

신경망 이론을 이용한 분석 eCRM 컴포넌트 설계 및 구현

강윤정, 최동운, 이용석

전북대학교 전산통계학과, *서남대학교 컴퓨터학과

yjkang66@jj.ac.kr, cdo@seonam.ac.kr, yslee@moak.chonbuk.ac.kr

Design and Implementation of Analytical eCRM Component using of Neural Network

Kang Yun Jeong, Choi Dong Oun, Lee Yong Seok

Department of Computer Science, Chonbuk National University,

* Department of Computer Science, Seonam University

요 약

CRM은 기존 고객을 잘 관리하면서 새로운 고객을 유치하는 마케팅 비용은 기존고객 유지비용의 몇 배라는 기본적인 원칙에 의한 접근이다. 물론, 데이터베이스 마케팅, 이메일 마케팅이 기존고객 유지의 수단일 수도 있다. 본 논문에서 개발한 신경망을 이용한 분석 eCRM 시스템의 컴포넌트를 설계 구현하였다. 이는 특성화된 컴포넌트 기반으로 개발되었으며, 기존 데이터 환경을 효율적으로 이용할 수 있는 모듈(module) 개발을 통하여 사용자들이 쉽게 이용할 수 있는 환경을 지원한다.

1. 서 론

새로운 고객을 유치하는 마케팅 비용은 기존 고객을 잘 관리하는 것이 유지비용의 몇 배가된다는 기본적인 원칙에 의한 접근으로 CRM(Customer Relationship Management)이 개발되었다[1].

총체적인 CRM시스템은 마케팅 자동화, 고객 서비스 등을 담당할 수 있는 운영CRM과 고객의 정보 수집하여 수집된 정보를 규칙 기반의 데이터 마이닝(data mining) 기법으로 분석하여 보고서를 생성할 수 있는 CRM엔진부분으로 구성된다[1,2,3]. 대부분의 대규모 데이터는 동질성을 갖지 않는 데이터들로 구성되어 있기 때문에 기존의 통계적 분석 방법으로 취급하는 것은 의미가 없음을 지적하면서, 데이터 시각화 기법의 활용을 강조하게 된다[4]. 또한, 데이터의 양이 증가되면 데이터를 분석하는 방법에 대해 완전히 다시 생각해야 한다고 지적하면서 데이터에 관련된 여러 기술들과 데이터 마이닝 분야에 통계학자들의 연구와 관심의 필요성을 주장하였다[5]. 현대의 통계학과 데이터 분석은 기존의 통계적 모델이 중요한 것이 아니며, 대규모 데이터에서 검정의 결과가 지니는 '유의성(Significance)'은 아무런 의미도 갖지 못함을 언급하면서 데이터의 궁극적 목적은 패턴을 발견하는 것이다[6, 7].

본 논문에서 제안된 분석 CRM시스템은 데이터 마이닝과 신경망 이론을 이용하여 데이터를 분석하는 컴포넌트로 구성되며, 컴포넌트 모듈을 독립적으로 구성함으로써 재사용이 용이하고, 사용자들이 이용하기에 편리하도록 인터페이스를 구성되었으며, 커스터마이징(customizing)이 용이하여 컴포넌트 모듈들의 유지보수가 쉽게 구성되었다.

2. 기존 CRM에 관한 연구

고객 지향적인 마케팅에서 가장 중요한 것은 기업이 개별 고객에 관한 정보를 수집하고 축적하여 개별 고객에게 적합한 정보를 적시에 효율적으로 제공하는 기술이다. 즉, CRM은 이미 있어왔던 기업의 활동을 고객에

대한 체계적이고 광범위하게 이해하는 사고로 전환하여 고객의 평생가치를 증대시키는 경영 개선 행동이다. CRM의 시작을 위해 데이터웨어하우징, LAP, 데이터 마이닝 등이 필요하다[1,2].

1991년부터 시작하여 미국의 Siebel Systems는 영업 부문의 효율성 극대화, 영업의 전 과정을 협업화, 투명한 관리 및 예측을 가능하게 하는 SFA(Sales Force Automation)응용을 개발하였으며, SFA는 Enterprise Application 부문에서의 새로운 시장 기회로 포착되어 급속히 발전하였다.

1997년경에는 고객을 대하는 부서가 사용하는 응용이라는 의미에서 Front Office Application이라 불렀다. 인터넷의 대중화로 인터넷이 고객에게 손쉽게 다가서고 서비스, 마케팅, 판매비용을 획기적으로 줄일 수 있는 매체로 인식되면서 이후 인터넷을 기반으로 한 어플리케이션의 개발은 CRM 응용의 역사 그 자체가 되었다.

E-commerce, E-marketing, E-service 응용들이 등장했고, 효율적인 웹서버 로그분석기, E-mail Marketing tool, Personalization 등의 제품들이 등장했다.

