이다[13][14]. k-means를 방대한 앱 집단에 적용하기 위해서는 분류 집단에 적용할 거리가 필수적이다. 거리 측정에 사용되는 기법들 중에 대표적인 것은 euclidean distance와 cosine coefficient이고, 본 연구에서는 euclidean distance를 활용하였다[15][16].

다차원 객체에 대한 euclidean distance는 식 1과 같다. 다차원 공간에 n개의 속성을 지닌 객체  $A_i$ 와 객체  $B_j$ 가 존재할 때, 객체  $A_i$ 와 객체  $B_j$ 의 거리  $Dist(A_i, B_j)$ 는 각 객체의 성분 거리를 이

용하여 측정한다.

$$Dist(A_i, B_j) = \sqrt{(A_1 - B_1)^2 + (A_2 - B_2)^2 + \dots + (A_n - B_n)^2} \quad (1)$$

객체 집합을 분할하여 군집화를 생성할 때는 다음과 같은 전체 조건을 만족하여야 한다. 객체들의 전체 집합을 S라 가정할 때, 집합 S는 공집합 아니어야 하고, 분할 집단들을 통합하면 전체 집합 S와 같아야 한다. 그러므로 객체를 군집화하기 위해서는 식 2를 만족하여야 한다. 식 2에서  $C_i$ 는 객체들의 군집을 의미한다.

$$S = \{C_1, C_2, C_3, \dots, C_i, \dots, C_n\} \text{ for } i \in N \quad (2)$$

각 군집들은 기준값을 갖으며, 모든 객체들은 군집 기준값에 의해 군집화를 이룬다. 군집에 대한 평균 기준값은 식 3으로 계산된다. 식 3에서 P는 군집 기준값을 나타내며,  $a_i$ 는 군집의 대표 속성이다.  $C_i$ 는 각 군집들을 나타내고, 모든 군집의 기준값을 통합하여 전체 군집 m으로 나눈 값이 평균 기준값 B가 된다. 이와 같은 평균 기준값을 활용하여 객체와 군집의 이산도를 측정할 수 있다.

$$B = \frac{\sum_{P \in C_i}^m P(a_1, a_2, \dots, a_j, \dots, a_n)}{m} \text{ for } i \in N, j \in N \quad (3)$$

k-means는 다음과 같은 단계로 진행된다. 모든 객체가 포함될 군집의 수를 결정하고 기준값을 생성한다. 기준값을 이용하여 객체와의 거리를 측정하고, 거리를 이용하여 객체를 군집에 할당한다. 이러한 과정을 반복하여 군집화가 완성되면 군집화를 종료한다.

같은 종류의 그림판 앱이라도 앱스토어에는 복잡한 기능을 소유한 앱과 간단한 기능만 소유한 앱이 다양하게 존재하며, 사용자는 개인의 성향에 따라 개인에게 적합한 앱을 선택한다. 그러므로 사용자의 앱에 대한 사용 성향을 분석하면 사용자에게 적합한 앱을 추천할 수 있다.

본 논문에서는 특성 유사도를 이용하여 사용자에게 적합한 앱을 추천하는 기법을 제안한다. 특성 유사도는 앱의 구성적 특성과 사용자의 행동 특성을 속성으로 이용하여, 사용자의 성향에

적합한 앱을 추천할 수 있는 장점이 있다. 본 연구에서 생성한 앱의 특성화 속성으로는 기능, 비용, 장르, 완성도, 인기도 등이 있다.

앱에 대한 군집  $C_i$ 와 사용자  $U_j$ 의 유사도 측정은 특성 유사도를 이용하며, 특성 유사도 공식을 식 4에 나타내었다. 식 4에서  $D(C_i, U_j)$ 는 군집  $C_i$ 와 사용자  $U_j$ 의 거리를 나타낸다.  $P(U_j|C_1, C_2, \dots, C_i, \dots, C_n)$ 는 사용자  $U_j$ 의 군집에 대한 소속확률을 의미한다.  $\xi_i$ 는 군집  $C_i$ 에서의 유사도 적용률을 의미한다.  $\alpha$ 는 사용자  $U_j$ 의 속성 총수이고,  $\beta$ 는 군집의 총수 있다.

