

CRM 기법의 전문도서관 적용 방안에 관한 연구

A Study on Customer Relationship Management in Special Libraries

박 여 원*

Yau-Won Park

차 례

- | | |
|----------------|--------------|
| 1. 서론 | 4. CRM 적용 방안 |
| 2. CRM과 정보기술 | 5. 결론 |
| 3. CRM 설계 및 수행 | • 참고문헌 |

초 록

요즘은 풍부한 고객관련 데이터를 다양한 정보기술을 활용하여 분석함으로써 기업과 고객간의 상호 교류를 관리하는 고객중심의 마케팅적 경영방식인 CRM 기법에 관심이 집중되고 있다. 이러한 CRM 기법을 도서관에 도입함으로써 고객중심의 운영이 궁극적 목표인 도서관은 자원의 특성과 목적에 맞게 CRM을 적용하고 각 채널에 적합한 서비스 계획을 구축할 수 있으며, 도서관 운영 전반을 고객중심으로 변화 시킬 수 있을 것으로 사료된다. 이를 위해 본 연구에서는 문헌조사를 바탕으로 CRM의 도서관 도입의 필요성과 CRM 수행에 필요한 정보기술의 이론적 배경을 고찰하였고, 실제로 경제학분야 전문도서관 한곳을 대상으로 고객중심의 장서관

* 한국디지털콘텐츠산업협회 연구원
(Researcher, Korea Digital Content Forum, alaphon@hanmail.net)
· 논문접수일자 : 2004년 3월 6일
· 게재확정일자 : 2004년 3월 15일

리 방안과 고객 세분화를 통한 고객관리 방안을 제시를 위하여 CRM을 수행하였다. 또한 데이터 마이닝 정보기술을 이용함으로써 대량의 데이터를 효율적으로 분석하고, 예측모형을 통해 객관적인 예측값을 얻고자 하였다. 그 결과로 고객 중심의 장서관리 방안과 고객세분화를 통한 고객관리 방안을 제시하였다.

키 워 드

도서관 마케팅, 도서관 평가, 이용연구, CRM, 데이터마이닝

ABSTRACT

Libraries have been made effect to satisfy customer by reflecting information need of customer on libraries. They have considered introducing the data mining techniques to analyze complicated and massive data of libraries and the Customer Relationship Management(CRM) to produce suitable services to each customer segmentation. The purpose of this study is to apply the CRM and data mining techniques to a library, ultimately intends to suggest rules for the collection management and the customer management.

KEYWORDS

Library Marketing, Library Statistics, User Study, CRM, Data Mining

1. 서 론

도서관은 이용연구를 통해 고객의 장서 이용을 측정하고 미래의 정보요구를 예측하여 도서관 장서관리에 반영함으로써 고객 만족을 극대화시키고자 노력하여 왔다. 1970년대 이후 도서관은“ 이용연구와 마케팅은 동전의 양면과 같다”는 Sumslon (2001a)의 말처럼, 이용자를 고객이라는 측면에서 바라보고 적극적으로 고객의 욕구를 파악하고 충족시키기 위한 마케팅 기법을 도입해 왔다. 그리고 1990년 후반 고객

가치를 최우선으로 하는 새로운 마케팅기법인 「고객관계관리(Customer Relationship Management : CRM)」가 등장했다. CRM은 풍부한 고객관련 데이터를 다양한 정보기술을 활용하여 분석함으로써 기업과 고객간의 상호 교류를 관리하는 고객중심의 마케팅적 경영방식을 말한다. 따라서 고객중심의 운영이 궁극적 목표인 도서관은 참고서비스, 인터넷서비스, 장서관리, 고객관리 등과 같이 다양한 분야에 CRM 기법을 도입할 필요성이 있다. 특히 CRM은 고객 중심의 장서관리와 정보서비스를 위한 이용

연구에 새로운 대안이 될 수 있으며, 적극적으로 고객의 이용을 유도할 수 있는 기초 자료도 제공할 수 있다.

CRM을 도서관에 도입하기 위해서는 기본적으로 풍부한 데이터와 데이터 분석을 위한 정보기술이 필요하다. 도서관과 이용자 상호관계 속에서 발생하는 데이터가 자동화 시스템에 의해 매일 상세하게 데이터베이스에 축적되고 있고, 대량의 데이터베이스를 분석하여 의미 있는 지식을 발견하는 다양한 정보기술도 발전되고 있어 CRM 도입의 기반을 제공하고 있다. 그 중에서도 데이터 마이닝(Data Mining) 정보기술은 대용량 데이터베이스 속에서 연구자가 의도하지 않았던 숨겨진 관계, 패턴, 흐름을 발견하는 능력을 가지고 있어, 도서관 데이터를 분석하여 도서관과 이용자, 장서와 이용자 사이에 잠재된 유용한 상관관계를 밝혀내는 역할을 할 수 있다.

따라서 도서관의 CRM 수행은, 첫째, 고객 중심의 도서관 운영과 개발에 필요한 의사결정을 위한 기초 자료를 제공할 수 있으며, 둘째, 고객 세분화를 통해 각 고객 집단에 적합한 맞춤 서비스를 제공할 수 있고, 셋째, 데이터 마이닝과 같은 최신 정보기술을 통해 대량의 데이터를 효율적으로 분석할 수 있으며, 마지막으로 고객의 요구를 적극적으로 충족시킴으로써 도서관의 최대 목적을 달성할 수 있을 것이라 고려된다.

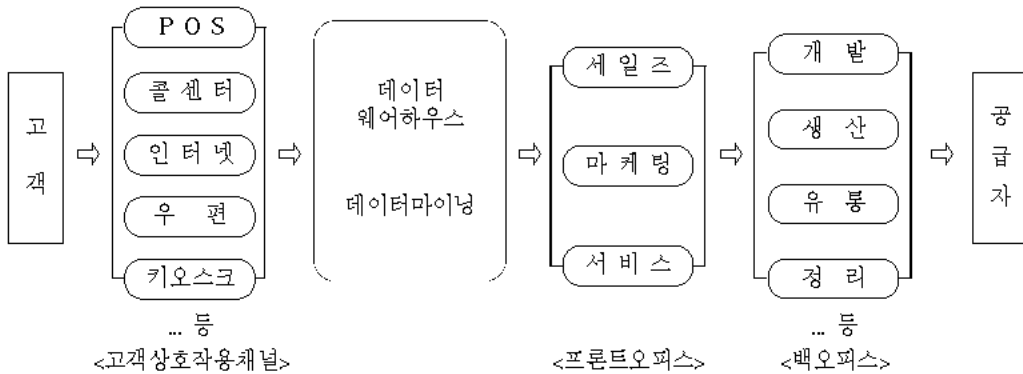
본고에서는 전문 도서관의 효율적인 장서관리와 고객관리를 위해 데이터 마이닝 정보기술을 이용하여 CRM을 수행하는 방안을 제안하고자 한다. CRM을 성공적으로

수행하기 위해서는 다양하고 풍부한 고객 관련 데이터가 가장 중요한 요인이 된다. 따라서 도서관에서 고객과 상호관계를 맺는 대출, 관내이용, 참고서비스, 전화문의, 인터넷사이트 등과 같은 다양한 접점채널(contact point : 기업이 고객과 상호작용하거나 고객과 접하게 되는 기회 혹은 고객이 회사를 접하게 될 수 있는 기회)에서 고객 데이터를 수집하여 연구를 수행하는 것이 바람직하다. 본고에서는 경제학분야 전문도서관 한곳을 대상기관으로 선정하고 대출기록을 접점채널로 선정하였기 때문에 장서를 통한 정보서비스에 국한되는 한계가 있다. 그러나 대출은 도서관 이용의 가장 기본적인 형태이며, 대상 자료의 범주화가 용이하고, 연구기간과 표본크기를 신속적으로 조정할 수 있으며, 정보가 비교적 정확하고 객관적이기 때문에 도서관 CRM 수행의 테스트베드로써 의의가 있다고 생각된다. 또한 도서관에서 비교적 간단하게 얻을 수 있는 대출기록을 대상으로 CRM 도입 방안을 제시함으로써, 이를 바탕으로 도서관에서도 손쉽게 CRM을 수행하고 적용할 수 있을 것이라 고려된다.

