

군집분석(Cluster Analysis)을 활용한 사용자 등급 기반의 서비스 수락 정책

박혜숙[†] · 백두권^{**}

요약

본 논문에서는 마케팅에서 주로 적용되는 CRM(Customer Relationship Management)의 군집분석 방법을 활용하여 콘텐츠 서비스를 이용하는 고객들을 서비스 패턴에 따라 세분화(Segmentation)하고, 군집별로 미디어 서버의 자원을 차별적으로 할당하여, 기업의 수익성을 높이기 위한 새로운 정책을 제시하였다. 새로운 서비스 정책의 구현을 위해 첫째, 고객 데이터에 대해 군집분석 중에서 K Means Method를 적용하여 고객들의 서비스 패턴 (총 사이트 방문 횟수, 서비스 종류, 서비스 이용 기간, 지불금액, 평균 서비스 시간, 사이트 방문 당 서비스 요금)과 수익에 대한 기여도 등을 분석하였다. 고객들은 수익에 대한 기여도에 따라 4개의 군집으로 분류되었다. 둘째, 군집별로 미디어 서버 자원을 할당하는 알고리즘인 CRFA(Client Request Filtering Algorithm)를 제시하였다. CRFA 는 고객이 속한 군집의 자원 한도 내에서 승인제어를 실시하는 기능을 수행하였다. 분석된 결과에 의하면 CRFA 를 적용하였을 때 기업의 네트워크 비용은 감소하고 기업의 수익에 기여도가 높은 군집에 속한 고객들의 수락률이 높아졌음을 알 수 있었다.

키워드 : 멀티미디어 콘텐츠, 승인제어, 군집분석

User-Class based Service Acceptance Policy using Cluster Analysis

Hea-Sook Park[†] · Doo-Kwon Baik^{**}

ABSTRACT

This paper suggests a new policy for consolidating a company's profits by segregating the clients using the contents service and allocating the media server's resources distinctively by clusters using the cluster analysis method of CRM, which is mainly applied to marketing. In this case, CRM refers to the strategy of consolidating a company's profits by efficiently managing the clients, providing them with a more effective, personalized service, and managing the resources more effectively.

For the realization of a new service policy, this paper analyzes the level of contribution vis-à-vis the clients' service pattern (total number of visits to the homepage, service type, service usage period, total payment, average service period, service charge per homepage visit) and profits through the cluster analysis of clients' data applying the K-Means Method. Clients were grouped into 4 clusters according to the contribution level in terms of profits. Likewise, the CRFA (Client Request Filtering Algorithm) was suggested per cluster to allocate media server resources. CRFA issues approval within the resource limit of the cluster where the client belongs. In addition, to evaluate the efficiency of CRFA within the Client/Server environment, the acceptance rate per class was determined, and an evaluation experiment on network traffic was conducted before and after applying CRFA. The results of the experiments showed that the application of CRFA led to the decrease in network expenses and growth of the acceptance rate of clients belonging to the cluster as well as the significant increase in the profits of the company.

Key Words : Multimedia Contents, Admission Control, Cluster Analysis

1. 서론

1.1 연구배경 및 필요성

최근 네트워크 기술의 발전과 인터넷 기술의 발달로 인하여 멀티미디어 콘텐츠 서비스 시장의 규모가 크게 증가하고

있다. 장애 요소이던 디스크 대역폭 문제, 버퍼 용량 문제, 네트워크 대역폭 문제들로 인해 발생하던 서비스 요청의 거절, 끊김 현상, 접속의 불안정 그리고 잦은 버퍼링 등의 현상들이 효과적인 승인제어 기법들과 자원할당 기법들이 제안되고 H/W 기술 및 인터넷 기술의 발전 덕분에 많이 해결되었다. 이에 따라 멀티미디어 콘텐츠를 이용하는 고객이 크게 증가하고 시장의 규모도 크게 증가하면서 기업 간의 경쟁도 심화되고 있다. 이러한 시점에서 기업들은 경

※ 본 연구는 2005년도 경인여대 연구비 지원에 의하여 수행되었음

† 정회원 : 경인여자대학 컴퓨터정보기술학부 조교수

** 종신회원 : 고려대학교 정보통신대학 컴퓨터학과 교수

논문접수 : 2004년 7월 12일, 심사완료 : 2004년 10월 2일

쟁력을 높이기 위해 수익성 높은 콘텐츠를 개발하는 것이 필요하고 또한 안정적인 수익 확보를 위한 시스템 운영 방안을 마련하는 것이 필요하다. 특히 기업은 현재 보유하고 있는 시스템 자원들의 활용도를 높이면서 고객들의 요구(satisfaction, of pleasure, usefulness)와 기업의 요구, 즉, 수익성을 제어하기 위한 방안이 필요하다. 왜냐하면 콘텐츠 서비스 사업은 시스템 구축 및 유지에 많은 비용이 들고 합리적인 자원의 확보를 위한 예측(forecasting)이 어렵기 때문이다. 또한 통신 시장처럼 콘텐츠 서비스 시장이 성숙기에 다다르게 되는 경우에 기업의 수익성은 신규 고객의 확보보다는 기존 고객, 특히 수익에 기여를 많이 하는 우량 고객들을 유지하는 것에 의해 결정되기 때문에 기업의 경쟁력을 높이기 위해서는 시스템과 연계된 고객관계 관리 방안도 필요하다.

본 논문에서는 이러한 이슈들에 대한 해결방안으로서 고객관계관리(CRM : Customer Relationship Management) 기법을 적용한 서비스 정책을 제안하고자 한다. 이 정책의 목표는 마케팅에서 주로 적용되는 CRM을 콘텐츠 서비스 분야의 고객에게 적용하여 고객의 서비스 패턴에 따라 서비스 요구를 제어하고자 하는 것이다.

1.2 연구 목적

본 논문의 목적은 기업의 수익성에 기초하여 고객 세분화를 통해 콘텐츠 서비스의 CRM 기반을 마련하는 것이다. 즉 수익성이 높은 고객 또는 수익성이 높아질 것이라 기대되는 고객에게 미디어 서버 자원을 이용할 기회를 더 많이 부여하고, 또한 모든 고객의 서비스 요구를 제어함으로써 네트워크 트래픽을 제어하는 정책을 제시하고자 한다. 기존의 서비스 정책은 모든 고객에게 똑같은 접속 기회를 제공하고, 접속이 된 후에는 똑같은 품질의 서비스가 제공되므로 고객의 등급에 따른 미디어 서버 자원의 차별화된 서비스는 가능하지 않았기 때문이다. 또한 고객들의 서비스 요청에 따른 네트워크 트래픽의 제어가 가능하지 않아 기업이 지불해야 할 네트워크 비용을 제어하지 못하였다.

본 논문의 목적은 구체적으로 크게 3가지이다.

