

데이터 마이닝 기술 동향

손재기* 박상현** 민수영***

◆ 목 차 ◆

- | | |
|---------------|-------------|
| 1. 서론 | 4. 최근 기술 동향 |
| 2. 데이터 마이닝 | 5. 결론 |
| 3. 데이터 마이닝 기법 | |

1. 서론

인터넷 및 네트워크 환경의 발전, 그리고 컴퓨팅 환경의 발전으로 인해, 엄청난 양의 정보가 홍수처럼 쏟아지고 있다. 또한 정보 기술의 발달로 인간의 데이터 수집 능력은 무한대에 가까워지고 있다. 수많은 데이터들을 저장하고, 필요한 데이터를 찾으려는 노력과 더불어 그들 데이터로부터 가치 있는 정보나 지식을 추출하고자 하는 욕구가 증대되고 있다. 단순한 형태의 자료나 지식보다는 인간이 이해하고 분석할 수 있는 형태의 정보가 더 중요하게 여겨지고 있다.

그러나 이러한 정보들은 대개 방대한 데이터 덩어리에 파묻혀 있어 지금까지 사용하던 DBMS 혹은 데이터베이스 기법으로는 좀처럼 발견하기 어려워지고 있다. 따라서 이러한 요구에 대한 솔루션으로 데이터로부터 무엇인가를 발굴한다는 뜻에서 '데이터 마이닝' 개념이 발생했다.

데이터 마이닝은 이전에는 발견되지 않았던 데이터들간의 상호관계를 분석하는 것이다. 예를 들어, 만약 특정 상품의 상품 판매기록이 충분히 분석되고 다른 시장 데이터와 관련 지어졌다면, 같은 회사의 다른 상품 구입과의 계절적인 상호관계를 밝힐 수 있다.

여러 서적이나 논문에 의하면 데이터 마이닝과 지

식 발견(KDD, Knowledge Discovery in Database)이라는 용어를 혼용해서 사용하는 경우가 많다. 개념이 소개되던 초창기에 데이터 마이닝이라는 용어는 특히 통계학자, 데이터베이스 연구자, 그리고 기업체에서 많이 사용한 반면, 지식 발견의 경우는 인공 지능이나 전문가 시스템 관련 연구에 주로 등장했다.

그러나 1995년 캐나다 몬트리올에서 개최된 지식 발견과 데이터 마이닝에 관한 국제 학술대회에서 '지식 발견은 데이터로부터 유용한 정보를 발견하는 프로세스의 전 과정'이라고 정의했고, 데이터 마이닝은 지식 발견 프로세스 중에서 데이터로부터 정보를 추출하기 위해서 기법을 적용하는 특정 단계라고 제안하였다.

본 고는 2장에서 데이터 마이닝에 대한 개요를, 3장에서는 데이터 마이닝 기법에 대해 기술한다. 4장에서는 데이터 마이닝의 최근 기술 동향에 대해 기술한다.

2. 데이터 마이닝

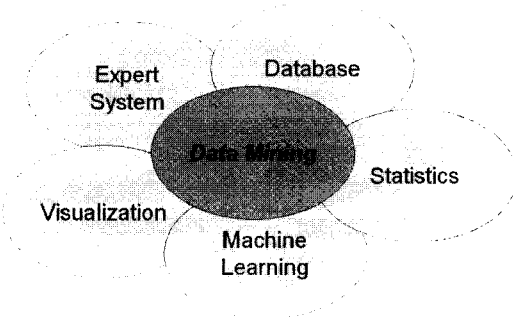
데이터 마이닝이란 대량의 데이터 사이에 묻혀있는 패턴을 발견하고 규칙을 추론함으로써, 의사결정을 지원하고 그 효과를 예측하기 위한 기법이다. 또한 데이터베이스 내의 데이터들간의 패턴을 발견하고 규칙을 추론한다. 이러한 패턴과 규칙을 의사결정을 지원하고 기업환경의 변화를 예측하는 데 사용될 수 있다.

