

# 유비쿼터스 데이터 마이닝, 도전과 과제

## Ubiquitous Data Mining, Challenge and Task

전성해

청주대학교 통계학과

Sung-Hae Jun

Dept. of Statistics, Cheongju University

E-mail: shjun@cju.ac.kr

### 요 약

21세기에 들어서면서 인터넷은 새로운 패러다임인 유비쿼터스 컴퓨팅 환경으로 빠르게 바뀌고 있다. 특히 2005년에 접어들면서 유비쿼터스는 정보기술 분야에서 건설, 의료, 교통, 안전, 교육 등 사회 각 분야에서 유비쿼터스 컴퓨팅의 도입을 추진하고 있다. 동시에 유비쿼터스 컴퓨팅이 각 분야에서 적용이 될 때에는 지능형 시스템에 의한 서비스가 이루어 져야 한다는 것에 대하여 모두가 공감하고 있다. 지능형 유비쿼터스 서비스가 이루어지기 위한 하나의 방법으로서 현재 인터넷의 지능형 서비스에서 활발하게 이루어지고 있는 데이터 마이닝 전략이 있다. 즉 유비쿼터스 컴퓨팅 환경에서 발생하는 엄청난 양의 데이터를 분석하여 지능형 유비쿼터스 서비스를 하기 위한 데이터 마이닝 분야가 바로 유비쿼터스 데이터 마이닝이다. 유비쿼터스 데이터 마이닝은 오프라인 데이터 마이닝, 웹 마이닝 등에 비해 여러 가지 다른 점들이 있다. 본 논문에서는 유비쿼터스 데이터 마이닝에 대한 소개와 기존의 데이터 마이닝 프로세스와의 차이점을 알아본다. 아울러 유비쿼터스 컴퓨팅 환경에서 이루어져야 할 데이터 마이닝 전략의 과제와 도전에 대한 이슈들을 살펴보고 몇 가지 모의실험을 통하여 이것들에 대한 확인을 하였다.

### 1. 서론

컴퓨팅 산업의 발전단계를 살펴보면 메인 프레임(mainframe) 기반의 초기 단순 데이터 저장 및 처리(simple data storage and processing)에서 시작하여 최근에는 개인용 컴퓨터 및 워크스테이션 서버 등을 이용한 웹 기반 대화형 컴퓨팅(web-based interactive computing)에서 이제는 유비쿼터스 컴퓨팅(ubiquitous computing) 장비들에 의한 지능형 컴퓨팅(intelligent computing) 서비스 단계로 진입하고 있다[1][2].

제록스(Xerox)사의 와이저(Weiser)는 유비쿼터스 컴퓨팅을 '실세계에 존재하는 모든 개체들을 기능적, 공간적으로 연결하여 사용자에게 필요한 정보나 서비스를 실시간으로 제공할 수 있는 기반 기술'로 정의하였다[15][16]. 유비쿼터스 컴퓨팅의 정의에서 모든 개체에는 물론 인간도 포함

된다. 인간은 자신이 가질 수 있는 컴퓨터(wearable computer)를 이용하여 언제든지 유비쿼터스 환경에 일원이 될 수 있다. 물론 생활 가전 기기의 작은 컴퓨터부터 고성능 워크스테이션까지 모든 컴퓨터들도 이에 해당 된다. 미래에는 크고 작은 다양한 컴퓨터들이 현실세계에서 무선 통신망으로 연결되어 사용자가 필요로 하는 정보를 즉시 제공할 수 있는 환경이 완벽하게 구현될 것이다. 현재는 개인 휴대단말기(PDA, personal digital assistant), 위치 측정시스템(GPS, global positioning system) 등과 같은 모바일 컴퓨터로서 유비쿼터스 환경에 대한 출발이 이루어지고 있다. 이를 위해서는 사용자가 시간적으로나 공간적으로 어떠한 거부감이나 불편함이 없이 편리하게 정보 서비스를 제공 받을 수 있는 기술 환경이 이루어져야 한다[3][4][5][15].