2. CRM과 정보기술

2.1 CRM의 기본개념

고객 가치에 대한 중요성을 인식하고 고객의 가치에 따른 차별화된 서비스를 제공하려는 노력은 기업경영 분야에서 마케팅기법을 필두로 시작되어, 1990년대 후반



<그림 1> CRM 개념도

출처 : <<http://www.digitalsme.com>>

부터는 고객중심 경영방식인 고객관계관리 (Customer Relationship Management : CRM) 기법이 활발히 연구되고 있다.

<그림 1>은 CRM 기법의 개념을 간단하게 나타낸 것이다. 기업은 매일 고객과 콜센터, 인터넷, 우편등과 같은 다양한 채널을 통해 계속적으로 대화를 해야하며 최신의 고객 및 시장정보를 획득하여야 한다. 그리고 수집된 다양한 데이터들은 데이터 마트와 데이터 마이닝 같은 정보기술에 의해 통합· 분석되어 고객관리에 사용된다. 이를 통해 고객의 기호, 가치, 요구사항, 문제점에 적절하고도 차별화 된 서비스를 지속적으로 제공받은 고객은 고객만족을 넘어 기업의 충실한 고객이 되고, 이 고객은 시간이 지남에 따라 더 많은 이익을 기업에 제공한다는 것이 CRM의 기본 골격이다.

CRM은 아직까지 명확한 학문적 개념이 정립되지 않은 상태이나, 연구가 활발히 진행되면서 최근 많은 학자들과 실무자들을 중심으로 다양한 정의가 내려지고 있다. 세계적인 컨설팅 회사인 Gattner Group은 신

규고객 획득, 기존 고객유지 및 고객 수익성을 증대시키기 위하여 지속적인 커뮤니케이션을 통해 고객 행동을 이해하고, 영향을 주기위한 광범위한 접근으로 CRM을 설명하고 있다. 시장조사 기관인 Meta Group은 고객에 대한 지식을 지속적으로 듣고, 추측하고, 대응하는 일련의 프로세스들로서 CRM은 기업이 고객의 요구, 기대치 및 행동을 더 잘 이해하게 하고, 이를 통해 사업 기회나 변화에 기민하게 대처할 수 있도록 도와주는 것이라 하였다(이화직 2000).

즉, CRM은 기업이 획득한 고객을 지속적으로 유지하고, 기업에 대한 고객의 가치를 증진시키기 위해 기업과 고객간의 상호 이익관계를 형성· 유지· 강화하려는 기업의 일련의 마케팅 과정으로 정의될 수 있다.

한편 도서관은 고객을 중심으로 장서를 개발하고 관리하여 고객의 정보요구를 만족시키기 위하여 고객의 정보요구, 정보이용패턴을 파악하는 이용연구를 수행해 왔다. 일반적으로 이용연구는 도서관 고객이

탐색하고 이용한 자료를 알아내기 위하여 개별자료, 일군의 자료, 혹은 어느 한 주제 분야 등을 조사하는 것으로서 도서관내의 자료에 초점을 맞추고 연구되어 왔다(양병훈 1991). 이용연구의 또 다른 특성은 순수한 학문적 호기심과함께 실용적인 목적, 즉 연구결과를 도서관 행정에 반영하고자 하는 실제적인 목적을 가지고 있다는 점이다(유재옥 1988). Lancaster(1990)도 이용연구의 목표는 현재의 이용패턴으로부터 장서의 강·약점을 식별하는 것이며, 그렇게 함으로써 고객의 요구에 대한 장서의 적합성을 증가시키도록 정책 변경을 유도할 수 있다고 고객중심 장서관리의 중요성과 이용연구의 실용성을 강조했다.

이제 도서관은 정보기술 발전에 발맞추고, 소장공간의 부족·예산의 감소·고객요구의 다양화와 같은 변화를 현명하게 극복하기 위해서는 고객중심의 도서관 운영으로 효율성을 높여야 한다. 다시 말해, 장서의 소장가치보다는 고객의 이용가치를 우선으로 하여 고객에게 더 많은 관심을 기울여야 한다. 고객들이 어떠한 사람들로 구성되어 있으며, 고객들이 어떠한 유형의 자료를 선호하는가를 파악함으로써, 사서들의 견해, 신념, 희망 등이 아닌 구체적인 증거를 기초로 장서를 관리하기 위한(Hernon and Altman 2001) 방안이 필요하다.

이를 위해 CRM 기법을 도서관에 적용한다면 도서관을 활발하게 이용하는 고객, 이용을 중단한 고객, 오랜 동안 이용하는 고객, 이용 가능성이 있는 고객 등의 행태를 파악함으로써 각 집단에 적합한 서비스를

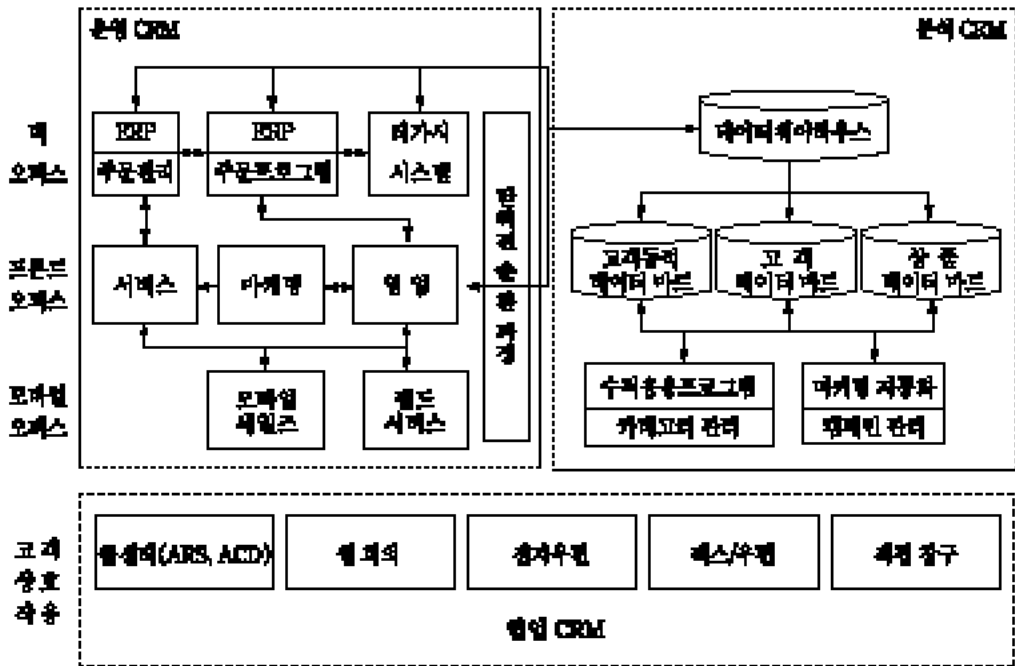
제공하기 위한 고객 세분화가 가능하다. 또한 고객 이용데이터를 자동화 분석기법을 이용하여 분석함으로써 현재 장서의 강·약점을 파악하고 미래의 이용 가능성을 객관적으로 예측함으로써 효율적인 장서관리가 가능하게 된다.

