

# 위치정보 및 사용자 경험을 반영하는 모바일 PA에이전트의 설계<sup>☆</sup>

## Designing mobile personal assistant agent based on users' experience and their position information

강 신 봉\*                      노 상 욱\*\*  
Shinbong Kang              Sanguk Noh

### 요 약

급변하는 모바일 환경에서 스마트폰을 비롯한 모바일 기기는 엔터테인먼트, 비즈니스, 정보서비스 등 사용자들의 삶의 방식을 직접적으로 변화시키는 핵심 도구로서의 역할을 하고 있다. 모바일 서비스 중 특히 사용자의 위치정보를 활용하여 서비스를 제공하는 위치기반 서비스(Location Based Service)는 검색, 증강현실, 모바일 SNS(Social Network Service), 게임 등의 다른 서비스 및 콘텐츠와 결합하여 사용자의 다양한 요구를 충족시키며 주요 서비스로 자리 잡아 가고 있다. 본 논문에서는 모바일 서비스가 갖는 잠재적 가능성을 이용하여 모바일 기기의 사용성을 증대시키며, 서비스의 복잡성을 해결하기 위하여 복잡한 태스크를 숨기고 사용자를 대신하여 프로세스를 수행시킬 수 있는 방안에 중점을 둔다. 사용자의 의도 혹은 선호도를 파악하여 사용자에게 개인화된 서비스를 제공하는 PA(Personal Assistant) 에이전트의 개념을 모바일 환경에 적용하기 위한 기법을 제시한다. 사용자의 선호도를 파악하고 개인화된 서비스를 제공하기 위하여 클러스터링 알고리즘과 데이터 분류 알고리즘을 사용하였다. 실험을 통하여 사용자 패턴별로 생성한 클러스터에 분류 알고리즘을 적용한 결과에 대한 분류정확도를 측정하였으며, 제안한 기법의 클러스터별 분류 정확도는 기존의 기법과 비교하여 17.42% 증가하였다.

### ABSTRACT

Mobile environments rapidly changing and digital convergence widely employed, mobile devices including smart phones have been playing a critical role that changes users' lifestyle in the areas of entertainments, businesses and information services. The various services using mobile devices are developing to meet the personal needs of users in the mobile environments. Especially, an LBS (Location-Based Service) is combined with other services and contents such as augmented reality, mobile SNS (Social Network Service), games, and searching, which can provide convenient and useful services to mobile users. In this paper, we design and implement the prototype of mobile personal assistant (PA) agents. Our personal assistant agent helps users do some tasks by hiding the complexity of difficult tasks, performing tasks on behalf of the users, and reflecting the preferences of users. To identify user's preferences and provide personalized services, clustering and classification algorithms of data mining are applied. The clusters of the log data using clustering algorithms are made by measuring the dissimilarity between two objects based on usage patterns. The classification algorithms produce user profiles within each cluster, which make it possible for PA agents to provide users with personalized services and contents. In the experiment, we measured the classification accuracy of user model clustered using clustering algorithms. It turned out that the classification accuracy using our method was increased by 17.42%, compared with that using other clustering algorithms.

☞ Keywords: PA(personal assistant) 에이전트, 기계학습, 모바일 사회망 서비스, personal assistant agent, machine learning algorithms, mobile social network service

\* 준 회 원 : 가톨릭대학교 대학원 컴퓨터공학전공 석사  
buena80@gmail.com

\*\* 정 회 원 : 가톨릭대학교 컴퓨터정보공학부 교수  
sunoh@catholic.ac.kr(교신저자)

[2010/10/01 투고 - 2010/10/11 심사 - 2011/01/18 심사완료]

☆ 이 논문은 2010년도 가톨릭대학교 교비연구비의 지원에 의하여 연구 되었음.

## 1. 서론

디지털 컨버전스(Digital Convergence)시대를 맞이하여 모바일 기기들은 가정용 PC와 같은 종합 멀티미디어 기기로 변모하고 있다. 모바일 기기의 대표주자인 휴대폰은 단순히 통신기능 이외에 엔터테인먼트, 비즈니스, 정보서비스 등 모바일 사용자들의 삶의 방식을 직접적으로 변화시키는 핵심 도구로 자리 잡고 있다.

지금까지 국내 환경에서는 이동통신사들의 Wi-Fi 이용 제한 정책 및 사용자 접근장벽이 높은 데이터 요금제 등으로 인하여 모바일 기반 서비스들이 활성화되지 못하였다. 하지만 애플사의 아이폰 상륙과 더불어 국내 이동통신사들의 활발한 안드로이드 플랫폼 휴대폰 출시 이후 국내 모바일 생태계는 Wi-Fi 이용 제한 해제, 3G 통신망을 통한 자유로운 인터넷 접속, 현실적인 데이터 요금제 등 급속한 변화를 맞고 있다. 이러한 환경의 변화 속에서 이동통신망 및 GPS를 통해 얻은 사용자의 위치정보를 활용하여 서비스를 제공하는 위치기반 서비스(Location Based Service)는 검색, 증강현실, 모바일 SNS(Social Network Service), 게임 등의 다른 서비스 및 콘텐츠와 결합하여 사용자의 다양한 요구를 충족시키며 모바일 환경에서의 주요 서비스로 자리 잡아 가고 있다 [1].

모바일 환경은 일반적인 유선 환경과 비교해 다음과 같은 차이점을 갖는다. 첫째, 시공간적인 제약이 없다. Wi-Fi 혹은 3G 통신망이 연결된 지역이라면 언제 어디서든 인터넷에 접속하여 필요한 정보를 검색하고 원하는 서비스를 사용할 수 있다. 둘째, 개인 지향적이다. 1인 1단말의 특성으로 소유자 개인의 취향과 목적에 따른 기기 이용 성향을 분석하고 이를 반영한 인터페이스와 서비스를 제공할 수 있다. 셋째, 입출력 공간이 협소하다. 휴대를 목적으로 하는 모바일 기기는 일반적인 PC 환경에 비해 한정된 디스플레이 화면과 입력 공간을 갖는다. 넷째, 정보 처리 능력과 메모리 자원의 한계를 갖는다. 모바일 기기에서 일반적인

데스크탑과 같은 처리 능력과 응답 속도, 메모리 자원 확보를 기대할 수는 없다.