$$Sim(C_i, U_j) = \frac{\xi_i D(U_j, C_i) + \beta}{P(U_j|C_1, C_2, \dots, C_i, \dots, C_n) + \alpha} \text{ for } i \in N, j \in N, k \in N \quad (4)$$

$$= \frac{\xi_i D(U_j, C_i) + \beta}{P(U_j, C_1) + P(U_j, C_2) + \dots + P(U_j, C_i) + \dots + P(U_j, C_n) + \alpha} \quad (5)$$

$$= \frac{\left( \sum_{j=1}^m C_i(U_j) \right) \cdot D(U_j, C_i) + \beta}{\sum_{i=1}^n \sum_{j=1}^m C_i(U_j) + \alpha} \quad (6)$$

$$= \frac{\left( \sum_{j=1}^m C_i(U_j) \right) \cdot D(U_j|c_{i1}, c_{i2}, \dots, c_{in}) + \beta}{P(U_j|c_{i1}, c_{i2}, \dots, c_{in}) + P(U_j|c_{21}, c_{22}, \dots, c_{2n}) + \dots + P(U_j|c_{n1}, c_{n2}, \dots, c_{nn}) + \alpha} \quad (7)$$

식 4를 각 군집으로 분해하여 표현하면 식 5와 같다.  $P(U_j, C_i)$ 는 군집  $C_i$ 에서 사용자  $U_j$ 의 발생확률을 의미한다. 발생확률을 이용하면 특정 군집  $C_i$ 에서 사용자  $U_j$  특성에 대한 출현빈도를 측정할 수 있다. 식 5를 유사도 적용률로 풀면 식 6과 같고, 식 6을 각 군집의 속성으로 풀면 식 7과 같다.  $D(U_j|c_{i1}, c_{i2}, \dots, c_{in})$ 는 군집  $C_i$ 의 속성과 사용자  $U_j$ 의 유사도이며,  $c_{i1}, c_{i2}, \dots, c_{in}$ 는 군집  $C_i$ 의 속성이다.

### 4. 실험결과

앱 추천에 대한 실험을 위해 실험용 데이터를 생성하였으며, 사용 특성의 다양성을 고려하여 다양한 데이터 속성을 생성하였다.

앱은 다양한 속성을 소유하고 있으며, 앱 사용에 대한 사용자 반응정보 또한 다양한 방식으로 표현된다. 앱스토어에 존재하는 이러한 정보에 대표적인 것이 기능, 비용, 장르, 완성도, 인기도 등이다. 본 연구에서는 앱스토어에 존재하는 대표적인 정보들을 활용하여 대분류 속성을 생성하였다. 실험에 사용된 앱에 대한 속성은 앱 속성의 확장성을 고려하여, 크게 대분류 속성 Da와 소분류 속성 Av로 나뉜다. 실험에서 Da의 속성 총수는 1에서 10의 범위에서 결정할 수 있으며, Av의 속성 총수는 1에서 5의 범위에서 결정할 수 있다. 소분류 속성은 대분류 속성의 등급으로 사용된다. 본 실험에서 사용한 대분류 속성은 5로 제한하였으며, 대분류 속성은 다시 3개의 소분류 속성으로 각각 나뉜다. 그러므로 본 실험에 사용한 속성 총수는 15개이고, 앱의 정보 구조를 표 1에 나타내었다. Number는 객체번호를 의미하고, Name과 Category는 객체이름과 객체번호를 의미한다. AttributeR1은 1차 속성 Da1을 의미하고, AttributeR1의 벡터 <Da1Av1, Da1Av2, Da1Av3>는 Da1에 대한 2차 속성으로 구성된다.

<표 1> 앱의 정보구조

Number	100
Name	App100_Title
Category	App100_Category
AttributeR1	<Da1Av1, Da1Av2, Da1Av3>
AttributeR2	<Da2Av1, Da2Av2, Da2Av3>
AttributeR3	<Da3Av1, Da3Av2, Da3Av3>
AttributeR4	<Da4Av1, Da4Av2, Da4Av3>
AttributeR5	<Da5Av1, Da5Av2, Da5Av3>