직면하는 현상을 기술하기 위해서는 모델이 필요하다. 전통적으로 모델의 구성은 현상을 관찰하여 가설을

* 전자부품연구원 정보시스템연구센터 전임연구원

** 전자부품연구원 정보시스템연구센터 연구원

*** 전자부품연구원 정보시스템연구센터 책임연구원



(그림 1) 데이터 마이닝 기술

설정하고, 예측치와 실제결과를 비교함으로써 가설을 검증하였다. 모델이 부적절한 경우, 즉 예측치와 관찰치가 다른 경우 모델은 수정되거나 재구축되었다. 이러한 모델을 구축하기 위해서는 모델링 하고자 하는 현상을 관찰할 필요가 있다. 많은 기업들은 데이터베이스를 가지고 있으며 데이터베이스 내에 모든 환경이 다 포함되어 있지는 않지만, 데이터베이스 내에는 많은 객체들이 존재하며 객체들 사이에는 많은 상호관계가 존재한다. 데이터 마이닝과 관련된 기술을 도식하면 그림 1과 같다.

모델을 구성하기 위해 다양한 기법들이 사용될 수 있다. 전통적으로 가장 대표적인 기법이 바로 통계 분석이다. 그러나 이러한 기법은 모델에 관한 기본적인 아이디어를 이미 가지고 있다는 전제에서 이루어진다. 즉 어떠한 변수가 가장 중요하며, 이러한 변수들 사이의 관계를 설명하기 위해서는 어떠한 함수가 사용되어야 하는지를 알고 있을 때 최상의 결과를 얻을 수 있다.

2.1 데이터 마이닝 정의

데이터 마이닝의 대표적인 정의들을 살펴보면 다음과 같다.

Frawley는 데이터 마이닝을 대량의 실제 데이터로부터 묵시적이고 전에는 알려지지 않았지만 잠재적으로 유용한 정보를 추출하는 것이라 정의하였다[2].

가트너 그룹은 데이터 마이닝을 대량의 데이터로부터 패턴 인식기술과 통계기법, 수학적 기법을 이용하여 의미 있는 새로운 상관관계, 패턴 그리고 추세

(trends)를 발견하는 과정이라고 정의하고 있다[3].

데이터 마이닝은 데이터를 여러 가지 다른 조작이나 변형을 통해 다음 2가지 목적으로 한다.

데이터 마이닝을 통해 얻을 수 있는 정보의 형태는 매우 다양하며, 이에 따라 다양한 기법이 존재한다. 가장 대표적인 데이터 마이닝 기법으로는 사건들의 연관성 탐사, 연속성 탐사, 분류 규칙 탐사와 군집 구분을 들 수 있다.

데이터 마이닝은 아래와 같이 많은 종류의 응용에 사용된다.

- * Marketing & retail
- * Banking
- * Finance
- * Insurance
- * Medicine & Health
- * Quality control
- * Transportation

3. 데이터 마이닝 기법

본 장에서는 데이터 마이닝의 일반적인 기법들에 대해 간단히 기술한다. 데이터 마이닝은 데이터에 내재하는 데이터 간의 관계, 패턴, 규칙을 발견하는 데 그 목적이 있다. 또한 작업유형은 데이터 마이닝을 통해 최종적으로 얻고자 하는 정보가 어떤 종류인가에 따라 구분될 수 있다.

3.1 데이터 마이닝 기법

데이터 마이닝에 관한 관심이 증대됨에 따라 여러 가지 데이터 마이닝 기법들이 개발되었으나, 실제계의 응용에서 효율적으로 적용되고 있는지에 관해서는 아직 제고해 볼 여지가 있다. 기본적으로 데이터 마이닝 알고리즘은 기존의 기계 학습 알고리즘에서 대규모의 데이터에 응용 가능하도록 변형되어 사용되는 경우가 많다. 이 밖에 통계학에서 사용되는 기법을 토대로 개발된 기법과 독자적으로 데이터베이스에서 패턴을 찾아내기 위해 개발된 기법들이 있다.

