데이터마이닝 기법을 이용한 지능형 기자재 관리 시스템 설계 및 구현

조영기 · 김상수 · 조주상 · 백성욱
세종대학교 컴퓨터공학부

요 약

본 논문에서는 디지털 콘텐츠 학과의 고가 장비들을 효율적으로 관리하기 위해 구축된 지능형 기자재 관리 시스템의 설계 및 구현사례를 제시한다. 시스템 운영을 지원 하기위해 사용자, 장비, 사용기록의 데이터를 기반으로 데이터마이닝을 수행했으며 데이터 마이닝 과정에서 나타난 여러 유용한 규칙들을 제시했다. 분석된 정보를 이용해 사용자 위주의 운영정책과 동적 인터페이스를 제공하기 위한 웹사이트의 개인화 계획을 제시하였다.

Design and Implementation of Intelligent Equipment Management System Using Data Mining

Yung-Ki Jo · Sang-Soo Kim · Ju-Sang Cho · Sung-Wook Baik

ABSTRACT

This paper presents a design and implementation example of intelligent equipment management system that is constructed to manage high price equipment of digital content department effectively. To support system operation we executed data mining and presented various rules that appear in data mining process based on data of user, equipment and using record. We presented personalization plan of web site to offer user dependent administration policy and dynamic interface using analyzed informatio.

1. 서 론

오늘날 디지털 정보기술의 발달로 정보관리와 활용에 대한 인식이 높아지면서 효과적인 정보관리와 정보 활용 방안에 대한 연구가 활발해지고 있다. 기업들은 다변화하는 경영환경, 고객, 기업 내부의 요구사항을 빠르게 소화하고 적응해야 한다. 이를 위해서는 신속하고 정확한 마케팅 전략과 여러 가지 상황에 대한 적절한 의사결정을 위한 의미 있는 고급 정보 혹은 지식들이 필요할 수밖에 없다. e-business 경쟁력을 강화하기 위해 기업들은 자사가 보유하고 있는 데이터를 활용해 제품이나 서비스에 대한 고객의 다양한 욕구를 충족시키기 위한 노력을 기울여야 한다. WWW의 규모와 사용의 지속적 증가는 온라인 정보서비스의 개발과 설계의 새로운 방법을 요구하고 있다. 대부분의 웹 구조는 크고 복잡하며 사용자들은 이런 웹 구조를 통해 항해할 때 종종 그들이 찾고자 하는 것을 찾지 못하는 경우가 많다. 웹상에서는 사이트 간에 이동이 쉽고 사용자들은 유동성이 강함으로 사용자들의 요구사항과 특징들을 잘 이해해서 사용자에 따른 개인화 방안을 강구하는 것은 매우 중요하다. 온라인에서 개인화를 활용하는 것은 실시간으로 개인화 기법을 적용하며 고객에게 새로운 것을 소개하고 고객의 잠재적 요구를 통해서 웹사이트의 내용을 변화시키는 것을 말한다. 이는 실시간으로 축적된 데이터를 데이터마이닝 기법으로 패턴을 분석함으로써 가능하다.[1]

데이터마이닝 기법을 이용한 지능형 개인화 엔진을 기자재 관리 시스템에 활용하기 위해 사용자들의 성향을 파악하여 사용자들에게 필요한 콘텐츠와 정보를 제공한다면 더욱 적극적인 관심을 보일 것이다.

본 논문에서는 지능형 기자재 관리 시스템의 설계 및 구현 사례를 제시한다. 기자재 관리 시스템을 통해 수집한 사용자들의 사용기록에 대한 데이터들과 학과 데이터베이스를 통해 얻은 사용자의 세부 정보들을 데이터마이닝 과정에 따라 분석하여 여러 유용한 패턴을 발견했다. 분석된 정보를 기반으로 사용자 위주의 운영정책과 동적 인터페이스를 제공하기 위한 웹 사이트 개인화(Personalization) 계획을 제시한다. 본 논문의 구성은 먼저 2장에서 시스템의 설계 및 구현 사례를 제시하고, 3장에서는 데이터마이닝의 목적과 수행과정에 대해 설명한다. 또한 4장은 데이터마이닝을 통해 획득한 정보를 시스템에 반영하기 위한 개인화 방안에 대해서 다룬다. 마지막으로 5장에서는 설계 및 구현 과정에서 드러난 문제점과 향후 연구방향을 제시한다.