시공간적인 제약 탈피, 개인지향성 등의 요인으로 일반적인 PC 환경의 서비스들이 모바일 환경으로 이전되며 다양한 잠재적 가능성을 갖게 되었지만, 모바일 기기의 물리적 제약과 시간, 장소 등 동적으로 변화하는 주변 환경으로 인해 서비스가 고도화될수록 복잡성은 증가하게 된다.

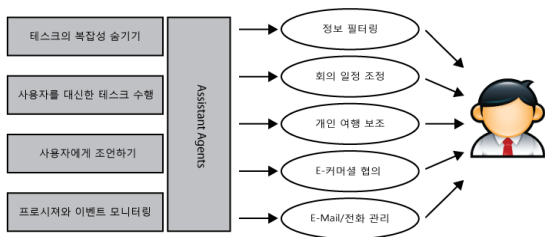
Pattie Maes [2]에 의해 소개된 PA(Personal Assistant) 에이전트의 개념에 따르면 태스크의 복잡성을 숨기고, 사용자를 대신하여 태스크를 수행 시키며, 사용자에게 조언함으로써 에이전트가 사용자와 협업하고 사용자의 개인 활동을 지원할 수 있다. 본 논문에서는 모바일 환경에 적합한 개인화 에이전트를 설계하기 위하여 PA 에이전트의 개념을 적용한 모바일 에이전트를 설계하고 그를 위한 기법들을 제시한다.

본 논문에서 제안된 시스템은 클러스터링 알고리즘을 이용하여 사용 로그 데이터 셋을 여러 개의 클러스터로 분할하고 분할된 각각의 데이터 셋에 대하여 분류 알고리즘을 적용하여 클러스터별 학습모델을 생성한다. 생성된 학습모델은 사용자 요청에 대하여 적합한 서비스 혹은 콘텐츠를 예측하여 사용자에게 추천한다. 위치정보를 속성으로 포함하는 사용 로그 데이터 셋에 분류 알고리즘을 적용하고 학습모델을 생성하기 위하여 위치 좌표에 따른 거리 값과 사용 패턴 상이도를 고려한 클러스터링 기법을 제시한다. 또한, 처리 능력과 메모리 자원의 한계를 갖는 모바일 기기에서 사용자의 선호도를 적시에 반영하기 위한 시스템 구조를 설계한다.

## 2. 관련 연구

PA(Personal Assistant) 에이전트의 아이디어는 1994년 Pattie Maes [2]에 의해 처음 소개되었는데, Maes는 ‘직접적인 조작(Direct Manipulation)’이라는 용어로 표현한 전통적인 컴퓨터 인터랙션 방

법을 대신하여 PA 에이전트의 개념을 제안하였다. 컴퓨터와 사용자에게 ‘직접적인 조작’을 필요로 하는 상호작용에서는 사용자가 직접 태스크를 수행시키고 이벤트에 대해 계속 모니터를 해야 한다. 반면에 PA 에이전트는 에이전트가 까다로운 태스크의 복잡성을 숨기고 사용자를 대신하여 태스크를 수행시키며, 사용자에게 조언하고 에이전트가 직접 태스크의 진행절차와 이벤트를 모니터함으로써 사용자와 협업한다. 사용자에게 중요하지 않은 정보의 필터링, 회의 일정 조정, 개인 여행 보조, 전자업무 협의, 개인 이메일/전화 관리 등이 PA 에이전트가 수행할 수 있는 다양한 어플리케이션의 대표적인 예이다. PA 에이전트의 어플리케이션과 사용자 지원을 위한 다양한 접근 방법은 그림 1과 같다.



(그림 1) PA 에이전트의 어플리케이션과 사용자 지원 방법.

Maes [2]가 언급한 바와 같이 PA 에이전트는 많은 어플리케이션에서 다양한 목적으로 적용되어 왔다. 웹에서의 정보 검색과 필터링이 그 대표적인 예이며 웹 에이전트는 다양한 형태로 진화하여 왔다. Syskill & Webert [3]는 웹에서의 정보 검색과 사용자가 원하는 정보를 추천하는 에이전트를 구현하였다. Personal WebWatcher [4] 프로젝트는 요청한 웹 문서와 관련있는 하이퍼링크를 제안해 주며, 이때 기계학습 알고리즘을 이용하여 사용자의 관심에 대한 학습모델을 생성하여 이를 기반으로 관련있는 웹 사이트를 추천해 주도록 한다. Remembrance Agent [5]와 Letizia [6] 등의 에이전트 또한 웹에서의 정보 검색을 도와주는 역할을 수행한다. Remembrance Agent는 사용자의

현재 내용과 관련있을 수 있는 웹 문서의 리스트를 사용자에게 제공하여 주는 소프트웨어 에이전트이며, Letizia는 사용자가 관심있을 것으로 기대하는 링크를 자동적으로 보여주는 에이전트이다.

또한, 사용자가 관심있는 항목을 자동적으로 추천하는 에이전트로서 IntelliShoper [7]와 content-based recommendation systems [8] 등이 있다. 사용자의 온라인 쇼핑을 도와주는 에이전트인 IntelliShoper는 사용자의 관심을 프로파일 형태로 저장하여 사용자에게 적합한 항목을 추천하여 준다. Pazzani [8]가 제안한 내용기반 추천 시스템도 사용자 프로파일을 기반으로 사용자가 관심있어 하는 항목을 추천한다. 예를 들면, 웹 사이트, 신문기사, 식당, TV 프로그램, 쇼핑 항목 등의 사용자가 원하는 대상을 자동적으로 추천해주게 된다.

최근에는 PA 에이전트의 활용 범위가 확대되고 있다. 사용자 작업공간에서 사용자 각각의 업무를 도와줄 수 있는 소프트웨어 에이전트의 개발도 이루어지고 있다 [9]. 사용자의 업무에 대한 과정을 기억하고 이를 기반으로 사용자 작업이 완성될 수 있도록 도와주는 역할을 수행한다. 협업설계 환경 [10]에서 각 개인 작업자의 모델을 생성하며 이를 기반으로 개인 작업자사이의 인터페이스 역할을 수행하는 PA 에이전트의 개발도 이루어지고 있다.

