

위치에 따른 연령대별 유용한 행동패턴 추출 기법

김혜란*, 이승철*, 김응모*

*성균관대학교 컴퓨터공학과

e-mail : {arimakim, eddie, umkim}@ece.skku.ac.kr

Efficient Mining of User Behavior patterns by classification of age based on location information

HyeRan Kim*, SeungCheol Lee*, and UngMo Kim*

*Dept. of Computer Engineering, SungKyunKwan University

요 약

통신기술의 발달로 무선단말기의 보급이 급증하고 무선 네트워크 사용이 일반화됨으로써, 최근 유비쿼터스 컴퓨팅 기술이 중요한 이슈가 되고 있다. 유비쿼터스 컴퓨팅은 시간과 장소의 한계를 넘어 사용자가 하고자 하는 일을 컴퓨팅 환경이 상황을 인지하여 돕는 것을 가능하게 한다. 상황인지를 위해 순차패턴과 시간 연관규칙 탐사를 이용하여 사용자의 행동패턴을 추출하는 연구가 활발히 진행되고 있다. 이러한 연구를 통한 행동패턴은 사용자의 특성을 간과하게 되며, 각 사용자에게 더욱 유용한 서비스를 제공하기 위해서는 사용자를 분류하는 것이 필요하다. 그러나 기존의 연구는 단지 통계적인 사용자의 빈발 행동패턴만을 추출하여 각 사용자의 관심사와는 무관한 서비스 제공이 이루어질 수 있다. 성별, 나이, 직업 등의 개인정보와 위치를 고려하여 사용자에게 더욱 더 효율적이고 유용한 서비스를 제공할 수 있도록 행동패턴을 유형별로 분류할 필요가 있다. 본 논문에서는 각 위치에 따른 사용자의 연령대별 유용한 행동패턴을 추출하여 정확한 서비스를 제공할 수 있는 마이닝 기법을 제안한다.

1. 서론

PDA 또는 모바일 폰 등의 무선단말기의 보급이 급증하고 무선 네트워크 사용이 일반화됨으로써, 언제 어디서나 컴퓨팅 서비스를 이용 가능하게 하는 유비쿼터스 컴퓨팅 기술이 중요한 이슈가 되고 있다. 유비쿼터스 컴퓨팅은 다양한 종류의 컴퓨터가 사람, 사물, 환경 속으로 스며들고 서로 연결되어, 언제 어디서나 컴퓨팅을 구현할 수 있는 환경을 말한다. 이러한 유비쿼터스 컴퓨팅 기술의 급속한 발전으로 우리 환경 도처에서 개개인의 데이터를 수집할 수 있게 되었다. 축적된 대량의 데이터로부터 사용자에게 적절한 서비스를 제공하기 위해서는 유용한 지식을 추출하는 효율적인 데이터 마이닝 기법이 필요하다. 그리하여 기존 서비스 제공 환경의 한계를 극복한 새로운 개념의 서비스 환경을 구축하기 위한 연구가 활발하게 진행되고 있다[5]. 유용한 서비스를 제공하기 위해 함께 고려되어야 하는 점은 다양한 사용자들의 서비스 요청에 각 사용자의 특성에 맞는 적절한 서비스를 어떻게 제공할 것이냐는 점이다. 이점은 특히 많은 수의 사용자들이 이용하고 여러 하위공간으로 이루어지며 다양한 종류의 서비스가 제공되는 캠퍼스, 멀티플렉스, 호텔 등의 복잡한 지역에서 더욱 심각하게 나타나는 문제이다. 즉, 입출력의 제한을 가진 모바일 사용자는 자신과 관련된 서비스를 쉽게 찾아 이용할 수 있어야 한다. 그러나 기존의 연구는 통계적인 사용자의 빈발

행동패턴만을 추출하여 각 사용자의 관심사와는 무관한 서비스 제공이 이루어질 수 있는 문제점이 있다. 사용자의 행동패턴을 성별, 나이, 직업, 취미 등의 개인정보를 이용하여 유형별 행동패턴으로 도출하면 각 사용자의 특성에 맞는 서비스를 제공할 수 있게 된다. 그리하여 본 논문에서는 위치에 따라 사용자의 연령을 분류하여 그룹별 행동패턴으로 정확한 서비스를 제공할 수 있는 마이닝 기법을 제안한다.