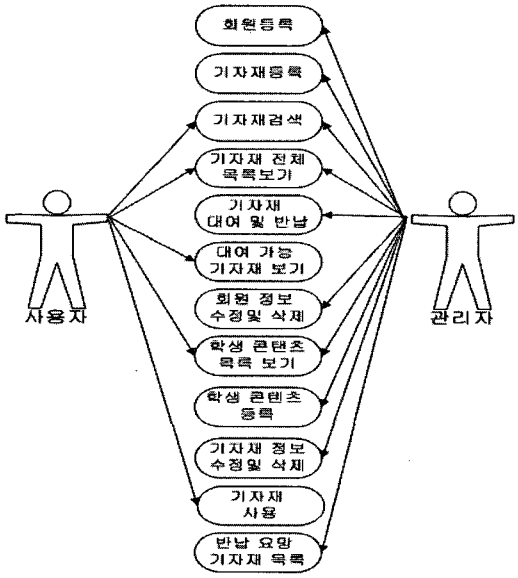
2. 시스템 설계 및 구현

2.1 지능형 기자재 관리 시스템의 개요

지능형 기자재 관리 시스템은 사용자들이 인터넷을 통해서 손쉽게 기자재 및 각종 콘텐츠를 접할 수 있도록 구현되었다. 콘텐츠 구성은 기자재 및 기자재를 이용해 만든 콘텐츠에 대한 문서와 기자재 이미지, 동영상 등 멀티미디어 콘텐츠로 되어 있다. 구현 환경은 Linux 7.2 운영체제 상에서 PHP와 MySql DB서버를 연동했으며 웹 서버로는 Apache 1.3을 사용했고 멀티미디어 콘텐츠의 가공 및 편집에는 Adobe Photoshop과 Premiere를 사용했다.

2.2 인터페이스

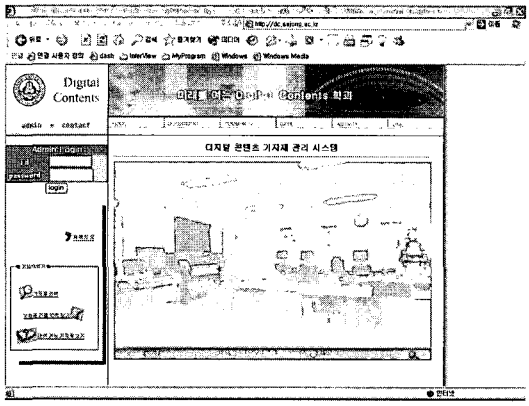
지능형 기자재 관리 시스템의 인터페이스는 기자재 콘텐츠와 기자재를 이용해 제작된 학생 콘텐츠를 관리하기 위해 구현된 것으로 관리자와 사용자 측면에서 기술되었다. [그림 1]은 사용자 인터페이스를 나타내는 Use Case Diagram이다. 관리자는 전체 시스템을 관리하기 위해 대부분의 기능을 수행하고, 사용자는 콘텐츠 정보를 받아들이는 기능을 수행한다.



(그림 1) 인터페이스 설계

2.3 기능

본 시스템의 주요구성은 기자재 관리, 사용자 정보관리, 학생 콘텐츠 관리, 대여 및 반납 관리, 기자재 검색 기능 등이 있다. 빠르고 정확한 검색 기능을 구현했고 편리한 입력양식을 기초로 한 인터페이스를 가지고 있다.



(그림 2) 기자재 관리 시스템 초기화면

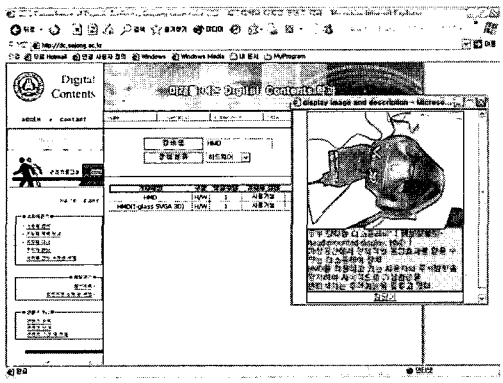
[그림 2]에 나타난 초기 화면은 세종대학교 디지털 콘텐츠 학과 기자재실의 내부 전경이다. 파노라마 카메라를 수직과 수평으로 이동시켜 모든 위치를 촬영한 후 Sticher를 사용하여 이미지 결합과 렌더링 작업을 했다. 최종 모습은 Photovista를 이용해 자바 애플릿으로 완성시켰다. 이를 통해 사용자는 3차원 공간상에서 원하는 각도에서 원하는 위치를 볼 수 있으며 확대와 축소가 가능해 장비의 자세한 모습까지도 볼 수 있다. 상단은 시스템 메뉴가 있고 좌측에 있는 인증을 위한 로그인 화면을 통해서 별도의 관리자 로그인 없이 관리자 계정을 입력하면 관리자 인증을 통해 로그인 할 수 있도록 구현했다. 또한, 브라우저를 프레임으로 나눠서 불필요한 페이지 이동을 줄임으로써 페이지의 빈번한 reload 없이 필요한 부분만 새롭게 볼 수 있도록 클릭에 대한 반응 속도를 향상시켰다.

2.3.1 키워드 기반 정보 검색

[그림 3]은 기자재 정보검색화면이다. 정보 검색엔진은 원하는 기자재 정보를 사용자가 빠르게 접근할 수 있도록 하기 위해 구현된 내용으로 키

워드를 통한 검색이다. 정보검색 알고리즘은 다음과 같다.

```
<? # equipment 테이블에 저장된 데이터를 불러온다. #
if ($_POST[classification]=='all') {
        $query = "SELECT distinct equip_name, s_number,
        cost, classification, description, image,
        count(*) as 'e_count' FROM equipment
        where equip_name LIKE '%$S_POST[equip_name]%'
        group by equip_name"; }
else { $query = "SELECT distinct equip_name, s_number,
        cost, classification, description, image,
        count(*) as 'e_count' FROM equipment
        WHERE equip_name LIKE '%$_POST[equip_name]%'
        AND classification='$_POST[classification]'
        group by equip_name"; }
$result = mysql_query($query); ?>
```

