

Data Mining Approach for Supporting Hoarding in Mobile Computing Environments

전성해¹⁾ 류재복²⁾ 이승주³⁾

요 약

본 논문에서는 낮은 대역폭, 높은 지연, 그리고 잦은 네트워크 단절로 인한 모바일 컴퓨팅 환경의 문제점들을 해결하기 위한 효과적인 캐시 적재 기법으로서 협업 추천 기반의 데이터 마이닝 전략을 제안하였다. 캐시 적재가 모바일 클라이언트의 이러한 문제점들을 해결하기 위한 효율적인 방법이 된다는 기존의 연구는 많이 진행되어 왔다. 하지만 모바일 컴퓨터의 요구에 대한 이력 정보만을 이용한 기존의 연구는 모바일 클라이언트가 필요로 하는 모든 정보 요구를 만족하지 못하였다. 특히 저장 공간의 제약을 갖는 모바일 컴퓨터의 한계 때문에 더욱 큰 어려움을 갖게 되었다. 본 연구에서는 모바일 클라이언트의 이력 정보에 대하여 데이터 마이닝 기법을 적용한 캐시 적재 기법을 제안하여 적은 캐시 용량만으로도 모바일 클라이언트의 요구를 만족할 수 있는 아이템들을 효과적으로 서비스할 수 있도록 하였다. CSIM Simulator를 이용하여 모의 데이터를 생성하여, 제안 모형의 성능 평가를 위한 실험을 수행하였다. Cache hit ratio를 이용한 객관적인 성능 평가를 통하여 제안된 모형이 모바일 클라이언트의 캐시 적재 기법으로서 우수한 성능을 보임이 확인되었다.

주요용어 : Data Mining, Supporting Hoarding, Mobile Computing.

1. 서 론

무선 통신(wireless communication) 장치(device)를 갖추고 있는 모바일 컴퓨터의 사용은 계속적으로 증가하고 있다. 모바일 컴퓨터의 특징 중의 하나는 네트워크와의 단절 기간 동안 캐시에 적재(hoarding)된 정보에 의해 수행되어지는 단절 중의 작업이다. 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. 이 논문에서는 모바일 클라이언트가 네트워크로부터 단절되기 전에 적재되어져야 할 아이템들이 무엇인지를 결정하기 위하여 통계적 학습 모형을 제안하였다. 네트워크와 단절되기 전에 모바일 컴퓨터에 적재되어야 하는 아이템들의 집합을 결정하기 위하여 데이터 마이닝 전략 중에서, 특히 협업 추천(collaborative filtering) 기법을 이용하여 자동 적재 규칙을 만들어 내었다. 기존의 제안된 규칙 기반(rule-based) 방법[7]과 제안하는 방법을 특별히 작은 캐시 용량을 위한 모바일 클라이언트의 캐시 hit 비율 측도를 이용하여 시스템의 성능 향상을 비교하였다. 즉 클라이언트의 아이템 요청(request) 패턴에 대한 이력(history)을 통해 미래에 단절될 때에 클라이언트가 요청하게 될 아이템을 예측하여 미리 적재

-
- 1) 청주대학교 통계학과, 전임강사 (360-764) 충청북도 청주시 상당구 내덕동 36번지
 - 2) 청주대학교 통계학과, 교수 (360-764) 충청북도 청주시 상당구 내덕동 36번지
 - 3) 청주대학교 통계학과, 부교수 (360-764) 충청북도 청주시 상당구 내덕동 36번지

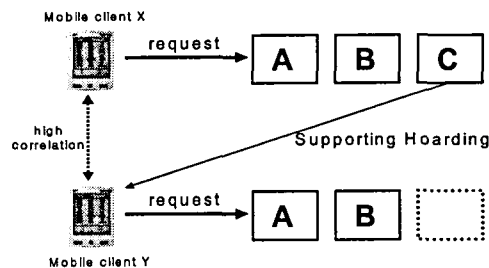
될 수 있도록 하는 프로세스에 대한 성능 비교이다.

2. 관련 연구

최근의 컴퓨터 하드웨어 기술의 발달에 의해 노트북, palmtop 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 계속적인 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 이루어졌다[1][2][3][4]. 사용자는 이동 중에 인터넷 서핑이나 이 메일 확인을 할 수 있다. 그러나 무선통신을 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. 단절될 때의 클라이언트 캐시에서 필요한 파일을 적재하는 방법에 대한 연구[5]와 사용자의 간섭 없이 자동으로 클라이언트의 캐시에 데이터를 적재해 주는 방법에 대한 연구[6]가 현재 수행되고 있다.

3. 데이터 마이닝 알고리즘을 이용한 캐시 적재 기법

데이터 마이닝 알고리즘을 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 다음의 [그림 1]과 같은 구조를 지니고 있다.