제안하는 기법은 모바일 사용자의 유용한 행동패턴을 모바일 에이전트 시스템에서 발견하며 사용자의 특성, 관심사를 고려한 서비스 제공을 목표로 한다. 예를 들어 유비쿼터스존에서 모바일 사용자가 특정 서비스를 이용하기 전 여가시간이 생겼을 때, 모바일 사용자가 해당하는 연령대의 행동패턴을 이용하면 사용자의 요구에 맞는 정확한 서비스 추천이 가능하게 된다. 유형별 행동패턴으로 사용자가 선호할만한 서비스를 예측하게 됨으로써 잠재적 고객을 이끌 수 있게 되는 것이다. 이렇듯 모바일 에이전트에서 사용자 행동패턴의 효과적인 모델링의 이익은 모바일 사용자뿐만 아니라 모바일 서비스를 제공하는 제공자에게도 많은 이익을 줄 수 있다.

본 논문의 구성을 다음과 같다. 2 장에서는 관련연구로 연관규칙 마이닝과 모바일 에이전트의 개념을 기술하고, 3 장은 시나리오, 4 장은 시스템 구성, 5 장은 위치에 따른 연령대별 유용한 행동패턴 추출기법을 설명하고, 마지막으로 6 장에서 결론을 맺는다.

2. 관련 연구

본 장에서는 연관규칙 마이닝과 모바일 에이전트를 살펴본다.

2.1 연관 규칙 마이닝

데이터 마이닝은 대량의 데이터베이스에서 알려지지 않은 데이터 규칙 및 패턴을 발견하는 데이터 분석 기술이다. 데이터 마이닝은 연관, 순차, 분류, 클러스터링, 의사결정, 사례기반 추론, 퍼지로지, 신경망, 유전적 알고리즘 등의 방법이 있다.

그 중 연관 규칙이란 어떤 사건(항목)이 일어나면 다른 사건이 일어나는 관련성을 의미하며, 이때 복수의 사건 사이의 관련성도 포함한다. 즉, 한 항목 그룹과 다른 항목 그룹 사이에 강한 연관성이 있음을 밝혀준다. 연관 규칙 탐사는 사용자에게 의해 적절하게 입력된 지지도(support), 신뢰도(confidence)라는 척도를 이용하여 데이터 상호간의 관련성을 파악할 수 있다. 이러한 비교기준은 수많은 품목들의 관계 속에서 의미 있는 관련성을 찾기 위해서 결과해석에 앞서 연관성의 내용이 일반화할 수 있는 내용인가를 판단하는데 있어 중요하다. 또한 데이터 마이닝에서의 연관성 분석 목적은 품목간의 연관관계를 수치로 정량화하기 위함이다.

2.1.1 연관 규칙 정의 및 성능척도

데이터베이스가 총 n 개의 트랜잭션 데이터로 구성되며 전체 m 개의 항목으로 구성된다고 하고 이를 I 라 하자. 연관규칙 R 은 조건부와 결과부로 구성되며 항목집합인 X 와 Y 에 대하여 ‘ X 가 일어나면 Y 도 일어난다’의 의미로 $R: X \Rightarrow Y$ 로 표현할 수 있다. 여기서 I 에 두 항목집합 X, Y 가 존재하고 ($X, Y \subseteq I$), X 와 Y 사이에는 어떠한 공통된 데이터가 존재하지 않을 경우 ($X \cap Y = \emptyset$) 이어야 한다.

연관규칙의 중요한 평가기준으로 지지도와 신뢰도가 있으며 다음과 같이 정의할 수 있다.

X 의 지지도 $support(X)$ 는 전체 트랜잭션에서 X 를 포함하는 트랜잭션의 비율을 나타내고, 규칙 R 에 대한 지지도인 $support(R) = support(X \cup Y)$ 은 전체 트랜잭션에서 X 또는 Y 에 있는 항목을 동시에 포함하는 트랜잭션수의 비율을 나타낸다. 지지도는 두 품목의 구매가 얼마나 자주 일어났는가를 측정하며 연관 규칙의 통계적 중요성을 의미한다.