첫째, CRM의 초기 단계인 고객 세분화 방법을 제시한다. 세분화의 정의와 적용 기준 그리고 세분화 방법들을 살펴보고 본 논문에 적합한 세분화 방법을 제시한다. 세분화를 위해 데이터베이스에 저장된 고객의 데이터를 활용하여 군집 분석을 실시한다. 군집 분석의 결과로는 고객들의 서비스 패턴과 기업에 대한 기여도 등이 산출된다.

둘째, 세분화 결과에 따라 미디어 서버 자원을 할당하고 고객 등급별 수락률을 제어하기 위한 알고리즘으로서 CRFA(Client Request Filtering Algorithm)를 제시한다. CRFA는 고객의 등급을 산출하는 모듈, 미디어 서버 자원의 가용량을 모니터링하는 모듈, 시스템 상태 정보를 모니터링하고 주기적으로 변경하는 모듈 등으로 구성된다.

셋째, 본 논문에서 제시하는 CRFA가 미디어 서버 자원의 활용도를 높이면서도 기업의 수익 증대에 기여할 수 있

음을 보여준다. 이를 위해 콘텐츠 서비스 업체의 고객 데이터를 이용하고 CRFA를 적용하기 전과 후의 등급별 수락률과 재가입률을 비교 평가한다.

1.3 논문의 구성

본 논문의 구성은 다음과 같다. 2장에서는 기존의 차별화된 서비스 제어 정책 및 시장 관점에서의 차별화된 자원 할당 기법들에 대해 살펴보고 관련되어 있는 여러 승인 제어 기법들에 대해 살펴본다. 또한 CRM의 정확한 개념과 여러 기법들에 대해 살펴본다. 3장에서는 CRFA의 구조와 각 구성 요소들의 기능들을 살펴보고 알고리즘에 대해 살펴본다. 4장에서는 CRFA의 성능 평가를 위해 기업의 고객 데이터를 적용하여 실험을 하고 그 결과를 분석한다. 5장에서 결론 및 향후 연구를 제시한다.

2. 관련 연구

본 논문은 차별화된 서비스 제어 정책과 관련되고 세부적으로는 승인제어 분야와 고객관계관리 분야와 관련이 된다. 2장에서는 본 연구와 관련이 있는 분야에 대한 관련 연구를 살펴본다.

2.1 차별화(Differentiated) 제어 모델(Control Model)에 대한 연구

[1]는 현재의 인터넷 기술이 주로 best effort 형태로 제공되고 있기 때문에 비즈니스 측면에서 안정된 서비스 품질과 가용 상태를 유지하면서 전송 대역을 보장하기 위해서는 현재의 상태로는 불가능하다고 지적하였다. 이러한 문제점을 해결하고자 하는 서비스로 차등 서비스(Diffserv : Differentiated Service) 서비스 또는 통합 서비스(Intserv : Integrated Service) 모델이 제안되었지만 실질적인 측면에서 적용되기 어렵다고 지적하였다. 차세대 인터넷은 기존의 기능과 더불어 새로운 비즈니스 모델을 수용하고, 사용자들의 서비스 요구를 제어하는 기술이 요구된다고 하였다. 특히 멀티미디어 콘텐츠 서비스를 위해 네트워크 자원을 효과적으로 이용하는 것은 매우 중요한 문제이고, 서로 다른 서비스 특성(네트워크는 데이터, 음성, 오디오 및 영상/방송 서비스)에 따라 차등적인 서비스 등급(CoS : Class of Service)을 나눌 필요가 있다고 제안하였다.

[2, 3]는 HTTP 프로토콜의 접속 관리 분야에 대한 연구에서 본 논문에서 제안한 차별화 제어 모델과 유사한 정책을 적용하였다. 이때 상위등급과 기본등급을 나누기 위해 히스토리 기반의 정책을 적용하였고 구현을 위해 LRFR(Latest Recently/Frequently Request) 알고리즘을 적용하였다. 위 연구들에서는 웹 서버 상에서의 지연시간 최소화를 위해 사용자 등급을 나누었다.

[4]는 시장관점(market based)에서 시스템의 유틸리티가 최적이 되도록 자원과 QoS 파라미터들이 설정될 수 있도록

하는 가격-형성 프로세스 모델(Price-Formation Process Model)을 제안하였다. 이 모델은 자원 할당 문제를 풀기 위하여 이익 함수와 자원 요구 함수를 모델링하고 이 함수들을 이용하여 최적의 자원 할당 해를 구하고자 하였다. [5]는 인터넷의 트래픽을 제어하는 메커니즘을 제안하였다. 네트워크 서비스 제공자가 유연성있게 여러 서비스 중에서 적합한 서비스를 선택할 수 있도록 하는 방안이다. 또한 서비스별로 가격과 과금 정책 등을 먼저 테스트하여 결과를 측정할 수 있도록 하였다.

[6]에서는 멀티미디어 콘텐츠 스트림 서비스를 이용하는 사용자들의 서비스 품질에 대한 요구를 만족시키기 위해 사용자 실적 또는 등급에 따라 자원을 배분하고 서비스 요청과 서비스 품질을 제어하기 위한 방안으로서 사용자 등급별 QoS 보장을 위한 서비스 요청 필터링 알고리즘(URFA: User Request Filtering Algorithm)을 제안한다. URFA는 기여도 및 등급 산출을 위해 지불정보, 서비스 시간 정보, 서비스 요청 빈도 정보를 이용하였다. 회원의 등급은 2 등급으로 제한하여 고객의 기여도를 계산하는 함수를 이용하였다.

본 논문은 기업에 대한 기여도를 나타내는 서비스 패턴을 임의로 정하는 것이 아니라 CRM 기반의 군집분석 도구를 이용하여 고객의 데이터를 분석하고 의미있는 서비스 패턴들과 등급의 수 등을 도출하였다.

2.2 CRM(Customer Relationship Management) & E-CRM (Electronic-Customer Relationship Management)

기업의 서비스 정책에 적용된 CRM과 E-CRM의 정확한 개념을 살펴보고, 서비스 분야에서 CRM 및 E-CRM이 적용된 사례들을 살펴봄으로써 본 논문에 적용할 수 있는 E-CRM 정책은 어떤 것인지 살펴본다.