## 2. 유비쿼터스 데이터 마이닝

유비쿼터스 환경에서 인간을 포함하여 네트워크에 연결된 클라이언트들의 정보 공유 및 서비스를 위한 자동화된 지능 시스템을 위한 데이터 마이닝은 접속고객의 소비성향, 사기유형 등의 비즈니스 지능화뿐만 아니라, 서버와의 네트워크 단절 기간에도 GPS와 같은 모바일 클라이언트가 마치 서버에 연결되어 있는 것처럼 정보 서비스를 받을 수 있도록 네트워크 단절 전에 미리 클라이언트 캐시에 필요한 정보를 방사(broadcasting)할 수 있는 규칙의 구축 등과 같은 것들은 모두 포함한다. 본 논문에서는 눈부시게 발전하고 있는 유비쿼터스 컴퓨팅 환경에서 데이터 마이닝의 역할 중에서 이동 컴퓨터를 위한 효과적인 캐시 적재에 대하여 알아보았다. 최근의 컴퓨터 하드웨어 기술의 발달에 의해 노트북, PDA 등 사용자가 이동하면서 사용할 수 있는 작은 컴퓨터의 생산이 가능해 졌다. 이러한 휴대용 컴퓨터들은 어느 위치에서도 전역적 데이터 서비스에 접속이 가능한 무선 통신 장치를 갖추고 있다. 하지만 네트워크 연결에 대한 비용문제 등 여러 가지 문제점 때문에 계속적으로 네트워크와 연결되지 못하고 단절되는 경우가 종종 발생하게 된다. 하지만 이때에도 모바일 클라이언트는 지속적인 서비스 요구가 발생하게 되고 이러한 것들을 클라이언트 캐시가 도와주어야 한다. 이 때 해당 사용자에게 필요한 아이템들이 해당 모바일 컴퓨터의 캐시에 적재되어 있어야 한다. 또한 모바일 컴퓨터가 정적이거나 이동 가능한 모바일 서버의 많은 공유 데이터베이스에 효율적으로 접속할 수 있게 하는 모바일 데이터베이스 시스템에 대한 연구가 최근까지 대단히 많이 이루어지고 있다. 특히 모바일 환경에서의 클라이언트-서버 컴퓨팅에 대한 연구도 이루어 졌다[6][7][8][10]. 사용자는 이동 중에 인터넷 서핑이나 이 메일 확인을 할 수 있다. 그러나 무선 통신을 하는 모바일 컴퓨터는 무선 통신의 비용과 무선 네트워크를 사용할 수 없는 경우의 발생 때문에 종종 네트워크와 단절된다. 가령 사용자가 모바일 컴퓨터를 가지고 비행기를 타게 되는 경우처럼 말이다. 단절 될 때의 클라이언트 캐시에서 필요한 파일을 적재하는 방법에 대한 연구와 사용자의 간섭 없이 자동으로 클라이언트의 캐시에 데이터를 적재해 주는 방법에 대한 연구가 현재 수행되고 있다[12][13]. 개인의 모바일 컴퓨터가 네트워크와 단절 기간 동안에도 사용자들은 단절 이전과 똑같은 정보 서비스를 받고 싶