근래 모바일 통신 시장이 급격히 성장함에 따라 PA 에이전트의 2세대 어플리케이션들이 부각되고 있다 [9, 11]. 특히 모바일 사용자를 지원하기 위해 설계된 PA 에이전트는 개인화된 모바일 서비스의 광범위한 요구를 충족시킬 수 있는 가능성을 갖고 있다. 이러한 에이전트들은 정보 검색 및 필터링과 같은 기존의 서비스를 제공할 뿐만 아니라 무선 환경에서 요구되는 새로운 타입의 서비스를 제공할 수 있으며, 본 논문은 기존의 PA 에이전트 역할을 무선 이동환경에 적용시키고자 한다.

### 3. 사용로그 데이터 클러스터링 및 선호도 추출

사용자의 모바일 기기 사용 로그로부터 사용자 선호도를 예측하고 이를 서비스 및 콘텐츠 추천에 응용하는 PA 에이전트의 기능을 제공하기 위해서는 우선 모바일 기기가 취득 할 수 있는 정보 중에서 클러스터링과 분류 등의 데이터마이닝 수행에 의미가 있는 속성을 추려내고 정의할 필요가 있다. 사용자는 모바일 기기를 이용하는 장소, 시각에 따라 다양한 종류의 서비스 혹은 어플리케이션을 사용하게 되고 각각의 서비스를 통해 검색하거나 요청하는 콘텐츠의 내용 역시 장소와 시각에 따라 변화하게 된다. 최종적으로 서비스의 종류와 콘텐츠 내용을 사용자에게 추천하기 위해서는 사용자의 요청이 발생한 지점 간의 물리적 거리를 구하기 위한 위치 정보가 필요하며 사용 패턴을 나타내는 사용 시각, 요일, 서비스 종류, 콘텐츠 내용이 사용 로그에 포함되어야 한다.

표 1은 사용 로그의 각 데이터를 구성하는 속성과 변수 타입을 정리한 것이다. 표 1과 같이 구간 변수의 쌍, 이항 변수, 항목 변수 등이 혼합되

어 표현된 데이터 집합에 일반적인 분류 알고리즘을 적용하여 학습 모델을 작성하게 되면 타당한 결과를 얻을 수 없다. 예를 들어, 사용자의 위치와 특정한 요일의 서비스 시각에 적절한 서비스와 그에 대한 콘텐츠가 의미 있다는 것이다.

따라서, 사용 로그의 속성에 따른 타당한 결과를 얻기 위해서는 각각의 속성이 갖는 의미적 중요도를 파악하고 속성의 변수 타입을 고려하여 그에 적합한 데이터 전처리와 데이터마이닝 과정을 수행할 필요가 있다. 이를 위해 다음 두 가지 단계를 진행하도록 하였는데 첫 번째 단계에서는 위치 좌표에 따른 거리 값과 사용패턴 상이도를 측정하여 로그 데이터 셋을 여러 개의 클러스터로 그룹화하고, 두 번째로 각각의 클러스터에 속한 데이터 객체들에 분류 알고리즘을 적용하여 클러스터별로 학습모델을 작성한다.

#### 3.1 위치정보와 사용패턴을 고려한 로그데이터 클러스터링

사용 로그 데이터를 관찰한 결과 사용자의 일정한 생활 패턴에 따라 로그 데이터 객체들 간의 밀집도가 높고 사용자의 한 번의 요청에 대하여

(표 1) 사용 로그 데이터 속성 및 변수 타입

| 속 성 | 변수 타입                            | 설 명   |
|-----|----------------------------------|---|
| 위 치 | Pair of Interval-Scaled variable | GPS로 수신한 경도, 위도 좌표  |
| 시 각 | Nominal                          | 서비스 사용 시각   |
| 요 일 | Symmetric Binary                 | 서비스 사용 요일   |
| 서비스 | Nominal                          | 전화, 메일, 음악 감상, 웹서핑, 날씨, 교통정보 등 사용자가 사용한 서비스 혹은 어플리케이션의 종류 |
| 콘텐츠 | Nominal                          | 서비스를 통해 검색하거나 제공을 요청한 콘텐츠의 내용                             |

하나의 데이터 객체를 생성하고 저장함으로써 데이터 집합 크기가 크다는 사실을 고려하여, 동일한 클러스터 안의 객체들이 잘 밀집되어 있을수록 좋은 결과를 내며, 알고리즘 복잡도가 낮아 큰 데이터 집합을 다루는데 상대적으로 효율적인 k-means 알고리즘이 사용 로그데이터 클러스터링에 사용되었다.

클러스터링을 적용하는 사용 로그 데이터는 크게 경도와 위도 좌표로 표시되는 위치 정보와 사용 시각, 요일, 서비스 종류와 콘텐츠 내용을 포함하는 사용 패턴 정보로 나눌 수 있는데 위치 정보와 사용 패턴의 의미적 중요성을 고려하여 두 객체간의 상이성을 측정하기 위해서 3장에서 소개한, 모든 변수를 [0.0, 1.0]의 범위를 갖도록 변경하는 일반적인 혼합형 변수 객체 상이성 척도가 아닌 유클리드 거리와 혼합형 변수 객체 상이성 척도를 결합한 함수를 사용하도록 한다.

