

개선된 데이터 마이닝 기술에 의한 웹 기반 지능형 추천시스템 구축

김 경 재* · 안 현 철**

Development of Web-based Intelligent Recommender Systems using Advanced Data Mining Techniques

Kyoung-Jae Kim* · Hyunchul Ahn**

Abstract

Product recommender system is one of the most popular techniques for customer relationship management. In addition, collaborative filtering (CF) has been known to be one of the most successful recommendation techniques in product recommender systems. However, CF has some limitations such as sparsity and scalability problems. This study proposes hybrid cluster analysis and case-based reasoning (CBR) to address these problems. CBR may relieve the sparsity problem because it recommends products using customer profile and transaction data, but it may still give rise to scalability problem. Thus, this study uses cluster analysis to reduce search space prior to CBR for scalability problem. For cluster analysis, this study employs hybrid genetic and K-Means algorithms to avoid possibility of convergence in local minima of typical cluster analyses. This study also develops a Web-based prototype system to test the superiority of the proposed model.

Keywords : Product Recommender System, Genetic Algorithm, Market Segmentation, Customer Relationship Management

논문접수일 : 2005년 5월 20일

논문게재확정일 : 2005년 7월 26일

※ 이 논문은 2003년도 한국학술진흥재단의 지원에 의하여 연구되었음(KRF-2003-041-B00181).

* 주저자, 동국대학교 경영대학 정보관리학과, (100-715)서울 중구 필동 3-26 동국대학교 경영대학 정보관리학과,
Tel : 02-2260-3324, Fax : 02-2260-3684, e-mail : kjkim@dongguk.edu

** 한국과학기술원 테크노경영대학원

1. 서론

전자상거래의 확산에 따라 인터넷 쇼핑물에서의 구매행위는 일반화된 현상이 되었다. 인터넷 쇼핑물의 고객은 시간적, 지역적인 제한 없이 언제, 어디서나 인터넷 환경이 제공되는 공간에서 실시간으로 쇼핑을 할 수 있다. 이러한 이유로 인터넷 쇼핑물의 고객들은 인터넷에 존재하는 수많은 쇼핑물에서의 비교구매를 통하여 보다 저렴한 가격에 양질의 상품을 제공하는 쇼핑물을 찾아다니게 되어 특정 쇼핑물에 대한 충성도가 낮은 “가격 추종자(price-butterfly)”의 특징을 가지게 되었다. 이런 새로운 구매행태의 출현은 인터넷 쇼핑물 운영자에게 오프라인 환경의 쇼핑물보다 차별화된 마케팅 전략을 요구하고 있다.

이러한 요구에 부응하기 위해 많은 인터넷 쇼핑물 운영자들이 데이터베이스 기술을 활용한 일대일 마케팅(one-to-one marketing) 전략을 이용하고 있으나 잘못된 정보기술의 활용은 오히려 고객들의 불만사항이 되기도 한다. 이의 대표적인 예가 차별화되지 않은 광고성 스팸메일의 발송이다. 고객들은 자신들의 구매의사와 관계없이 수신되는 스팸메일에 오히려 해당 인터넷 쇼핑물에 대한 불만과 불신을 가지게 되기도 한다. 이러한 불만을 최소화하면서 고객에게 판매촉진을 하기 위해서는 고객 별로 개인화된 마케팅 전략이 필요하다. 따라서 인터넷 쇼핑물의 운영자에게 있어서 각 고객의 구매성향을 파악하고 고객 개인별로 차별화된 서비스를 제공하는 것은 매우 중요한 과제로 인식되고 있다.

고객의 구매성향을 파악하는 대표적인 방법으로는 구매성향이 유사한 고객들을 하나의 집단으로 군집화하는 시장세분화가 있다[Smith, 1956]. 시장세분화를 통하여 운영자는 무엇이 고객의 구매의사를 자극하는지, 각각의 고객들

과 어떻게 연결될 수 있는지를 파악할 수 있고 이에 의해 고객의 충성도를 제고함으로써 고객의 생명가치(lifetime value)를 증가시킬 수 있다. 시장세분화는 오프라인 환경에서도 데이터베이스 마케팅의 기본적인 수단으로 이용되고 있으며, 이에 따라 오프라인 환경에서의 시장세분화에 대한 연구는 많이 진행되어 왔다. 그러나 온라인 환경에서의 시장세분화에 대한 연구는 상대적으로 적게 이루어진 것이 현실이다.

인터넷 쇼핑물은 개인 컴퓨터 화면을 통해 자연적으로 대인 서비스를 제공하므로 개인화된 마케팅 서비스의 효과는 매우 높을 가능성이 있다. 온라인 환경에서의 시장세분화는 쇼핑물의 개인화(personalization)를 위한 기초자료가 될 수 있으며 일대일 마케팅의 기본적인 수단이 될 수 있으므로 오프라인 환경에서의 시장세분화보다 더 중요한 의미를 가질 수 있다. 한편, 온라인 환경에서는 오프라인 환경의 쇼핑물 고객과 달리 구매의사가 없이 단순히 방문하는 고객의 경우에도 회원가입과 로그인 과정을 통해 인적사항 등 인구통계학적인 자료를 수집할 수 있으므로 보다 정교한 시장세분화를 할 수 있는 특징이 있다.

시장세분화를 통해 유사한 구매성향을 가지는 고객군을 생성하면 이에 따라 개인화된 마케팅 서비스를 제공할 수 있다. 시장세분화를 통해 얻어진 각각의 고객군의 구매성향에 따른 선호상품목록은 유사한 구매성향을 가진 다른 고객의 과거 구매이력을 통해 생성할 수 있으며, 이에 따라 개인화 서비스를 제공할 수 있는 것이다.

본 연구에서는 이러한 요구에 부응할 수 있는 개인화된 마케팅 서비스로서 개선된 상품추천 기법을 제안할 것이다. 시장세분화와 상품추천 기법에 있어서는 전통적으로 군집분석과 협업 필터링(collaborative filtering)이 많이 사용되어

왔다. 그러나 시장세분화에 사용되어온 전통적인 군집분석기법은 각 군집의 중심점을 찾는 과정에서 국부탐색기법을 이용하기에 최적의 군집중심점을 찾지 못할 가능성이 중요한 한계점으로 지적되어 왔다[Selim & Iamail, 1984; Murthy & Chowdhury, 1996; Pena et al., 1999; Bandyopadhyay & Maulik, 2002]. 한편, 대표적인 상품추천기법인 협업필터링의 경우에는 희박성(sparsity)과 확장성(scalability)의 문제가 한계점으로 지적되고 있다[김재경 등, 2002; Cho et al., 2002; Kim et al., 2002; 김재경 등, 2003; 김종우 등, 2004; 조운호 등, 2004; Cho & Kim, 2004; 김재경 등, 2005].