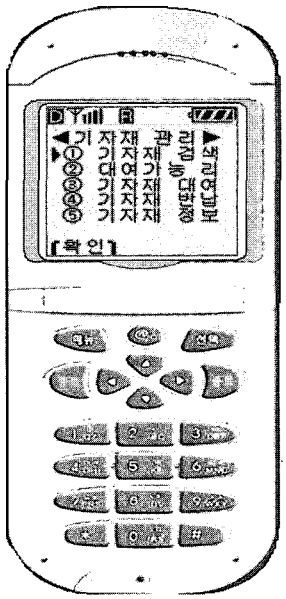


(그림 3) 기자재 정보검색화면

2.3.2 모바일 연동 기능

모바일 연동은 브라우저가 내장된 휴대폰을 이용해 웹 서버에 접속할 수 있는 기능으로 언제 어디서나 접속이 가능한 이동성이 가장 큰 장점으로 부각 된다. 이런 흐름에 따라 제한적이지만 기자재 관리 시스템을 모바일을 통해 접속할 수 있도록 구축했다. [그림 4]는 시뮬레이터를 이용해 웹 서버에 접속한 화면이다.

무선 인터넷 서비스를 구축하는데 있어 특별한 웹 서버를 필요로 하는 것은 아니다. 무선인터넷용 웹 서버를 Windows98/NT, Linux, Unix 등 기존 웹 서비스용으로 구축된 서버에 MIME type이 지원되도록 설정을 추가해 줌으로써 무선 인터넷 웹 서버로서 역할을 수행할 수 있으며 Server가 받은 Request에 대한 응답으로 데이터를 보낼 때에는 MIME type도 같이 보낸다. 브라우저가 내장된 휴대폰은 인터넷 서비스에 있어서 PC에 해당된다고 볼 수 있다. 휴대폰의 액정 화면이 매우 작고, PC나 노트북의 CPU와 메모리의 용량에 비해 사양이 매우 열악하므로 작은 사이즈의 데이터를 처리할 수밖에 없어 HTML이 아닌 모바일을 위해 mHTML을 사용했다.

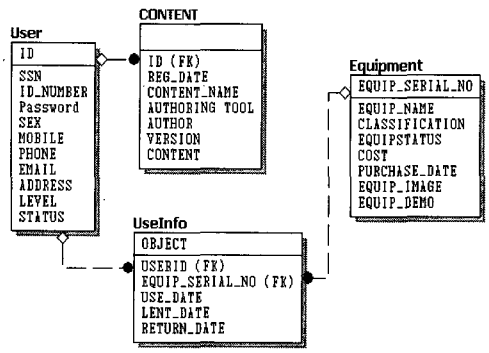


(그림 4) 시뮬레이터를 통한 연동

2.4 데이터베이스 설계

데이터베이스의 설계는 사용자 정보, 기자재 정보, 사용과 대여 정보 등을 각각의 테이블로 나타냈으며 기자재의 활용분야별로 일련번호를 구분했다.

특성에 따라 분류된 필드를 보면 기자재들은 'H/W', 'S/W', 'H/W&S/W', 'ETC'로 분류했으며 사용자들은 '교수', '직원', '대학원생', '학부생'으로 분류했다. 기자재의 사용목적에 따라서는 'Lecture', 'Research', 'Practice', 'Seminar', 'Development'로 분류했고, 기자재의 상태에 따라서는 '사용가능', '대여가능', '사용불가', '사용중', '고장', '수리중'으로 분류했다. 데이터베이스는 서로 다른 패턴을 보이는 사용자들의 성향과 신용도, 사용목적에 따른 기자재 이용 실태를 평가하기 쉽게 설계하려고 노력했다.



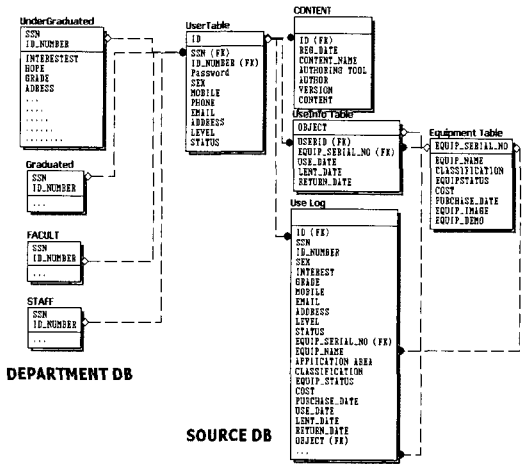
(그림 5) 데이터베이스 모델링

[그림 5]는 부분적인 데이터베이스 모델링을 나타낸 것 이다. User 테이블은 사용자 프로필을, Equipment 테이블은 기자재 정보를, UseInfo 테이블은 사용과 대여 및 반납 정보를, CONTENT 테이블은 기자재를 활용해 제작된 콘텐츠 정보 및 콘텐츠를 나타냈다. 각 테이블은 시스템 운영에 필요한 데이터들이 관리되는 필수 테이블이다.