먼저 데이터베이스에서의 지식탐사로서 구체화 된 데이터 마이닝 기법인 연관규칙, 클러스터링에 대해 기술한다. 기계 학습 알고리즘에서 개발되어 데이터 마이닝에 효율적으로 적용되고 있는 결정트리, 신경망, 베이저안 네트워크, 메모리 기반 추론 기법을 기술한 다음, 마지막으로 유전자 기법을 기술한다.

3.1.1 연관규칙

연관규칙은 항목 집합으로 표현된 트랜잭션에서 각 항목간의 연관성을 반영하는 규칙이다. 데이터베이스 검색 회수를 줄이거나, 주기억 장치의 한계를 없애는 등의 발전된 알고리즘들이 발표되어 왔다. 연관규칙을 탐사하는 문제는 기본적으로 미리 결정된 최소 지지도 이상의 트랜잭션 지지도를 가지는 항목집합들의 모든 집합들인 빈발항목집합들(large itemsets)을 찾아 내어 연관규칙을 생성하는 단계로 이루어진다[6].

최근에는 기본적인 연관규칙을 응용한 순차 패턴, 순회 패턴, 주기적인 연관성 등에 연구가 이루어져 이를 활용하는 소프트웨어 개발에도 연구가 집중되고 있다. 최근의 연구 경향은 여러 단계에 걸쳐 최소지지도를 만족하는 연관규칙을 찾는 알고리즘[7], 병렬 알고리즘으로는 후보항목집합을 효과적으로 찾기 위한 확장 가능한 알고리즘의 연구가 있으며[8], Classification Hierarchy를 가진 일반적인 연관규칙을 찾는 H-HPGM이 있다[9]. 가장 최근의 멀티미디어 데이터에서 연관규칙을 찾기 위한 알고리즘으로 MaxOccur에 관한 연구가 진행중이다[10].

3.1.2 클러스터링

클러스터링(Clustering)이란 주어진 데이터 셋을 서로 유사성을 가지는 몇 개의 클러스터로 분할해 내기는 과정으로, 하나의 클러스터에 속하는 데이터 점들 간에는 서로 다른 클러스터 내의 점들과는 구분되는 유사성을 갖게 된다[16]. 데이터 마이닝에서 클러스터링 방법은 기존의 통계학, 기계 학습, 패턴 인식에서 쓰이던 방법에 부가적으로 데이터베이스 지향적인 제약 사항들을 첨가 시킨 것으로써, 최근의 멀티미디어 데이터와 같이 혼합되고 다양한 다차원 데이터를 효율적으로 분류해 나가기 위한 방안으로 연구 되고 있다. 클러스터링 방법은 크게 Partitioning approach와 Hierarchical approach

로 나눌 수 있다.

클러스터링 알고리즘은 다량의 데이터 셋에 대해 효율성을 증진시키는 방법으로 여러 가지 샘플링기법이나 경계 최적화 기법, 인덱스 기법, 집중화 기법 등을 사용하고 있으며, 향후 대상 데이터 집합의 특성과 클러스터링의 목적에 따른 최상의 알고리즘 선택 기준에 대해 지속적인 연구가 필요하다. 응용 영역은 사용자 의도에 따른 원본 이미지 필터링이나 특성 인식, 티슈 분할 등의 영상 이미지 분석 분야이다.

3.1.3 결정 트리

결정트리는 데이터 마이닝의 분류작업에 주로 사용되는 기법으로, 과거에 수집된 데이터의 레코드들을 분석하여 이들 사이에 존재하는 패턴, 즉 부류별 특성을 속성의 조합으로 나타내고 분류모형을 트리의 형태로 만드는 것이다. 이렇게 만들어진 분류모형은 새로운 레코드를 분류하고 해당 부류의 값을 예측하는데 사용되며, 분석의 정확도보다는 분석과정의 설명이 필요한 경우에 더 유용하게 사용된다. 결정트리 분석을 수행하기 위한 다양한 방법들과 분리기준, 정지규칙, 가지치기 방법들이 제안되고, 이들을 어떻게 결합하느냐에 따라서 서로 다른 결정트리 형성방법이 만들어진다. 또한 정확하고 빠르게 결정트리를 형성하기 위해서 다양한 알고리즘이 연구되고 있고, 보다 개선된 알고리즘들이 계속 연구되어 발표되고 있다.