본 연구에서는 선행연구에서 사용된 방법들의 한계점을 보완할 수 있는 새로운 추천기법을 제안한다. 본 연구에서 제안하는 방법은 2단계로 이루어져 있는데, 1단계에서는 협업필터링의 확장성 문제를 완화하며 동시에 시장세분화의 효과를 얻을 수 있는 군집분석을 수행하고, 2단계에서는 각 군집 내에서 사례기반추론을 활용하여 상품추천을 하는 방법이다. 이 방법은 전통적으로 추천시스템에 활용되어 온 협업필터링에 비해 희소성과 확장성의 문제를 완화하여 추천의 정교성과 현실적용 가능성을 제고할 수 있을 것으로 기대된다. 한편, 군집분석의 방법에 있어서는 전통적으로 많이 활용되어 온 K-Means 군집분석의 한계점인 군집 중심점의 국부 최적해 수렴가능성을 보완한 방법으로 유전자 알고리즘에 기반한 K-Means 군집분석(GA K-Means 군집분석)을 제안하여 시장세분화의 정교성을 더욱 제고하고자 한다. 본 연구에서 제안하는 기법은 전통적으로 시장세분화에 사용되어온 군집분석기법의 한계점과 추천상품목록의 생성에 많이 사용되어온 협업필터링의 한계점을 보완할 수 있을 것으로 기대된다.

본 연구에서는 제안하는 방법론의 유용성을

확인하기 위해 전통적으로 오프라인 환경에서 시장세분화에 많이 사용되어 온 K-Means 군집분석에 기반한 시장세분화와 최근에 데이터 마이닝 분야에서 시장세분화의 한 수단으로 많이 활용되고 있는 자기조직화지도(self-organizing map)에 의한 군집분석을 수행하고, 제안한 유전자 알고리즘에 기반한 K-Means 군집분석을 이용한 시장세분화 방법과 비교하여 분석할 것이다. 또 제안한 방법에 의해 구성된 군집 내에서 사례기반추론을 수행하여 상품추천을 할 것이다. 본 연구에서는 가상의 인터넷 쇼핑몰을 구축하여, 제안하는 방법을 상품추천시스템 내 추천기법으로 이용하고 이를 통해 고객의 만족도를 확인하는 과정을 통해 제안하는 방법의 유용성을 확인할 것이다.

본 연구는 다음과 같이 구성되어 있다. 먼저, 2장에서는 선행연구에서 사용된 상품추천기법들과 시장세분화 방법들을 살펴보고, 그 장단점을 확인한다. 3장에서는 확인된 장단점을 기반으로 단점을 보완할 수 있는 2단계 상품추천시스템을 제안한다. 4장에서는 먼저 실험에 사용할 데이터에 대해 설명하고 제안추천기법의 유용성을 확인할 수 있는 웹 기반의 프로토타입 시스템을 설계한다. 5장에서는 제안 추천시스템의 유용성 확인 결과를 살펴보고, 마지막 6장에서는 연구의 결론과 한계점을 제시한다.

2. 연구배경

2.1 기존 추천기법의 한계점

인터넷 쇼핑몰에서의 상품추천시스템은 실생활에서도 활발하게 이용되고 있는데, 그 대표적인 예로는 아마존(Amazon), 무비렌즈(Movie-Lens), 시디나우(CD Now), 제이씨 페니(JC Penney) 등이 있다. 상품추천시스템의 가장 핵심적인 부분은 추천알고리즘이라 할 수 있으며, 대

표적인 추천알고리즘은 협업필터링과 내용기반 필터링 등의 방법이 있다. 두 방법은 각각 장단점을 가지고 있으나 현실적으로는 협업필터링이 선행연구에서 더 활발하게 이용되고 있다[김재경 등, 2002; 김중우 등, 2004; 조운호 등, 2004; 김재경 등, 2005 참고].

협업필터링은 사용자 사이의 연관성을 기반으로, 선호도 또는 구매 패턴이 유사한 고객군을 분류하고, 유사한 고객에 속하는 다른 사람들이 선호하는 상품을 추천하는 방법이다[Funakoshi and Ohguro, 2000]. 협업필터링에 대한 초기 연구로는 Tapestry[Goldberg et al., 1992], GroupLens[Resnick et al., 1994] 등의 사례가 대표적이며, Ringo & Video Recommender 등과 같은 이메일과 웹 기반의 협업필터링 기법에 의한 추천시스템 등이 있다[Sarwar et al., 2000]. 협업필터링은 일반적으로 고객들이 동질적인 평가결과를 보이는 상품군에 대해 상대적으로 높은 예측력을 보이며, 데이터가 충분한 경우에는 다른 기법에 비해 상대적으로 높은 예측력을 보이는 장점을 가지고 있다[Konstan et al., 1997; Pazzani, 1999]. 이에 따라 협업필터링은 상품추천시스템 관련 연구에서 활발하게 이용되고 있으나 아래와 같은 한계점도 가지고 있다.

협업필터링은 기본적으로 상품에 대한 고객의 선호도 또는 구매이력자료를 바탕으로 추천을 하게 되므로 구매이력을 많이 보유하고 있는 대형 인터넷 쇼핑몰에서는 유용하지만, 구매이력이 상대적으로 부족한 중소 인터넷 쇼핑몰이나 사업 초기단계의 인터넷 쇼핑몰의 경우에는 적용 가능성이 떨어진다고 할 수 있다[안현철과 한인구, 2002]. 즉, 협업필터링의 속성상 구매이력이 부족한 경우에는 추천의 성과가 떨어질 수 밖에 없으며 이런 점은 이미 선행연구에 의해 협업필터링의 가장 중요한 문제점 중 하나로 지적되고 있다[김재경 등, 2002; Cho et

al., 2002; Kim et al., 2002; 김재경 등, 2003; 김중우 등, 2004; 조운호 등, 2004; Cho & Kim, 2004; 김재경 등, 2005]. 이런 문제점을 흔히 희박성(sparsity) 문제라고 하며, 희박성 문제를 완화하기 위해 선행연구에서는 웹 로그 정보를 활용하여 간접적으로 선호도 데이터를 보충하고자 하였다[Cho et al., 2002; Kim et al., 2002; 김재경 등, 2003; Cho & Kim, 2004; 김재경 등, 2005]. 그러나 웹 로그 정보는 일반적으로 대용량이며 정제되지 않은 형태이어서 고객의 선호도 점수를 직접 취득하는 것만큼 전처리에 많은 시간과 비용이 소요된다는 단점이 있다.