로그 데이터가 갖는 속성의 집합이  $A = (g, h, w, s, c)$  일 때 ( $g$ =위치,  $h$ =시각,  $w$ =요일,  $s$ =서비스,  $c$ =콘텐츠),  $i = (x_{ig}, x_{ih}, x_{iw}, x_{is}, x_{ic})$  와  $j = (x_{jg}, x_{jh}, x_{jw}, x_{js}, x_{jc})$  두 객체 간의 상이도  $d(i, j)$ 는 다음과 같이 정의된다.

$$d(i, j) = (|x_{ig} - x_{jg}|^2 + |x_{ih} - x_{jh}|^2 + |x_{iw} - x_{jw}|^2 + |x_{is} - x_{js}|^2 + |x_{ic} - x_{jc}|^2)^{1/2} \times \frac{\sum_{l \in A-g} \delta_{ij}^{(l)} d_{ij}^{(l)}}{\sum_{l \in A-g} \delta_{ij}^{(l)}} \quad (\text{수식 1})$$

위 함수는 객체 간의 위치 정보 속성 위도, 경도 값의 유클리드 거리와 위치정보 속성을 제외한 나머지 사용 패턴과 관련된 속성의 값들로 측정된 상이도의 곱으로 표현되었다.

- $x_{ig}, X$ 와  $x_{ij}, Y$ 는 각각 객체  $i$ 의 위치정보 속성 값  $x_{ig}$ 에 포함된 위도, 경도 값을 의미한다.
- $x_{il}$ 나  $x_{jl}$ 의 값을 측정하지 못하였을 때, 혹은  $x_{il} = x_{jl} = 0$ 이고 변수  $l$ 이 비대칭 이항 변수일 때 지시값  $\delta_{ij}^{(l)} = 0$ 이며, 그 외의 경우

$\delta_{ij}^{(l)} = 1$ 이다.

- 변수  $l$ 의 두 객체  $i, j$ 의 상이성에 대한 기여도  $d_{ij}^{(l)}$ 은  $x_{il} = x_{jl}$ 이면 0, 그 외의 경우 1이다.

위 함수를 사용하여 두 객체 사이의 거리를 구하고 k-means 알고리즘을 적용하여 사용 로그 데이터 셋을 k개의 클러스터로 나눈다.

### 3.2 분류 알고리즘을 이용한 선호도 예측 학습 모델 생성

여러 개의 클러스터로 분할된 사용 로그 데이터 셋의 학습과 분류에 적합한 알고리즘을 선택하기 위해 다음과 같은 사항을 고려한다. 첫째, 데이터 셋의 크기가 크다. 둘째, 속성의 개수가 많지 않으며 속성과 클래스 레이블 모두 이항(binary) 혹은 항목(nominal) 변수 타입을 갖는다. 셋째, 모바일 기기의 특성상 일반 데스크탑 컴퓨터에 비해 계산 능력과 메모리 자원의 한계를 갖는다. 즉 대용량 데이터에 대하여도 보장된 결과를 내어주며 학습모델 생성을 위한 학습 과정과 생성된 학습모델을 이용한 분류 과정의 복잡도가 크지 않아야 한다. 도메인의 이러한 특성들을 고려하여 개별 클러스터에 적용할 분류 알고리즘은 대용량 데이터에 대하여도 보장된 결과를 내어주며 학습 모델 생성을 위한 학습 과정과 생성된 학습모델을 이용한 분류 과정의 복잡도가 크지 않은 의사 결정 트리(Decision Tree) 알고리즘을 적용하여 개별 클러스터에 속한 데이터 셋의 객체들을 대상으로 분류(Classification)를 수행한다.

분류 과정은 각각의 클러스터에 독립적으로 수행되며 결과적으로 각 클러스터는 개별적인 학습 모델을 보유한다. 또한, 사용자 요청에 따라 서비스 혹은 콘텐츠를 예측하고 추천하기 위해서 서비스 속성을 클래스 레이블로 하는 하나의 학습 모델과 콘텐츠 속성을 클래스 레이블로 하는 학습 모델 그룹을 생성한다.

첫 번째, 서비스 속성을 클래스 레이블로 하는

학습 과정을 살펴보면, 이미 동일한 클러스터에 속한 데이터 객체 간의 위치 속성 값은 데이터 분류에 큰 의미가 없으므로 학습 과정에 고려하는 속성 목록에서 제외하고, 콘텐츠 속성 값의 경우 동일한 튜플의 서비스 속성 값에 종속적이며 서비스 값에 따라 전혀 다른 범주와 특성을 가지므로 (예를 들어, 전화 서비스에서 사용자가 요청하는 콘텐츠는 전화번호이며 음악 감상 서비스의 경우 노래 혹은 노래 목록이 저장된 플레이리스트이다.) 콘텐츠 속성 역시 학습 과정에서 제외된다.

두 번째, 콘텐츠 속성을 클래스 레이블로 하는 학습 과정을 살펴보면, 첫 번째와 마찬가지로 위치 속성은 학습 과정에서 고려하는 속성 목록에서 제외된다. 이 역시 학습과정에서 제외한다.

첫 번째에서 언급했듯이 콘텐츠 속성 값은 동일 튜플의 서비스 속성 값에 종속적이다. 따라서 서비스 속성 값이 다른 튜플들을 콘텐츠 값 예측을 위한 학습 과정의 트레이닝 데이터 셋에 포함시키는 것은 아무런 의미가 없으며 학습과정의 복잡도만 증가시킬 뿐이다. 같은 서비스 속성 값을 갖는 튜플들을 트레이닝 데이터 셋으로 하여 학습 과정을 진행하고 학습모델들을 생성한다.

위의 과정을 거치면 각각의 클러스터는 서비스 값을 예측하기 위한 한 개의 학습 모델과 콘텐츠 값을 예측하기 위한 서비스 개수와 동일한 개수의 학습 모델을 보유하게 된다.

결과적으로  $k$ 개의 클러스터가 존재하고 서비스 속성 값의 종류가  $n$ 개일 때 에이전트가 보유하는 총 학습모델 개수는  $k \times (1+n)$ 개가 된다.

서비스 혹은 콘텐츠의 클래스 레이블이 정해지지 않은 새로운 사용자 요청은  $x_{es} = null$  and  $x_{ec} = null$  혹은  $x_{ec} = null$  인 사용자 로그 데이터 객체  $e = (x_{eg}, x_{eh}, x_{ew}, x_{es}, x_{ec})$  로 정의된다.