2.2 CRM 구조 및 과정

CRM에 대한 분류는 일반적으로 분석(Analytical) CRM, 운영(Operational) CRM, 협업(Collaborative) CRM으로 구성된다. 도서관의 장서개발, 선정, 평가의 각 단계가 독립적인 과정이 아닌 것과 같이, <그림 2>처럼 CRM도 상호 밀접하게 연관되는 순환구조로 구성되어 있다.

첫째, 분석 CRM은 영업/마케팅/서비스 측면에서 고객 정보를 활용하기 위해 고객 데이터를 추출, 분석하고 고객의 행동을 예측하기 위한 시스템이다. 분석 CRM을 구성하는 대표적인 요소로서 여러 운영계 시스템에 분산되어 있는 고객 관련 정보를 추출하여 저장하는 데이터 웨어하우스나 데이터 마트가 필요하다. 분석 CRM을 구성하는 분석도구로서는 온라인 분석처리(On-Line Analytical Processing : OLAP)와 데이터 마이닝 등이 있다.

둘째, 운영 CRM이란 CRM의 구체적인 실행을 지원하는 시스템으로 기업과 고객간의 관계 향상을 위해 기업의 전방위 업무인 마케팅, 영업, 서비스 기능을 지원하는 것이다. <그림 2>에서 보듯이 이 영역에서는 분석적 CRM에서 도출된 마케팅 및 영업에 대한 아이디어를 구체화하여 사업



<그림 2> CRM 구조

출처 : <<http://www.metagroup.com>>

전략에 활용하여 실행하는 역할을 한다.

셋째, 협업 CRM이란 1990년대 후반부터 인터넷을 기반으로 한 비즈니스의 성장 및 오프라인 기업의 온라인화가 가속화되면서 인터넷에 대응하는 신 개념의 CRM이다. 고객의 서비스 요청에 대한 처리를 지원함과 동시에 고객을 상대로 지속적인 마케팅 및 영업 활동을 하기 위한 채널로서의 접점(콜센터, 인터넷, 전자우편, 메일 등)을 관리하는 역할을 한다(김정희 2002).

본 연구는 특히 분석 CRM에 초점을 맞추어 진행하였으나, 앞으로 분석 CRM의 결과를 토대로 도서관의 다양한 분야에 적용하기 위한 운영 CRM과 온라인 원문서비스나 E-mail 참고서비스 같은 인터넷 환

경으로 점점채널을 확장한 협업 CRM에 관한 연구도 필요하다.

성공적인 CRM의 수행을 위해서는 고객과 접촉하는 순간부터 고객을 유지하는 단계까지 전 과정에 있어서 고객 데이터를 획득하고 활용하는 아래의 3단계가 필요하다.

① CRM 전략수립

- 고객데이터 확보 방안 수립
- 외부데이터 활용 방안 수립
- CRM 구축/활용 방안 수립

② 데이터 웨어하우스 구축을 위한 기반 정비

- 데이터 웨어하우스 기반 구축

- 고객 DB 구축
- ③ CRM 시스템 구축
 - 영업 및 마케팅 지원을 위한 분석 지원 시스템 구축 (잠재 고객 확보, 제품 서비스의 차별화, 고객성향, 고객 행태분석)
 - 데이터 마이닝을 이용한 지능형 분석 시스템 구축
 - 적절한 채널에 대한 캠페인 전개 시스템 구축(이화직 2000)

2.3 고객정보의 재구축

CRM을 성공적으로 수행하기 위해서는 높은 이익 잠재력을 가진 고객집단을 먼저 정의해야 한다. 그러나 고객이 무의미하게 세분화되어 있는 것이 아니라 애매모호한 상태로 나뉘어 있기 때문에, 고객의 세분화(segmentation)는 이미 단순한 속성(연령, 성별, 직업 등)으로는 판별과 식별이 어려우며, 상품과 그 밖의 서비스 구매 형태를 본 후에야 비로소 알 수 있게 된다(Berson 1999). 따라서 대상 고객집단을 알기 위해서는 전화, 참고문의, 대출 등과 같은 기존의 고객접점 채널뿐만 아니라 온라인 로그인 파일이나 이메일, 인터넷 게시판과 같은 최신 정보통신을 이용한 접점채널을 통해서 가능한 많은 정보를 수집해야 한다.

계속적으로 변화하는 시장과 고객 요구에 대응하기 위하여 실시해야 할 고객정보 수집 4단계가 있다.

첫 번째는 충실한 고객 정보 수집의 단계이다. 고객 데이터는 비용을 들여서 모으는 것이라는 인식이 필요할 만큼 중요하다.

이미 유통업계 등에서는 포인트 카드로 할인을 제공하는 대신에 고객으로부터 정밀도가 높은 회원 정보를 수집하고 있다. 도서관도 이용자파일 뿐만 아니라 SDI 서비스, 상호대차 신청서, 희망도서 신청서 등 다양한 경로를 통해서 적극적으로 고객의 정보를 수집할 수 있는 방안을 마련해야 한다.

두 번째는 변화와의 전쟁이다. 성별이나 연령 같은 기본 속성뿐만 아니라 행동 유형에 따른 세분화 등에서 나온 특징적인 고객 정보가 CRM의 가장 중요한 요소가 된다. 따라서 도서관은 고객의 대출, 참고문의 등과 같은 과거이력을 고객과 관련하여 정리, 수집하여야 한다.

세 번째는 고객 정보를 통합하는 단계이다. 기존 시스템의 대부분은 특정 업무를 수행하기 위한 것이며 거래 혹은 상품 중심으로 설계되어 있다. 그러나 고객의 관점에서 보기 위해서는 데이터의 2차 가공이 필요한 경우가 많다. 도서관도 마찬가지로 데이터베이스내의 수서파일, 서지파일, 대출파일 등이 대부분 장서를 위주로 구조화되어 있다. 따라서 도서관 CRM 기법을 도입하기 위해서는 수집된 데이터들을 고객을 중심으로 재구성하여 데이터 웨어하우스나 데이터 마트로 통합해야 한다.

마지막으로 분석 및 평가가 필요하다. 정리 통합된 고객 정보를 기반으로 개개의 고객 요구와 정보의 변화를 알기 위해서는 다양한 관점에서 분석 및 평가할 필요가 있다. 그리고 개개의 고객에 대한 대응·개선 방안을 세워야 한다(앤더슨 권설텡 1999).