협업필터링의 중요한 한계점 중 다른 하나는 고객과 거래 데이터가 증가함에 따라 유사한 고객군을 찾기 위한 연산량이 기하급수적으로 증가하는 현상이 발생할 수 있다는 것이다[김재경 등, 2002; Cho et al., 2002; Kim et al., 2002; 김재경 등, 2003; 조운호 등, 2004; Cho & Kim, 2004; 김재경 등, 2005]. 선행연구에서는 이런 문제점을 확장성(scalability) 문제라 하였는데, 이는 해결해야 하는 문제가 제시된 이후에야 추론을 시작하는 ‘게으른 학습방법(lazy learning technique)’의 일반적인 특징으로, 신속한 응답을 원하는 인터넷 사용자의 특성을 감안할 때 고객의 이탈을 유도할 수 있는 치명적인 한계점이다. 선행연구에서는 이 문제점을 보완하기 위해서 아래와 같은 여러 방법들을 제안하였다.

김재경 등[2002], Cho et al.[2002], Kim et al.[2002], 조운호 등[2004], Cho & Kim[2004], 김재경 등[2005] 등은 확장성과 희박성의 문제를 보완하기 위하여 상품계층도(product taxonomy)를 활용하는 방법을 제안하였으나, 여전히 하나의 상품계층도의 각 상품계층군 안에서는 각 고객의 선호도가 제대로 반영되지 않아서 추천의 성과가 떨어지는 경우가 발생할 수 있

다. 또, 상품계층도의 작성이 상품추천의 성과에 큰 영향을 미칠 수 있는데 선행연구에서는 전문가의 주관적인 판단을 참고하는 방식으로 연구를 진행하였으나 이 점 역시 한계점이 될 수 있다.

확장성 문제를 완화하기 위한 또 다른 노력으로는 군집분석을 협업필터링의 사전과정으로 수행하여 탐색공간을 줄이는 방법이 있다. 김재경 등[2003]은 K-Means 군집분석을 협업필터링 사전단계로 활용하여 탐색공간을 축소하였고, Roh et al.[2003]과 강부식[2003]은 군집분석 기법의 하나인 자기조직화지도를 활용하여 사례탐색공간을 축소하였다. Kim & Han[2001]은 협업필터링과 함께 '게으른 학습방법'의 하나인 사례기반추론에서 자기조직화지도 분석을 추론 이전 단계에 활용하여 분석 데이터의 양을 줄이고자 하였다. 그러나 이 같은 선행연구들은 '게으른 학습방법'에서 확장성 문제의 심각성을 인지하고 이에 대한 보완방법으로 군집분석의 방법을 이용하는 것이 유용하다는 주장만 하였을 뿐, 제안되었던 상품추천시스템의 성능을 향상시키는데 결정적인 요인이 될 수 있는 군집분석의 방법론적 개선에는 큰 관심을 두지 않았다.

선행연구들의 결과를 종합해 보면 협업필터링이 상품추천에 있어서 주요한 방법이지만, 희박성과 확장성의 문제가 한계점이며, 이 한계점을 보다 정교하게 보완하여 희박성과 확장성의 정도를 완화시키고 이에 따라 추천의 성능을 제고할 수 있는 방안에 대한 연구가 필요하다고 할 수 있다.

2.2 기존 시장세분화 방법의 한계점

전통적으로 K-Means 군집분석과 같은 군집분석 방법은 시장세분화의 한 수단으로 이용되어 왔다[Gehrt & Shim, 1998 ; Kuo et al., 2002a ;

Kuo et al., 2002b]. 최근에는 무감독 인공신경망 기법의 하나인 자기조직화지도(self-organizing maps)도 시장세분화의 수단으로 활용되고 있다 [Vellido, 1999 ; 송수섭, 이의훈, 2001 ; 하성호, 이재신, 2003]. Kuo et al.[2002a ; 2002b]에서는 자기조직화지도와 K-Means 군집분석을 결합한 이단계 군집분석을 시장세분화의 방법으로 제시하기도 하였다. 이런 선행연구들은 군집분석이 시장세분화의 수단으로써 유용함을 주장하였으나, K-Means 군집분석과 자기조직화지도와 같은 일반적인 군집분석의 한계점에 대한 고려를 하지 않았다. 본 연구에서 제안하는 군집분석의 장점을 설명하기 위해 상기 군집분석 방법들을 간단히 설명하면 다음과 같다.

군집분석기법에는 크게 계층적 군집분석과 비계층적 군집분석이 있으며[Anderberg, 1973], 비계층적 군집분석기법 중 K-Means 군집분석이 가장 널리 활용되고 있다[Murthy & Chowdhury, 1996]. K-Means 군집분석은 사전에 정해진 군집의 수(k)를 기반으로 임의의 군집 중심점을 정하고, 이후에는 k개의 시작점을 중심으로 유사성에 근거하여 군집화를 수행하는 방법이다. K-Means 방법에서는 한번 군집이 만들어 질 때마다 각 군집 별로 그 군집의 평균값을 구하여 군집 내 대상들 간의 유클리드 거리의 합을 구하게 되며 이 값이 더 이상 작아지지 않을 때까지 군집화를 수행하게 된다.

K-Means 군집분석은 여러 가지 군집분석 방법들 중 가장 대표적인 방법이지만 중요한 한계점들을 가지고 있다. 대표적인 한계점들은 최종 군집의 수인 K를 미리 지정해야 하므로 데이터에 대한 사전지식이 충분히 확보되어야 하는 점, 최종 군집분석 결과가 초기 군집중심점에 의존하게 되는 점, 각 군집의 중심점을 탐색해 나가는데 있어서 지역 최적화 알고리즘(local

optimization algorithm)에 기반하고 있어서 지역 최적해(local minima)에 수렴할 수 있다는 점 등이다[Selim & Ismail, 1984 ; Babu & Murty, 1993 ; Murthy & Chowdhury, 1996 ; Pena et al., 1999 ; Bandyopadhyay & Maulik, 2002 ; Kuo et al., 2005]. 따라서 K-Means 군집분석에 기반한 시장세분화의 결과는 궁극적으로 전역 최적해에 도달하지 못할 수 있는 한계점이 있으며 이에 따라 시장세분화의 결과가 왜곡될 수 있다.