s_number	equip_name	classification	cost	purchase_date	equip_status
a3001001	모션추적장치(기계식)	H/W	88110000	2001-10-01	사용가능
a1001001	스테레오프로젝터/편광스크린/안	H/W	58000000	2001-10-01	사용가능
a1002001	원격강의용 화이트보드 영상PC일	H/W	1350000	2001-10-01	사용가능
a1003001	Cyberglove	H/W	1980000	2001-10-01	사용가능
a2001001	사운드샘플러	H/W	820000	2001-10-01	사용가능
a1004001	사운드 신디사이저	H/W	1980000	2001-10-01	사용가능
a1005001	사운드 시퀀서	H/W	2530000	2001-10-01	사용가능
a1020001	VCR(Sony DSR-45)	H/W	5830000	2002-08-01	사용가능
a2002001	RapidForm(2001,2002)	S/W	9900000	2001-10-01	대여가능
a1021001	표준 모니터(Sony 14'')	H/W	1452000	2002-08-01	사용가능
a2015001	편집 S/W(Avid, DV Express)	S/W	3267000	2002-08-01	사용가능
a1006001	3D스캐너	H/W	39600000	2001-10-01	사용가능
a4002001	모션 마이크	Etc	1815000	2002-08-01	사용가능
a1007001	NoteBook	H/W	2520000	2001-10-01	사용가능
a3002001	편집 슬루션(DC-2000+PC)	H/W & S/W	9790000	2002-08-01	사용가능
a3003001	편집 슬루션(DD-200)	H/W & S/W	660000	2002-08-01	사용가능
a2003001	3DsMax	S/W	11A0000	2001-10-01	대여가능
a4003001	조명 set	Etc	1991000	2002-08-01	사용가능
a1022001	카메라 set(Sony,PD-150)	H/W	5360000	2002-08-01	사용가능
a2016001	PREMIERE 5.0	S/W	275000	2002-08-01	대여가능

(그림 6) 기자재 테이블

[그림 6]은 기자재 테이블을 나타낸 것이다. 기자재의 사용과 대여가 발생할 때마다 기자재의 상태가 새로 업데이트 되도록 했으며 사용자별로 데이터베이스 접근의 제한을 다르게 했다.



(그림 7) 데이터마이닝을 위한 데이터베이스 모델링

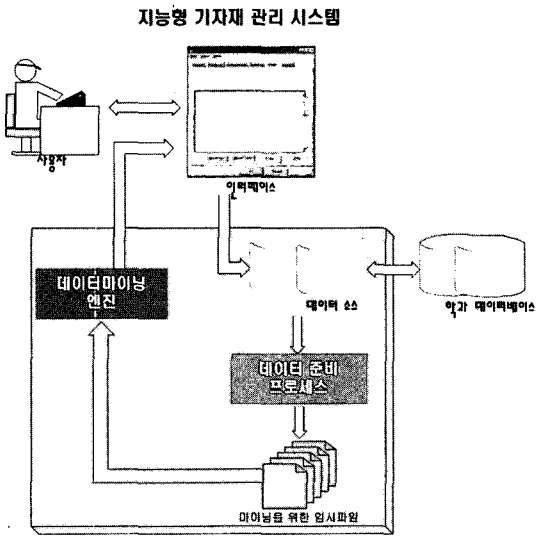
[그림 7]은 데이터마이닝을 위한 데이터베이스 모델링 사례이다. 사용자별 세부 정보를 얻기 위해 학과 데이터베이스에 있는 교수, 대학원생, 학

부생, 직원의 테이블을, 사용자 테이블과 조인했다. UserLog 테이블은 여러 정보들을 나타내기 위해 많은 필드로 구성했으며 기자재가 이용될 때마다 사용기록, 사용자 세부정보, 사용된 기자재 세부정보가 기록된다. 기록된 데이터는 데이터마이닝 기법을 적용하기 위해 여러 단계를 거치게 된다.

3. 데이터마이닝을 통한 규칙발견

데이터마이닝은 방대한 데이터를 기반으로 데이터웨어하우스나 데이터마트안에 저장되어 있는 대량의 데이터로부터 쉽게 드러나지 않는 유용한 정보들을 추출하는 과정을 말한다. 대량의 데이터는 분석없이 그 자체만으로는 의미가 없다. 그 이유는 그 데이터의 용량이 너무 방대하고 복잡해서 보통 사람들이 그 데이터를 이해하기란 매우 힘들기 때문이다. 데이터 분석가들도 그들이 필요한 부분만을 선택하여 심층 분석을 한다. 그러므로 그 이외에 사용하지 않은 데이터들로부터의 의미있는 정보나 지식 등을 놓치는 경우가 발생하게 된다. 이러한 지식을 유출해내는 방법은 어떤 특정 기법과 그 기술 자체만을 의미하는 것이 아니고, 비즈니스 문제나 과학용 데이터들을 분석하는 문제 등에서 주어진 상황을 이해하고 그 특정한 문제를 해결하기 위하여 여러 학문 분야의 방법들을 적용하는 포괄적인 과정을 의미한다. 인터넷과 분산 데이터베이스의 발달로 인한 온라인 데이터들의 급증으로 데이터마이닝의 필요성은 더욱 절실해졌다. 데이터마이닝 기법은 여러 분야에서 활용되고 있다. 비즈니스 분야에서 CRM(Customer Relationship Management:고객 관계 관리 시스템)의 핵심적인 기술로 사용되고 있으며 바이오테크놀로지의 발달로 인해 그 분야의 방대한 데이터를 분석하는 기술로 부각되었다.[2,3] 성공적인 데이터마이닝 시스템 구축을 위해서는 목표에 대한 정확한 이해가 필요하며 구체적인 활용방안이 제시되어야 한다. 또한 레코드의 충실도를 고려해야 하고 도출된 결과에 대한 올바른 해석이 중요하다.[4]

지능형 기자재 관리 시스템을 통해서 얻어진 데이터를 기반으로 각각의 패턴을 분석해서 기자재 관리의 효율성 증진을 위한 데이터마이닝 사례를 제시한다.