[그림 1] Supporting Hoarding을 위한 협업 추천 구조

[그림 1]은 모바일 클라이언트의 캐시 적재를 위한 협동 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 정보

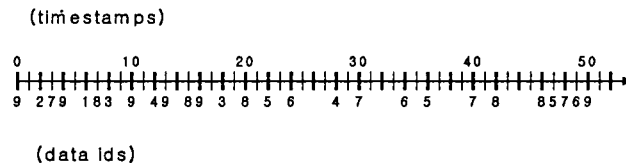
뿐만 아니라 모바일 클라이언트의 위치 정보까지 고려하였다. 기존의 연구에서는 모바일 클라이언트의 요청 아이템에 대한 예측 모형을 위한 학습 데이터로서 현재까지의 클라이언트의 요구 이력 정보만을 사용하였다[7]. 본 논문에서는 각 클라이언트의 이력 정보뿐만 아니라 클라이언트의 현재의 위치 정보까지를 결합한 학습 데이터를 통하여 미래에 단절되는 시점에서의 모바일 컴퓨터가 필요로 하는 아이템을 예측하는 모형을 만들었다. 각 클라이언트의 현재의 위치 정보는 GPS 센서 등을 이용하여 얻을 수 있다. 본 논문에서는 모바일 클라이언트가 위치하게 되는 전체 네트워크를 몇 개의 셀로 나누고 해당 셀의 특성 정보를 갖는 아이템들을 결정하였다. 또한 각 클라이언트는 현재 셀에서 다음 셀로 이동하는 방향을 8개의 각도로 이동할 수 있도록 가정하였다. 즉 전체 360도에서 각 45도 씩의 각도를 분할하여 클라이언트의 이동 방향을 결정할 수 있도록 하였다.

4. 실험 및 결과

본 논문의 실험을 위하여 두 가지 사전 작업을 수행했다. 첫 번째로는 실험을 위한 시뮬레이션 데이터의 생성을 위한 사전 가정이다. 우선 클라이언트의 필요로 하는 아이템의 전체 집합은 다음 식과 같이 정의하였다.

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

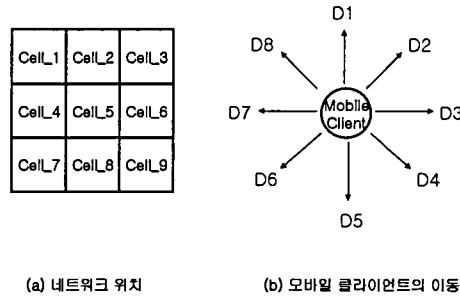
본 논문에서는 N의 값을 9로 하였다. 따라서 제안 알고리즘의 성능 평가를 위한 아이템 집합을 9개로 결정하였다. 그리고 각 아이템들은 자기 고유한 데이터 생성 분포를 갖는다. [그림 2]는 이러한 이력 정보 데이터의 구조이다.



[그림 2] 9개의 Supporting Hoarding 아이템들의 이력 데이터 구조

위 그림은 9개의 캐시 적재 아이템들의 이력 데이터의 구조이다. 시간에 따라서 모바일 컴퓨터가 필요로 했던 데이터 아이템들이다. 이러한 데이터를 이용한 협업 추천 모형을 통하여 모바일 클라이언트의 hoarding set을 결정할 수 있게 된다. 모형을 구축하는 일정 시점이전까지의 클라이언트의 request 아이템들의 history(client request log)를 이용하여 미래 시점에서 모바일 클라이언트의 캐시에 hoarding 할 아이템들을 결정하게 된다. 이 때 client request log 데이터는 연결 기간과 단절 기간(disconnected and connected periods)으로 나뉘어져 있는 구조이다. 또한 본 논문에서는 모바일 컴퓨터의 이동성(mobility) 정보를 클라이언트 캐시의 hoarding set 결정에 반영하기 위하여 모바일 컴퓨터의 위치 정보(각도, 방향, 속도 등)와 모바일 컴퓨터가

현재 위치하고 있는 주변환경의 정보 등을 고려하여 논문에서 제안하는 기법을 이용하여 모바일 컴퓨터의 캐시에 적재하는 아이템들을 예측하여 서버가 broadcasting 할 수 있도록 한다. 그림 3은 클라이언트의 위치 정보에 대한 개념이다.



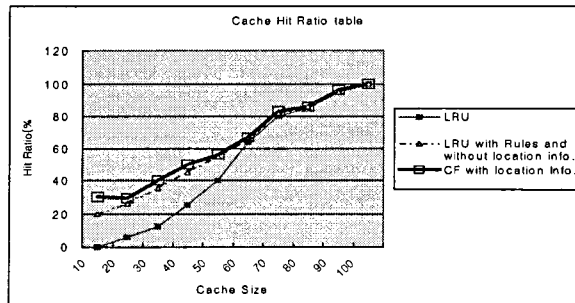
[그림 3] 모바일 클라이언트의 위치 정보의 구조

[그림 3]의 (a)는 이 실험에서의 전체 네트워크를 9개의 셀로 나누었다. 각 셀을 대표하는 아이템은 다음의 표 1과 같이 결정하였다.