<Table 1> The information structure of app

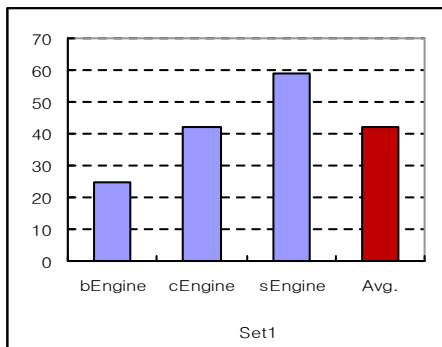
앱 데이터의 크기에 따른 다양한 실험을 위해 실험에 사용한 데이터는 총 4개이며, 4개의 데이터는 크게 대규모 데이터와 소규모 데이터로 나뉜다. 소규모 데이터는 100,000개(Set1), 200,000개(Set2)로 객체를 한정하였으며, 대규모 데이터는 1,000,000개(Set3), 2,000,000개(Set4)로 객체를 한정하였다. 앱 데이터를 생성할 때는 동일 종류의 앱 다양성을 표현하기 위해, 각 실험 집단의 1%에서 2%의 유형 범위에서 랜덤하게 앱 데이터를 생성하였다. 사용자의 앱 사용정보를 생성할 때는 동일 앱에 대해 1명에서 100명까지의 정보를 랜덤하게 축적하였다.

정확도 측정에 사용된 테스트 집합은 실험 집합의 20%를 사용하였으며, 실험 방법은 5-fold cross validation을 사용하였다. 사용자 집단을 추출할 때는 kNN 기법을 사용하였으며, k값은 10으로 제한하였다. 정확도 측정을 위해 blind test를 적용하였으며, 정확도 계산에는 적중률을 사용하였다. blind test를 적용하기 위해 테스트 집단에 소속된 사용자의 앱 사용정보 1개를 제거하고, kNN을 적용하여 제거된 앱 사용정보를 예측하도록 하였다.

앱 추천에 사용한 엔진은 크게 3가지로 나뉜다. 첫 번째 엔진(bEngine)은 일반적인 앱스토어에서 사용하는 내용 기반에 의한 앱 추천 엔진이고, 두 번째 엔진(cEngine)은 객체 군집화를 수행하여 앱을 추천하는 엔진이며, 세 번째 엔진(sEngine)은 본 논문에서 제안한 특성 유사도를 적용한 엔진이다.

(Figure 1)에 Set1의 실험결과를 나타냈다. Set1에서 bEngine은 25%의 적중률을 나타냈고, cEngine은 42%의 적중률을 나타내어 cEngine이 bEngine보다 17% 높게 나타났다. cEngine은 앱 군집화가 적용된 상태에서 앱을 추천하기 때문에, 앱스토어의 기본적인 앱 추천 방식을 취하는 bEngine에 비해 적중률이 높게 나타났다. 군집화를 이용한 sEngine은 59%의 적중률을 나타내어 bEngine의 적중률에 비해 34% 높게 나타났고, cEngine의 적중률보다는 17% 높게 나타났다. 특성 유사도가 적용된 sEngine은 사용자의 앱 사용 특성을 고려하여 앱을 추천하기 때문에 다른 엔진에 비해 높은 적중률을 나타냈다.

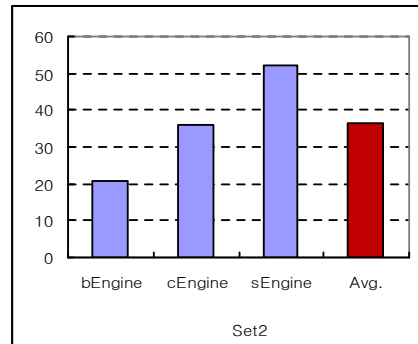
(그림 1) Set1을 이용한 추천 정확도



(Figure 1) The accuracy of recommendations using Set1

(Figure 2)에 Set2의 실험결과를 나타냈다. Set2에서 bEngine은 21%의 적중률을 나타냈고, cEngine은 36%의 적중률을 나타내어 cEngine이 bEngine보다 15% 높게 나타났다. Set1보다 앱 수량이 증가된 Set2에서도 cEngine의 적중률이 bEngine의 적중률보다 높게 나타났으며, 내용 기반 방식보다 앱 군집화 기반 방식이 앱 추천에 효과적이라는 것을 두 가지 실험에서 확인하였다. Set2에서 평균 적중률은 36%로 나타났고, Set1에서 평균 적중률은 42%로 나타났다. Set2의 평균 적중률이 Set1의 평균 적중률보다 낮게 나타난 이유는 수량 증가에 따른 추출 혼잡도가 높아졌기 때문이다. sEngine은 52%의 적중률을 나타내어 bEngine의 적중률과 cEngine의 적중률보다 각각 31%와 16% 높게 나타났다. Set2에서 sEngine이 bEngine과 cEngine에 비해 우수한 성능을 나타낸 이유는 특성 유사도가 개인화 특성을 고려하기 때문이다.