(그림 8) 데이터마이닝을 위한 시스템 구성

3.1 문제 정의 및 목표설정

본 사이트는(http://dc.sejong.ac.kr) 학교 내의 수많은 사용자들이 방문을 하며, 이들 대부분은 고가장비의 사용이나 대여에 큰 관심을 가진 이들이다.[5] 등록된 수많은 고가장비들은 대여 및 사용에 매우 신중한 판단이 요구되고 있어서 제한적이고 편향적인 관리가 이뤄지기 쉽다. 따라

서 각각의 사용자 특성에 맞는 사용자위주의 적극적이고 공정한 관리시스템 운영이 필요하다. 제시할 데이터마이닝 사례의 목적은 사용자마다 개인화된 인터페이스를 제공하기 위해 관심분야와 사용기록을 토대로한 추천 시스템을 구현함으로써 효과적인 기자재 활용 방안을 제시하는 것과 새로운 기자재 구입 계획에 사용자의 관심도, 사용 성향, 요구사항등 분석한 내용을 반영함으로써 기자재가 사용 목적에 맞게 사용될 수 있도록 하는 것이다.

데이터마이닝을 통해 얻고자 하는 정보는 사용기록을 토대로 한 사용자 성향과 사용패턴, 사용자 특성에 따른 관심분야 예측 등이다. 이를 위해 의사결정나무 기법을 적용했다.

3.2 데이터마이닝 수행과정

데이터마이닝의 수행 과정은 그 목적에 따라 다양한 방법들이 존재한다. 유용한 패턴과 같은 지식들을 추출하기 전에 데이터를 선택하거나 여과, 통합, 샘플링, Noise 제거, 변형등과 같은 과정을 통해 업무량을 줄이거나 분석의 효과를 높인다.[6]

3.2.1 표본 선정 및 추출(Sampling/Selection)

방대한 양의 데이터로부터 적은 양의 표본을 추출하는 과정으로, 얻고자 하는 정보를 정의하고 활용방안을 수립하며 마이닝하고자 하는 데이터의 종류를 파악하는 과정이다. 데이터마이닝에 적용될 데이터의 선택은 앞서 설명한 것처럼 질의를 통해 UserLog 테이블에 기록했다. 그 결과 비교적 많은 양의 레코드를 갖는 테이블을 얻게 되었다.

3.2.2 데이터 정제 및 전처리(Data Cleaning/Pre Processing)

데이터베이스에서 결손값, 잡음 등을 제거하고 데이터 범위를 벗어난 데이터 및 특이값을 제거하는 단계이다. UseLog 테이블에 적재하는 과정에서 발생한 결손값을 제거하기 앞서, 사용자 세부정보와 프로필 정보가 일치하지 않은 레코드를 먼저 선별했다. 이 과정에서 상당양의 레코드가 줄어들었다. 결손값을 가지는 원인은 많은 필드가 필수 항목이 아닌 경우 임시로 값을 입력하거나 입력을 뒤로 미루면서 발생하는 경우가 많았고, 학과 데이터베이스에 존재하는 사용자 세부정보와 시스템 데이터베이스의 사용자 프로필 정보가 일치하지 않는 경우도 많은 것으로 나타났다.

3.2.3 탐색 및 변형 (Exploration/Transformation)

탐색 단계는 본격적으로 구축된 데이터베이스에서 데이터를 살펴보는 작업을 수행하게 된다. 탐색과정에서 이미 알고 있는 사실들을 수치화하는 작업을 시작으로 수많은 변수들의 관계를 살펴보는 단계이다. 데이터의 변형단계에서는 데이터에 포함된 불필요한 레코드와 항목을 삭제하는 작업이 우선적으로 시행된다. 결손값을 가지는 레코드를 제거했다고 해도 선택항목으로 입력된 데이터는 부정확하고 마이닝의 유용성을 떨어뜨릴 확률이 높다. 먼저 주소, 기자재명, 관심분야 등 텍스트형식의 필드를 범주형 값을 갖는 필드로 변환했고 다른 범주형 값을 갖는 데이터와 함께 수치화했다. 비교적 정확한 값을 갖는 주민등록번호 필드를 이용해서 나이와 성별을 추출했다. UseInfo 테이블, User 테이블, Equipment 테이블을 기반으로 사용자별 대여 및 사용 현황 파악을 위해서 대여 및 사용횟수, 반납 및 사용기일 초과 횟

수 필드를 추가했다.