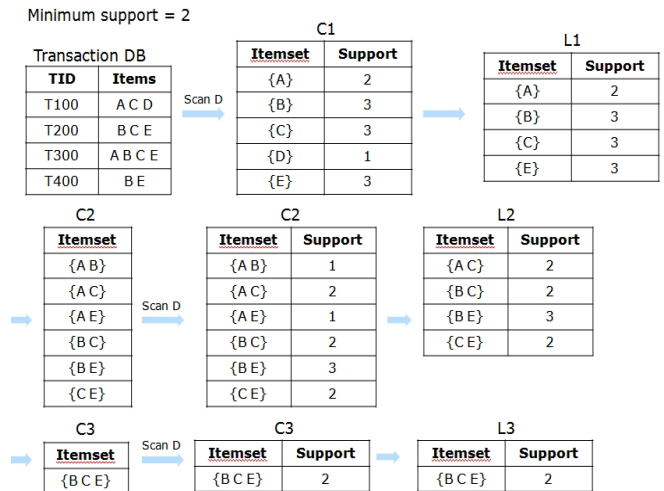
규칙 $X \rightarrow Y$ 의 신뢰도 $conf(X \rightarrow Y)$ 는 X 를 포함하는 트랜잭션에서 Y 를 동시에 포함하는 트랜잭션의 비율로 규칙의 강도를 의미하는 척도로 연관 규칙 마이닝에 이용된다. 즉, 품목 A 가 구매되었을 때 품목 B 가 추가로 구매될 확률을 의미한다[2,3,4].

2.1.2 Apriori 알고리즘

연관 규칙 탐사 방법 중의 하나인 Apriori 는 지지도를 이용하여 동시에 자주 나타나는 항목(빈발 항목 집합)들을 정제하고 빈발 항목 집합에서 생성된 규칙

들은 신뢰도를 이용하여 정제하는 방식이다.

Apriori 는 후보 항목 집합에서 각각의 지지도를 계산한 후 사용자가 정의한 지지도보다 크거나 같은 조건을 만족하는 데이터로 빈발 항목 집합을 구성한다. 그리고 후보 항목 집합은 전 단계의 빈발항목 집합의 조인연산을 통해 구성된다[6]. Apriori 알고리즘은 AprioriTid, AprioriHybrid 등과 같이 확장되어 연구되고 있다. (그림 1)은 Apriori 알고리즘이 빈발항목집합을 구성하는 과정을 나타낸다.



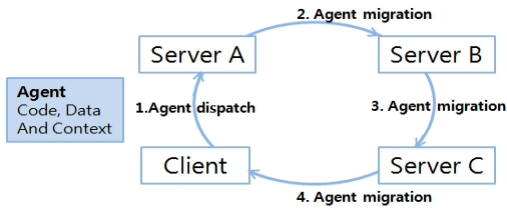
(그림 1) Apriori Algorithm 예시

2.2 모바일 에이전트의 개념

모바일 에이전트(Mobile Agent)는 하나의 독립적인 소프트웨어 모듈로써 자율성을 가지고 비동기적으로 수행이 가능한 개체이다. 선택적으로 지능을 가지고 특정 작업을 수행하기 위해 각 호스트와 호스트 사이, 여러 피어와 피어 사이의 네트워크 영역을 이동할 수 있는 개체라고 정의할 수 있다. 네이밍 서비스 기법을 이용하여 각 서비스 자원에 해당하는 유니크한 Name 을 찾기도 하고, 등록하기도 하며, 각각의 위치에 관한 정보를 전파하는 P2P 네트워크 형성자로써의 역할도 하게 된다. 또한 메시지 전달의 메커니즘도 포함한다.