표 1. 각 셀과 해당되는 대표 아이템

cell	Cell_1	Cell_2	Cell_3	Cell_4	Cell_5	Cell_6	Cell_7	Cell_8	Cell_9
item	1	2	3	4	5	6	7	8	9

표 1에 의해 만약 특정 클라이언트가 Cell_3에 위치하게 되면 아이템 3이 추가로 더 캐시에 적재되어야 한다. 그리고 [그림 3]의 (b)는 특정 모바일 클라이언트가 다음 셀로 이동하게 되는 방향을 D1부터 45도 간격으로 D8까지 8방향으로 이동할 수 있도록 결정하였다. 이들 정보가 [그림 1]의 이력 정보에 의한 협업 추천 모형의 결과와 결합하여 단절되기 전에 모바일 컴퓨터의 캐시에 필요한 아이템들을 적재할 수 있게 한다. 이 장의 실험에서의 성능 평가 방안은 Cache Hit Ratio를 이용한다[7]. 이 방법은 단절 기간 동안 클라이언트가 요구가 얼마나 자동화 적재(automatic hoarding) 기법에 의한 캐시 적재 아이템에 있는지를 비교하게 된다. 본 논문의 실험을 위한 시뮬레이션 데이터는 XMM-Newton SOC의 CSIM Simulator 소프트웨어를 이용하여 생성하였다[8]. 전체 생성 데이터의 80%는 학습(training)에 이용하고 나머지 20%는 모형의 타당성을 검증하는 테스트(testing) 데이터로 이용하였다.



[그림 4] 3개의 모형간의 Cache hit ratio의 비교 결과

[그림 4]에 의하면 제안하는 방법이 다른 방법에 비해 작은 캐시 용량에서 보다 큰 Cache hit ratio 값을 갖게 됨을 알 수 있었다. 따라서 모바일 클라이언트의 요구 이력 정보만을 이용하는 것 보다 위치정보를 이용하고 또한 본 논문에서 제안하는 협업추천 모형을 이용한 전략이 더 좋은 성능을 나타내고 있음을 알 수 있었다.

5. 결 론

본 논문에서 제안하는 모바일 클라이언트의 캐시 전략은 크게 두 가지이다. 첫째는 데이터 마이닝 기법인 협업 추천 모형이다. 이 모형은 다수의 데이터 마이닝 모형 중에서 서로 다른 사용자의 과거 이력 정보를 이용한 추천 모형들 중에서 성능이 우수한 결과를 제공하여 주는 장점을 갖고 있다. 두 번째는 학습 데이터에서 고려되는 요소를 기존의 클라이언트의 요구 이력 정보만을 이용하지 않고 클라이언트의 위치 정보까지를 이용한 것이다. 이러한 전략을 통하여 적은 캐시 용량만으로도 미래의 네트워크 단절에서 클라이언트가 필요로 하는 아이템을 더 많이 제공할 수 있었다. 따라서 현재와 같이 제한된 자원만을 이용해야 하는 모바일 컴퓨팅 환경에서 제안 방법은 우수한 정보 서비스 결과를 제공할 수 있게 될 것이다. 향후 연구과제로는 모바일 컴퓨터의 이동 속도 등과 같은 더욱 정교한 위치 정보의 고려와 클라이언트의 프로파일 등을 이용하고 베이지안 학습을 포함한 통계적 학습 이론을 사용하게 되면 더욱 정확한 캐시 적재 예측 모형을 구축할 수 있게 될 것이다.

참 고 문 헌

- [1] R. Alonso and H. F. Korth (1993), Database system issues in nomadic computing, In Proceedings of the ACM SIGMOD International Conference on the Management of Data.
- [2] P. K. Chrysanthis (1993), Transaction processing in mobile computing environment, In Proceedings of the IEEE Workshop on Advances in Parallel and Distributed Systems.
- [3] M. H. Dunham and A. S. Helal (1995), Mobile computing and databases: Anything new?, SIGMOD Record, 24(4):5-9.
- [4] J. Jing, A. Helal and A. K. Elmagarmid (1999), Client server computing in mobile environments, ACM Computing Surveys.
- [5] J. J. Kistler and M. Satyanarayanan (1992), Disconnected operation in the coda file system, ACM Transactions on Computer Systems, 10(1):3-25.
- [6] G. Kuenning and G. Popek (1997), Automated hoarding for mobile computers, In Proceedings of the ACM Symposium on Operating Systems Principles, St. Malo, France.
- [7] Y. Saygin, O. Ulusoy and A. Elmagarmid (2000), Association Rules for Supporting Hoarding in Mobile Computing Environments, IEEE 10th International Workshop on Research Issues on Data Engineering (RIDE'2000).
- [8] <http://xmm.vilspa.esa.es/>