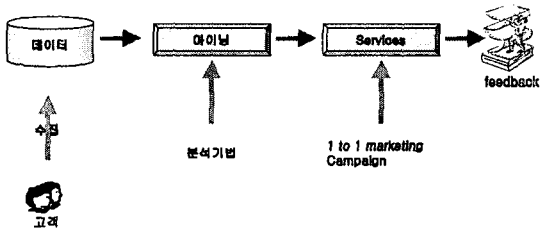
최근에 Siebel은 Native Web Architecture로 code를 다시 써서 Siebel 7을 출시하였다. Siebel의 우수함은 최고의 Business 분석의 전문가 집단에 의한 제품 디자인, 이를 실현시키는 실리콘 밸리 최고의 엔지니어들의 결합체이며, 100%고객만족도를 강조한다.

새로운 동향은 Realtime analysis나 분석적인 기능을 이용하여 운영CRM에 접목해 운용 CRM의 기능을 극대화하려는 시도가 보이고, Siebel, Oracle, PeopleSoft 등의 제품에 보완적인 역할을 하고 있다.

3. 신경망 이론을 이용한 분석 eCRM의 구현

신경망 이론을 도입한 분석 eCRM에서는 웹 로그 데이터를 포함하여 웹상에서 발생하는 다양한 데이터를 분석 대상으로 삼으며, 연구자에 따라 여러 가지 방법으로 데이터를 구분하고 있다. 본 논문에서는 웹 데이터를 데이터가 얻어지는 특성에 따라서 로그(log) 데이터,

고객(customer) 데이터, 마케팅(marketing) 관련 데이터로 구분한다. 특히 로그 데이터는 그 데이터의 특성상 분석하기 전에 사전처리(preprocessing)가 해야 하고, 데이터의 분석을 위해 사용자들의 방문 정보를 각각의 그룹으로 분류하는 세션구분(session identification) 작업이 수행된다.



(그림 1) CRM의 과정

(그림 1)는 CRM의 과정으로 고객으로부터 데이터를 수집하고, 데이터 마이닝 과정을 통하여 고객의 특성을 파악하며, 고객에게 효율적인 서비스를 제공하고, 피드백(feedback)하는 과정으로 이루어진다. 따라서 CRM은 데이터의 수집, 저장 및 관리를 위한 데이터 웨어하우스를 구축하면 데이터 마이닝 과정을 통하여 고객 특성에 대한 정보를 추출한다.

본 논문에서 제안된 eCRM 시스템은 수집된 데이터를 여러 분석 기법을 적용하여 분석하고, 분석된 정보를 제공하는 측면에 주안점을 두고 있다. 대용량 데이터의 계산 시간의 효율성을 위해서 요약 테이블(summary table)을 이용하였다. 요약 테이블의 원래의 목적은 통계 요약 정보를 2차원 테이블 형태로 관리하기 위한 것이며, 본 논문에서 이용하는 요약 테이블은 로그 데이터를 분석할 때 자주 이용되는 순차적인 정보들을 미리 계산해 저장하는 구조로 활용된다.

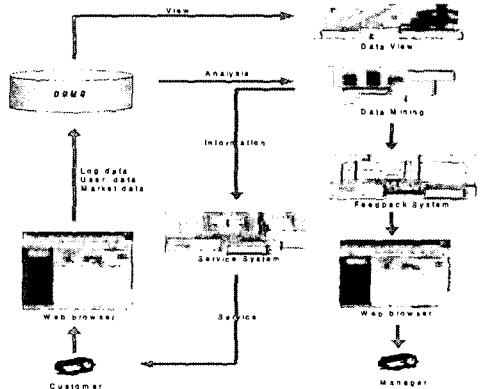
3.1 시스템 구조

eCRM시스템의 구조로 데이터 수집과 분석과정을 통하여 고객에게는 서비스를, 기업 관리자에게는 분석 정보와 피드백 정보를 제공하는 과정은 (그림 2)와 같다.

eCRM 시스템은 고객에 대한 다양한 데이터를 수집하고, 데이터 웨어하우스는 데이터의 저장과 데이터를 활용하기 위한 준비 단계로 무엇을 분석하기 위해서 어떤 데이터를 어떤 형태로 준비할 것인지를 먼저 고려한 후에 구축하게 된다. 데이터 준비에 대한 이러한 개념은 eCRM, OLAP 시스템과 같이 데이터베이스 안에 저장되어 있는 데이터를 분석하기 위한 시스템 설계 및 구현에 있어 매우 중요하며, 다른 데이터 분석용 어플리케이션과 차이점이라 할 수 있다.