2.2.1 CRM과 E-CRM

CRM이란 고객들에게 더 좋은 서비스를 제공하고 고객들과의 상호작용을 효율적으로 관리함으로써 기업이 그들의 가장 바람직한 고객에게 좀더 효과적인 차별화(개별적 서비스)를 제시하고, 자원을 효과적으로 배치하도록 하여 기업의 수익을 증대시키기 위한 전략이다. 즉 (1) The right offer (2) To the right person (3) At the right time (4) Through the right channel이라 할 수 있다[7]. E-CRM이란 인터넷에서 상품을 구매하는 고객이나 서비스를 구매하는 고객을 대상으로 인터넷 기반 기술을 이용하여 신규 고객의 획득, 기존 고객의 유지 등을 통하여 기업의 수익성을 최적화 하고자 하는 전략이다. 이미 제조업체, 의료서비스 업체 그리고 통신회사 등은 고객의 이탈 방지와 수익 증대를 위해 CRM과 E-CRM을 기업 운영에 적극 적용하고 있다. 본 논문과 관련된 분야는 E-CRM이 적용 가능한 분야로서, 고객들의 데이터베이스를 활용하고, 고객 행동 분석을 위한 적절한 세분화 기법을 적용하면 어떤 고객 층이 수익에 큰 영향을 줄 것인지를 파악할 수 있다. 또한 고객에 대한 정보와 지

식을 바탕으로 고객 개개인이 필요로 하는 맞춤 서비스도 가능하여 고객의 만족도를 높일 수도 있다.

2.2.2 고객 세분화(Client Segmentation)와 군집분석(Cluster Analysis)

고객 세분화[8]는 많은 고객들을 서로 비슷한 소비자들끼리 분류하는 작업을 말한다. 고객의 성별 또는 나이, 직업 등에 따라 나누는 것도 세분화 작업이다. 세분화는 전체 데이터베이스를 바라볼 수 있도록 해주거나, 기업이 서로 다른 등급에 있는 고객들을 서로 다르게 대우할 수 있도록 해 주고 등급에 맞는 적절한 마케팅이 가능하게 해준다. 데이터 기반의 세분화는 다양한 통계적 기법과 데이터마이닝[9]을 사용하여 수행되어 지는데 대부분은 의사결정 나무(Decision Tree)와 군집분석 그리고 신경망(Neural Network)을 이용한 기법 등 세 개로 분류된다.

본 논문에서는 군집분석을 적용하고자 한다. 본 논문의 대상이 되는 고객 데이터가 비계층적 구조이고 대량의 고객 데이터를 빠른 속도로 세분화하기에는 비계층적 군집분석 방법이 적합하기 때문이다. 비계층적 군집분석[10]은 개체를 K개 군집으로 세분화하는데 이용되는데, 이때 군집의 수는 미리 규정되거나 군집화 절차의 한 부분으로 결정될 수 있다. 군집화 단계마다 거리(유사성) 행렬을 다시 구하거나 또는 컴퓨터 작업 수행 중에 기본 자료를 저장할 필요가 없기 때문에 비계층적 군집분석은 대규모 자료에 적합하다. 본 논문에서는 군집분석 방법 중에서 가장 많이 사용되는 K-Means-Method을 적용하고자 한다. K-Means-method[9]는 가장 가까운 중심점을 갖는 군집에 각 항목을 할당하는 과정을 반복하여 K개의 군집으로 항목들을 나누는 것이다. K-Means-Method의 알고리즘은 크게 3 단계로 구성된다. 첫번째 단계, 항목들은 K 개의 초기 군집들로 분할한다. 두 번째 단계, 각 항목을 가장 가까운 중심점을 갖는 군집에 할당한다. 표준화 또는 비표준화된 유클리드 거리를 주로 사용한다. 새로 할당된 항목을 군집과 그 항목이 빠진 군집에 대해 군집 중심점을 다시 계산한다. 세 번째 단계, 두 번째 단계를 반복하여 더 이상 할당이 이루어지지 않을 때까지 계속한다.

2.2.3 사례연구

[11, 12]는 무선 통신 분야에서 CRM과 E-CRM 적용을 위해 고객들의 사용 내역을 분석하여 고객 세분화를 실시하는 방법을 제시하였다. 분석 대상의 회사는 국내 통신 업체 중에서 가장 높은 매출을 보이는 회사이다. 이 회사는 고객의 이탈 방지 및 수익 강화를 위한 마케팅 전략을 세우기 위해 통신 서비스 분야에 CRM을 적용하였다. 세분화에 의해 고객이 4개의 등급(최우량고객, 우량고객, 기본고객, 이탈고객)으로 분류되었고, 수익에 영향을 끼치는 주요 변수들을 분석되었다. 등급에 따라 고객들을 분류하고 차별화된 마케팅을 하고있다. 이동 통신서비스는 본 논문에서 관심을 갖

는 분야와 유사한 분야이기 때문에 사례 연구를 통해 [11, 12]에서 적용한 세분화 방법론을 본 논문에도 적용하고자 한다.

[13]는 의료 서비스 분야에서 CRM적용을 위해 외래 환자의 진료내역 데이터 분석을 통한 세분화 방법을 제시하였다. 고객의 병원 이용 빈도와 진료 수익성에 기초하여 군집 분석 기법을 적용하여 분석하였다. 연구 결과 4개의 환자군(우량환자, 잠재우량환자, 잠재이탈환자, 이탈환자)로 세분화하였고 진료 수익관점에서 영향을 끼치는 주요 변수들을 도출하였다.

3. CRFA(Client Request Filtering Algorithm)

3.1 CRFA

본 논문에서 제안한 CRFA는 웹 서버의 안에서 동작하는 알고리즘으로서 접속하는 고객들의 수익에 대한 기여도(Contribution Value)를 계산하고 이에 맞는 등급을 산출하는 부분과 등급에 따라 미디어 서버의 자원을 할당하는 부분으로 크게 나누어진다. CRFA는 미디어 서버에서 동작하는 승인제어와는 다르게 웹 서버 상에서 작동하도록 설계되어 구현이 용이하다. (그림 1)은 CRFA의 구현을 위한 시스템의 구조를 보여준다.

(그림 1)의 Client Certification은 웹 서버에 접속하여 서비스를 요청하는 고객에 대한 기본적인 인증 과정을 수행하는 모듈이다. 인증을 거친 고객의 서비스 요청만이 CRFA 과정을 수행하게 된다. CRFA는 Client Segmentation Manager, Client Request Filter, Resource Monitoring Manager, Client Information Manager 그리고 Service Provider로 구성된다. 각 구성요소의 주요 기능은 다음과 같다.

3.1.1 Client Segmentation Manager

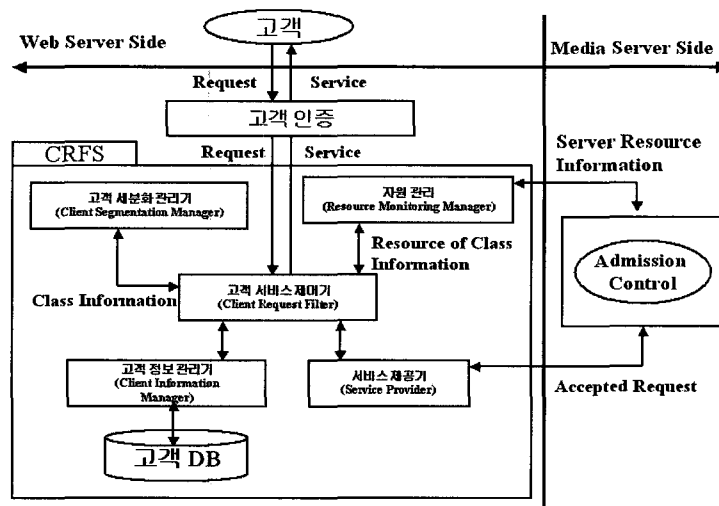
주기적으로 고객 데이터베이스에 저장된 데이터를 이용하여 세분화를 실시한다. 분석 기간동안의 고객의 서비스 빈

도, 서비스 이용기간, 금액 등의 데이터를 활용한다. 세분화 결과로써 등급의 수와 종류, 등급별 고객 비율 등을 산출한다. 이러한 정보들은 Client Request Filter를 통해 Resource Monitoring Manager로 전송된다.