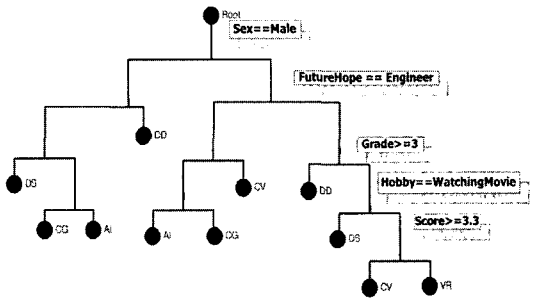
3.2.4 모형화(Modeling)

모형화 단계는 데이터마이닝 과정중 가장 중요한 단계로서 선행 단계에서 선정된 주요 변수를 사용하여 다양한 모형을 적용해 보는 단계이다. 모형화 단계에서는 의사결정나무(Decision Tree)모형을 적용했다. 의사결정나무는 추출된 모델이나 지식들을 아주 잘 표현할 수 있는 장점이 있기 때문에 데이터마이닝의 방법론 중에 가장 많이 활용되는 방법 중 하나이다. 의사결정나무 기법을 통해 얻어진 규칙정보를 통해 사용자의 특성에 따른 기자재 사용패턴, 관심분야, 기자재 특성에 관한 여러 규칙들을 얻었다.

학부생의 세부정보를 변수로 이용하여 얻은 규칙을 보면 다음과 같다.

```
IF (Sex==Male)
  AND (FutureHope==″Engineer″)
  AND (Grade〉=3)
  AND (Hobby==″WatchingMovie″)
  AND (Score〉=3.3)
RESULT-〉 세부 관심분야==′VR′
```

디지털 디자인(DD), 디지털 사운드(DS), 컴퓨터 그래픽스(CG), 컴퓨터 비전(CV), 인공지능(AI)을 각각의 클래스로 분류했다. 장래희망은 Programmer, Engineer, Graphic Designer등으로 분류했고, 취미는 웹 서핑, 영화감상, 스포츠, 게임 등으로 분류했다.



(그림 9)의사결정나무 분류과정

이 같은 패턴을 가지는 것으로 나타난 학생들의 경우 관심분야가 가상현실이므로 로그인시 가상현실에 관련된 장비를 추천하여 사용자로 하여금 관련 장비를 사용할 수 있게 하거나 관심을 가질 수 있도록 유도할 수 있다.

기자재 세부 정보를 바탕으로 사용자의 관심도를 분류한 규칙을 보면 다음과 같다.

```
IF(Classification==′H/W′)
  AND(EquipmentCost〉=100)
  AND(EquipmentStatus==′repair′)
  AND(EquipmentAppArea==′Digital
Sound′)
  AND(UseObject== ′Development′)
  AND(PurchaseDate〉=3)
RESULT-〉 기자재 관심도==Low
```

주로 디지털 사운드 제작에 활용되는 하드웨어로 고장 발생률이 높으며 구입일이 3년 이상 지난 노후화된 장비에 대해 낮은 관심도를 보이는 사례가 많은 것으로 나타났다. 이를 토대로 활용가치는 높지만 노후화로 인해 사용에 대한 관심도가 낮은 장비의 패턴을 발견했으며 추가 구입 장비의 판별에 도움을 줄수 있는 데이터를 얻었다.

이외에도 사용자의 신용도나 사용성향 등에 관한 몇 가지 신뢰할 만한 규칙들이 발견됐다.

3.2.5 가시화(Visualization)

데이터마이닝의 수행 결과를 사용자들이 보기 편하고 이해하기 쉬운 형태로 제공하기 위해서 각종 차트 형태로 보여주는 것을 말한다. [그림 10]은 모형화 과정에서 얻어진 규칙(rule)들을 사용자에게 효과적으로 전달하기 위해 InferView를 이용해 가시화(Visualization)한 예이다. 그림에서 나타난 규칙은 앞서 언급한 분류조건을 만족하는 규칙들이다. InferView에서는 3차원 구를 통해 분류조건을 만족하는 데이터의 군집을 표현 했으며 구를 클릭 했을때 나타난 3차원 막대그래프는 분류조건을 만족하는 모든 데이터분포를 나타낸다. Visual Pruning(가지치기) 기능을 이용해서 신뢰성을 가지기에는 군집이 미세한 규칙들을 제거할 수 있다.[7]

(그림 10) InferView를 이용한 가시화(Visualization)

4. 개인화(Personalization)

개인화는 콘텐츠와 서비스를 현재 사용자들과 미래 사용자들에게 선택적으로 전달하는 것이다. 즉, 개인화는 사용자들의 현재와 과거의 관심, 접촉, 행동 등을 통해 개인화된 스타일에 응답하는 능력을 말한다. 많은 인터넷 전자 상거래 회사들이 개인화 솔루션 구축을 위한 노력을 기울이고 있다. 이들은 개인화를 통해서 고객과 좀더 친밀한 관계를 가지려고 노력하며 개인화를 적용한 웹사이트를 개발하려 하고 있다. 지능형 기자재 관리 시스템의 개인화를 위해 앞서 분석한 내용을 토대로 웹 사이트 개인화를 위한 시스템 구성과 데이터마이닝 그리고 사용자마다 개인화된 인터페이스를 제공하는 방안을 제시한다.

4.1 개인화의 활용

온라인에서 개인화를 활용하는 것은 실시간으로 개인화 기법을 적용하며 웹사이트의 사용자에게 새로운 것을 소개하고 사용자들의 잠재적 요구를 통해서 웹사이트의 내용을 변화시키는 것을 말한다. 이는 실시간으로 축적된 데이터의 패턴을 분석함으로써 가능하며 이를 위해 실시간으로 데이터 패턴을 분석하는 지능형 개인화 엔진이 필요하다. 개인화 엔진은 인구통계, 구입된 정보, 등록 데이터, 다양한 유형의 방문객들이 웹사이트와 상호작용으로 얻은 사용자 데이터 등을 활용한 웹 마이닝 기법을 사용한다. 개인화를 통해 제공되는 것을 보면 다음과 같다.