(그림 2) Set2를 이용한 추천 정확도

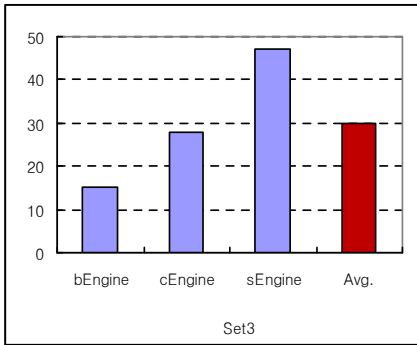


(Figure 2) The accuracy of recommendations using Set2

(Figure 3)에 Set3의 결과를 나타냈고, Set3는 Set1과 Set2와는 다른 대용량 집합이다. Set3에서 bEngine은 15%의 적중률을 나타냈고, cEngine은 28%의 적중률을 나타내어 cEngine이 bEngine보다 13% 높게 나타났다. sEngine은 47%의 적중률을 나타내어 bEngine의 적중률에 비해 32% 높게 나타났고, cEngine의 적중률보다는 19% 높게 나타났다. 소규모 데이터 집합인 Set1과 Set2의 적중률과 Set3의 적중률을 비교할 때, Set3에서 모든 Engine이 낮은 성능을 나타냈다. Set3는 대용량 집합으로 Set1과 Set2에

비해 추출 혼잡도가 매우 높은 집합이기 때문에, 소규모 데이터 집합에서보다 낮은 적중률을 나타낸 것이다. cEngine은 Set3에서도 우수한 추천 성능을 나타냈으며, 이러한 결과를 보더라도 특성 유사도 기법이 다른 기법에 비해 앱이 광범위하게 존재하는 집합에서 우수하게 적용한다는 것을 확인할 수 있다.

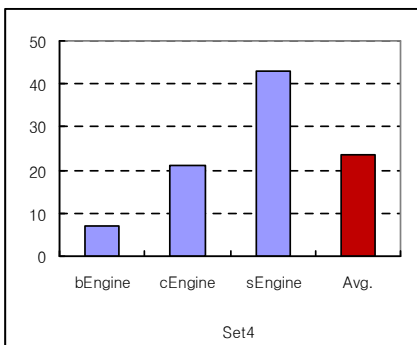
(그림 3) Set3을 이용한 추천 정확도



(Figure 3) The accuracy of recommendations using Set3

Set4에서 bEngine의 적중률이 7%로 나타났으며, cEngine과 sEngine의 적중률은 각각 21%와 43%로 나타났다. 여러 실험결과를 보더라도 앱 수량이 증가하면 추천 성능이 낮아진다는 것을 확인할 수 있다. 특히 내용 기반 기법을 이용한 bEngine은 방대한 앱이 존재하는 앱스토어에서는 효과적인 기법일 수 없다는 것을 여러 실험을 통해 확인하였다.

(그림 4) Set4를 이용한 추천 정확도



(Figure 4) The accuracy of recommendations using Set4

## 5. 결론

일반적으로 앱스토어에서 앱을 추천할 때는 내용 기반 질의를 사용한다. 내용 기반 방식은 내용정보가 풍부한 객체에는 효과적일 수 있으나, 앱에는 내용정보가 풍부하지 않아 효과적으로 적용될 수 없는 방식이다.

본 논문에서 제안한 특성 유사도 기반 앱 추천 기법은 앱에 나타난 일반화 속성을 사용하여 앱을 군집화한 후, 사용자의 성향에 맞는 앱을 추천하는 방식을 취한다. 본 논문에서 제안한 기법은 내용정보가 풍부하지 않은 앱을 효과적으로 추천할 수 있는 장점이 있다.