분류나 예측의 근거를 알려주기 때문에 이해하기가 쉽다. 데이터를 구성하는 속성의 수가 불필요하게 많은 경우에도 모형 구축시 분류에 영향을 미치지 않는 속성들을 자동으로 제외 시키기 때문에 데이터 선정이 용이하다. 또한, 연속형이나 명목형 데이터 값들을 기록된 그대로 처리할 수 있기 때문에 지식발견 프로세스 중 데이터의 변환단계에서 소요되는 시간과 노력을 단축시킨다. 마지막으로, 어떠한 속성들이 각각의 부류 값에 결정적인 영향을 주는가를 쉽게 파악할 수 있으며, 모형구축에 소요되는 시간이 짧다. 이에 반하여 결정트리의 단점은 나이나 소득등과 같은 연속형 데이터를 처리하는 능력이 신경망이나 통계기법에 비해 떨어지며 그 결과 예측력도 감소한다는 것이다. 또한, 부류가 연속형 변수의 형태를 취하며, 이것을 예측하는 모형을 구축하는 것이 목적인 경우에는 적합하지 않으며,

모형을 구축하는데 사용되는 표본의 크기에 지나치게 민감하다. 결정트리는 의학 진단 및 연구, 고객 관리를 위한 작업, 날씨 예측, 사기 행위 발견 등 분류를 필요로 하는 여러 영역에서 효율적으로 적용되고 있다.

3.1.4 신경망

신경망은 예측 모델을 만들어내는 매우 강력한 도구로서 데이터 마이닝에 적용되는 중요한 기법 중 하나이다[11]. 신경망의 경우 결과에 대해 어떻게 계산되었는지를 설명하기란 어렵다. 마치 블랙박스와 같아서 사용자들에게 결과를 납득시키기 힘들다는 문제점을 낳고 있다. 이를 해결하기 위해 최근에는 하이브리드 학습으로 퍼지 로직이나 유전자 알고리즘을 이용하여 사용되는 확장된 개념의 신경망들이 사용되고 있다. 신경망의 적용분야는 모델의 특성에 따라 아주 다양하지만, 공통된 장점은 입력 선택이나 트레이닝 없이 다양한 입력 데이터를 다룰 수 있고 훌륭한 예측 모델이라는 것이다.

신경망의 유연한 비선형 회귀 모델 그리고 분류 모델은 통계학 또는 마이닝 툴 등에 유용하게 쓰이고 있으며, 분류, 클러스터링, 모델링, 타임시리즈 예측의 영역에 적용된다. 신경망 알고리즘의 예측 모델로 데이터베이스 마케팅, 고객 관계 유지, 사기행위 적발, 파산 예측 등 그 응용 분야가 매우 다양하다[12].

3.1.5 베이지안 네트워크

베이지안 네트워크는 흥미 있는 변수들 사이의 확률적인 관계를 표현하는 그래프 모델이다[13]. 베이지안 네트워크의 특징은 다음과 같다. 첫째, 불완전한 데이터 집합을 처리할 수 있다. 모든 경우에 대한 모든 변수를 측정하여 분류를 수행할 경우, 입력 변수 중 한 개라도 관찰되지 않으면, 대부분의 모델은 정확한 예측을 수행하지 못한다. 베이지안 네트워크는 그러한 의존성을 표현하는 자연스러운 방법을 제공한다. 둘째로, 베이지안 네트워크는 인과 관계에 대한 학습이 가능하다. 인과 관계가 중요한 이유는, 우리가 문제 도메인에 대하여 이해를 할 때 유용하고, 또한 간섭에 생길 경우에도 예측을 가능하게 한다는 것이다. 셋째, 베이지안 네트워크는 도메인 지식과 데이터를 결합하여 사용한다. 지식 발견을 위한 대부분의 테크닉들이 단지