- 사용자의 관심사를 기억해서 제공한다.
- 사용자의 과거의 행동을 기반으로 현재의 행동을 예측해서 그에 맞는 콘텐츠를 보여준다.
- 웹사이트에 처음 방문하는 사용자들을 충성도가 높 은 사용자로 바꿔준다.
- 기자재 활용과 기술교육에 도움을 준다.

가장 잘 알려진 예는 Amazon.com으로 웹사이트에서 고객에게 널리 제품을 판매하기 위해 개인화 기법을 사용했다. 개인화는 판매, 마케팅, 고객 서비스이 세 분야에서 실시간으로 고객 상호작용을 지원한다. 따라서 데이터마이닝 어플리케이션과 시스템의 기능적인 연결은 매우 중요하다.[8]

4.2 개인화 방법

4.2.1 규칙 기반 필터링

개인화 방법 중 가장 기초적이고 일반적인 방법이다. 즉, 초기 사용자에게 신상에 대한 몇 가지 정보를 질문하는 것이 일반적이다. 질문을 통해 얻은 사용자 신상에 관한 정보를 토대로 사용자들을 구분한다. 수집된 인구통계학적 정보와 심리학적 정보를 이용해 사용자 선호도에 알맞은 정보를 제공하는 것을 말한다.

4.2.2 협업 필터링

사용자가 제공한 기초정보와 사용자들의 선호도, 관심 표현 등을 바탕으로 비슷한 선호도를 가진 집단으로 나누어 그 집단 내에서 서로에게 추천하는 방식을 말한다.

4.2.3 학습 에이전트

학습 에이전트(Learning agent)는 사용자의 입력에 의존하지 않고 사용자들의 행동을 기준으로 정보를 수집하게 된다. 특정한 페이지를 보는 시간, 인쇄한 페이지, 대여한 기자재와 대여 예약 목록에 넣은 기자재 등의 로그파일로 저장된 정보를 분석하는 웹 마이닝 기술이 활용된다. 즉 실시간으로 축적된 데이터의 패턴을 고려하여 사용자에게 알맞은 콘텐츠를 전달한다. 여기에 선행되어야 할 것은 웹 페이지의 내용에 대한 정보가 일정한 기준에 따라 분류되어 있어야 한다는 것이다.[9]

4.3 개인화 방안

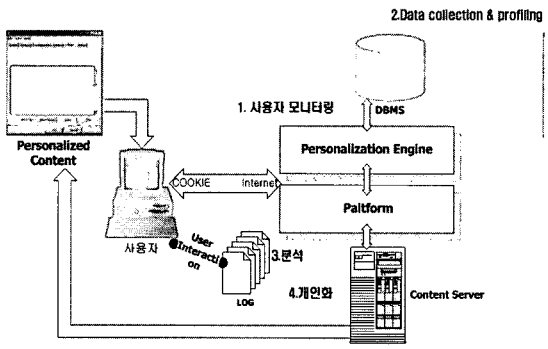
지능형 기자재 관리 시스템의 개인화 방안은 사용자별로 서로 다른 인터페이스를 제공하며 이를 기반으로 한 추천 시스템(Recommendation System)을 구현하기 위한 설계를 제시한다. 예를 들면, 사용자의 신뢰도에 따라 미반납 기자재가 있는 경우 기자재의 사용 및 대여에 관한 메뉴를 없애고 반납 독촉 메시지를 출력한다. 모든 사용자 정보는 데이터베이스에 기록되기 때문에 사용자의 대여정보를 통해서 현재 사용 중인 장비가 있는 경우와 그렇지 않은 경우로 구분하고 사용자별 신용도를 적용하여 간단히 구현할 수 있다. 사용자의 관심도에 따라 사용자가 실제로 관심을 가지거나 잠재적으로 관심을 가질 것으로 예측되는 기자재 목록을 예측하는 추천 시스템을 구현하기 위해서는 앞서 분석된 내용을 활용해야한다. 즉, 사용자 특성에 따라 분류된 관심분야와 데이터베이스에 기록된 실제 사용기록을 토대로 사용자 그룹을 정의한다. 정의된 사용자 그룹별로 관심 기자재를 추천하도록 구현한다. 또 다른 방법은 서버의 로그 파일을 분석하는 것이다. 즉, 이미 분석해서 저장되어 있는 자료를 활용하는 방법이다. 사용자의 클릭을 분석해 사용자별 관심 기자재를 알아낸다.

개인화를 웹 사이트에 통합하기 위해 상위레벨에서 필요한 것들을 보면 다음과 같다.

- 규칙 데이터베이스 : 여기에는 어떤 조건에 어떤 일을 하도록 하는 정보 뿐 아니라 각종 규칙들도 넣는다.
- 이미 정의된 사용자 그룹 : 몇 가지 영역으로 분류한 사용자 그룹을 저장해 둔다.
- 클라이언트 커뮤니케이터(Client Communicator) : 사용자가 하는 일에 따라 적절한 말을 해준다.
- 컴포넌트 아키텍처(Component Architecture) : 서버 프로세싱을 하다 보면 위의 세 가지 방

법을 구현하기 위해 컴포넌트 아키텍쳐가 필요하다.[10]

[그림 11]은 개인화 과정과 개인화를 위한 시스템 구조를 나타낸 것이다.