실험에서 bEngine은 평균 17%의 적중률을 나타냈고, cEngine은 평균 32%의 적중률을 나타내어 cEngine이 15% 높게 나타났다. 이 결과를 보더라도 앱 군집화는 앱 추천에 효과적일 수 있다는 것을 확인할 수 있다. sEngine은 평균 50%의 적중률을 나타내어 bEngine보다 33% 높게 나타났고, cEngine보다 18% 높게 나타났다. 다양한 실험에서 본 논문에서 제안한 기법이 앱 추천에 효과적으로 사용될 수 있다는 것을 확인하였다.

향후 연구로는 속성의 연결성을 이용한 다차원 속성에 대한 연구와 효과적으로 앱 군집화를 수행할 수 있는 거리 측정에 대한 연구가 필요하다.

## References

- [1] T. Chen, W. Han, H. Wang, Y. Zhou, B. Xu, and B. Zang, "Content Recommendation System Based on Private Dynamic User Profile," International Conference on Machine Learning and Cybernetics, pp.2112-2118, 2007.
- [2] C. Jian, Y. Jian, H. Jin, "Automatic content-based recommendation in e-commerce," The 2005 IEEE International Conference on e-Technology, e-Commerce and e-Service, pp.748-753, 2005.
- [3] S. K. Ray and S. Singh, "Blog content based recommendation framework using WordNet and multiple Ontologies," 2010 International Conference on Computer Information Systems and Industrial Management Applications, pp.432-437, 2010.

[4] M. N. Uddin, J. Shrestha, and G. Jo, "Enhanced Content-Based Filtering Using Diverse Collaborative Prediction for Movie Recommendation," 2009 First Asian Conference on Intelligent Information and Database Systems, pp.132-137, 2009.

[5] T. Chen and L. He, "Collaborative Filtering Based on Demographic Attribute Vector," International Conference on Future Computer and Communication, pp.225-229, 2009.

[6] M. Pazzani, "A Framework for Collaborative, Content-Based, and Demographic Filtering," Artificial Intelligence Review, pp.393-408, 1999.

[7] L. Sebastia, I. Garcia, E. Onaindia, and C. Guzman, "e-Tourism: A Tourist Recommendation and Planning Application," International Conference on Tools with Artificial Intelligence, 2008.

[8] C. Stiller, F. Ross, and C. Ament, "Demographic recommendations for WEITBLICK, an assistance system for elderly," International Symposium on Communications and Information Technologies, pp.406-411, 2010.

[9] M. Pazzani, J. Muramatsu, and D. Billsus, "Syskill & Webert: Identifying interesting web sites," Proceedings of the 13th National Conference on Artificial Intelligence, 1996.

[10] B. Krulwich, "Lifestyle Finder: Intelligent user profiling using large-scale demographic data," Artificial Intelligence Magazine, Vol.18, No.2, 1997.

[11] D. Billsus and M. Pazzani, "Learning Collaborative Information Filters," Proceedings of the 15th International Conference on Machine Learning, 1998.

[12] U. Shardanand and P. Maes, "Social information filtering: Algorithms for automating 'word of mouth'," Proceedings of the Conference on Human Factors in Computing Systems, 1995.

[13] J. Han and M. Kamber, Data Mining Concepts and Techniques, Morgan Kaufmann, 2001.

[14] J. Z. Huang, M. K. Ng, H. Rong, and Z. Li, "Automated variable weighting in k-means type clustering," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 27, No. 5, pp.657-668, May 2005.

[15] S. H. Al-Harbi, "Adapting k-means for supervised clustering," Applied Intelligence, Vol. 24, No. 3, pp. 219-226, Jun. 2006.

[16] P. N. Tan, M. Steinbach, and V. Kumar, Introduction to Data Mining, Addison Wesley, 2006.



**김형일**

1999년 : 동국대학교 대학원 (공학 석사)

2004년 : 동국대학교 대학원 (공학 박사-인공지능)

1996년~1998년: (주)경기은행

2005년~2006년: 동국대학교 컴퓨터공학과 IT교수

2007년~현재: 나사렛대학교 멀티미디어학과 교수

관심분야 : 지능형시스템(Intelligent System), 임베디드 시스템(Embedded System), 기계학습(Machine Learning), 추천시스템(Recommendation System) 등