서버 컴퓨터의 운영 환경은 Windows NT, 웹 서버로 IIS를, 데이터를 저장, 관리하기 위한 데이터베이스 관리 시스템은 SQL을 사용하였으며, 사용자들과의 인터페이스는 웹 브라우저를 이용한다. 데이터 분석에 필요한 모듈들은 Digital Visual Fortran과 IMSL, MS Visual C++가 제공하는 ATL(Active Template Library) 프로그래밍 기법을 이용하여 DLL형태인 COM(Component Object Model)으로 제작하여 사용하였다. 동적 웹 페이지

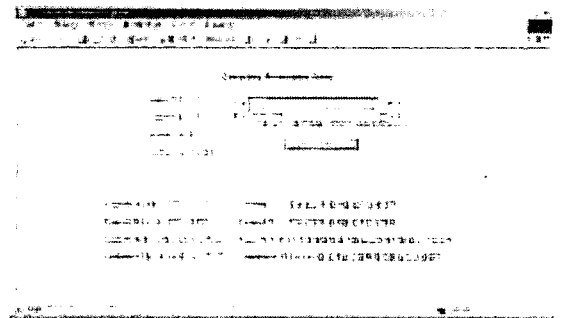
개발은 ASP, Dynamic HTML 등을 이용하였고, Chart를 보여주기 위해 ChartFX를 이용하였다. 또 고객들에게 메일을 이용한 서비스를 제공하기 위하여 SMTP를 이용하였다.



(그림 2) eCRM 시스템 구조

3.2 분석 eCRM 컴포넌트

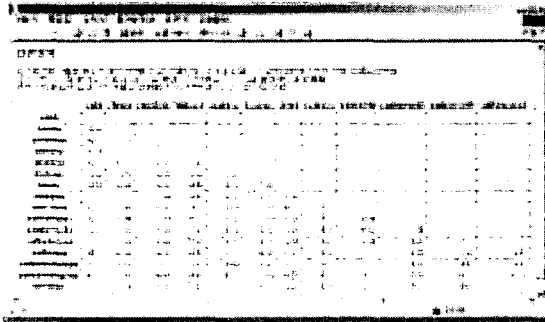
분석 eCRM 시스템은 운영적 측면에서 수집된 데이터를 분석 과정을 통하여 가치 있는 정보를 추출하는 컴포넌트로 구성된다. 과거에 알려지지 않은 가치 있는 정보들이 이 단계에서 얻어지고 이러한 정보를 바탕으로 고객의 요구를 충족시켜 줄 수 있다. 따라서 고객 관리의 수준도 높아지고 마케팅 전략도 효과적으로 수행할 수 있다. 분석 eCRM 부분에서는 데이터 마이닝 기법을 이용하여 데이터를 분석하기 위한 컴포넌트를 개발한다. 분석 CRM은 본 논문의 핵심적인 부분으로 패턴 발견을 위한 컴포넌트, 제품 및 고객 분석을 위한 컴포넌트, 군집화를 위한 컴포넌트, 분류 및 고객 세분화를 위한 컴포넌트, 고객 가치 평가를 위한 컴포넌트 등으로 구성된다.



(그림 3) 연관규칙

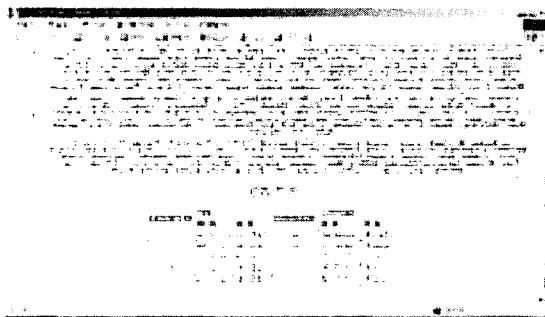
(그림 3)은 살펴보고자 하는 특정 웹 페이지를 입력하면 연관규칙을 탐사하여 그 결과를 보여주는 컴포넌트이다. 또한, 지지도(support)와 신뢰도(confidence)를 설정하고 이 조건에 맞는 모든 규칙들을 추출할 수 있다. 웹 사이트를 방문한 고객들의 순차패턴은 apriori All algorithm을 이용하여 추출하고 새로운 고객의 방문 패턴을 분류한다. 새로운 고객의 분류는 신경망 기법의 하나의 홉필드 네트워크(Hopfield network) 알고리즘을

이용하였다. 방문 패턴에 따른 이러한 고객 분류는 일대일 마케팅과 개인화된 서비스를 제공하기 위한 중요한 정보로 CRM에서 매우 의미 있게 사용된다. 분석된 데이터를 통해 전체 제품의 판매현황에 대한 정보를 확인할 수 있다.