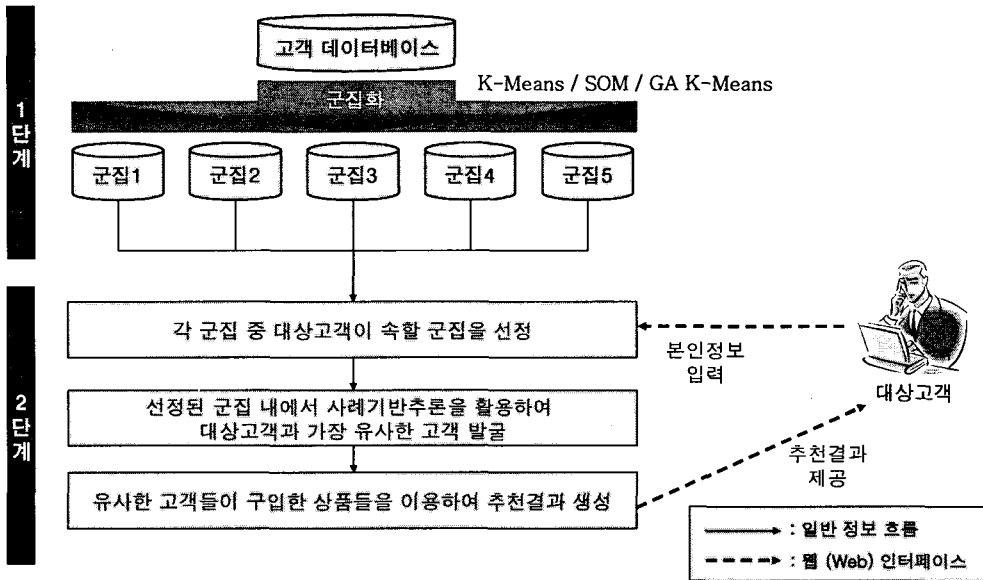
한편, 최근에는 K-Means 군집분석 이외에 신경망 모형의 일종인 자기조직화지도[Kohonen, 1995]가 많이 활용되고 있다. 자기조직화지도는 일반적으로 입력층과 출력층, 두 개의 층으로 구성되어 있으며, 입력층 뉴런(neuron)은 각 출력층 뉴런에 입력패턴을 배열시키는데, 유사한 패턴을 가진 입력층 뉴런은 동일한 출력층 뉴런으로 배열된다. 자율학습을 통하여 입력패턴과 가장 유사한 연결강도를 갖는 출력층 뉴런j가 워너(winner)뉴런이 된다. 그리고 이 워너 뉴런을 중심으로 반경(radius)을 설정하면, 이 범위 내의 모든 뉴런들은 출력층에서 유사한 연결강도를 갖는 입력패턴의 하위집합이 되고, 이 집합이 입력공간의 지도를 형성하게 되는데, 이것이 자기조직화지도이다. 유사도 측정에는 주로 유클리드 거리의 개념을 사용하며 규정된 반복 횟수만큼 학습이 진행되면 반경과 학습률을 감소시킨 후 다음 학습과정을 반복하게 된다. 자기조직화지도는 인공신경망의 정교한 연산능력을 이용한 군집화를 가능하게 하므로 시장 세분화에 적합한 방법으로 제안되고 있다[Vellido et al., 1999 ; 송수섭, 이의훈, 2001 참고]. 그러나 자기조직화지도 역시 군집 중심점을 찾는 과정이 지역 최적화 알고리즘에 기반함으로써 궁극적으로 전역 최적해에 도달하지 못할 가능성이 있다는 한계점이 있다.

3. 연구내용 및 방법

본 연구에서는 선행연구의 한계점을 보완하기 위해 2단계로 구성된 상품추천시스템을 제안한다. 구체적으로 본 연구에서 제안하는 상품추천시스템은 협업시스템의 희박성 문제를 완화하기 위해 고객 정보에 기반한 사례기반추론 방식의 상품추천시스템을 제안한다. 그런데, 사례기반추론 방식도 협업필터링의 경우와 마찬가지로 확장성의 문제가 나타나게 되므로 이에 대한 보완방안으로 군집분석을 사례기반추론 이전 단계에 적용한다. 또한 제안 시스템의 성능을 평가하기 위해 가상의 인터넷 쇼핑몰을 구축하고 이에 의해 고객 만족도를 조사하는 방식을 이용한다. 본 연구에서 제안하는 2단계 상품추천시스템의 전체적인 구성도는 <그림 1>과 같다.

1단계 : 확장성 문제 완화와 시장세분화를 위한 개선된 군집분석의 적용

첫 번째 단계에서는 사례기반추론 이전에 군집분석을 수행한다. 군집분석은 선행연구에서도 협업필터링과 사례기반추론에서의 확장성 문제를 해결하기 위해 사용되어 왔다[Kim & Han, 2001 ; 강부식, 2003 ; 김재경 등, 2003 ; Roh et al., 2003]. 군집분석이 사례기반추론이나 협업필터링을 수행하기 이전에 수행되면, 유사한 속성을 가지는 군집 내에서만 사례기반추론이나 협업필터링을 수행할 수 있으므로, 데이터의 증가에 따라 연산의 양이 기하급수적으로 증가하는 확장성의 문제를 감소시킬 수 있으며, 또한 유사한 군집 내에서 유사성이 높은 이웃 사례를 탐색하므로 모든 데이터를 탐색해야 하는 기존의 방식보다 유사성이 높은 이웃 사례를 찾을 가능성이 높다. 한편, 군집분석은 상품추천 이전에 고객 데이터를 대상으로 수행되므로 유사한 속성을 가지는 고객군을 같은 군집으로 묶어 주는 시장세분화의 효과도 동시에 기대할 수 있다.



〈그림 1〉 추천시스템의 구성도

본 연구에서는 군집분석을 수행함에 있어서 연구배경에서 지적인 기존 기법들의 한계점을 인지하고 이를 보완할 수 있는 유전자 알고리즘 (genetic algorithms) 기반의 K-Means 군집분석을 제안한다. 본 연구에서는 이 기법을 GA K-Means 군집분석이라고 한다. 이 방법은 K-Means 군집분석과 자기조직화지도의 단점인 군집분석의 결과가 지역 최적해에 수렴할 가능성을 최소화하고 전역 또는 유사전역 최적해에 도달할 수 있도록 하기 위하여 군집중심점을 탐색하는 알고리즘으로 전역 최적화 방법인 유전자 알고리즘을 이용한다. 선행연구에서도 유전자 알고리즘 기반의 군집분석방법이 기존 군집분석기법의 지역 최적해 수렴가능성을 줄이고 군집화의 성과를 제고하는 것으로 보고하고 있다[Babu & Murty, 1993 ; Murthy & Chowdhury, 1996 ; Pena et al., 1999 ; Bandyopadhyay & Maulik, 2002 ; Tsai & Chiu, 2004 ; Kuo et al., 2005].