2.4 정보기술 요소

데이터 마트(data mart)의 기본적인 구성은 데이터 웨어하우스와 유사하다. 그러나 데이터 마트는 데이터 웨어하우스 등으로부터 유도될 수 있는 데이터의 저장소로서 특별히 계획된 목적을 위해 접근의 용이성과 유용성을 강조한 것이다. 이들을 구축하는 방법에는 ‘하향식(top-down)과 ‘상향식(bottom-up) 방식이 있다. 하향식방식은 전사적 차원의 데이터 웨어하우스를 먼저 구축한 후 필요에 따라 하나의 주제 또는 부서 중심의 데이터 마트를 구축하는 방법이고, 상향식은 반대로 데이터 마트를 먼저 구축한 후 점차적으로 이들 데이터 마트를 통합하고 확장하는 방법이다(장남식 1999). 즉, 데이터 마트는 특정한 목적의 사용자 그룹의 활용을 위해 만들어진 통합된 데이터의 집합이다.

하지만 데이터 마트나 데이터 웨어하우스는 의미 있는 정보를 추출할 뿐 새로운 정보를 제공하지 못한다는 점에서 보다 적극적인 지식발견 수단인 데이터 마이닝과 대별된다. 그리고 데이터 마이닝을 위해서 데이터 마트가 반드시 존재해야 하는 것은 아니지만, 이 과정을 거치게 되면 이미 데이터는 가공처리 되기 때문에 데이터 마이닝의 전처리 과정을 경감시켜 주고 성공가능성도 높여 준다.

데이터 마이닝은 전 세계적으로 1990년대부터 활용되기 시작한 지능형 분석기법으로 CRM 실현의 기술 요소로 주목받고 있다. 데이터 마이닝은 수집된 데이터에서 아직 발견되지 않은 가치 있는 패턴을 찾

는 것을 목적으로 하는 기술의 집합을 지칭한다. 데이터 마이닝의 목적은 과거 행위의 분석을 바탕으로 향후 행위를 예측하는 의사결정 지원을 위한 모델을 만드는 것이다(도기운 2002). 기업이 보유하고 있는 일일거래 데이터, 고객 데이터, 상품 데이터 혹은 각종 마케팅 활동에 있어서의 고객반응 데이터 등과 그 밖의 외부 데이터를 포함하는 모든 사용 가능한 근원 데이터를 기반으로 감춰진 지식, 기대하지 못했던 경향 또는 새로운 규칙 등을 발견할 수 있어 CRM 데이터분석에 널리 이용되고 있다.

지금까지 통계는 미리 가설을 세우고 데이터를 통해 그 가설을 확인하는 방식이었다. 이러한 방식은 적절한 가설을 세울 수 없는 상황에는 적용할 수 없으며, 새로운 정보가 거의 생성되지 않는다. 그러나 데이터 마이닝은 지능형 데이터 분석기법으로 사용자가 미리 가설을 세우지 않아도 시스템이 스스로 가설을 생성·검증하여 연구자가 의도하지 않았던 새로운 정보를 제공할 수 있다.

데이터 마이닝이란 표층에 드러난 것보다 더 많은 지식이 데이터에 은닉되어 있는 측면과 마찬가지로 하나의 기법으로 이루어진 것이 아니다. 이러한 관점에서 보면 데이터 마이닝은 실제 ‘어떠한 기법도 관련될 수 있는’ 것이다. 데이터에서 드러난 것 이상을 추출할 수 있는 기법들은 모두 적용 가능하나, 현재 주목되는 기법들은 질의도구(Query Tools), 시각화(Visualization), OLAP(On-Line Analytical Processing), 통계적 기법(Statistics Technique), 의사결정

나무(Decision Tree), 신경망 (Neural Networks), 연관규칙(Association Rule), 발생 알고리즘(Genetic Algorithm), 군집분석(Clustering)이 있다(Adriaans and Zantinge 1998).

기법을 선정할 때는 데이터 분석의 목적, 수집된 데이터의 성격, 산출되는 정보의 설명력, 사용의 용이성 등 여러 가지 기준을 고려하여야 한다. 초기 데이터 분석은 주로 데이터의 관계나 연관성을 살펴 보기 위해서 주로 연관규칙 분석, 회귀 분석, 군집 분석을 중심으로 실시하게 된다. 이를 토대로 의사결정 나무 분석이나 신경망 분석을 실시하게 되는데, 예를 들어 결과의 정확성만을 중요시하는 경우는 신경망 분석을, 결과의 해석이 요구되는 경우는 의사결정나무 분석을 적용하는 것이 적합하다. 본 연구에서는 해석이 용이하고 데이터간의 상관관계 분석이 가능한 장점을

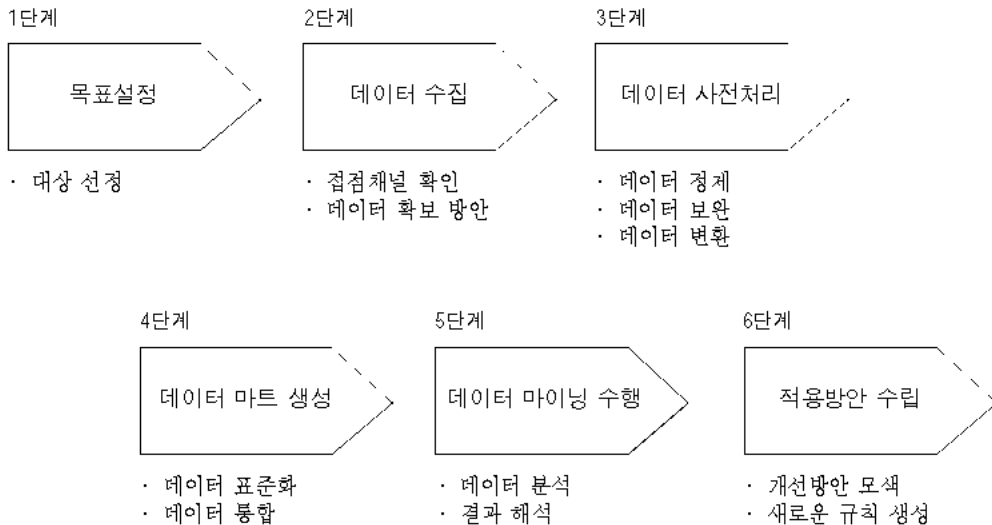
을 가지고 있기 때문에 연관규칙 분석, 군집분석, 의사결정나무 분석 기법을 선택하였다.

3. CRM 설계 및 수행

CRM은 고객중심의 도서관 운영을 궁극적인 목표로 '장서관리를 위한 CRM', '고객관리를 위한 CRM', '참고서비스를 위한 CRM', '온라인 서비스를 위한 CRM' 등 도서관의 다양한 분야에서 적용가능하다. 본 연구는 전문도서관을 대상으로 장서관리와 고객관리를 위한 CRM을 수행하기 위하여 <그림 3>과 같이 6단계로 진행하였다.

3.1 목표설정 및 대상선정

목표설정의 단계로 본 연구에서는 고객



<그림 3> CRM수행 단계

관리와 장서관리를 위한 규칙을 생성하고, 고객의 세분화를 통한 정보서비스 방안을 마련하는 것을 목표로 설정하였다.

3.2 데이터 수집

연구대상 정보자료실은 경제연구소인 모 기관의 개원과 함께 연구업무지원을 목적으로 설립·운영되고 있으며, 국내 경제학 분야 전문도서관으로서 다양한 정보를 검색·수집하여 제공하는 역할을 담당하고 있다.

정보자료실 장서의 대부분이 경제학과 관련된 분야로써 11만 여권을 소장하고 있으며, 1998년부터 2002년까지 대출기록을 보면 매년 6천여 권 정도 대출이 이뤄지고 있다. 그리고 2002년에는 대출정책의 변화로 인해 대출이 1.5배 이상 증가하여 9천여 권이 넘는 것을 볼 수 있다.