3.1.2 Client Request Filter

고객으로부터 서비스 요구가 도착하면 고객의 데이터를 분석하여 서비스 허용 여부를 결정한다. 이때의 과정은 다음과 같다.

- Step 1 : Client Information Manager에게 고객의 기여도 산출을 요구한다.
- Step 2 : Resource Monitoring Manager에게 고객이 속한 등급의 미디어 서버의 자원 상태 정보를 요구한다.
- Step 3 : Client Request Filter는 step 1과 step 2에서 얻은 정보를 종합하여 고객의 서비스 요구에 대한 허용 여부를 결정한다.
- Step 4 : 만약 서비스가 허용되면 이 결과는 Service Provider에게 전송되고 미디어 서버로 고객의 서비스 요구가 전달되어 서비스가 개시된다.
- Step 5 : 서비스가 개시되면 Client Request Filter는 Resource Monitoring Manager에게 자원 상태 정보의 갱신을 요구한다.
- Step 6 : 서비스가 완료되면 Client Request Filter는 Client Information Manager에게 고객의 정보 갱신을 요구하고 Resource Monitoring Manager에게 자원 상태 정보의 갱신을 다시 요구한다. 만약 고객의 서비스 요청이 거절되는 경우에는 일정 간격으로 다시 위와 같은 프로세스를 반복한다. 이때 일정시간이 지난 후에도 서비스 요청이 거절되는 경우 고객에게 메시지를 전달하고 프로세스를 종료한다. 만약 일정 시간 안에 서비스 허용이 된다면 Step 5부터 과정을 수행한다.



(그림 1) CRFS의 구조

3.1.3 Client Information Manager

Client Request Filter의 요구에 의해 서비스를 요청한 고객의 기여도를 산출한다. 기여도 산출을 위해 사이트 방문 시간 동안 평균 서비스 시간, 가입기간 동안 평균 서비스 시간, 고객의 매출이 전체 매출에서 차지하는 비율 등을 사용한다. 산출된 기여도는 Client Request Filter에게 전달된다.

3.1.4 Resource Monitoring Manager

Resource Monitoring Manager 는 미디어 서버 자원 상태를 모니터링하고 Client Request Filter의 요구에 대해 필요한 정보를 전달한다. 새로운 서비스가 개시되거나 서비스가 종료되면 미디어 서버의 자원 정보에 대한 갱신작업을 수행한다. 주기적으로 Client Segmentation Manager로부터 전달된 등급 관련 정보를 이용하여 등급별로 자원을 할당한다.

3.1.5 Service Provider

Service Provider는 Client Request Filter로부터 서비스 개시 요구가 도착하면 미디어 서버에게 서비스 요구를 전달하여 서비스가 개시되도록 한다. 또한 서비스 시작과 종료 때마다 고객의 정보를 Client Request Filter로 전달한다.

3.2 고객 세분화 (Client Segmentation)

Client Segmentation Manager는 세분화를 위해 군집분석 방법인 K-Means Method 를 적용하여 데이터베이스에 저장된 고객 데이터를 분석한다. 고객의 일반적 특성(이름, 성별, 거주지), 서비스 특성(서비스 종류, 가입일, 관심종목) 그리고 서비스 이용 패턴(총 사이트 방문 횟수, 서비스 종류, 서비스 이용 기간, 총 지불금액, 평균 서비스 시간, 사이트 방문 당 서비스 요금)을 분석한다. 서비스 이용 패턴 분석을 위해 DB마케팅에서 세분화 기준으로 많이 사용되고 있는 RFM(Recently, Frequency; Monetary)를 적용하였는데, 총 사이트 방문 횟수 및 서비스 받은 간격 등은 사용빈도(Frequency)를 나타내고, 처음 서비스 받은 날로부터 최종으로 서비스 받은 날까지의 기간은 최근성(Recently), 평균 서비스 금액과 가입기간 동안의 평균 서비스 금액은 수익성(Monetary)를 나타내는 지표로 삼았다. 분석 결과는 등급의 수와 종류, 등급별 고객의 수와 비율, 등급별 전체 수익에서 차지하는 비율 등이다. 추출된 결과 값들은 Client Request Filter와 Resource Monitoring Manager에게 전달되어 미디어 자원의 할당을 위해 사용된다.

3.2.1 등급별 미디어 서버의 자원 할당

Resource Monitoring Manager는 세분화 결과를 이용하여 등급에 따라 미디어 서버 자원을 할당한다. 이때 자원은 미디어 서버가 허용할 수 있는 스트림 수가 되고, 등급별 전체 수익에서 차지하는 비율에 따라 할당된다. 미디어 서버가 허용할 수 있는 스트림 수를 MaxVal 이라 하고, 등급 i 의 수익률 P_i 는 로 표현한다면 등급 i 에 할당되는 자원을 나타내는 $Q_i = MaxVal * P_i$ 로 표현할 수 있다.

3.2.2 고객의 기여도(Contribution Value) 산출

Client Information Manager는 고객이 서비스 요청을 하였을 때 고객이 어떤 등급에 속하는 지를 결정하기 위하여 기여도를 계산한다. 본 논문에서는 기여도 계산을 위해 참고문헌[12]에서 적용한 방법을 적용하였다. Table 1은 기여도 산출을 위해 필요한 변수들과 변수들의 의미를 나타낸 것이다.

- $SALE_i$: 변수들 중에서 가장 중요한 변수로서 고객의 매출액을 점수로 환산한 값을 나타낸다. 매출액이 많을 수록 고객이 기업의 수익에 기여하는 바가 크다는 점을 반영하였다.
- $Period_i$: 고객의 가입기간을 일(day) 단위로 하여 점수로 환산한 것이다. 장기 가입 고객을 확보하기 위한 차원에서 가입기간이 길수록 높은 점수를 받도록 하였다.
- EST_i : 고객이 방문할 때마다 서비스받은 시간을 반영하는 값으로써 서비스를 오래 받을수록 우량 고객일 확률이 높고 향후 수익을 많이 낼 수 있는 고객으로 판단할 수 있으므로 높은 점수를 받도록 하였다.
- $ESTT_i$: 고객이 가입기간 동안 얼마나 많은 서비스를 받았는지를 반영하는 값으로써 서비스를 많이 받은 고객일수록 앞으로도 계속적으로 서비스를 받을 확률이 높은 고객으로 분석되었기 때문에 높은 점수를 받도록 하였다.