어 한다. 어떤 경우에는 사용자는 현재 자신의 모바일 컴퓨터가 네트워크와 단절된 상황인지 아닌지 모를 정도로 정보 서비스를 계속적으로 받았으면 한다. 이러한 지속적인 정보 서비스가 가능하기 위해서는 각자의 모바일 컴퓨터의 캐시에 적재된 정보에 의해 수행되어지는 단절 중의 작업을 위한 효과적인 캐시 적재 모형의 필요성이다. 캐시 적재 프로세스는 미래에 단절되기에 앞서 클라이언트 캐시에서 필요로 하는 데이터 아이템들의 적재하는 것이다. 그러나 사용자가 미래에 어떤 파일을 필요로 하게 될지를 서버가 아는 것은 매우 어렵다. 또한 이러한 작업은 자동화 되어져야 한다. 자동화 적재는 사용자의 어떠한 간섭이 없이도 적재 집합을 예측하여 만들어 내는 프로세스이다. 효과적인 캐시 적재를 위하여 데이터 마이닝 전략을 사용할 수 있다. 데이터 마이닝 기법을 이용하면 대량의 트랜잭션 데이터로부터 숨겨진 패턴을 찾아내어 새로운 현상이나 미래의 행위를 예측하는 모델링 작업을 수행할 수 있다. 모바일 컴퓨터가 네트워크와 단절이 발생했을 때 이 클라이언트의 미래의 요청에 대한 아이템 예측을 위한 규칙을 생성하기 위하여 본 논문에서는 모바일 클라이언트의 서비스 요청(request) 이력(history) 데이터에 협업 추천(collaborative filtering) 모형의 적용을 제안하였다. 협업 추천은 기본적으로 클라이언트들이 아이템의 요구에 대한 정보를 기반으로 하여 특정 클라이언트가 특정 아이템에 대한 유용성(utility) 혹은 선호도(preference)를 예측해 내는 것을 목적으로 한다. 본 논문에서 제안 적용하는 협업 추천은 위 그림과 같은 구조를 지니고 있다. (그림 2)는 모바일 클라이언트의 캐시 적재를 위한 협업 추천 시스템의 구조를 나타내고 있다. 즉 서로 유사한 모바일 클라이언트의 아이템 요구 이력 정보를 이용하여, 미래의 네트워크 단절에 대비한 캐시 적재를 수행하게 된다. 특히 본 논문에서는 클라이언트의 요구 이력 정보뿐만 아니라 모바일 클라이언트의 위치 정보까지 고려하였다. 기존의 연구에서는 모바일 클라이언트의 요청 아이템에 대한 예측 모형을 위한 학습 데이터로서 현재까지의 클라이언트의 요구 이력 정보를 사용하였다[14].

## 3. 실제 적용 예

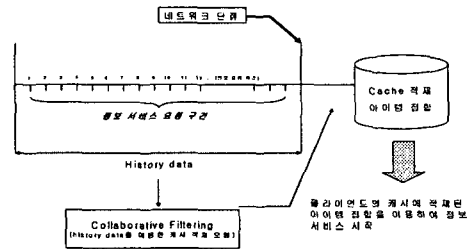
유비쿼터스 컴퓨팅 환경에서 데이터 마이닝의 적용 예를 살펴보기 위하여 본 논문에서는 협업 추천 모형을 이용한 모바일 클라이언트의 효과적인 캐시 적재 모형을 구축하였고, 실험을 통하여

기존 모형들의 성능비교를 수행하였다. 기본적인 유비쿼터스 컴퓨팅 환경인 현재의 무선 네트워크에서는 지역적 통신 두절, 시스템적 대역폭의 축소 등 여러 가지 이유로 잦은 네트워크 단절이 발생한다. 이로 인한 이동 컴퓨팅의 문제점들을 해결하기 위한 하나의 전략으로서 네트워크 단절을 대비하여 사전에 네트워크에 연결된 모바일 클라이언트의 캐시에 향후에 필요로 할 정보를 미리 적재한다. 현재 주로 사용되고 있는 적재 방법은 각 클라이언트 컴퓨터가 가장 최근에 서버에게 요청했던 정보를 서버로부터 클라이언트에 방사(broadcasting)하는 LRU(least recently used)방식이다. 하지만 이러한 캐시 적재 전략은 실시간으로 변화하는 네트워크 환경과 사용자의 정보 요구에 대한 변화를 잘 반영하지 못하여 결국 네트워크 단절 상황에서 사용자에게 꼭 필요한 정보를 효과적으로 제공해 주지 못하게 된다. 본 논문에서는 이러한 상황에서 정보 아이템들 간의 상호 의존성에 기반한 협업추천 모형을 적용하여 기존의 방법에 비해 효과적인 정보 서비스가 가능하도록 하였다. 특정 모바일 클라이언트에 대한 필요 정보만 찾아내어 해당 클라이언트 컴퓨터에 적재하면 이 모바일 컴퓨터를 사용하는 사용자에게는 만족할 만한 정보 서비스를 기대할 수 있게 된다. 아울러 저장 공간의 제약을 갖는 모바일 컴퓨터의 한계도 극복할 수 있게 된다. 본 논문의 적용 실험에서는 CSIM Simulator를 이용하여 이동 컴퓨터의 과거 정보 요구에 대한 이력 데이터를 모의 생성하여 사용하였다[17]. Cache hit ratio를 이용한 객관적인 성능 평가를 통하여 제안 기법을 적용한 캐시 적재 성능이 기존의 것에 비해 우수함을 보였다[14]. 비교 측도는 사용자가 필요로 하는 정보가 모형에 의해 만들어진 적재 데이터 집합에 얼마나 많이 있는지를 측정하는 것이다. 실험을 위하여 우선 모바일 클라이언트가 필요로 하는 정보 아이템의 전체 집합은 다음 식과 같이 정의하였다.  $I_N$ 은 전체  $N$ 개의 원소로 이루어진 아이템 집합이다.