해당 데이터의 클래스 레이블 값을 예측하기 위해서는 유사한 클러스터를 확인할 필요가 있으며 이를 위해 수식 1의 상이도 함수를 이용하여  $1 \leq i \leq k$ 인 클러스터  $C_i$ 의 평균  $m_i$ 과의 상이도를 측정하고 그 값이 최소가 되는 클러스터의

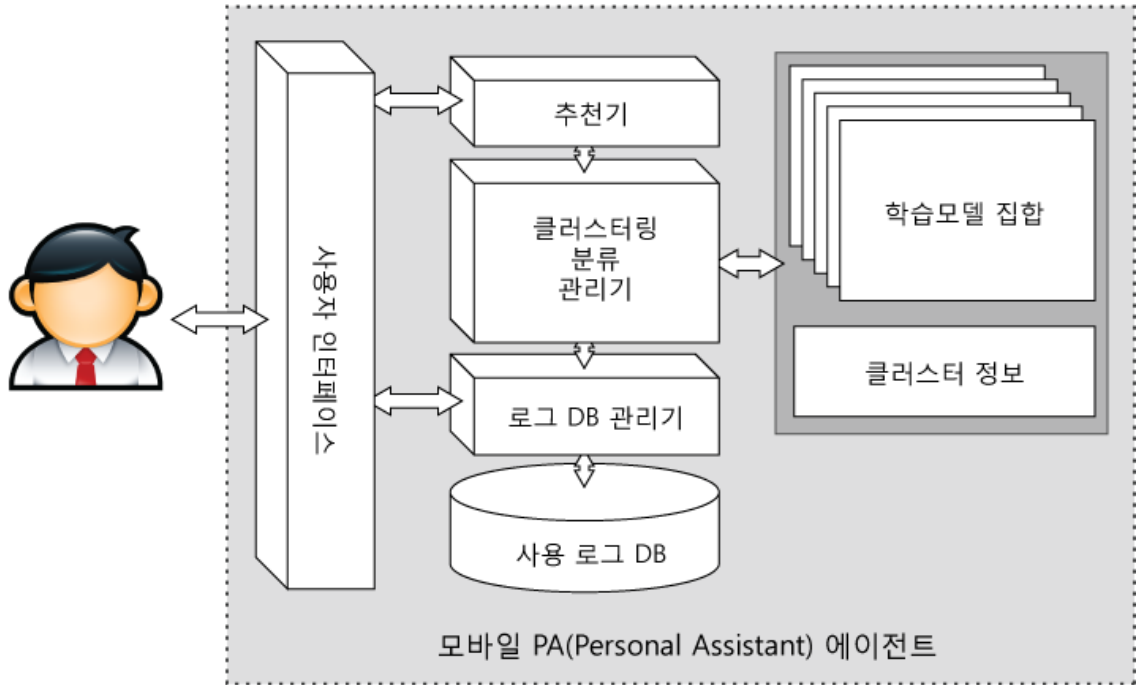
학습모델을 적용하여 서비스 혹은 콘텐츠의 클래스 레이블 값을 예측한다.

#### 4. 모바일 PA에이전트 구조

모바일 환경에서 PA 에이전트가 사용자의 위치정보와 선호도를 반영하여 잠재적 요구와 필요에 부합하는 지원을 제공하기 위해서는 사용 로그 데이터를 관리하며 사용 로그 데이터로부터 의미 있는 속성들을 추출해 분류 알고리즘을 적용하고 작성된 학습모델을 토대로 클래스를 예측하여 사용자가 서비스 혹은 어플리케이션을 이용할 때 적합한 서비스 혹은 콘텐츠를 추천해주어야 한다. 위와 같은 각각의 기능들을 수행하는 컴포넌트들은 상호 의사소통을 통해 최종적으로 개인 적응형 지원이라는 PA 에이전트의 목적을 달성한다. 각 기능을 구현하는 컴포넌트들로 구성된 모바일 PA 에이전트의 구조는 아래 그림 2와 같다.

각각의 컴포넌트들은 아래와 같이 주어진 기능을 수행하고 다른 컴포넌트와 상호 의사소통한다.

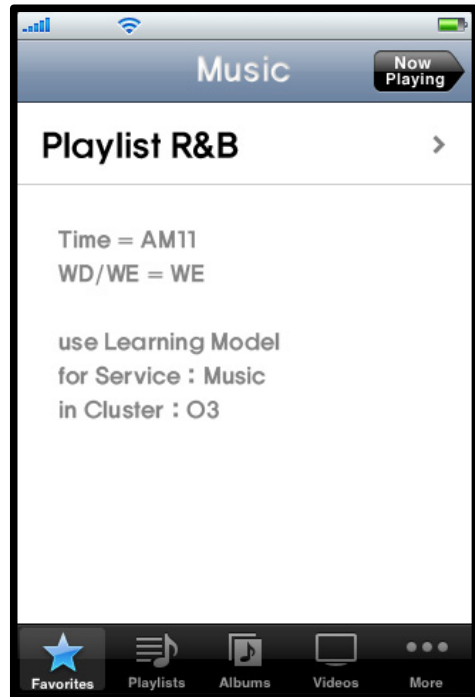
- 사용 로그 데이터베이스(Usage Log Database) : 사용 로그 데이터 집합이 저장된 스토리지
- 로그 DB 관리자(Log DB Manager) : 사용자의 사용 로그를 사용 로그 데이터베이스에 저장하고, 클러스터링/분류 관리기의 요청에 따라 필요한 형태와 종류의 데이터를 제공한다.
- 클러스터링 분류 관리자(Clustering and Classification Manager) : 로그 DB 관리기로부터 제공받은 데이터를 이용하여 클러스터링 및 분류를 수행하고 그 결과를 토대로 클러스터 정보와 학습모델을 작성하거나 수정한다. 계산 능력과 메모리 자원의 제약이 있는 모바일 환경에서 사용자 선호도를 학습모델에 반영하기 위해서 클러스터링 및 분류를 이용한 학습 과정은 정해진 주기로 수행된다. 또한 추천기의 요청에 따라 서비스 혹은 콘텐츠의 값을 예측하고 추천기에 반환한다.