4가지 변수들을 종합해보면 가입기간이 길고 자주 그리고 많이 서비스를 받은 고객은 기업의 수익성 측면에서 가장 우수한 고객이 되는 것이다. 반대로 가입기간은 오래되었지만 거의 서비스를 받지 않거나 짧은 시간 서비스를 받은 고객들은 낮은 등급의 고객에 속하게 된다. 그리고 최근에 회원 가입을 하고 서비스를 많이 받지 않은 고객은 4가지 변수들 중에서 $SALE_i$ 와 $Period_i$ 는 값이 작지만 EST_i 와 $ESTT_i$ 는 가입기간만 길고 거의 서비스를 받지 않은 고객에 비해 높은 값을 가질 수 있으므로 중간 등급에 속하게 된다. 그러므로 4가지 변수들은 서로 독립적인 관계라고 판단할 수 있다.

<표 1> 고객 기여도 산출을 위한 속성

Attributes	Descriptions
$SALE_i$	고객 i 의 매출액, 10000원을 1점으로 환산한다.
$Period_i$	고객 i 의 가입기간, 일(day)일을 1점으로 환산한다.
EST_i	고객 i 의 1회 방문했을 때 서비스 받은 시간, 1분당 1점으로 환산한다.
$ESTT_i$	고객 i 의 가입 기간동안 서비스 받은 시간, 1분당 1점으로 환산한다.

3.3 CRFA(Clients Request Filtering Algorithm)

<표 2>은 CRFA의 핵심적인 부분만을 간략하게 표현한 것이다. 이 알고리즘은 고객의 등급을 3개로 분류한다고 가

〈표 2〉 CRFA (Clients Request Filtering Algorithm)

```

(1-1) while ()
(1-2) For (j = 1; j <= 3; j++)  $Q_j = MaxVal * P_j$ ;
      Level_Value_1 = Find_LowerBound_First( );
      Level_Value_2 = Find_LowerBound_Second( );
      Level_Value_3 = Find_LowerBound_Third( );
(1-3)  $EST_i = Scoring\_EST(i)$ ;  $ESTT_i = Scoring\_ESTT(i)$ ;
(1-4)  $SALE_i = Scoring\_SALE(i)$ ;  $Period_i = Scoring\_Period(i)$ ;
(1-5)  $CV_i = DET(EST_i, ESTT_i, SALE_i, Period_i)$ 
(1-6) If Level_Value_1 <=  $CV_i$  <= 100.0 Then
      class_value = 1;
(1-7) If sReqNum_First_Class <  $Q_1$  Then
      Allocate_Resource(i, class_value); Calculate_Acceptance_Rate_Class(class_value);
      Update_System_Resource_Parameter(class_value); Update_Client_Information(i, class_value);
      sReqNum_First_Class++;
(1-8) Else
      Reject(i); Update_Client_Information(i, class_value);
(1-9) Else if Level_Value_2 <=  $CV_i$  < Level_Value_1 Then
      class_value = 2;
(1-10) If sReqNum_Second_Class <  $Q_2$  Then
      Allocate_Resource(i, class_value); Calculate_Acceptance_Rate_Class(class_value);
      Update_System_Resource_Parameter(class_value); Update_Client_Information(i, class_value);
      sReqNum_Second_Class++;
(1-11) Else
      Reject(i); Update_Client_Information(i, class_value);
(1-12) Else
      class_value = 3;
(1-13) If sReqNum_Third_Class <  $Q_3$  Then
      Allocate_Resource(i, class_value); Calculate_Acceptance_Rate_Class(class_value);
      Update_System_Resource_Parameter(class_value); Update_Client_Information(i, class_value);
      sReqNum_Third_Class++;
(1-14) Else
      Reject(i); Update_Client_Information(i, class_value);
// while end

```

정하고 등급 결정 과정을 나타낸 것으로서 실제로 등급이 N 등급으로 나누어 진다면 N 등급에 맞게 수정되는 것이다.

식 (1-2)에서 MaxVal은 미디어 서버의 허용 가능한 최대 스트림 수를 의미하는 것이고, P_j 는 j 등급이 전체 수익에서 차지하는 비율을 의미한다. j 등급이 할당 받을 수 있는 자원의 수를 Q_j 로 나타내면 이 값은 P_j 와 MaxVal의 곱으로 표현이 된다. Level_Value_1는 First Class에 속하기 위한 CV_i 의 하한값을 나타내는 것으로서 4.2 절의 세분화 결과에 의해 결정된 First Class속한 고객들의 CV_i 값 중에서 가장 낮은 값으로 한다. Find_LowerBound_First()는 Level_Value_1 값을 결정하기 위한 함수이다. 같은 방법으로 Level_Value_2, Level_Value_3를 결정한다.

식 (1-3)의 Scoring_EST(i)는 고객 i가 사이트 방문 때마다 서비스 받은 시간을 이용하여 사이트 방문 당 평균 서비스 받은 시간을 점수로 환산하는 함수이다. 환산된 결과값이 EST_i 로 표현된다.

식 (1-3)의 Scoring_ESTT(i)는 가입기간동안 서비스 받은 시간을 점수로 환산하는 함수로써 결과값은 $ESTT_i$ 로 표현된다.

식 (1-4)의 Scoring_SALE(i)는 고객 i의 매출액을 점수로 환산한 것으로 본 알고리즘에서는 국내 화폐 단위를 사용하여 10,000원을 1점으로 환산하였다. 환산된 결과값이

SALE_i로 표현된다. 식 (1-4)의 Scoring_Period(i)는 고객 i의 가입 기간을 점수로 환산한 것으로서, 1일을 1점으로 환산한 것이다. 환산된 결과값이 Period_i로 표현된다.

식 (1-5)의 DET()는 고객 i의 기여도를 산출하기 위한 식으로 (1-3)과 (1-4)에서 구한 값들을 인자로 하여 DET() 함수에 의해 100.0부터 0.0까지의 값으로 환산한 것이다. 예를 들면 First Class 속할 수 있는 값이 100.0부터 180.0 사이라고 가정했을 때 만약 고객 i의 CV_i 값이 90.0 이라면 고객 i는 First Class에 속한다고 판단하도록 하는 것이다.