제안된 방법의 장점을 설명하기 위해 유전자 알고리즘의 작동원리를 간단히 설명하면 아래

와 같다. 유전자 알고리즘은 재생산, 교차, 돌연변이 등의 생물의 진화과정을 모사하여 적응적으로 탐색공간을 탐색하여 최적 또는 유사 최적해를 찾아내는 전역 최적화 알고리즘이다[Goldberg, 1989 ; Adeli and Hung, 1995]. 보다 구체적인 과정을 살펴보면, 유전자 알고리즘은 우선 임의의 값을 가진 초기화된 개체집단을 생성한다. 즉, 전체 탐색공간 내에서 임의의 n 개의 개체들을 선택하여 개체집단(population)을 형성하게 된다. 이렇게 생성된 개체집단이 문제에 얼마나 적합한 지를 적합도 함수(fitness function)를 이용하여 평가한다. 다음 세대의 생성은 평가된 개체집단을 확률적으로 선택(selection)하거나, 교배(crossover)와 돌연변이(mutation) 등의 과정을 통해 이전 세대와 다른 새로운 개체를 생성한다. 다음 세대의 개체집단은 다시 적합도 함수에 의해 평가를 하여 종결조건에 도달할 때까지 즉, 목표로 한 적합도 수준에 도달하거나 미리 정해진 최대 진화 수에 도달할 때까지 반복적으로 다음 세대를 생성한다[조희연, 김영민, 2003]. 모든 세대에 대한 평가가 끝나면

모든 세대의 적합도 정도를 파악하고 가장 최적의 적합 정도를 나타낸 세대를 선택하게 된다. 유전자 알고리즘을 수행하기 위해서는 탐색공간 내의 다양한 변수 집합을 염색체(chromosome)라고 불리는 선형 스트링(string)에 사상(mapping)하는 과정이 필요하다.

유전자 알고리즘은 지역 최적화 알고리즘을 기반으로 한 역전파 인공신경망(back-propagation neural networks)의 단점을 보완하기 위해 인공신경망과 결합되어 사용되기도 하고, 기타 인공지능기법 또는 데이터 마이닝 기법에서 지역 최적해를 찾기 위해 결합되어 사용되기도 한다[Maniezzo, 1994 ; Ignizio and Soltys, 1996 ; Sexton et al., 1998 ; Sexton et al., 1999]. 본 연구에서는 시장세분화를 위한 군집 중심점을 탐색하는 데에 유전자 알고리즘을 이용함으로써 군집분석의 결과가 지역 최적해에 수렴하지 않고 전역 최적해를 찾아 낼 수 있는 개선된 데이터 마이닝 기술을 제안한다.

2단계 : 추천목록 생성과 희박성 문제를 완화하기 위한 사례기반추론의 이용

1단계에서 GA K-Means 군집분석을 통해 시장세분화를 수행하고 이에 의해 도출된 결과는 2단계에서 상품추천에 활용된다. 전술한 바와 같이 상품추천의 선행연구에서 많이 활용되어 왔던 협업필터링은 희박성의 문제가 발생할 수 있으므로 본 연구에서는 사례기반추론 방식의 상품추천시스템을 이용한다.

사례기반추론은 사례의 유사성을 평가하여 유사한 이웃 사례를 찾아내고 이웃 사례의 결과를 이용하여 새로운 사례에 대한 대응을 할 수 있다는 점에서 협업필터링과 유사하다. 선행연구에서는 사례기반추론을 협업필터링의 한 방법으로 분류하고 있는 경우도 있다[Berry & Linoff, 2004]. 그러나 사례기반추론은 고객의 인구통계학적 속성만으로도 유사한 고객을 추출할 수 있

으며, 이에 따라 선호도가 입력되지 않은 상품에 대해서도 추천이 가능한 장점이 있다.

4. 실험데이터와 실험설계

4.1 실험데이터

제한한 연구모형의 유용성을 확인하기 위해, 본 연구에서는 실제 상거래에서 추출한 데이터를 모형에 적용하여 그 결과를 도출한다. 분석에 사용되는 데이터는 2000년 7월에 개업한 국내 G 인터넷 쇼핑몰의 거래데이터로서, 이 쇼핑몰은 국내 최대 규모의 다이어트 상품전문 인터넷 쇼핑몰이다. 오프라인에서 이미 다이어트 분야에 확고한 입지를 가지고 있는 P사가 운영하고 있는 G 사이트는 방대한 콘텐츠와 우수한 서비스, 그리고 운영사의 신뢰도 높은 브랜드 이미지로 인해 국내 온라인 다이어트 분야에서 선두를 달리고 있다.

G 사이트로부터 확보한 데이터는 2001년 5월부터 8월 사이에 상품을 구매한 총 3298명에 대한 고객 정보 및 구매 내역 데이터로서, 이 중, 자료의 밀도가 충실한 총 2633명을 대상으로 분석한다. 데이터의 변수는 전처리 이후를 기준으로 해서 총 42개인데, 이 중 37개의 변수가 명목형(nominal) 변수이고, 나머지 5개가 비율(ratio)척도로 된 변수이다. 각 변수들은 대상 사이트가 다이어트 분야의 전문 쇼핑몰이므로, 구매자의 연령, 체중, 키, 기타 건강 정보 등 고객의 개인적인 특성과 관련된 정보가 대부분을 이루고 있다. 또한 전체 대상 데이터 중에서 가장 많은 상품을 구매한 고객이 총 8개의 상품을 구입한 점을 감안해, 각 레코드마다 총 8개의 제품구매내역 변수를 설정하고, 각 고객이 구입한 모든 상품의 ID코드를 이곳에 저장하여, 차후 추천 엔진에서 이 정보를 활용할 수 있도록 하였다. 분석에 사용된 42개 변수의 이름과 간단한 설명은 다음의 <표 1>과 같다.

〈표 1〉 분석에 사용된 변수 내역

변수명	내 용
add0	“서울 지역” 거주 여부
add1	“대도시 지역” 거주 여부
add2	“기타 지역” 거주 여부
age	만 나이
bmi	비만도(키와 체중의 비율)
d1	“부종” 경험 여부
d2	“변비” 경험 여부
d3	“지방간, 고혈압, 고지혈증 등 성인병” 경험 여부
d4	“빈혈 또는 영양실조” 경험 여부
e01	“식사대용식 다이어트 전문 제품(식사대신)”의 경험 여부
e02	“기능성 다이어트 전문 제품”의 경험 여부
e03	“이뇨제, 설사제 등 약을 이용한 다이어트”의 경험 여부
e04	“무조건 굶기”의 경험 여부
e05	“원푸드 다이어트, 덴마크 다이어크 등 민간요법”의 경험 여부
e06	“단식원”의 경험 여부
e07	“비만클리닉”의 경험 여부
e08	“한방다이어트”의 경험 여부
e09	“토닝 시스템”의 경험 여부
e10	“기타 다이어트 방법”의 경험 여부
height	키
loss1	배 부위의 살을 빼고 싶은지 여부
loss2	엉덩이 부위의 살을 빼고 싶은지 여부
loss3	허리, 옆구리 부위의 살을 빼고 싶은지 여부
loss4	얼굴 부위의 살을 빼고 싶은지 여부
loss5	등 부위의 살을 빼고 싶은지 여부
loss6	다리, 허벅지 부위의 살을 빼고 싶은지 여부
loss7	팔 부위의 살을 빼고 싶은지 여부
loss8	가슴 부위의 살을 빼고 싶은지 여부
loss9	어깨 부위의 살을 빼고 싶은지 여부
married	결혼 여부
obj	감량 목표
occu0	직업이 “회사원”인지의 여부
occu1	직업이 “주부”인지의 여부
occu2	직업이 “학생”인지의 여부
occu3	직업이 “자영업”인지의 여부
occu4	직업이 “기타”인지의 여부
pur0	다이어트의 목적이 ‘미용’인지의 여부
pur1	다이어트의 목적이 ‘취업, 결혼 등 급한 이유’인지의 여부
pur2	다이어트의 목적이 ‘기타’인지의 여부
sex	성별
waist	허리사이즈
weight	체중(현재 시점에서의 체중)