(그림 2) 모바일 PA(Personal Assistant) 에이전트의 구조

- 추천기(Recommender) : 사용자가 서비스 혹은 어플리케이션을 사용할 때 사용자 위치정보와 선호도를 반영한 서비스 혹은 콘텐츠를 추천하기 위해 클러스터링/분류 관리기에게 예측을 요청하고, 반환받은 예측 값을 UI를 통해 사용자에게 추천한다.
- 클러스터 정보(Cluster Information) : 사용자 로그 데이터 셋 클러스터링을 토대로 생성된 클러스터들의 정보를 갖는다.
- 학습모델 집합(Learning Model Set) : 분류 알고리즘의 학습 과정에서 작성한 개별 클러스터들의 학습 모델 집합.

구현한 모바일 PA 에이전트가 동작하는 GUI 화면에 대한 예시는 그림 3과 같다. 그림 3은 주말 오전 11시 특정한 지역(클러스터 O3)에서 사용자의 위치정보와 선호도를 반영하여 음악 서비스를 하고 있는 화면을 나타낸다.



(그림 3) 모바일 PA(Personal Assistant) 에이전트 GUI 예시

## 5. 실험

실험에서는 위치 좌표에 따른 거리 값과 사용 패턴 상이도를 측정하여 로그 데이터 셋을 여러 개의 클러스터로 그룹화하고, 각각의 클러스터에 속한 데이터 객체들에 대하여 분류 알고리즘을 적용한 정확도를 측정하였다.

### 5.1 상이도 측정 함수에 따른 클러스터링 결과 비교

최적의 클러스터 개수  $k$ 를 찾기 위해 일반적인 타입 혼합형 상이도 측정 함수를 사용한 클러스터링 결과와 제시한 상이도 측정 함수를 사용한 클러스터링 결과를 일반적으로 널리 사용되는 다음의 세 가지 평가 지수 [12]로 평가한다. 클러스터링 적용을 위해 실제 모바일 기기 사용자의 어플리케이션 및 콘텐츠 이용 내역을 2달간 기록하여 1,276개의 인스턴스를 갖는 로그 데이터를 생성하였으며, PC 환경에서 클러스터링 알고리즘을 구현하여 실험하였다.

#### 1) Harigan

$k$ -means 클러스터링 알고리즘에서 최적의 클러스터 개수  $k$ 를 찾기 위해 J. A. Harigan에 의해 제안되었다 [13].

$$H(k) = \gamma(k) \frac{W(k) - W(k+1)}{W(k+1)}, \gamma(k) = N - k - 1 \quad (\text{수식 2})$$

$W(k)$ 는 클러스터 중심과 클러스터에 속한 객체들의 제곱 거리의 합으로 정의되며, 파라미터  $\gamma$ 은  $k$ 값 증가에 따른 단조 증가를 피하기 위해 제시되었다. Harigan에 따르면 최적의 클러스터 개수는  $H(k) \leq \eta$  (일반적으로  $\eta = 10$ )인  $k$  중 최소값으로 정한다.

#### 2) Davies Bouldin 지수

Davies Bouldin 지수는 조밀하고 하고 잘 분리된 클러스터들을 찾기 위해 제안되었으며 다음과 같이 계산한다 [14].

$$DB(k) = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \frac{\Delta(C_i) + \Delta(C_j)}{\delta(C_i, C_j)} \quad (\text{수식 3})$$

$\Delta(C_i)$ 은 클러스터의  $C_i$ 의 모든 객체와 클러스터 중심과의 평균 거리를 나타내며  $\delta(C_i, C_j)$ 는 클러스터  $C_i$ 와 클러스터  $C_j$ 의 거리 (클러스터 중심 사이의 거리를 나타낸다. 최적의 클러스터 개수는  $DB(k)$ 의 값이 최소가 되는  $k$ 이다.

#### 3) Silhouette

이 방법은 각 객체의 퀄리티 지표를 나타내는 Silhouette 너비에 기반하며 다음과 같이 정의된다 [15].

$$Sil(x_i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (\text{수식 4})$$

$a(i)$ 는 객체  $i$ 와 같은 클러스터에 속한 모든 객체들 사이의 거리 평균을 나타내며  $b(i)$ 는 객체  $i$ 와 가장 클러스터의 객체들 사이의 거리 평균을 나타낸다. 각 클러스터의 Silhouette 값은 클러스터에 속한 객체들 Silhouette 값 평균을 이용하여 구하며,

$$sil(x_i) = \frac{1}{|C_j|} \sum_{x_i \in C_j} sil(x_i) \quad (\text{수식 5})$$

모든 클러스터의 Silhouette 값 평균으로 전체적인 Silhouette 값을 구한다.

$$sil(k) = \frac{1}{k} \sum_{r=1}^k sil(C_r) \quad (\text{수식 6})$$



$sil(k)$  값을 최대로 하는  $k$ 가 최적의 클러스터 개수이다.

표 2의 실험 결과에서 괄호 안의 값은 평가지수별 순위를 의미한다. 세 가지 지수를 종합적으로 살펴보면 일반적인 타입 혼합형 상이도 측정 함수를 사용한 경우 최적의 클러스터 개수  $k$ 는 4로 측정되었으며 (각각  $H(4)$ : 44.9 (2),  $DB(4)$ : 0.708 (1),  $sil(4)$ : 0.969 (1)), 제안한 상이도 측정 함수를 사용한 경우  $k$ 는 8로 측정되었다 (각각  $H(8)$ : 22.7 (1),  $DB(8)$ : 0.723 (3),  $sil(8)$ : 0.872 (1)).

### 5.2 상이도 함수에 따른 클러스터별 분류 정확도 비교 및 분석

위 과정의 클러스터 평가 결과에 따라 일반적인 타입 혼합형 상이도 함수를 사용하여 클러스터링 한 경우와 제시한 상이도 측정 함수를 사용하여 클러스터링 한 경우 각각의 최적 클러스터 개수를 정하고 클러스터 그룹을 생성한다. 각각의 클러스터 그룹에 대하여 클러스터별로 분류 알고리즘을 수행하여 학습모델을 작성하고 분류 정확도를 비교한다. 분류 정확도는 가장 일반적인 10-fold cross validation 방식으로 측정하였다.