식 (1-6)의 class_value는 고객이 속한 등급을 나타내는 변수이다. 식 (1-6)부터 식 (1-8)는 만약 고객 i가 First Class(최우량고객, 즉, 가장 수익을 많이 내는 등급)에 속한다고 판단되면 할당 가능한 자원이 있는지를 확인하여 자원을 할당한다. 이때 남은 자원이 없다면 고객의 요청을 거절하고 각종 고객 정보 등을 변경한다. 알고리즘 설계 과정 중에 최고 등급에 속한 고객의 요청을 충분히 허용할 수 있도록 많은 자원을 할당하도록 하였다. 이러한 경우에는 자원을 미리 제약(reserved)함으로써 자원 낭비의 위험 요소가 존재하지만 이 등급에 속한 고객들은 기업의 수익에 가장 큰 기여를 하는 고객이므로 안전한 정책을 적용하는 것이다. 식 (1-7)의 sReqNum_First_Class는 서비스 중인 고객들의 수를 의미하는 것으로 이 값이 Q_1 보다 작은 경우에만 자원

할당이 가능하게 된다. 서비스가 종료하게 되면 할당된 자원이 회수되면서 *sReqNum_First_Class* 값도 변경되는 것이다. *Allocate_Resource(i, class_value)* 는 자원을 할당하도록 하는 함수이다. 또한 *Calculate_Acceptance_Rate_Class(class_value)* 는 등급에 따른 서비스 수락률을 계산하는 함수이다. *Update_System_Resource_Parameter(class_value)* 는 등급에 따른 시스템 자원 관련 값을 변경시키는 함수이다. *Update_Client_Information(i, class_value)* 는 고객의 정보를 변경하는 함수이다. 이러한 함수들은 모두 데이터베이스와 연동이 되는 함수들이다. 식 (1-8)의 *Reject()* 함수는 허용 가능한 자원이 없을 경우에 고객의 요청을 거절하도록 하는 함수이다.

식(1-9) 부터 식(1-11) 그리고 식(1-12) 부터 식(1-14)도 각각 위와 같은 방식으로 수행된다. 이 알고리즘에는 표현되어 있지 않지만 이 이외에도 서비스가 종료된 경우에는 자원을 회수하고 고객의 정보를 변경하는 부분과 시스템의 자원 상태를 수정하는 부분 등이 포함되어 있다.

4. 실험 및 결과 분석

4.1 실험 설계

본 논문에서 제시한 CRFA의 성능을 검증하기 위해 클라이언트/서버 환경 하에서 CRFA 적용 전과 후의 등급별 수락률과 네트워크 트래픽을 비교 평가하고자 한다. 등급별 수락률을 평가하고자 하는 이유는 CRFA를 적용하기 전과 후에 수락률이 어떻게 변했는지를 분석하고 수익을 많이 내는 등급의 수락률이 다른 등급에 비해서 더 높은 수락률을 확보하는지를 확인하기 위해서이다. 이 실험은 특정 등급의 수락률을 조절하기 위하여 자원 할당 정책을 변경할 수 있도록 하는 실험이기도 하다. 예를 들면 가장 높은 등급의 수락률을 100%로 하고자 한다면 기존의 자원 할당 정책이 어떻게 변경해야 하는지를 파악할 수 있게 하는 실험이다.

4.2 데이터 수집 및 세분화

알고리즘을 적용하기 위한 대상은 콘텐츠 서비스 회사인 *iteaching(www.iteaching.co.kr)*이다. 이 회사는 교육용 콘텐츠를 서비스 하는 회사로써 년 매출이 약10억원 년 회원수가 약 3만 정도인 회사이다. 고객 데이터 중에서 2003년의 고객 5000 명의 데이터를 샘플로 하여 고객의 데이터를 분

석하였다. 데이터를 분석하기 위하여 군집분석 도구(TOOL)인 SAS사의 Enterprise Miner 4.0을 이용하였다. section 2.2.3의 사례 연구를 분석하여 초기 군집의 수를 4로 정하였다. 세분화를 위하여 1년 동안의 고객 데이터 중에서 가입 기간, 총 사이트 방문 횟수, 서비스 종류, 서비스 이용 기간, 요금, 평균 서비스 시간, 평균 서비스 요금 등에 대한 분석을 실시하여 <표 3>와 같은 결과를 얻을 수 있었다.

세분화된 고객의 특성을 살펴보면 다음과 같다. First Class는 수익성도 높으면서 서비스를 지속적으로 이용하고 있는 등급이다. 전체 고객 중에 차지하는 비율이 약 9.8%이고 수익률은 전체 중에서 약 48.7%를 차지한다. 이들 고객은 기업 입장에서 볼 때 가장 적극적으로 서비스를 이용하는 등급으로 판단되며, 집중적인 관리가 필요한 등급으로 판단되므로 미디어 서버의 자원을 가장 많이 할당하여 항상 높은 수락률을 보장해야 하는 등급이다. Second Class는 First Class보다는 수익률 면에서 작지만 지속적으로 서비스를 받는 집단으로 판단되므로 관리를 잘하면 언제든지 First Class에 속할 확률이 높은 등급으로 판단할 수 있다. 수익률도 약 30.8%를 차지함으로써 수익에 많은 기여를 하는 집단이다. 위의 두 등급은 기업 입장에서 볼 때 전체 수익의 약 80%를 차지하기 때문에 하나의 등급으로 볼 수도 있다. Third Class는 4개의 등급 중에서 기본 등급으로 분류할 수 있다. 이 등급은 전체 고객 중에서 가장 많은 고객이 속한 등급으로서 가입기간이 길지 않지만 서비스를 자주 이용하는 고객들이 속하는 것으로 판단된다. 회원가입 후에 처음으로 서비스를 받는 고객들이 여기에 속하는 것으로 분석되었다. Fourth Class는 가입기간이 기본등급보다 길지만 서비스를 거의 이용하지 않거나 서비스 받은 시간도 짧은 고객들이 여기에 속한다. 이러한 고객들은 다른 회사의 서비스를 이용하는 것으로 추정되거나 다른 사이트로 이탈할 확률이 매우 높은 고객으로 분석되므로 자원을 할당할 때에도 많은 자원이 할당되지 않도록 해야 한다.

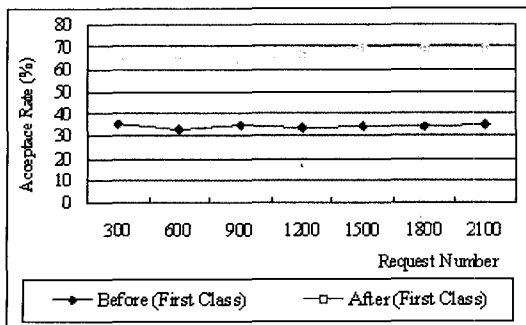
4.3 성능 평가 실험

4.3.1 실험 1 : CRFA 적용 전과 후의 등급별 수락률 평가
실험 1에서는 수락률을 구하기 위하여 식(2)을 적용한다.