$$I_N = \{i_1, i_2, \dots, i_N\}, \quad i_j \in I_N \quad (4.1)$$

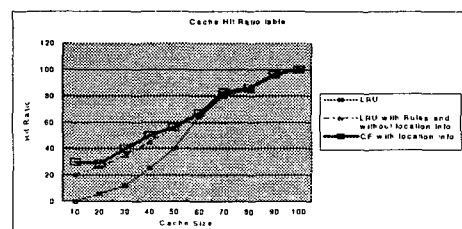
위 식에서  $i_j$ 는  $j$ 번째에 해당하는 아이템을 나타낸다. 본 실험에서는 편의상 전체 아이템 집합의 크기  $N$ 을 9로 하였다. (그림 1)은 이력 정보 데이터(history data)의 구조로서 캐시 적재 아이템들의 이력 데이터의 구조와 캐시 정보 서비스의 과정을 나타내고 있다. 이력 데이터는 모바일

클라이언트가 네트워크에 연결되어 있는 동안 서버에게 요청했던 정보 아이템들을 가지고 있다. 이것을 이용하여 협업 추천 모형을 통한 모바일 클라이언트의 캐시 적재 데이터 집합(hoarding data set)을 결정할 수 있게 된다.



(그림 1) 이력 정보의 데이터 구조

클라이언트를 구축하는 일정 시점 이전까지의 클라이언트의 요청(request) 아이템들의 이력을 이용하여 미래 시점에서 모바일 클라이언트의 캐시에 적재할 아이템들을 결정하게 된다. 이 때 클라이언트 요청 로그(client request log) 데이터는 연결 기간과 단절 기간으로 나뉘어져 있는 구조가 된다. 추가적으로 본 논문에서는 모바일 컴퓨터의 이동성(mobility) 정보를 클라이언트 캐시의 적재 집합 결정에 반영하기 위하여 모바일 컴퓨터의 위치 정보(각도, 방향, 속도 등)와 모바일 컴퓨터가 현재 위치하는 주변 환경의 정보 등을 고려하여 제안 기법을 이용하여 모바일 컴퓨터의 캐시에 적재하는 아이템들을 예측하여 서버가 방사(broadcasting)할 수 있도록 하였다. 앞에서 언급한 바와 같이 성능 평가는 캐시 적중률을 이용하였다. 즉 단절 기간 동안 클라이언트가 요구가 얼마나 자동화 적재(automatic hoarding) 기법에 의한 캐시 적재 아이템 집합에 있는지를 비교한다. 본 논문의 실험을 위한 시뮬레이션 데이터는 XMM-Newton SOC의 CSIM Simulator 소프트웨어를 이용하여 생성하였다[17]. 구체적인 협업 추천을 위한 방법으로는 통계적 상관 계수와 코사인 유사도(cosine similarity)를 이용하였다[9][11]. 다음 그림은 비교 모형들 간의 적중률의 평균값을 시각적으로 표현하였다.