(그림 11) 개인화 과정 및 시스템 구조

사용자 로그인시 사용자 특성에 따라 메뉴의 구성과 링크를 바꿔주도록 하기 위해서는 홈페이지의 메뉴구조, 디자인요소, 실제 제공되는 콘텐츠 등을 분리해서 관리해야 한다. 즉, 서로 다른 디자인 템플릿을 적용해 동일한 콘텐츠를 여러 메뉴에 제공할 수 있어야 하고 상이한 채널을 통해 출판할 수 있도록 해야 하며 사용성(Usability)을 고려한 인터페이스 관리가 필요하다.

5. 결론 및 향후 연구방향

인터넷 환경의 급속한 변화로 언제 어디서나 사용자들에게 원하는 정보를 원하는 형태로 제공해야 하며, 사용자들의 요구와 특징을 다각도에서 실시간으로 파악하여 그에 맞는 서비스를 제공하는 것은 매우 중요하다.

본 논문에서 제시한 지능형 기자재 관리 시스템의 설계 및 구현 사례는 시스템을 통해서 얻은 기자재 사용 및 대여 기록, 학과 DB에서 얻은 사용자의 세부 정보를 분석했다. 추출된 데이터를 기반으로 의사결정나무 기법을 적용해 몇 가지 중요한 패턴을 발견했으며 이를 통한 개인화 방안을 제시했다. 그러나 사용자의 접속 기록이나 방문페이지 등을 분석하는 웹 마이닝 기법을 사용하지 않았으며 데이터마이닝 기법도 의사결정나무 만을 사용했다. 보다 정확하고 신뢰도 높은 분석을 위해서는 데이터마이닝의 목적이나 데이터의 특징에 맞는 기법들이 선택되어야 하므로 신경망, 클러스터링 등 여러 기법들을 적용하기 위한 노력이 필요하다. 본 논문에서 개인화를 위해 진행된 데이터마이닝 과정을 통해 사용자별 특성, 신용도, 기자재별 관심도 분포를 측정했으며, 추천 시스템 구축을 위해 관심분야 분석 작업으로 개인화 시스템 설계를 제시했다. 이를 토대로 실시간 데이터마이닝을 이용한 개인화 솔루션을 구현하기 위한 연구와 개발이 필요하다.

참고문헌

[1] Lingras. P, "Rough set clustering for Web mining", FUZZ-IEEE Proceedings of the 2002 IEEE International Conference, vol.2, pp.1039 - 1044, 2002

[2] 백성욱, "데이터마이닝 분야의 연구, 개발 및 활용", KOSEN(www,kosen21.org), 2001

[3] 백성욱,"데이터마이닝의 활용에 관한 기술조사", KOSEN(www,kosen21.org), 2000

[4] Alex Berson and Stephen Smith, "Building Data Mining Application for CRM", Kert Therling, 1999

[5] http://dc.sejong.ac.kr, (Id:guest, Password: guest)

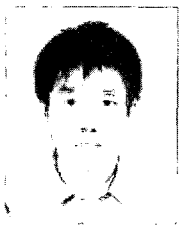
[6] Simoudis. E, "Reality check for data mining" IEEE Vol. 11, issue. 5, pp. 26 - 33, 1996

[7] Bala J.,Baik S.,Gutta S., Hadjarian A.,Mannucci M.,and Pachowicz P, "InferView:An Integrated System For Knowledge Acquisition And Visualization", proceedings of the Fedral Data Mining Symposium & Exposition 99, McLean, Virginia, 1999

[8] "Technology Forecast:2002-2004", vol. 1:Navigating The Future of Software

[9] Magdalini Eirinaki, Michalis Vazirgiannis, "Web mining for web personalization" ACM Transactions on Internet Technology (TOIT), Vol. 3, issue. 1, 2003

[10] Jiawei Han, Chang, K.C.-C, "Data Mining for Web Intelligence", Computer, Vol. 35, Issue. 11, pp. 64 - 70, 2002

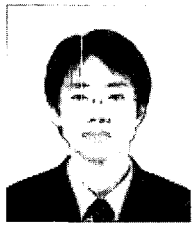


조 영 기

2004 세종대학교 컴퓨터공학과(공학사)

현재 세종대학교 디지털콘텐츠학과 석사과정

관심분야 : 컴퓨터 비젼, 콘텐츠관리시스템 , 데이터마이닝



김 상 수

2004 세종대학교 컴퓨터공학과(공학사)

현재 세종대학교 디지털콘텐츠학과 석사과정

관심분야 : 이미지 프로세싱, 무선 인터넷, 전자상거래

조 주 상

2004 세종대학교 컴퓨터공학과(공학사)

현재 세종대학교 디지털콘텐츠학과 석사과정

관심분야 : 유비쿼터스컴퓨팅, 멀티미디어 데이터베이스, 분산형 데이터마이닝

백 성 욱

1987년 서울대 계산통계학과(학사)

1992년 미국 Northern Illinois University (석사)

1999년 미국 George Mason University (박사)

1997~2002년 미국 워싱턴 DC 소재 Datamat Systems Research Inc.

2002~현재 세종대학교 컴퓨터 공학부 디지털콘텐츠학과 교수 Email: sbaik@sejong.ac.kr