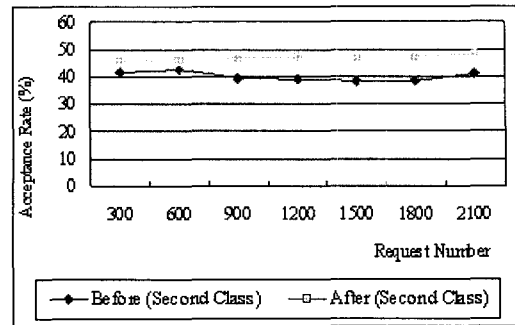
$$Acceptance_Rate(i) = \frac{number_of_acceptance_of_class(j)}{total_number_of_requests} \quad (2)$$

<표 3> 고객 세분화 결과

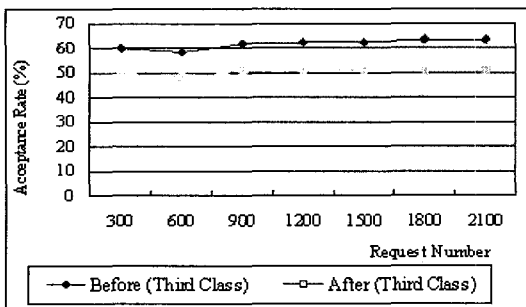
	First Class(최우량고객)	Second Class(우량고객)	Third Class(기본고객)	Fourth Class(이탈고객)
총 사이트 방문 횟수(회)	338.4	228.5	105.4	52.5
이용기간(일)	320.5	202.3	120.8	170.5
이용기간동안 평균 서비스 시간(일)	196.8	120.3	74.0	38.7
이용기간동안 평균 서비스 금액(원)	787,800	534,900	274,600	113,000
고객 수 (%)	9.8	15.3	46.5	28.4
수익률 (%)	48.7	30.8	15.5	5.0



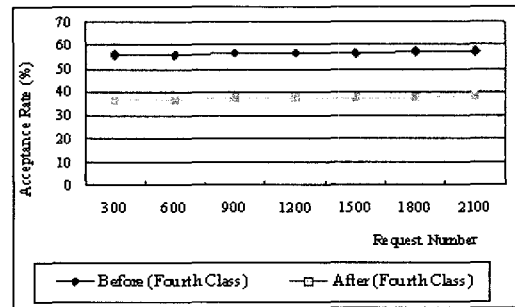
(a) First Class



(b) Second Class



(c) Third Class



(d) Fourth Class

(그림 2) CRFA 적용 전과 후의 수락률 비교

식(2)는 회원들의 총 서비스 요구에 대해 등급에 따른 수락률을 구하는 식이다. 이때 데이터 패킷의 전송속도인 TransRate = 300Kbps로, 동시접속이 가능한 자원의 양 즉, MaxVal = 1000으로 하였다. 고객들이 요구하는 서비스는 콘텐츠의 내용만 다를 뿐 다른 모든 조건은 동일하다고 가정하였다. Table4의 등급별 수락률에 기초하여 자원을 할당하면 $Q_1 = 487, Q_2 = 308, Q_3 = 155, Q_4 = 50$ 가 된다.

(그림 2)은 실험 1의 결과를 그래프로 나타낸 것이다. 클라이언트/서버 환경 하에서의 CRFA를 적용하기 전의 등급별(First, Second, Third, Fourth Class) 수락률과 CRFA를 적용한 후의 등급별 수락률(First, Second, Third, Fourth Class)을 측정하였다. 결과를 분석하면 다음과 같은 결과를 얻을 수 있다. 첫째, (그림 2)의 (a)는 First Class에 속한 고객들의 수락률을 비교한 것이다.

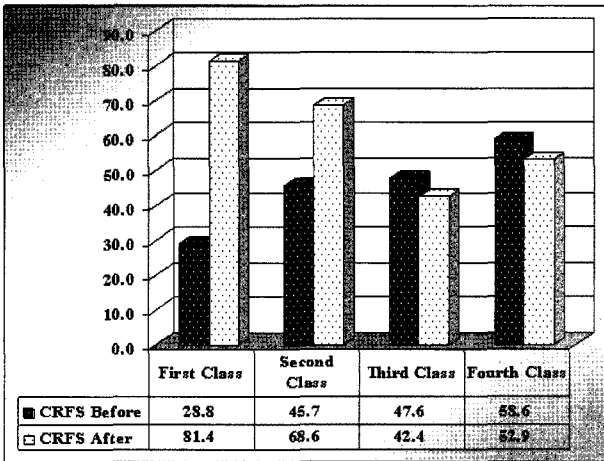
CRFA를 적용하기 전의 First Class에 속한 고객의 평균 수락률이 34.16 %를 나타내었지만 CRFA를 적용하였을 때 약 32.55% 정도 증가하여 66.71 %를 나타내었다. 가장 많은 자원이 할당되었기 때문이라고 분석된다. 둘째, (그림 2)의 (b)는 Second Class에 속한 고객들의 수락률을 비교한 것이다. Second Class에 속한 고객의 수락률이 40.06%에서 46.93%로 증가하였다. 기존보다 더 많은 자원이 할당되었기 때문이다. 셋째, (그림 2)의 (c)는 Third Class에 속한 고객들의 수락률을 비교한 것이다. 가장 많은 고객이 속해있는 등급인 Third Class 경우에는 CRFA를 적용하기 전의 수락률이 61.54%를 나타내었지만 CRFA를 적용하였을 때는 50.26%로 감소하였다. 수락률이 감소한 이유는 기존보다 더 적은 자원 할당이 이루어졌기 때문이지만 신규 고객들이 이

등급에 속하기 때문에 고객 관리 차원에서 약 50% 정도의 수락률을 유지할 수 있도록 할당하였기 때문이다. 넷째, (그림 2)의 (d)는 Fourth Class에 속한 고객들의 수락률을 비교한 것이다. 수락률이 가장 작은 등급의 경우에는 기존에 할당되었던 자원보다 더 적은 자원을 할당하게 되므로 수락률이 56.66 %에서 37.44 %로 감소하였다. 이 등급에 속한 고객들은 이탈 예상 고객들이기 때문에 가장 적은 자원을 할당하기 때문이다. 이 실험을 통해 CRFA를 이용하면 등급별 수락률을 제어할 수 있다는 것을 알 수 있었다.

4.3.2 Experiment 2 : 고객의 재가입률 평가

본 논문에서 제안한 방법이 적용된 기업들의 요금 적용 정책은 선불 방식으로서 일정기간동안 서비스를 이용하고자 미리 요금을 지불하고 서비스를 이용하는 것이므로 전체적인 네트워크 트래픽의 감소가 기업 이익의 감소로 이어진다고 판단하기는 어렵다. 반면에 과금 기간 동안 발생한 트래픽 중에서 최대 트래픽에 대해 비용을 지불해야 하는 경우에는 이 비용이 감소하므로 기업이 지불해야 하는 비용이 감소하는 것을 알 수 있다. 실험 4에서는 전체 네트워크 트래픽은 줄었지만 기업의 이익이 증가하는 것을 보이고자 우수 등급에 속한 고객의 재가입률을 측정하는 실험을 하였다. (그림 3)은 사용 기간이 만료된 고객들(각 190명, 187명)을 대상으로 20일간의 재가입률을 평가한 것이다. 실험 결과로부터 우수고객의 만족도가 높아졌다는 것을 알 수 있고 매출의 80%를 차지하는 고객들의 재가입률이 높아져서 기업의 매출도 증가하였음을 알 수 있었다. 증가한 매출액은 고객별 매출을 합산하여 계산할 수 있었다.