네트워크 상의 호스트는 모바일 에이전트가 실행할 수 있는 환경을 제공하여 이질적인 하드웨어와 운영체제로 구성된 분산 컴퓨팅 환경에 동적인 이식을 가능케 하여(Portability) 서로 이질적인 분산 환경에서도 에이전트가 실행할 수 있는 환경을 제공하여 주는 에이전트 시스템의 역할을 수행한다. 이러한 에이전트는 특정 작업을 수행하기 위해 자신을 복제(Clone)하여 작업 부하를 분산시킬 수 있으며 자신의 현재 실행 상태를 저장하여 네트워크를 통해 이동하며(Code mobility) 상태를 다시 복원함으로써 해당 작업에 대한 실행을 재개할 수 있어 과부하 된 호스트의 부하를 줄여 수행의 효율성을 높이는 기능을 제공할 수 있다[1]. 대표적 모바일 에이전트의 특성으로는 이동성, 협동성, 자동성, 적응성 등이 있다.

다음 그림에서 모바일 에이전트의 기본 동작원리를 나타냈다.



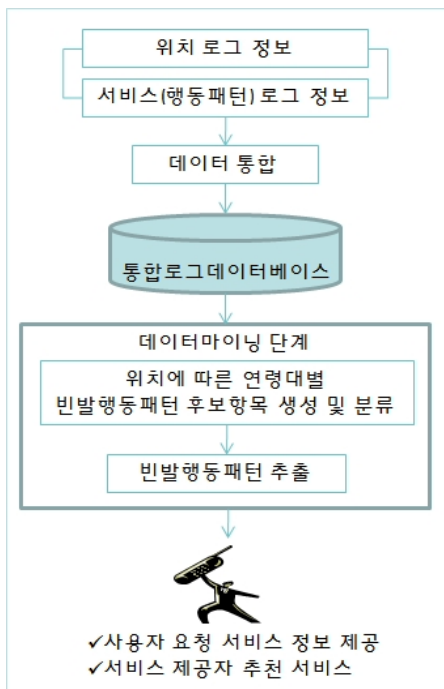
(그림 2) 모바일 에이전트

3. 시나리오

모바일 사용자는 연령대별, 성별 등으로 관심사가 다를 것이며 취미, 결혼유무 등의 부가적인 정보도 추가하여 분류하면 더 정확하게 선호되는 서비스를 제공할 수 있을 것이다. 초등학교생의 행동패턴과 대학생의 행동패턴은 다를 수 밖에 없다. 그러므로 위치에 따른 연령대별 행동패턴을 이용 시 모바일 사용자가 선호하는 서비스 정보를 적절하게 제공하여 잠재적 고객을 효과적으로 이끌 수 있다. 예를 들어, 유비쿼터스존인 위치 A 에 모바일 사용자가 인식됐을 시 사용자의 위치와 연령에 해당하는 추출된 빈발행동패턴을 이용하여 서비스 정보를 제공한다. 모바일 사용자가 20 대라면 통합로그데이터베이스의 데이터 마이닝 결과에서 A 위치의 20 대 행동패턴을 선택해서 정보를 제공하는 것이다. 이렇듯 해당 사용자 정보와 일치하는 유형의 서비스를 추천할 수 있게 된다.

4. 시스템 구성

본 논문에서 제안하는 모바일 에이전트 기반의 위치에 따른 연령대별 행동패턴을 이용한 마이닝 기법은 유비쿼터스 모바일 컴퓨팅 환경에서 상황인식 데이터를 대상으로 하여 효율적이며 확장성이 용이한 상황인식 처리를 가능하게 한다.



(그림 3) 시스템 흐름도

(그림 3)은 세부적인 데이터 마이닝 모듈 구조도를 나타내고 있다. 위치에 따른 연령대별 유용한 행동패턴 추출기법은 세 단계를 거친다. 첫째, 데이터 통합 단계로 위치에 따라 요구되었던 서비스 정보를 데이터베이스에 저장하는 단계이다. 이는 기존 위치정보에 따라 빈발하게 요구되었던 서비스 정보를 추출하는 것이다. 둘째, 데이터 마이닝 단계에서는 연관규칙을 사용하여 빈발행동패턴을 추출한다. 우선 서비스가 요청된 위치에 해당하는 트랜잭션을 추출한다. 최소지지도를 만족하는 트랜잭션만을 남긴다. 이러한 트랜잭션 데이터베이스에서 연령대별로 빈발행동패턴을 추출한다. 마지막으로 모바일 사용자가 해당 연령대의 빈발행동패턴 결과에 따라 서비스 정보를 제공한다.