(그림 2) 3개의 모형간의 Cache hit ratio의 비교 결과

(그림 2)는 모바일 클라이언트의 캐시 저장 공간의 크기에 따른 필요 아이템들의 적중률을 보여주고 있다. 가로 축의 캐시 저장 공간의 크기가 100이 되면 필요한 모든 아이템들을 모두 저장할 수 있는 공간이 되기 때문에 모형간의 비교에 대한 의미가 없게 된다. 하지만 10, 20, 그리고 30과 같이 개시의 저장 공간이 제한되어 있는 경우에는 각 모형에 의해 결정되어진 서로 다른 아이템들이 저장되고 이들로부터 적중률을 계산하면 서로 다르게 된다. (그림 2)에 의하면 제안 방법이 기존 방법들에 비해 보다 큰 캐시 적중률 값을 갖게 됨을 알 수 있었다.

#### 4. 결론 및 향후과제

본 논문에서는 유비쿼터스 컴퓨팅 환경에서 다양한 u-서비스를 위한 데이터 마이닝 전략에 대한 소개를 하였다. 인터넷 이후 새로운 패러다임인 유비쿼터스 컴퓨팅 환경에서 네트워크에 연결된 모든 컴퓨터들을 효과적으로 관리하고 발생하는 데이터를 분석하여 최적의 서비스 제공을 위한 지능화된 유비쿼터스 컴퓨팅 환경을 구축하기 위하여 데이터 마이닝은 중요한 역할을 할 것이며 향후 이 부분에 대한 지속적인 연구 및 개발이 이루어져야 할 것이다.

#### 참고문헌

[1] 백성욱, "유비쿼터스 컴퓨팅 환경을 위한 지능형 미디어 기술", 정보과학회지, 21권 5호, pp. 36-42, (2003).

[2] 유준재, "유비쿼터스 컴퓨팅 플랫폼", 충북 IT 신기술연구회 워크샵 자료집, pp. 9-73, (2004).

[3] 장세이, 우운택, "유비쿼터스 컴퓨팅 환경을 위한 센싱 기술과 컨택스트-인식 기술의 연구 동향" 정보과학회지, 21권 5호, pp. 18-28, (2003).

[4] 전성해, 류제복, 이승주, "Data Mining Approach to Supporting Hoarding in Mobile Computing Environments", 한국통계학회 2003 춘계 학술대회 논문집, pp. 13-18, (2003).

[5] 전성해, 전홍석, "Ubiquitous Computing and Statistics; What's the Connection?", The Korean Communications in Statistics Vol. 11 No. 2, pp. 287-295, (2004)

[6] R. Alonso, H. F. Korth, "Database system issues in nomadic computing", Proceedings of ACM SIGMOD International Conference on the Management of Data, (1993).

[7] P. K. Chrysanthis, "Transaction processing in mobile computing environment", Proceedings of the IEEE Workshop on Advances in Parallel and Distributed Systems, (1993).

[8] M. H. Dunham, A. S. Helal, "Mobile computing and databases: Anything new?", SIGMOD Record, Vol. 24, No. 4, pp. 5-9, (1995).

[9] M. Gaber, S. Krishnaswamy, S. A. Zaslavsky, "Ubiquitous Data Stream Mining, Current Research and Future Directions", Proceedings of The Eighth Pacific-Asia Conference on Knowledge Discovery and Data Mining, (2004).

[10] R. Hunter, "World Without Secrets: Business, Crime, and Privacy in The Age of Ubiquitous Computing", John Wiley and Sons, (2002).

[11] J. Jing, A. Helal, A. K. Elmagarmid, "Client server computing in mobile environments", ACM Computing Surveys, (1999)

[12] J. J. Kistler, M. Satyanarayanan, "Disconnected operation in the coda file system", ACM Transactions on Computer Systems, Vol. 10, No. 1, pp. 3-25, (1992).

[13] G. Kuenning, G. Popek, "Automated hoarding for mobile computers", Proceedings of the ACM Symposium on Operating Systems Principles, (1997).

[14] Y. Saygin, O. Ulusoy, A. Elmagarmid, "Association Rules for Supporting Hoarding in Mobile Computing Environments", IEEE 10th International Workshop on Research Issues on Data Engineering, (2000).

[15] M. Weiser, "The computer for 21st Century", Scientific American, Vol. 265, No. 3, pp. 94-104, (1991).

[16] <http://www.ubiq.com/weiser>.

[17] <http://www.atl.lmco.com/proj/csim>.