	CRFS Before		CRFS After	
	Number of Clients	재가입 고객의 수	Number of Clients	재가입 고객의 수
First Class	52	15	43	35
Second Class	46	21	51	35
Third Class	63	30	59	25
Fourth Class	29	17	34	18



(그림 3) 재가입률의 비교 평가

5. 결론과 향후 연구

본 논문에서는 기업의 수익성 기초하여 고객 세분화를 통해 멀티미디어 콘텐츠 서비스의 CRM 정책을 제안하고 이를 위해 고객의 서비스 요청을 제어하는 알고리즘을 제안하였다. 즉 수익성이 높은 고객 고객에게 미디어 서버 자원을 이용할 기회를 더 많이 부여하고, 또한 모든 고객의 서비스 요구를 제어함으로써 네트워크 자원의 사용을 제어할 수 있는 정책을 제시하였다. 알고리즘 구현을 위해 첫째, 고객의 데이터를 분석하기 위하여 군집분석을 적용하였다. 분석결과에 의해 대상이 되는 고객들을 4개의 집단(First Class, Second Class, Third Class, Fourth Class)으로 분류할 수 있었다. 둘째, 알고리즘을 구성하는 각 구성 요소들의 기능을 모델링하고 구현하였다. CRFA은 Client Segmentation Manger, Client Request Filter, Resource Monitoring Manager, Client Information Manager 그리고 Service Provider로 구성된다. 셋째, 알고리즘의 성능 평가를 위해 두 가지 실험을 실시하였다. 첫번째 실험에서는 알고리즘 적용 전과 후의 등급별 수락률을 비교 평가하였다. 기업의 수익에 중요한 영향을 끼치는 First Class와 Second Class의 수락률이 크게 향상됨을 알 수 있었다. 두번째 실험에서는 고객의 재가입률을 비교 평가하였다. 새로운 알고리즘을 적용하였을 경우에 재가입률이 높아져 기업의 매출도 증가하였다. 향후 연구과제로는 기업의 이윤과 등급별 수락률 사이의 관계를 모델링하여 기업의 이익을 극대화 하기 위해 등급별 최적의 수락률을 구

하는 과정에 대한 연구가 필요하다.

참 고 문 헌

- [1] Jun-kyun Choi, "Technology of the Assurance of QoS for Next Generation Network", Journal of KISS, Vol.21, No.8, pp.51-66, 2003. 08.
- [2] E. S. Hyun, Y. J. Rhee, and T. Y. Kim, "Differentiated-HTTP for Differentiated Web Service", Journal of KISS, Vol.28, No.1, pp.126-135, 2001. 03.
- [3] Y. J. Lee, E. S. Hyun, T. Y. Kim, "Connection Management for QoS Service on the Web", Journal of Network Computer Applications, Vol.25. No.1, 2002. 10.
- [4] Won-jun Lee, Jaideep Srivastava, "A Market based Resource Management and QoS Support Framework for Distributed Multimedia System", Conference on Information and Knowledge Management, Vol.1, No.1, pp.472-479, 2000.
- [5] Wooyoung Kim, S. Graupner, A. Sahai, D. Lenkov, C. Chudasama, S. Whedbee, Yuhua Luo, B. Desai, H. Mullings, Pui Wong, "Web E-speak: facilitating Web-based e-services", IEEE Multimedia, Vol.9, No.1, pp.43-55, 2002. 01.
- [6] Hea-Sook Park, Doo-kwon Baik, "User Request Filtering Algorithm for QoS based on Class Priority", Journal of KISS, Vol.10, No.5, pp.487-492, 2003. 10.
- [7] Alex Berson, Stephen Smith, Kurt Thearing, "Building Data Mining Applications for CRM", McGraw-Hill, pp.4-14, ISBN 0-07-134444-6
- [8] W. Kamakula, "A Least Squares Procedure for Benefit Segmentation with Conjoint Experiments", Journal of Marketing Research, pp.157-167, 1998. 5.
- [9] Kye-sun An, Se-Jin Go, Jun Jiong, Phill-Kue, Rhee, "Generator of Dynamic User Profiles Based on Web Usage Mining", Journal of Korea Information Processing Society, Vol.9-B, No.4, pp.389-398, 2002. 8.
- [10] M. Wedel, W. Kamakula, "Market Segmentation: Conceptual and Methodological Foundation, Kluwer Academic Publisher, 2000.
- [11] Tae Hyup Roh, Ingoo Han, "Customer Relationship Management Under the Environment of Internet Business", Telecommunications Review, Vol.12, No.1, pp.50-60, 2002. 2.
- [12] Moon-Koo Kim, Dong-Heon Jeong, Kyoung-Sook Park, "An Empirical Study on the Determinants of Customer Loyalty in Mobile Telecommunication Services", Telecommunications Reviews, Vol.12, No.6, pp.970-984, 2002. 12.
- [13] Sang-Hee Rou, Su-Kyung Baik, "Market Segmentation of the Clients for CRM of Health Service", Journal of Health Care Marketing, Vol.3, pp.22-34, 2002.



박혜숙

e-mail : edpsphs@kic.ac.kr
1991년 고려대학교 산업공학과(학사)
1993년 고려대학교 대학원 산업공학과
(석사)
1999년~현재 고려대학교 대학원 컴퓨터
학과 박사과정

2002년~현재 경인여자대학 컴퓨터정보기술학부 조교수
관심분야 : 멀티미디어 콘텐츠, 데이터베이스, 소프트웨어공학



백두권

e-mail : baik@software.korea.ac.kr
1974년 고려대학교 수학과(학사)
1977년 고려대학교 대학원 산업공학과
(석사)
1983년 Wayne State Univ.(전산학 석사)
1985년 Wayne State Univ.(전산학 박사)

1986년~현재 고려대학교 컴퓨터학과 교수
1989년~현재 한국정보과학회 이사/평의원
1991년~현재 한국시물레이션학회 이사/부회장/감사
1991년~현재 iso/iec jtc1/sc32 국내위원회 위원장
2002년~현재 고려대학교 정보통신대학 컴퓨터학과 교수
관심분야 : 데이터베이스, 소프트웨어공학, 시물레이션,
메타데이터, 정보통합, 정보보호 등