가장 큰 장서비율을 차지하고 있는 자료유형은 단행본(4만8천여 권)으로 <표 1>과 같이 대출비율도 가장 높다. 또한 2002년에는 단행본의 대출비율이 2배 가까이 증가한 것을 볼 수 있다. 다음으로 장서비율이 높은 연구보고서(4만3천여 권)는 대부분이 「IMF」, 「World Bank」, 「KDI」와

같은 국내의 중요 경제기구의 연구보고서들로 이뤄져 있다. 2002년 단행본과 통계·연감류의 대출비율이 증가한 것에 비해 연구보고서의 경우 그 차이가 크지 않았다. 이것은 정보자료실 사서와 면담한 결과 연구보고서의 대부분이 원문서비스로 온라인에서 제공되고 있기 때문이라고 한다. 세 번째로 통계와 연감류(1만5천여 권)의 경우 장서비율 대비 대출비율(약 23%)이 가장 높은 것으로 나타나 경제학 연구에서 중요한 자료유형 중 하나라고 볼 수 있다. 마지막 경기간행물은 1,131종을 구독하고 있다. 경기간행물은 2002년부터 대출이 가능해졌고, 1년 동안 대출기록을 보면 1,244건으로 전체 대출의 약 10%를 차지하고 있다. 경기간행물의 이용이 높은 전문도서관의 특성에 반해 대출비율이 낮은 이유는 연구보고서와 마찬가지로 연구원들이 온라인을 통해 원문서비스를 주로 이용하고 있기 때문이다.

본 연구에서는 서지데이터·대출데이터·고객데이터·연구물데이터를 대상으로 추출하였다. 정보자료실 데이터베이스에서 서지 데이터(총 52,022권, 단행본 36,724권, 통계·연감 14,918권), 고객 데이터(총 456명), 대출기록 데이터(1998년 ~ 2002년. 총

<표 1> 연도별 대출 빈도

연 도 \ 자료유형	1998	1999	2000	2001	2002	계
단행본	2,598	2,421	2,543	2,357	4,571	14,490
통계·연감류	1,262	1,542	1,775	1,368	2,241	8,188
연구보고서	2,052	1,748	1,861	1,674	2,386	9,721
계	6,181	6,134	6,561	5,796	9,456	34,128

17,613건)를 추출하였다. 정기간행물과 연구보고서는 사서와 면담결과 이용자들이 주로 원문서비스를 이용하고 있기 때문에 대출기록 분석이 무의미한 것으로 판단되어 제외되었다. 고객 데이터는 5년 동안 정보자료실을 이용한 기록을 모두 분석하기 위하여 현재 직원(211명)과 퇴직하거나 이직한 직원까지 포함하였다. 그리고 인터넷 사이트의 원문서비스를 통해 모기관의 주요 기관지인 「정책연구」와 「연구보고서」, 「정책연구시리즈」에 실린 1998년부터 2002년까지 연구보고서의 연구자(총 106명)와 주제분야(총 435건)를 추출하였다. 외부 연구원을 제외하고, 이중 소속 연구원 72명이 연구업무와 대출의 상관관계 분석 대상이 되었다.

3.3 데이터 사전처리

데이터 사전처리 단계에서는 데이터를 정제·보완하고, 파생데이터를 생성하는 과정을 수행하였다. 그 내용은 다음과 같다.

첫째, 고객파일의 주민번호 속성에서 연령과 성별을 추출하였고, 직위와 연구팀 속성은 데이터분석이 용이하도록 문자형에서 범주형으로 변환하였다.

둘째, 서지파일에서 출판년도가 없는 레코드는 서명이나 권·호 정보에서 출판년도를 추출하여 최대한 데이터 결실을 줄이는 데이터 보완 작업을 수행하였다.

셋째, 서지파일에서 같은 발행기관이지만 다르게 기록되어 있는 경우는 하나로 통일하였다. 예를 들어 발행기관이 대학인 경우에 'Up'와 'University Press'로 혼용

되고 있어 'University Press'로 모두 수정하였다. 다양하게 기록되어 있는 발행기관은 가장 많이 사용되는 발행기관명을 기준으로 통일하였다.

넷째, 주제별 분석을 위해서 서지파일의 LC 분류번호에서 강목까지 데이터를 추출하였다.

다섯째, 기관지에서 추출한 연구물과 장서의 주제별 연관성을 살펴보기 위해서, 이화여대 경제학과 석사과정 학생 2명의 도움을 받아 연구물 주제분야(51개)를 LC분류(9개)로 재분류하였다.

여섯째, 각 파일에서 테스트용으로 등록된 레코드, 결실이 많은 레코드를 제거하였다.

일곱째, 데이터 예비분석을 통해서 필요하다고 판단되는 파생데이터 11개 요소를 생산하였다. 먼저 자료별로 5년 동안의 대출 횟수를 계산한 값으로 '대출빈도'를 생산했고, 미래의 이용가능성을 결정하기 위하여 마지막 대출부터 조사시점까지의 기간을 '대기기간'으로 계산하였다. 그리고 예비분석 중 발행기관 분석결과 발행기관과 대출의 관련성을 발견할 수 있었으나 3,900여개의 발행기관의 문자형 기관명을 그대로 이용할 수 없으므로, 각 발행기관별 회전율과 가중치 값을 구하여 데이터 분석에 대신 이용하였다. 그리고 연구업무와 장서와의 상관관계를 분석하기 위해 고객별 총 대출빈도를 계산하였고, 자료유형과 언어, 주제에 따른 대출빈도를 계산하여 자료유형 선호도, 언어 선호도, 주제 선호도 값을 구하였다. 출판년도 선호도는 대출된 장서의 출판년도의 평균을 계산한 값이다. 그

리고 3종의 기관지에서 추출한 데이터를 이용하여 총 연구빈도와 주제별 연구빈도를 구하였다.

3.4 데이터 마트 생성

데이터 마트 생성 단계에서는 사전처리가 된 장서관련 데이터와 고객관련 데이터를 「SAS v.8 Data Analysis」를 이용하여 <그림 4>와 같이 고객 중심으로 통합하였다.

장서관련 데이터는 서지파일과 대출파일 그리고 파생데이터를 통합하여 14개의 속성과 52,022개 레코드로 구성된 테이블을 생성하였다. Print_accesion_no(제어번호)를 아이디로, Bad(대출여부)를 목표변수(Target)로 설정하였다.