데이터에만 의존하지만, 베이지안 네트워크에서는 데이터 뿐만 아니라 전문가로부터 온 전문가의 지식도 함께 사용한다. 즉 여기서의 지식 표현은 우리가 전문가 영역 지식과 통계적 데이터를 혼합한 새로운 지식을 발견할 수 있도록 한다는 것이 큰 특징이다. 넷째, 베이지안 방법은 데이터의 over-fitting을 피할 수 있는 효율적이고, 원칙적인 방법을 제공한다. 테스트를 위한 유용한 데이터 중 몇몇을 유지할 필요가 없고, 베이지안 방법에서는, 모델이 모든 유용한 데이터를 트레이닝을 위해 사용할 수 있도록 한다. 베이지안 네트워크에서의 지식 발견은 신경망을 사용하는 것과 유사하지만, 중요한 장점이 있다. 그것은 베이지안 네트워크에서는 전문가 지식을 쉽게 인코딩할 수 있고 지식 발견의 정확성과 효율성을 증가시키기 위해 이 지식을 사용할 수 있다는 것이다.

베이지안 네트워크의 효율적 응용 영역으로는 DNA 구조와 같은 생물학적 데이터의 자동 패턴 분류나 불확실한 자료에 대한 추론이 요구되는 여러 영역이다.

3.1.6 메모리 기반 추론(MBR: Memory-Based Reasoning)

메모리 기반 추론은 사례 기반 추론의 한 방법이다. 전통적인 사례 기반 추론 방법은 주어진 문제와 가장 유사한 문제를 찾아, 미리 풀어진 문제의 해답을 적용함으로써 해답을 구하고자 하는 것이다[15]. 여기서 메모리 기반 추론은 메모리에 저장된 사례들 중 가장 가까운 사례를 선정하여 문제 해결에 적용한다.

메모리 기반 추론이 현재 구축되어있는 데이터베이스 내에서 동작한다는 점은 많은 다른 학습기반 접근 방식에서도 사용되어진다. 또한 매우 일반적인 메카니즘, 규칙, 인식 추론을 단일 프레임워크 안에서 가능하게 한다. 수백만 건의 카탈로그에 관심 있어 하는 수백만의 사람들을 분류할 때와 데이터 포맷을 구별하기 어려운 free-text라 할지라도 잘 동작한다. 결과를 쉽게 이해할 수 있고, 학습 알고리즘으로 새로운 사례가 가능하게 되자마자 초과되는 계산 비용 없이 학습이 수행되어 트레이닝 집합을 유지하기에 최소의 노력이 필요하다는 강점이 있다. 현재 메모리 기반 추론의 구현에서 수정이 요구되는 몇 가지의 약점으로는, 대표적으로 트레이닝 집합을 위한 많은 공간이 필요

하다는 것과, 표현에 대한 의존도가 높고, 아직 그것을 표현하는데 잘 수정할 만한 수단이 없다는 것이다.

데이터 마이닝에서의 메모리 기반 추론은 축적된 과거 자료들로부터 유사한 사례들을 바탕으로 문제를 해결하는 기법으로 사용되고 있다. 특히 분류와 예측에 활용되어지고 있어 비계획적으로 수집된 대용량의 데이터를 다루는데 유용하며, 신경망 알고리즘과 유전자 알고리즘과 결합하여 다양한 범주로의 분류나 수치값의 예측을 수행한다. 데이터 마이닝에서 사례 기반 기법은 기존의 기계학습, 패턴인식에서 쓰이던 방법에 부가적으로 데이터베이스 지향적인 제약 사항들을 첨가시킨 것으로써, 최근의 멀티미디어 데이터와 같이 다양한 데이터를 효율적으로 분류해 나가기 위한 방안으로 연구되고 있다.