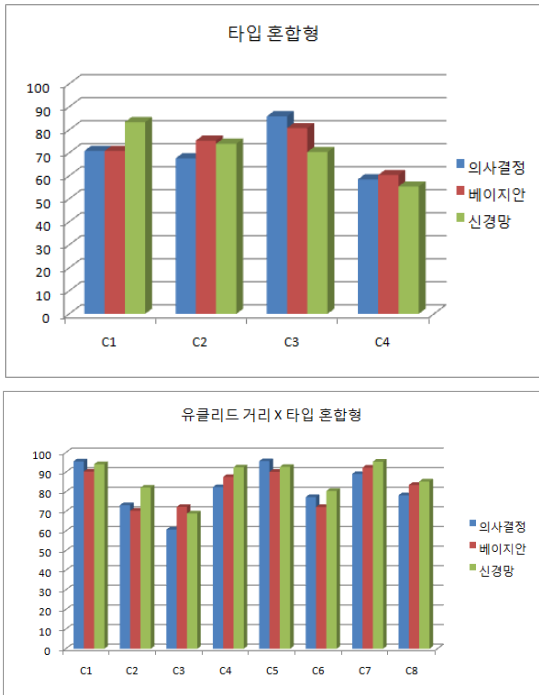
(표 2) 상이도 측정 함수별 클러스터링 결과

| 상이도 함수 | 타입 혼합형 |          |           | 유클리드 거리 × 타입 혼합형 |          |           |           |
|--------|--------|----------|-----------|------------------|----------|-----------|-----------|
|        | $H(k)$ | $DB(k)$  | $Sil(k)$  | $H(k)$           | $DB(k)$  | $Sil(k)$  |           |
| k      | 2      | 242.4    | 0.806 (2) | 0.707            | 236.7    | 0.717 (2) | 0.751 (3) |
|        | 3      | 89.7     | 0.908     | 0.59             | 261.2    | 0.745     | 0.643     |
|        | 4      | 44.9 (2) | 0.708 (1) | 0.969 (1)        | 103.1    | 0.654 (1) | 0.722     |
|        | 5      | 108.5    | 1.671     | 0.707            | 125.1    | 0.784     | 0.793 (2) |
|        | 6      | 56.7     | 2.208     | 0.717            | 97.6 (3) | 0.810     | 0.668     |
|        | 7      | 55.6     | 1.508     | 0.737 (3)        | 32.8 (2) | 0.791     | 0.706     |
|        | 8      | 168.7    | 0.88 (3)  | 0.838 (2)        | 22.7 (1) | 0.723 (3) | 0.872 (1) |
|        | 9      | 52.4 (3) | 1.327     | 0.720            | 167.8    | 0.944     | 0.716     |
|        | 10     | 33.5 (1) | 1.615     | 0.702            | 99.8     | 0.982     | 0.677     |

(표 3) 상이도 함수에 따른 클러스터별 분류 정확도

| 상이도 함수           | 클러스터  | 분류 알고리즘 |       |       |
|------------------|-------|---------|-------|-------|
|                  |       | 의사결정 트리 | 베이지안  | 신경망   |
| 타입 혼합형           | $C_1$ | 70.84   | 70.84 | 83.33 |
|                  | $C_2$ | 67.58   | 75.32 | 73.95 |
|                  | $C_3$ | 85.87   | 80.73 | 70.36 |
|                  | $C_4$ | 58.47   | 60.36 | 55.36 |
|                  | avg.  | 70.69   | 71.81 | 70.75 |
| 유클리드 거리 × 타입 혼합형 | $C_1$ | 95.52   | 90.34 | 94.12 |
|                  | $C_2$ | 73.24   | 70.38 | 82.17 |
|                  | $C_3$ | 60.75   | 72.32 | 68.96 |
|                  | $C_4$ | 82.43   | 87.58 | 92.52 |
|                  | $C_5$ | 95.67   | 90.25 | 92.75 |
|                  | $C_6$ | 77.32   | 72.27 | 80.46 |
|                  | $C_7$ | 89.12   | 92.34 | 95.38 |
|                  | $C_8$ | 78.26   | 83.54 | 85.32 |
|                  | avg.  | 81.54   | 82.38 | 86.46 |

표 3을 그래프 양식으로 정리한 결과는 그림 4와 같다. 생성된 클러스터들의 분류정확도의 평균을 살펴보면 일반적인 타입혼합형 상이도 측정 함수를 사용한 경우 분류 알고리즘 별로 의사결정 트리, 베이지안 분류기, 신경망 알고리즘이 각각 70.69, 71.81, 70.75의 결과를 나타내었으며, 이에 대한 평균 분류 정확도는 71.08이다. 제안한 상이도 측정 함수를 사용한 경우 각각 81.54, 82.38, 86.46의 결과를 나타내었으며, 평균 분류 정확도는 83.46이다. 따라서, 일반적인 타입혼합형 상이도 측정 함수를 사용한 경우 평균적인 분류 정확도 보다 본 논문에서 제안한 상이도 측정 함수를 사용한 경우 17.42%의 성능향상을 나타냄을 알 수 있었다. 이러한 결과는 클러스터를 구성하는 각 인스턴스의 군집력을 강화한 대상으로 분류 정확도를 측정하는 것이 보다 일관적일 수 있음을 반영한다.



(그림 4) 상이도 함수별 결과 비교

## 6. 결론

본 논문에서는 모바일 환경의 여러 가지 특성들을 고려하여 이에 적합한 PA(Personal Assistant) 에이전트를 설계하고 각 컴포넌트의 기능을 구현하기 위한 기법을 제안하였다. 위치정보가 포함된 사용 로그 데이터에 분류 알고리즘을 적용하기 위해 위치 좌표에 따른 거리 값과 사용 패턴 상이도를 고려한 클러스터링을 제안하였다. 클러스터링을 수행하기 위하여 모바일 환경의 사용자 로그 데이터에서 두 객체간의 상이도를 측정하는 함수를 제시하였다. 모바일 기기가 취득할 수 있는 사용 내역에 관한 다양한 정보 중에서 학습 모델 생성을 위한 학습 과정에서 의미적 중요도가 높은 속성들을 추출하고 정의하였으며, 클러스터링을 통해 생성한 클러스터별로 기계학습을 적용하여 학습모델을 작성하고 이용함으로써 위치에 따른 사용자의 선호도를 반영하는 서비스와 콘텐츠

츠의 추천이 가능하도록 하였다.