(그림 4) 제품간의 상관관계

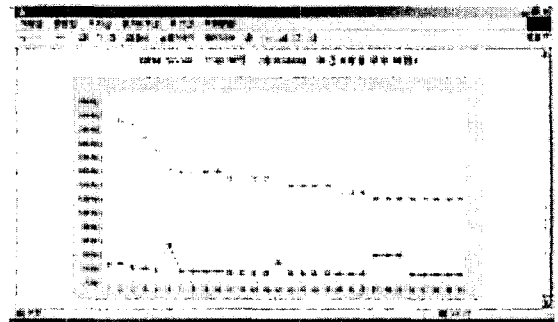
또한 제품간의 상관관계를 (그림 4)과 같이 나타내며 상관분석에 사용된 변수는 많이 이용되는 제품을 자동으로 검색하여 이용한다. 상관관계에 대한 분석은 제품들의 판매 연관성을 추출함으로써 제품 홍보 및 개인화된 서비스 등에 이용될 수 있다.



(그림 5) 고객 군집화

(그림 5)은 고객들을 특성에 따라 군집화하는 컴포넌트이다. 군집화 기법은 일대일 마케팅, 개인화된 서비스를 제공하기 위해 필수적으로 요구되는 분석 기법으로 eCRM에서 이 기법은 주로 고객 군집화와 웹 페이지 군집화에 이용된다. 고객에 대한 군집화는 일반적으로 성별, 나이, 수입 정도 등과 같은 몇몇 정형화된 변수(주로 사용자 데이터)를 이용한다. eCRM에서는 이러한 변수뿐만 아니라 로그 데이터나 Market 데이터도 이용된다. 예를 들어 사용자가 웹 페이지를 방문한 시간 또는 제품을 구입하는 순서 등의 정보를 이용할 수 있으며, K-평균 군집화 기법을 이용하였다.

eCRM의 목적은 고객에 관한 데이터 획득에서 시작하여 고객에 대한 정보를 추출하고 이를 바탕으로 고객과의 관계를 강하고 깊게 함으로써 기업의 이윤을 최대화하는 것이다. 이를 위해서는 고객 개인의 경제적 가치 파악이 필수적으로 요구된다. 고객들의 경제적 가치는 매우 다양하게 나타나며 수익성이 있는 고객도 있고 그렇지 못한 고객도 있다.



(그림 6) R-F-M 점수의 그래프

(그림 6)는 최근 6개월간의 우수 고객들의 R-F-M 점수를 그래프로 표현한 컴포넌트이다.

4. 결론 및 향후연구

eCRM 시스템에서 가장 중요한 역할을 하는 것은 정보를 추출하기 위한 데이터 분석 기법과 추출된 정보를 적시에 효율적으로 제공하는 정보 기술의 활용이라 할 수 있다. 따라서 데이터 분석 기법의 개발과 정보 기술의 효율적 활용 방법에 관한 연구는 매우 중요하다. 본 연구를 통하여 개발된 분석 eCRM시스템은 고객 분석 능력의 향상, 관리 지원 능력 향상, 마케팅 수행 능력 등을 향상 등에 도움을 줄 수 있으며, 쇼핑 사이트와 같은 실제 웹 사이트에서 이용하고자할 때 간단한 커스터마이징 절차를 거치면 쉽게 활용할 수 있고, 마케팅 효과를 높이는데 기여할 수 있을 것이다.

구현한 시스템은 웹 데이터의 분석 결과를 어떤 포맷으로 보고 또 어떻게 활용할 것인가의 문제는 기업의 환경을 고려해야 한다. 향후 연구로는 다양한 분석 기법을 활용할 수 있는 컴포넌트의 개발과 실제로 기업 환경에 적용함으로써 더 효율적인 시스템을 개발하기 위한 세부적인 전략이 추가해야 한다.

참고문헌

- [1] David Bradshaw, "eCRM/What is eCRM?", Ovum, 2001.
- [2] B. Eisenfeld, W. Close, "Benefits to be gained from eCRM", Gartner, December, pp. 1-6, 2000.
- [3] Arthur M. Hughes, "Strategic Database Marketing", McGraw-Hill, Second Edition, pp. 3-53.
- [4] Huber, P. J., "Huge Data Sets", COMPSTAT(Proceedings in Computational Statistics), pp. 3-13, 1994.
- [5] Friedman, J. H., "Data Mining and Statistics : What's the connection?", Proceedings of the International Conference on the Interface : Computing Science and Statistics, 1997, <http://www.stat.rice.edu/interface97.html>.
- [6] Hand, D. J., "Intelligence Data Analysis : Issues and Opportunities, Intelligence Data Analysis, Vol. 2, No. 2, pp. 1-14, 1998.
- [7] Famili, A., Shen, W. M., Weber, R. and Simoudis, E., "Data Preprocessing and Intelligent Data Analysis", Intelligent Data Analysis, Vol. 1, No. 1, 1997, <http://www-east.elsevier.com/ida/browse/voll.htm>.