3.1.7 유전자 알고리즘

유전자 알고리즘은 자연계의 진화현상으로부터 유도된 기계학습의 한 모델이다[34]. “자연은 자연선택과 적자생존을 통해 적응한 유기체를 골라낸다.”는 다윈의 이론을 따온 이 알고리즘은 1975년 미시건 대학의 존 홀랜드가 인공지능을 연구하는 과정에서 개발되었다. 유전자 알고리즘의 수행과정은 다음과 같다. 첫번째 단계로, 만들기 쉽고 다루기 쉬우며 다양한 정보를 쉽게 부호화할 수 있는 일정한 길이의 이진 문자열로 개체를 만들고 초기값으로 할 세대를 생성한다. 다음으로 가능한 해법들의 집합 또는 가설들의 집합인 개체군에서 모든 개체들의 적합도를 산출한다. 또한, 개체에서 교배, 적합도 비례 재생성, 변이 등과 같은 유전적 연산들을 수행함으로써 새로운 세대 생성한다. 마지막으로, 새로운 개체군을 사용하여 위의 과정을 반복하며 이러한 반복 순환을 한 세대라고 말한다. 따라서 유전자 알고리즘은 적합한 해법을 얻어내는 방법으로서, 대규모의 탐색공간에서 최선의 적합도의 해를 찾는 일반적인 최적화 과정이다.

유전자 알고리즘이 데이터 마이닝 영역에 적용될 때에는 다른 알고리즘들의 부분 작업들을 최적화하는 작업으로 사용된다.

유전자 알고리즘의 강점으로는 최종세대의 결과값이 곧 최적해를 나타냄으로 결과에 대한 별도의 해석이 필요없다는 점과 결과값을 적용하기가 쉽다는 것,

그리고 다양한 데이터 형태에 적용가능하며 넓은 영역의 데이터를 조절할 수 있으며 많은 최적화 문제의 응용에 적용된다는 것이다. 단점은 많은 문제들을 일정한 길이의 지놈으로 인코딩하기가 어렵다는 점과 최적화에 대한 보장을 할 수 없고, 계산비용이 높으며 현재까지 적은 수의 상업패키지에서만 유용하다는 것이다. 따라서 유전자 알고리즘은 데이터 마이닝의 영역에서 결과를 최적화 시키는데 있어 효과적인 내장 엔진으로써 적용될 수 있다.

4. 최근 기술 동향

최근 데이터 마이닝 분야에서 가장 주된 관심사는 CRM과 관련된 어플리케이션이다.

가트너 그룹에서는 데이터 마이닝의 최근 기술 동향을 다음과 같이 분석하고 있다.

오라클이나 IBM, 그리고 마이크로소프트는 자사의 데이터베이스 제품에 데이터 마이닝 기능을 추가하고 있다. 또한 정규화되지 않은 데이터 마이닝 기술이 있다. 정규화되지 않은 데이터를 분류 및 분석, 검색할 수 있는 전반적인 기술이다. 이러한 기술들은 자연어 처리, 정보 분석, 패턴 인식 등이다. 많은 메이저 벤더들이 정규화되지 않은 데이터 마이닝 기능을 갖는 컴포넌트들을 추가하고 있다.

실시간 엔터프라이즈급 및 상호 운용성에 관한 분야이다. 데이터베이스 분야의 다양한 프로토콜들이 상호 운용성을 위해 연구되고 있다. 브로커나 어댑터 기술의 미들웨어를 통해 통합되어 나가는 움직임이 있다.

또 다른 동향으로 하이브리드 데이터 마이닝에 관한 연구이다. XML을 기반으로 한 데이터베이스 시스템에서 멀티미디어 데이터 마이닝과 같은 새로운 레벨의 서비스가 연구되고 있다. 최근에는 데이터 마이닝 모델에서 좀 더 간편화된 그리고 자동분석이 가능하도록 연구되고 있다. 신경망과 같은 기법을 통해 자동분석이 가능하다.