4.2 실험 설계

본 연구에서 제안한 추천시스템의 유용성을 실증적으로 확인하기 위하여 다음과 같이 실험 설계를 한다.

4.2.1 고객 데이터베이스의 군집화

본 단계에서는 전통적으로 시장세분화에 많이 활용되어 온 K-Means 군집분석과 최근에 시장세분화 연구에서 많이 활용되고 있는 자기조직화지도에 의한 군집분석을 수행하고 그 결과를 본 연구에서 제안하는 GA K-Means 군집분석 결과와 비교한다. 자기조직화지도는 상용 S/W인 Neuroshell 2 Release 4.0을 이용해 실험하며, 이 때 학습율(α)은 0.5로 설정한다. K-Means나 GA K-Means의 경우, MS-EXCEL과 VBA(Visual Basic for Application)로 실험용 모형을 구현하고, GA K-Means의 유전자 알고리즘 부분은 Palisade Software社의 Evolver 4.06을 활용해 개발한다.

실험설계를 위한 GA K-means 군집분석의 절차는 크게 다음의 4단계로 이루어진다.

- 1) 궁극적으로 GA K-Means의 목적은 군집화해야 할 대상 중에서 군집을 최적으로 구분할 수 있도록 해주는 최적의 군집중심점을 찾는 것이다. 이러한 작업을 위해, 각 레코드는 자신만의 고유한 ID번호를 갖게 되며, GA K-Means에서는 구분하고자 하는 군집의 개수(K)만큼의 ID들을 선별해 하나의 검색체에 사상시켜, 이를 최적화 하게 된다. 본격적인 최적화 작업에 앞서, 1단계에서는 이러한 검색체 내의 유전자 속성값을 임의의 값으로 초기화 하는 작업이 이루어지게 된다.
- 2) 2단계에서는 주어진 검색체 값에 따라 파라미터들을 정하고, 그 값들을 기반으로 K-Means 군집분석을 수행하게 된다. 즉, 주어

진 검색체에서 규정하고 있는 초기 군집중심점에 따라, 각 레코드들을 각각의 군집으로 세분화 하는 것이다. 세분화가 완료되면, 군집화가 얼마나 잘 이루어졌는지를 확인하기 위해, 군집간 관성지표(Intraclass inertia)라는 지표를 계산하게 된다. 군집간 관성지표는 다차원 공간에서 각 군집들의 조밀성 정도를 나타내는 척도이며, 각 군집 내에서 개체와 군집중심점 간의 거리의 합으로 계산된다 [Michaud, 1997; Laflamme & Eilert-Petersson, 1998]. 이 지표는 Michaud[1997]와 Shin & Sohn[2004] 등의 연구에서 군집분석의 성능을 확인하는 척도로 사용되었다. N개의 군집으로 분류한 경우의 군집간 관성지표는 아래의 식 (1)에 의해 계산된다.

$$F(N) = \frac{1}{n} \sum_{K=1}^N I_K \quad (1)$$

$$= \frac{1}{n} \sum_{K=1}^N \sum_{i=1}^{n_K} \sum_{P=1}^M (X_{iP} - \bar{X}_{KP})^2$$

(n : 전체 레코드수, n_K : K번째 군집의 레코드수, M : 전체 속성변수의 개수, P : 개별 레코드를 구성하는 속성 변수, 즉, i 번째 레코드 $X_i = (X_{i1}, X_{i2}, \dots, X_{iP}, \dots, X_{iM})$, \bar{X}_{KP} : K번째 군집에 대한 변수 P의 중앙값으로서, 다음의 식에 의해 계산.

$$\bar{X}_{KP} = \frac{1}{n_K} \sum_{i=1}^{n_K} X_{iP}$$

식 (1)에서 볼 수 있듯이, 이 지표는 같은 군집 내 개체끼리 서로 잘 응집되어 있을수록 작은 값을 갖게 된다. 즉, 이 지표값이 작으면 작을수록 군집화는 더 효과적으로 이루어진 것이라고 할 수 있다.

- 3) 3단계에서는 앞서 계산한 군집간 관성지표 값이 더 작아질 수 있도록, 대상 검색체에 유전자 알고리즘의 각종 연산을 적용하는 과정이

이루어지게 된다. 즉, 유전자 알고리즘의 적함도 함수를 군집간 관성지표의 최소화로 설정하고 이 함수의 값이 종료 조건을 만족시킬 때까지 선택, 교배, 돌연변이 등의 과정을 통해 계속적으로 군집중심점을 변동시키게 된다.

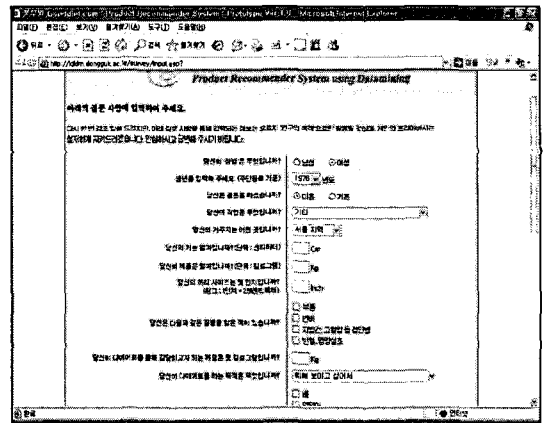
- 4) 새로운 염색체들이 생성되면, 다시 이 염색체를 기준으로 하여, 2), 3)단계 작업을 반복하게 된다. 이러한 작업은 종료 조건이 만족될 때까지 계속 반복되는데, 종료조건은 약 15~20세대(generation) 동안 연산을 반복하는 것으로 설정하는 것이 일반적이다.