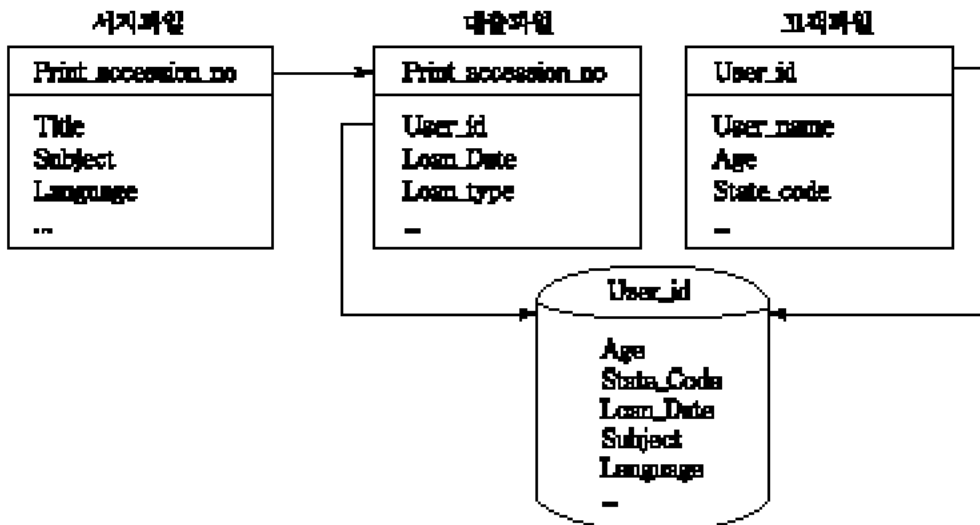
고객관련 정보를 분석하기 위해서 2개의 테이블을 생성하였다. 첫 번째는 고객별 장

서이용패턴을 보기 위해 서지파일, 대출파일, 고객파일, 그리고 파생데이터를 결합한 고객관련 테이블로, 19개의 속성을 가진 17613개의 레코드로 구성되었고, User_id(고객 아이디) 속성을 아이디로, Bad(대출유무)를 목표변수로 설정하였다. 두 번째는 고객의 연구업무와 장서이용의 상관관계를 살펴보기 위해 생성된 테이블로써 31개 속성과 442개의 레코드로 구성되었다.

3.5 데이터마이닝 수행

데이터 분석과 결과해석을 위해서 생성된 데이터 마트를 「SAS E-Miner 4.0」 데이터 마이닝 도구를 이용하여 연관규칙분석, 군집분석, 의사결정나무 분석을 실시하였다.

첫째, 고객의 장서이용 연관규칙 분석결과, 핵심 장서주제는 경제사·경기분야와



<그림 4> 고객 데이터 마트 생성

<표 2> 주제별 이용지수와 연구지원지수

주제분야	평균출판년도 (년)	장서수 (권)	대출빈도 (회)	예상대출율 (%)	실제대출율 (%)	이용 지수	연구율 (%)	연구지원 지수
통 계	1980.41	2,498	637	4.80	3.63	0.76	13.85	0.35
경제이론	1981.08	5,825	1,616	11.20	9.20	0.82	16.34	0.69
경제사· 경기	1984.24	10,767	4,479	20.70	25.50	1.23	14.96	1.38
산 업	1983.10	14,479	4,469	27.83	25.44	0.91	11.08	2.51
교통· 통신	1985.38	1,002	441	1.93	2.51	1.30	6.65	0.29
통 상	1982.42	5,077	1,596	9.76	9.09	0.93	8.86	1.10
재 경	1984.17	5,761	2,548	11.07	14.51	1.31	13.30	0.83
국가재정	1983.47	3,435	1,199	6.60	6.83	1.03	6.37	1.04
사회과학	1980.94	3,178	580	6.11	3.30	0.54	8.59	0.71
계		52,022	17,565	100	100		100	

산업분야이며, 특히 산업분야 영어 단행본이 가장 많이 이용되는 장서 유형이었다.

둘째, 주제별로 장서가 얼마나 이용되고 있고, 연구를 지원하는지 알아보기 위해서 <표 2>와 같이 상대 이용지수와 연구지원지수를 산출하였다. 상대이용이란 Jain이 제시한 대출 분석 방법으로써, 예상되는 이용과 실제 이용의 비율을 이용지수로 나타내어 각 주제별 장서의 이용수준을 평가하는 하나의 지표로 사용하는 것을 말한다(유재욱 1988). 고객의 연구업무와 장서 이

용의 상관관계를 분석한 결과 주제별 장서비율과 장서의 최신성, 그리고 연구주제와의 적합성이 고객의 장서 이용에 영향을 미치는 것을 확인할 수 있었다.

셋째, 정보자료실 고객을 세분화하여 각 집단에 적합한 정보서비스를 제공하기 위하여 고객데이터를 대상으로 군집분석을 실시하였다.

그 결과 <표 3>과 같이 5개의 고객 군집으로 형성되었으며, 각 집단을 구별하는데 발행기관의 회전율, 소장기간, 연구비율, 선

<표 3> 고객 군집분석 결과

군 집	발행기관의 회전율	소장기간 (년)	연구빈도 (개)	선호출판년도	평균연령 (세)	평균대출빈도 (회)
1	18.03	6.06	3.87	1995.00	41.62	178.33
2	2.36	10.44	2.72	1990.60	50.09	203.85
3	3.06	9.22	2.59	1991.75	36.19	149.00
4	29.57	4.72	0.69	1995.86	32.32	107.48
5	0.81	21.82	0.00	1979.57	38.58	66.72
계	3.68	9.94	2.60	1991.07	39.33	160.91

<표 4> 소장률에 비해 대출빈도가 높은 발행기관 리스트

발행기관	장서수(권)	대출수(회)	장서비율(%)	대출비율(%)	회전율	가중치
醫療保險聯合會	7	29	0.01376	0.1660	12.06	2
예산청	6	57	0.01180	0.3262	27.65	4
蒼蒼社	6	42	0.01180	0.2404	20.37	4
Federal Deposit Insurance Corporation	6	23	0.01180	0.1316	11.16	2
World Economic Forum	5	46	0.00983	0.2633	26.78	4
한국증권업협회	5	27	0.00983	0.1545	15.72	2

<표 5> 소장률에 비해 회전율이 저조한 발행기관 리스트

발행기관	장서수(권)	대출수(회)	장서비율(%)	대출비율(%)	회전율	가중치
A. M. Kelley, Bookseller	549	36	1.07956	0.2060	0.19	3
Macmillan	546	4	1.07366	0.0229	0.02	1
Praeger	513	53	1.00877	0.3033	0.30	4
Wiley	411	37	0.80820	0.2118	0.26	3
韓國貿易協會	385	41	0.75707	0.2347	0.31	4
Greenwood Press	303	18	0.59582	0.1030	0.17	2

호출판년도, 평균연령, 대출빈도의 속성이 가장 큰 구분점이 되었다. 그리고 발행기관 회전율이 높은 장서를 이용할수록 최신의 장서를 선호한다는 것을 발견할 수 있었다. 또 인구통계학적인 연령과 성별, 직위와 같은 속성보다는 고객과 도서관의 상호작용 속에서 발생하는 고객 이용데이터가 고객의 특성을 더 잘 대변해 주는 것을 알 수 있었다. 따라서 일반적으로 수행되어 온 연령이나 직위에 따른 일괄적인 정보 서비스 보다는, 각 그룹의 특성에 맞는 정보서비스를 수행한다면 고객만족을 더욱 향상시킬 수 있을 것이다.

넷째, 발행기관 데이터분석에 이용하기 위하여 각 발행기관의 장서수, 대출빈도,

장서비율, 대출비율을 계산하였고, 대출비율을 중심으로 10단계의 가중치를 각 발행기관에 부여하였다. 분석결과 179개 발행기관(4.5%)이 전체 대출의 80%를 담당하고 있어 발행기관과 고객의 장서이용에 상관관계가 있음을 발견하였다. 이를 바탕으로 <표 4>, <표 5>와 같이 “소장률에 비해 대출빈도가 높은 발행기관 리스트”와 “소장률에 비해 대출빈도가 저조한 발행기관 리스트”를 추출하였는데, 이를 통해 도서를 선정하는 사서의 예측과 실제 고객의 정보요구의 차이를 발견할 수 있었다.