5. 위치에 따른 연령대별 유용한 행동패턴 추출 기법

<표 1>은 샘플 통합 로그 데이터베이스로, 이를 기반으로 하여 위치에 따른 연령대별 유용한 행동패턴을 추출해볼 것이다.

<표 1> 통합 로그 데이터베이스

integrated log DB					
사용자ID	성별	나이	취미	위치정보	행동패턴
1	남	35	온라인 게임	C	PC방, 당구장
2	남	22	영화	A	백화점, 영화관, PC방
3	여	24	-	A	커피숍, 영화관, 레스토랑
4	여	43	미술	C	전시회, 커피숍
5	여	17	-	B	PC방, 레스토랑
6	여	27	-	C	커피숍, 영화관
7	남	33	드라이브	B	당구장, 백화점
8	남	42	-	A	전시회, 레스토랑, 골프장
9	남	25	-	A	커피숍, 식당
10	남	23	-	B	PC방, 레스토랑
11	남	19	온라인 게임	C	PC방, 레스토랑, 영화관
12	여	26	영화	A	백화점, 커피숍, 영화관, 레스토랑
13	남	36	-	A	당구장, 레스토랑, 백화점
14	여	48	-	A	전시회, 커피숍, 골프장
15	여	35	당구	A	당구장, 커피숍, 백화점
16	남	51	-	A	골프장, 레스토랑

<표 2>는 통합 로그 데이터베이스에서의 위치정보가 A 인 사용자의 트랜잭션을 추출한 것이다.

<표 2> A 위치 사용자 분석표

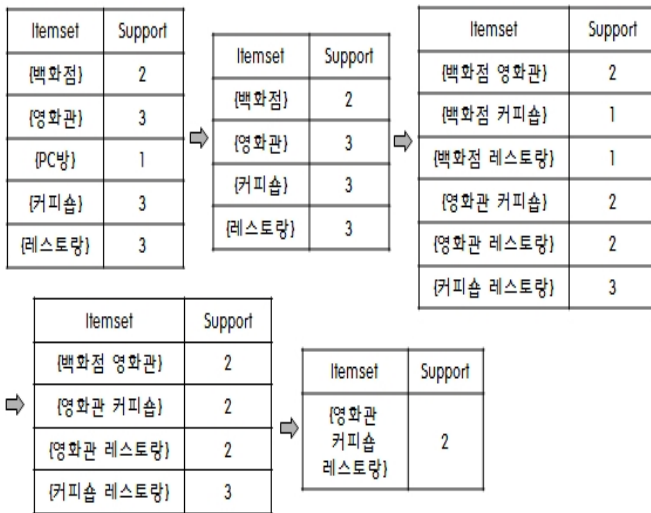
A 위치의 연령대	
UserID	연령
2	22
3	24
8	42
9	25
12	26
13	36
14	48
15	35
16	51

<표 2>에서 최소 지지도가 2 이상인 연령만 추출한다. 그러므로 ID 가 16 인 사용자의 연령은 51 세로 최소 지지도에 만족하지 않으므로 제거된다. 연령순으로 정렬한 <표 3>을 기반으로, 연령대별 빈발 행동패턴을 추출하게 된다.