5. 결 론

정보기술 산업의 발전과 더불어 데이터의 양은 폭발적으로 증가하고 있다. 또한 대량의 데이터들을 모

두 분석하여 활용하기에는 한계가 있기 때문에 새로운 개념의 데이터 분석기법이 필요하다. 이에 따라 대용량의 데이터들간의 상호 관계를 분석하는 일이 매우 중요한 문제로 부각되고 있다.

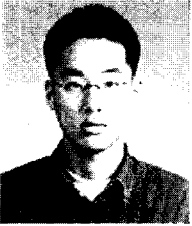
데이터 마이닝은 이전에 발견되지 않았던 데이터들간의 상호관계, 패턴 등을 찾아내는 기법으로 데이터 분석에 활용된다.

본 고에서는 데이터 마이닝의 최근 기술 동향에 대해 고찰해 보았다. 최근 많은 분야에서 데이터 마이닝 기술을 적용하려고 시도하고 있으며 기존의 많은 방법들이 혼합되어 시도되고 있다.

참 고 문 헌

- [1] ACM SIGKDD, <http://www.acm.org/sigkdd>
- [2] Frawley, W., Batheus, C. Knowledge Discovery in Databases: An Overview, 1991.
- [3] Gartner Group, <http://www.gartner.com>
- [4] J. Han and M. Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann, 2001.
- [5] 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, Washington, D.C., May 1993.
- [6] R.Agrawal and R. Srikant, "Fast algorithms for mining association rules", In Proceedings of the 20th VLDB Conference, Santiago, Chile, Sept, 1994.
- [7] Bing Liu, Wynne hsu and Yiming Ma, "mining Association Rules with Multiple minimum Supports. In proceedings". of ACM KDD-99, 1999.
- [8] Eui-Hong Han, George Kerypis, Vipin Kunmar, "Scalable Parallel Data mining for Association Rules", In Proceedings of the SIGMOD, 1997.
- [9] Takahiko Shimtani, Masaru Kitsuregawa, "Parallel Mining Algorithms for Generalized Association Rules with Classification Hierarchy", In Proceedings of SIGMOD '98, 1998.
- [10] Osmar R. Waiane, Jiawei Han, Hua Zhu, "Mining Recurrent Items in Multimedia with Progressive Resolution Refinement", In Proceedings of Data Engineering Conference, San Diego, California, March 2000.
- [11] William J. E. Potts, "Generalized Additive Neural Networks", SAS Institute Inc., ACM KDD-99, 1999.
- [12] J.P.Bigus, Data Mining with Neural Networks, Mc Graw-Hill, 1996.
- [13] David Heckerman, "A Tutorial on Learning With Bayesian Networks", 1995.
- [14] David Heckerman, "Bayesian Networks for Knowledge Discovery", Advances in knowledge discovery and data mining, pp. 273-305, 1996.
- [15] Farhi Marir and Ian Watson Case-Based Reasoning: A Categorized Bibliography.
- [16] Alexander Hinneburg, Daniel A.Keim, "Clustering Method for Large Data Sets", SIGMOD99, 1999.
- [17] IT World, <http://www.itworld.co.kr/>.

◎ 저자 소개 ◎



손 재 기

1998년 경기대학교 전자계산학과(학사)

2001년 경기대학교 대학원 전자계산학과(석사)

2001년~현재 : 전자부품연구원 정보시스템 연구센터 전임 연구원

관심분야 : 자료저장 시스템, 리눅스 프로그래밍, 객체지향 프로그래밍



박 상 현

2000년 한국외국어대학교 컴퓨터공학과(학사)

2002년 한국외국어대학교 대학원 컴퓨터공학과(석사)

2002년~현재 : 전자부품연구원 정보시스템 연구센터 연구원

관심분야 : 임베디드 시스템, 자료저장 시스템, 웹 프로그래밍



민 수 영

1987년 인하대학교 전자공학과(학사)

1992년 고려시스템 기술 연구소 주임 연구원

1993년~현재 : 전자부품연구원 정보시스템 연구센터 책임 연구원

관심분야 : 자료저장 시스템, 임베디드 시스템, 네트워크 프로토콜