다섯째, 장서의 폐기와 수서정책에 사용할 규칙을 발견하기 위해서 의사결정나무 분석을 수행하였다. 먼저 장서데이터에서

<표 6> 자료유형별 군집분석 결과

군 집		대출 (%)		발행기관의 회전율	소장기간 (년)	대기기간 (월)	평균 대출빈도(회)
		무	유				
단 행 본	1	41.07	58.93	2.22	6.18	27.77	1.33
	2	87.58	12.42	0.98	12.36	41.98	0.14
	3	98.07	1.93	0.64	21.04	31.33	0.02
	4	100.00		0.32	37.07	.	0.00
	평 균			0.87	21.52	31.00	0.25
통 계 · 연 감	1	30.30	69.70	2.90	5.92	21.61	2.32
	2	82.00	18.00	1.39	11.44	26.87	0.22
	3	96.28	3.72	0.88	18.95	25.64	0.04
	4	99.53	0.47	0.77	30.18	27.00	0.00
	평 균			1.43	16.50	23.08	0.57

중요한 속성들을 찾아내기 위한 예비분석으로 표본데이터 2,000개를 대상으로 군집분석을 실시하였다. 아래 <표 6>와 같이 군집분석을 통해 찾아낸 고객의 장서 이용에 영향을 미치는 속성(발행기관 회전율과, 소장기간, 마지막 대기기간, 대출빈도)을 의사결정나무 분석에 사용하였다.

의사결정나무 분석을 수행하여 얻은 대출가능성 모형을 수서정책에 직접 활용하기 위해서는 점수화(Scoring)과정이 필요하다. 점수화과정은 예측모델링의 마무리 단계로 대출가능 예상 리스트를 추출하는 것으로 새로운 장서에 대한 대출 빈도를 예상할 수 있다.

여섯째, 주제별 이용분석을 통해 얻은 결과를 바탕으로 의사결정나무 대출가능성 모형을 확장·변형하여, 실제 수서와 폐기 정책에 적용하기 위해 장서의 주제분야, 자료유형, 형태, 발행기관별 대출 예상치를 아래 <그림 5>와 같이 수치화하였다. 그러나 이것은 어디까지나 과거의 이용에 근거

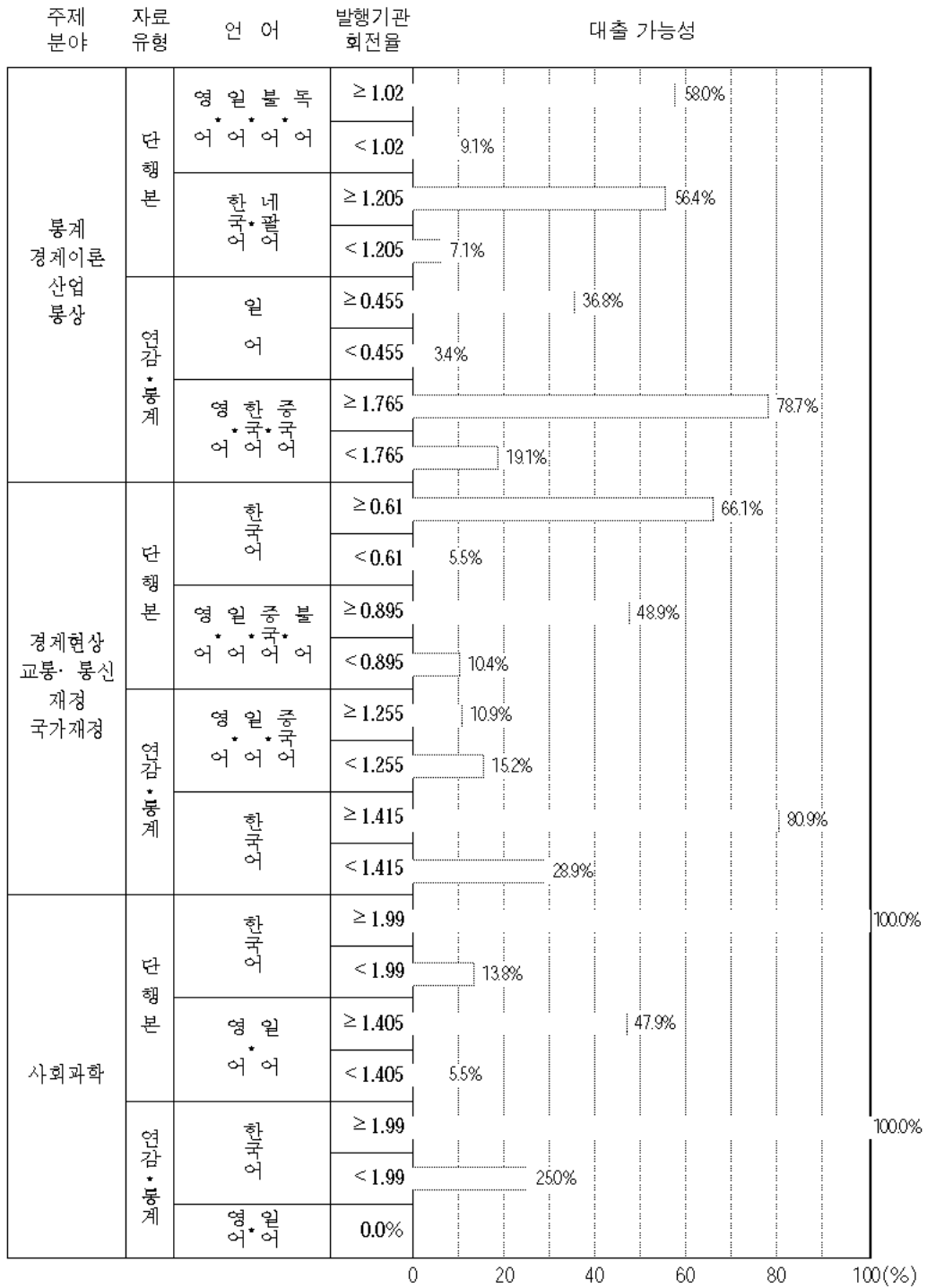
한 객관적 수치이기 때문에, 사서의 의사결정시 장서의 본질적 가치(독자와 관계없이 저자의 저술목적 및 내용에서 높은 수준을 유지함으로써 평가받는 장서자체의 절대적 가치)에 대한 판단이 함께 필요하다고 고려된다.

4. 적용방안 수립

데이터 마이닝 분석 결과를 통한 새로운 규칙을 생성하고 이를 바탕으로 장서관리와 고객관리를 위한 각 채널에 적합한 적용방안을 모색하였다.

첫째, 각 연구팀에 정보서비스를 수행할 때, 모든 분야보다는 각 연구팀에 적합한 분야를 선택하여 제공함으로써 고객의 흥미를 유지할 수 있다.

둘째, 고객의 군집분석과 점수화 결과를 바탕으로 고객을 4개의 그룹으로 세분화하였고, 그에 따라 각 그룹의 특성에 맞는 정



<그림 5> 대출가능성 모형 규칙

보서비스를 제공해야 한다. 우수고객에게는 특별한 서비스를 받고 있다는 인상을 주어 흥미를 계속 유지시키는 것이 중요하고, 잠재고객은 최우선 정보획득의 장소로 정보자료실의 존재를 인식시키는 일이 가장 필요하다. 그리고 신규고객에게는 정보자료실에서 수행하고 있는 모든 서비스를 적극적으로 홍보하기 위한 정보교육 기회를 제공해야 한다. 마지막으로 장기고객은 오랜 경력의 연구원들로 단행본 위주의 보서비스를 통해 관계를 지속시켜야 한다.