본 연구에서는 GA K-Means의 유전자 알고리즘 파라미터로서, 모집단의 개체수를 50개로 설정하고, 교배율과 돌연변이율을 각각 0.7, 0.1로 설정하고 실험을 수행한다. 한편, 비계층적 군집분석방법은 대부분 최종 군집의 개수를 분석 이전에 미리 설정하여야 되는데, 최적의 군집수를 미리 파악하기 어려우므로 본 연구에서는 수행된 세가지 종류의 군집분석 모두에 대해서 최종 군집의 개수를 5개로 설정하고 실험을 진행한다.

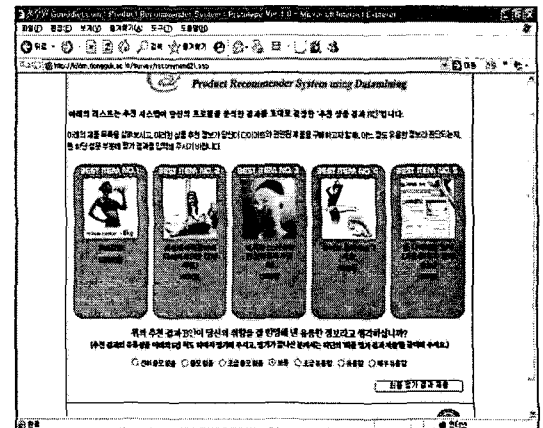
4.2.2 사례기반추론을 활용한 상품추천시스템 구축

축적된 고객데이터베이스의 시장세분화가 완성되면, 사례기반추론을 이용한 상품추천이 이루어지게 된다. 즉, 새로운 고객이 상품 추천을 받기 위해 사이트에 들어오게 되면, 시스템은 전 단계에서 구분된 5개의 군집 중 해당 고객에게 적합한 군집을 배정하고, 사례기반추론을 이용해 각 군집 내에서 해당 고객과 가장 유사한 고객들을 찾아 이들이 과거에 구입했던 상품들을 추천하게 되는 것이다. 군집의 배정방법은 최종 군집의 군집중심점과 새 고객데이터와의 유클리드 거리를 계산하여 가장 거리가 가까운 군집에 배정하게 된다. 사례기반추론의 구현방법으로는 여러 가지 방법이 있을 수 있는데, 그

중에서 본 연구에서는 다수(K개)의 유사한 사례를 찾아, 이들의 추천결과를 결합하는 K-NN (Nearest Neighbor) 방법을 사용한다. 실제 데이터베이스에 사례기반추론을 적용해, 추천할 상품을 결정하는 핵심 엔진은 비주얼 베이직 (Visual Basic) 언어로 프로그래밍하여, 웹 기반 언어 중 하나인 ASP(Active Server Pages)와의 상호 연동을 통해 웹 서비스 형태의 프로토타입 시스템에 활용될 수 있도록 한다. 개발된 프로토타입 시스템의 정보입력 화면이 <그림 2>에, 추천결과 제시 화면이 <그림 3>에 제시되어 있다.



<그림 2> 프로토타입 시스템의 정보입력 화면



<그림 3> 프로토타입 시스템의 추천결과 제시 화면

본 연구에서 구축한 프로토타입 시스템은 사용자가 <그림 2>와 같은 입력화면을 통해 본인의 정보를 입력하게 되면, <그림 3>과 같이 추천 결과를 보여주고, 만족도를 물어보는 형태로 구성되어 있다. 이 때, 본 연구에서는 제안하고자 하는 모형, 즉 GA K-Means 기반의 군집분석과 사례기반추론이 결합된 모형이 실질적으로 고객들에게 유익한 추천결과를 제시하는지를 보다 정밀하게 검증하기 위해, 결과를 두 차례로 나누어 보여주는 방법을 활용한다. 즉, 한번은 '모형을 통해 도출된 실제 추천 시스템의 결과'를 보여주고, 나머지 한 번은 '무작위로 골라진 추천 결과'를 보여 주어, 각 경우에 대해 고객 만족도의 차이가 있는지 검증하고자 하는 것이다. 이 때 어떤 결과를 먼저 보여줄지에 대해서는, 시스템이 무작위로 결정하게 하여 이른바 주시험효과(main testing effect)의 문제를 상쇄시키고자 하였다.

5. 실험결과

앞 장에서 언급했듯이, 본 연구에서 우리는 군집분석의 효과를 측정하는 지표로서 군집간 관성지표라는 지표를 사용하였다. 본 연구의 1 단계에서 적용해 본 자기조직화지도, K-Means, GA K-Means의 3가지 군집분석 기법의 최종 결과에 대해, 군집간 관성지표를 계산해 본 결과, <표 2>와 같이 GA K-Means가 가장 우수

한 결과를 보이고 있는 것을 확인할 수 있다.

앞서 소개했듯이 군집간 관성지표와는 별도로, 과연 GA K-Means로 고객 데이터베이스를 군집화하여 사례기반추론을 적용하는 경우, 고객에게 의미 있는 추천결과를 제공할 수 있는지를 현실적으로 확인하기 위해, 본 연구에서는 웹 기반의 프로토타입 시스템을 구축하여, 사용자에게 대한 간단한 만족도 설문조사를 수행하였다. 설문은 각 개인별로 무작위 추천결과를 생성하는 경우와 본 연구에서 제안하는 모형에 의거해 추천결과를 생성하는 경우에 대해 사용자의 추천 만족도를 물어보도록 설계되었으며, 2005년 4월 11일부터 22일까지 약 12일 동안 100건의 응답을 확보할 수 있었다. 만족도는 7점 리커트 척도로 측정되었는데, 각 경우에 대해 <표 3>과 같은 결과를 얻을 수 있었다.

<표 3>에서 볼 수 있듯이, 응답자들은 평균적으로 본 연구에서 제공하는 추천 결과에 대해 약 4.5점, 즉, '어느 정도 유용하다'는 평가를 내리고 있으며, 무작위로 선택되는 추천 결과에 대해서는 약 3.8점, 즉 '그냥 그렇다'는 평가를 내리고 있음을 알 수 있다. 본 연구에서는 이러한 만족도의 차이가 통계적으로 유의한지를 검증해 보기 위해, 추가적인 분석을 수행하였다. 검정에 사용한 통계량은 대응 데이터의 모평균 차이를 검정할 수 있는 '대응표본 T검정 통계량'을 적용하였다. 이 검정을 위해 세운 귀무가설과 대립가설은 다음과 같다.