기존의 일반적인 상이도 함수를 적용한 클러스터링 결과와 제시한 상이도 함수를 적용한 클러스터링 결과에 대하여 세 가지 평가지수를 이용하여 평가를 시행하였다. 각각 최적의 클러스터 개수(k)를 찾았으며 이를 바탕으로 생성된 각 클러스터의 인스턴스들을 이용하여 클러스터별 학습모델을 생성하여 그 분류정확도를 측정하였다. 그 결과 제시한 상이도 함수로 클러스터링을 수행하였을 경우의 클러스터별 학습모델 분류정확도가 상대적으로 높음을 알 수 있었다.

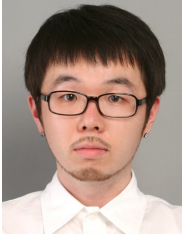
앞으로 계속적인 연구를 통하여 사용자의 위치와 선호도를 반영하는 PA 에이전트를 스마트폰 환경에서 구현할 것이다. 개발한 PA 에이전트가 사용자의 선호도를 어느 정도 반영하는가를 사용자의 만족도를 통하여 측정할 것이다. 또한, 궁극적으로 사용자의 학습모델이 사용자에 맞도록 적용할 수 있는 점진적인 학습모델을 구현 할 것이다. 사용자의 경험을 추적하여 이를 선호도로 반영할 수 있는 지능적인 PA 에이전트의 구현은 이성적인 소프트웨어 에이전트(rational software agent)의 개인화된 프로토타입이 될 것이며, 앞으로 모바일 스마트폰이 언제 어디서나 개인활동을 지능적으로 지원하는 개체로써 자리 잡을 수 있도록 할 것이다.

## 참고 문헌

- [1] 유성호, "위치기반(LBS) 모바일게임의 활성화 방안 - 국내게임과 해외게임의 사례연구를 중심으로", 한국콘텐츠학회논문지, 제8권 제11호, pp. 74-84, 2008.
- [2] P. Maes, "Agents that reduce work and information overload," Communication of the ACM, Vol. 37, No. 7, pp. 31-40, July 1994.
- [3] M. J. Pazzani and D. Billsus, "Learning and Revising User Profiles: The Identification of Interesting Web Sites," Machine Learning, Vol.

- 27, pp. 313-331, 1997.
- [4] D. Mladenic, "Machine learning used by Personal WebWatcher," In Proceedings of ACAI-99 Workshop on Machine Learning and Intelligent Agents, 1999.
- [5] B. J. Rhodes and T. Starner, "Remembrance Agent: A continuously running automated information retrieval system," In Proceedings of the First International Conference on the Practical Application of Intelligent Agents and Multi Agent Technology (PAAM '96), pp. 487-495, 1996.
- [6] H. Lieberman, "Letizia: An Agent that Assists Web Browsing," In Proceedings of 14th International Joint Conference on Artificial Intelligence (IJCAI-95), pp. 924-929, 1995.
- [7] F. Menczer, A. E. Monge and W. N. Street, "Adaptive Assistants for Customized E-Shopping," IEEE Intelligent Systems, Vol. 17, no. 6, pp. 12-19, 2002.
- [8] M. J. Pazzani and D. Billsus, "Content-based recommendation systems," The Adaptive Web, Lecture notes in computer science, Springer-Verlag, Berlin, Germany, Vol. 4321, pp. 325-341, May 2007.
- [9] Ke-Jia Chen and J. P. Barthes, "Enhancing Intelligence of Personal Assistant Agent Using Memory Mechanism," In Proceedings of the 11th International Conference on Computer Supported Cooperative Work in Design, pp. 361-365, April 2007.
- [10] S. Wu, H. Ghenniwa, Y. Zhang and Weiming Shen, "Personal assistant agents for collaborative design environments," Journal of Computers in Industry, Vol. 57, Issues 8-9, pp. 732-739, December 2006.
- [11] T. Rahwan, T. Rahwan, I. Rahwan and R. Ashri, "Agent-Based Support for Mobile Users Using AgentSpeak(L)," In P. Giorgini, B. Henderson-Sellers, and M. Winikoff, editors, Agent-Oriented Information Systems AAMAS-04 Workshop, Lecture Notes in Artificial Intelligence, Vol. 3030, pp. 47-62, Springer-Verlag, 2004.
- [12] M. Anderberg, Cluster Analysis for Applications, New York: Academic, 1973.
- [13] J. Hartigan, Clustering Algorithms, New York, USA: Wiley, 1975.
- [14] D. Davies and D. Bouldin, "A cluster separation measure," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 1, No. 2, pp. 224-227, 1979.
- [15] P. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," Journal of Computational and Applied Mathematics, Vol. 20, No. 1, pp. 53-65, 1987.

## ◎ 저 자 소개 ◎



### 강 신 봉

2008년 가톨릭대학교 컴퓨터정보공학부 컴퓨터공학전공 졸업(학사)

2010년 가톨릭대학교 대학원 컴퓨터공학과 졸업(석사)

관심분야 : 데이터마이닝, 인공지능, 개인화 에이전트

E-mail : buena80@gmail.com



### 노 상 옥

1987년 서강대학교 생명과학과 졸업(학사)

1989년 서강대학교 대학원 컴퓨터공학과 졸업(석사)

1999년 텍사스주립대학교(Arlington) 대학원 컴퓨터공학과 졸업(박사)

2000~2002 미조리주립대학교(Rolla) 컴퓨터학과 교수

2002~현재 가톨릭대학교 컴퓨터정보공학부 교수

관심분야 : 지능형 에이전트 시스템, 기계학습, 데이터마이닝, 인공지능, 등

E-mail : sunoh@catholic.ac.kr