셋째, '소장률에 비해 대출빈도가 높은 발행기관' 리스트와 '소장률에 비해 대출빈도가 저조한 발행기관' 리스트를 참고하여, 필요한 장서 구입을 실패할 확률과 불필요한 장서를 소장할 확률을 줄인다.

넷째, 대출가능성 모형을 토대로 수서정책을 위한 규칙을 생성하였다. 새로운 장서의 입수를 결정할 때, 간단한 서지정보(주제분야, 자료유형, 언어, 발행기관)를 대입함으로써 객관적인 대출가능성 예측값을 얻을 수 있다.

5. 결 론

도서관의 최고의 목적은 고객의 정보요구의 만족에 있다. 이를 위해 도서관은 이용연구를 통해 고객의 장서 이용을 측정하고 미래의 정보요구를 예측하여 도서관의 장서관리와 정보서비스에 반영하고자 노력해 왔다. 요즘은 최신 정보기술을 이용하여 복잡해지고 방대해진 도서관 데이터를 정확하고 효율적으로 분석하기 위한 데이터

마이닝 기법과 고객의 세분화를 통해 각 집단에 적합한 서비스를 제공하기 위한 CRM 도입을 고려하고 있다. 따라서 본 연구에서는 CRM과 데이터 마이닝 정보기술을 도서관에 적용해보고, 그 결과로 고객 중심의 장서관리 방안과 고객세분화를 통한 고객관리 방안을 제시하였다.

이를 위해 문헌조사를 바탕으로 CRM의 도서관 도입의 필요성과 CRM 수행에 필요한 정보기술의 이론적 배경을 고찰하였다. 그리고 도서관 한 곳을 선정하여 CRM을 적용해 보았다.

이상과 같은 도서관의 CRM 수행을 통해 다음과 같은 장점을 찾을 수 있었다.

첫째, 고객 중심의 장서관리와 개발에 필요한 의사결정을 위한 기초 자료를 제공할 수 있다. 더욱이 대출기록과 상용 데이터 마이닝 프로그램만 있으면 자관에 맞는 규칙을 쉽게 도출하여 적용할 수 있으며, 폐기·이관 규칙의 평가 결과 효율성도 높은 것으로 나타났다. 또한 CRM은 고객의 이용을 바탕으로 하기 때문에 사서의 예측과 고객의 정보요구의 차이를 효과적으로 줄일 수 있다.

둘째, 고객 세분화를 통해 각 고객집단에 적합한 맞춤 서비스를 제공할 수 있다.

지금까지 도서관 서비스들은 대부분이 고객의 연령이나 성별, 직업에 따라 구분되어 수행되어 왔지만, CRM 수행결과 고객과 도서관의 상호작용 속에서 발생하는 데이터가 고객의 특성을 잘 반영하고 있었다. 예를 들어, 우수고객에 속하는 연구원에게 단행본의 질 좋은 장서보다는 최신의 통계 자료를 제공함으로써 고객 만족을 향상시

킬 수 있다. 본 연구에서는 단행본과 통계·연감류의 대출기록만 분석하였기 때문에 장서를 통한 정보서비스에 국한되는 한계가 있지만, 온라인 서비스와 정기간행물, 연구보고서와 같은 다른 자료유형에 확대 적용한다면 더욱 정확한 고객 세분화를 통해 고객관리를 일반화할 수 있을 것이다.

셋째, 데이터 마이닝과 같은 최신 정보 기술을 통해 대량의 데이터를 효율적으로 분석할 수 있다. 특히 데이터 마이닝 최대 장점인 예측모형을 통해 현재의 이용분석 뿐만 아니라 의사결정을 위한 객관적인 예측값을 얻을 수 있었다. 또한 연구자가 의도하지 않았던 변수간의 상관관계를 분석할 수 있어 발행기관과 대출의 관계, 자료유형과 언어와의 관계, 그리고 고객의 장서 이용패턴을 발견할 수 있다.

앞으로 도서관 CRM 도입의 적합성을 뒷받침하기 위해서 장서구성과 고객구성이 보다 다양하고 대규모인 대학도서관이나 공공도서관에 CRM을 적용한 연구가 필요하다. 도서관의 다양한 분야에 CRM 적용 가능성을 모색하기 위해서 도서관과 고객간에 상호관계를 맺는 더 많은 접점채널들(관내이용, 참고서비스, 전화문의, E-mail 문의, 온라인 로그인 등)을 대상으로 확장한 연구가 뒤따라야 할 것이다.

참고문헌

강현철, 한상태, 최종후, 김은석, 김미경. 2001. 『데이터마이닝 - 방법론 및 활용 -』. 제2판. 데이터마이닝 총서, 1.

서울 : 자유아카데미.

김정희. 2002. 『CRM(Customer Relationship Management) 성장 발전 단계 모델 개발에 관한 연구』. 석사학위논문. 이화여자대학교 대학교, 경영학과.

도기윤. 2002. 『데이터마이닝 기법을 이용한 은행계 신용카드 연체고객 분석』. 석사학위논문, 고려대학교 경영대학원, 경영과학 및 MIS 전공.

앤더슨 컨설팅. 1999. 『CRM : 고객관계관리』 권태경, 양경미 옮김. 서울 : 대청.

梁炳訓. 1991. 大學圖書館 藏書評價方法에 관한 考察 - 藏書開發 側面을 중심으로 -. 『국회도서관보』, 28(4) : 42-58.

유재욱. 1988. 장서평가에 관한 소고 - 특히 이용조사를 중심으로 -. 『도서관학』, 15 : 175-195.

이화직. 2000. 『CRM의 이론과 사례』. 서울 : 한국능률협회.

장남식, 홍성완, 장재호. 1999. 『데이터마이닝』. 서울 : 대청.

중소기업정보은행. 『CRM이란 무엇인가』. [Online]. [인용 일자 2003. 3. 27]. <<http://www.digitalsme.com/smdb/jsp/nguide/infor/04guid/01crm.htm>>

Adriaans, Pieter and Dolf Zantinge. 1998. 『데이터마이닝』 용환승 역. 서울 : 그린. 원전 : *Data Mining* Harlow : Addison Wesley Longman, 1996.

Berson, Alex, Stephen Smith, and Kert Therling. 1999. *Building Data Mining Applications for CRM*

- Hernon, Peter and Ellen Altman. 2001. 『도서관 서비스 품질 관리론』 이은철 역. 서울 : 한국도서관협회. 원저 : *Assessing Service Quality*. Chicago : American Library Association, 1998.
- Lancaster, F. W. 1990. 『도서관 서비스 평가론』 장혜란 역. 서울 : 구미무역. 원저 : *If you want to evaluate your library*. London : Library Association, 1988.
- Needamangala, Ashwin. 2000. *A Library Decision Support System Built on Data Warehousing and Data Mining Concepts and Techniques* Master s Thesis. University of Florida.
- Sumston, John. 2001a. "Library Statistics for Marketing." *IFLA Journal* 27 : 221-231.
- _____. 2001b. "Library Statistics to Enjoy - Measuring Success!" *IFLA Journal* 27 : 211-214.