<표 3> 연령대별 행동패턴

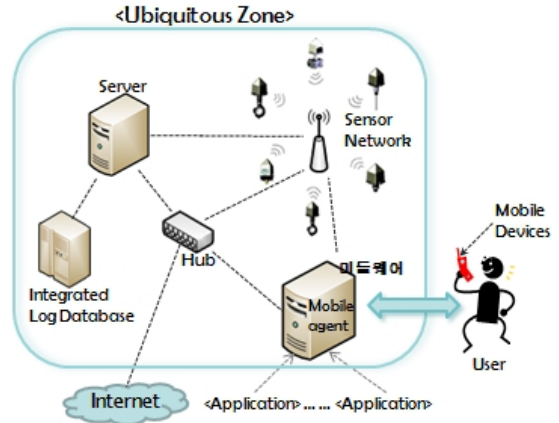
UserID	연령	행동패턴
2	22	백화점, 영화관, PC방
3	24	커피숍, 영화관, 레스토랑
9	25	커피숍, 레스토랑
12	26	백화점, 커피숍, 영화관, 레스토랑
15	35	당구장, 커피숍, 백화점
13	36	당구장, 레스토랑, 백화점
8	42	전시회, 레스토랑, 골프장
14	48	전시회, 커피숍, 골프장

20 대, 30 대, 40 대로 분류하여 행동패턴을 추출할 것이다.



(그림 4) 20 대의 빈발행동패턴 추출 과정

(그림 4)는 20 대의 빈발행동패턴 추출 과정을 나타낸 것이다. 최소 지지도는 2 이상으로, 추출된 빈발행동패턴은 영화관, 커피숍, 레스토랑이다. 같은 방법으로 행동패턴을 추출하여 30 대는 당구장, 백화점이 빈발하고 40 대는 전시회, 골프장이 빈발한 것으로 나왔다. 이 행동패턴을 이용하면, 20 대의 모바일 사용자가 해당 유비쿼터스존에 진입했을 시 서비스 제공자는 영화관, 커피숍, 레스토랑을 추천해줄 수 있다. 또한, 모바일 사용자가 영화관에서 티켓을 예매했을 시 서비스 제공자는 근처 커피숍이나 레스토랑을 추천할 수 있을 것이다. 이러한 상황을 아래 (그림 5)에서 표현하고 있다. 사용자 중심 서비스로 모바일 에이전트를 이용하여 사용자에게 정보를 제공한다. 추가로 영화관, 커피숍, 레스토랑과 같은 서비스 정보를 상점명으로 수집하면 더욱 정확한 정보를 제공할 수 있을 것이다.



(그림 5) 모바일 에이전트를 이용한 사용자 중심 서비스

6. 결론 및 향후 연구

이 논문에서 제안한 위치에 따른 연령대별 유용한 행동패턴 추출기법은 모바일 사용자가 선호할 만한 서비스 정보를 효율적으로 도출하도록 하였다. 사용자의 행동패턴을 성별, 나이, 직업, 취미 등의 개인정보와 위치정보, 서비스정보를 이용하여 행동패턴을 추출하면 각 사용자의 특성에 맞는 서비스를 제공할 수 있었다. 이 기법은 빠르고 정확하게 운영되며 의사결정 시스템에 효율적이고, 확장성이 높은 의사결정을 할 수 있도록 하였다.

7. 감사의 글

본 연구는 21 세기 프론티어 연구개발사업의 일환으로 추진되고 있는 정보통신부의 유비쿼터스컴퓨팅및네트워크원천기술개발사업의 지원에 의한 것임.

참고문헌

- [1] D. Chess, C. Harrison, A. Kershenbaum. "Mobile agents : Are they a good idea?, In Mobile Object Systems: Towards the Programmable Internet", Vol.1222 of Lecture Notes in Computer Science, Springer-Verlag, 1997.
- [2] J. Han, M. Kamber "Data Mining : Concepts and Techniques" Academic Press 2000.
- [3] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," In Proc. of the 20th International Conference on Very Large DataBases (VLDB94), pp. 487-499, Santiago, Chile, September 1994.
- [4] R. Agrawal, T. Imielinski, and A. Swami. "Mining association rules between sets of items in large databases," In Proc. of the ACM SIGMOD Conference on Management of Data, pp. 207-216, Washington, D.C., May 1993.
- [5] 박정규, 서승호, 김양남, 이궁해 "Personal Kiosk : 유비쿼터스 컴퓨팅 환경을 위한 모바일 서비스 모델", 정보과학회논문지, 제 12 권, 제 3 호, 2006 년 6 월.
- [6] 박종수, 유원경, 홍기형, "연관규칙탐사와 그 응용", 정보과학회지, 98 년 9 월.