<표 2> 각 군집분석기법의 군집간 관성지표 측정 결과

군집분석기법	K-means	SOM	GA K-means
군집간 관성지표(Intraclass inertia)	5800.78	5618.80	5612.55

<표 3> 만족도 설문 결과

추천 알고리즘	무작위에 의한 추천	모형에 의한 추천
평균	3.76	4.51
표준편차	1.342	1.010

〈표 4〉 대응표본 T검정 적용 결과

대응차					t	자유도	유의확률 (양쪽)
평균	표준편차	평균의 표준오차	차이의 95% 신뢰구간				
			하한	상한			
0.75	1.604	0.1604	0.432	1.068	4.675	99	0.0000

귀무가설 H_0 : '모형에 의한 추천결과'와 '무작위에 의한 추천결과' 만족도의 평균은 같다.

대립가설 H_1 : '모형에 의한 추천결과'와 '무작위에 의한 추천결과' 만족도의 평균은 다르다.

통계도구인 SPSS에 위 검정을 적용해 본 결과 <표 4>와 같은 결과를 얻을 수 있었다.

<표 4>에서 볼 수 있듯이, '모형에 의한 추천결과'의 만족도가 '무작위에 의한 추천결과'의 만족도보다 99% 신뢰수준 하에서 통계적으로 유의한 차이를 보이고 있음을 알 수 있다.

6. 결 론

상품추천시스템은 인터넷 쇼핑몰에서 고객관계를 관리하기 위한 가장 중요한 마케팅 수단 중의 하나라고 할 수 있다. 전자상거래의 확산으로 인터넷 쇼핑몰이 범람하면서 기존의 상품추천시스템과는 차별화된 서비스에 대한 요구가 많아지고 있으며, 개선된 형태의 상품추천은 정확성과 신속성이 보장되어야 할 것이다.

본 연구에서는 이러한 요구를 인식하고 개선된 형태의 상품추천시스템을 제안하고자 하였다. 본 연구에서 제안한 2단계의 상품추천시스템은 먼저 보다 정교한 상품추천을 위해 시장세분화를 수행하였고, 다음 단계로 하나의 세분군집 속에서 사례기반추론을 활용하여 유사 사례를 찾아 추천을 하는 방식이다. 시장세분화를 위한 군집분석에 있어서는 최종 군집중심점을 찾아 가는 과정이 지역 최적화 알고리즘에 의하

기에 분석의 최종 결과가 지역 최적화에 수렴할 수 있는 단점을 가진 전통적인 군집분석기법 대신, 전역 최적화 알고리즘인 유전자 알고리즘에 기반한 GA K-Means 군집분석을 이용하여 전통적인 기법의 한계점을 보완하고자 하였다. 또 상품추천목록 생성을 위해 기존의 협업필터링 대신 사례기반추론을 활용함으로써 희박성의 문제를 완화하였으며, 사례기반추론 이전에 군집분석을 수행하여 사례탐색공간을 한 군집 내로 한정함으로써 기존 상품추천시스템의 단점인 확장성의 문제를 보완하였다. 제안한 모형의 유용성을 확인하기 위해 프로토타입 시스템을 구축하고 이의 사용자에게 상품추천에 대한 만족도를 조사한 결과, 제안한 상품추천시스템에서의 추천결과가 무작위 방식의 추천결과보다 유용한 것으로 나타났다.

본 연구는 기존 상품추천시스템의 한계점인 희박성, 확장성 문제와 시장세분화를 위한 군집분석의 지역 최적해 수렴현상을 보완한 개선된 상품추천시스템을 제시하였다는 의의가 있으나 웹 로그 데이터와 같이 고객의 행태에 관한 기타의 자료를 충분히 활용하지 못하였다는 한계점도 가지고 있다. 또 하나의 인터넷 쇼핑몰 고객에 대한 자료를 이용하여 분석함으로써 일반화의 한계도 큰 한계점이라 할 수 있다.

향후 연구에서는 고객과 거래에 대한 자료 이외에 고객의 구매행태에 관한 다른 정보를 제공해 줄 수 있는 자료를 충분히 활용한 연구모형의 개발이 요구되며, 연구의 결과를 보다 더 일반화할 수 있도록 많은 자료원을 이용한 분석이 필요하다.

참 고 문 헌

- [1] 강부식, "자기 조직화 신경망을 이용한 협력적 여과 기법의 웹 개인화 시스템에 대한 연구", *한국지능정보시스템학회논문지*, 제9권 제3호, 2003, pp. 117-135.
- [2] 김재경, 서지혜, 안도현, 조운호, "A personalized recommendation methodology based on collaborative filtering", *한국지능정보시스템학회논문지*, 제8권 제2호, 2002, pp. 139-157.
- [3] 김재경, 안도현, 조운호, "Development of a personalized recommendation procedure based on data mining techniques for internet shopping malls", *한국지능정보시스템학회논문지*, 제9권, 제3호, pp. 177-191.
- [4] 김재경, 안도현, 조운호, "개인별 상품추천시스템, WebCF-PT : 웹마이닝과 상품계층도를 이용한 협업필터링", *경영정보학연구*, 제15권 제1호, 2005, pp. 63-79.
- [5] 김종우, 배세진, 이홍주, "협업 필터링 기반 개인화 추천에서의 평가자료의 희소 정도의 영향", *경영정보학연구*, 제14권 제2호, 2004, pp. 131-149.
- [6] 송수섭, 이의훈, "인공신경망을 이용한 소비자 선택 예측에 관한 연구", *한국경영과학회지*, 제26권 제4호, 2001, pp. 55-70.
- [7] 안현철, 한인규, "데이터 마이닝을 활용한 인터넷 쇼핑몰의 상품 추천 시스템 개발", *한국경영정보학회 춘계학술대회 논문집*, 2002, pp. 739-748.
- [8] 조운호, 박수경, 안도현, 김재경, "재구성된 제품 계층도를 이용한 협업 추천 방법론 및 그 평가", *한국경영과학회지*, 제29권 제2호, 2004, pp. 59-75.
- [9] 조희연, 김영민, "유전자 알고리즘을 이용한 주식투자 수익률 향상에 관한 연구", *한국정보시스템학회지*, 제12권 제2호, 2003, pp. 1-20.
- [10] 하성호, 이재신, "데이터마이닝을 활용한 동적인 고객분석에 따른 고객관계관리기법", *한국지능정보시스템학회논문지*, 제9권 제3호, 2003, pp. 23-47.
- [11] Adeli, H. and Hung, S., *Machine learning : neural networks, genetic algorithms, and fuzzy systems*, New York, Wiley, 1995.
- [12] Anderberg, M.R., *Cluster Analysis for Application*, Academic Press, New York, 1973.
- [13] Babu, G.P. and Murty, M.N., "A near-optimal initial seed value selection in K-Means algorithm using a genetic algorithm", *Pattern Recognition Letters*, Vol. 14, No. 10, 1993, pp. 763-769.
- [14] Bandyopadhyay, S. and Maulik, U., "An evolutionary technique based on K-Means algorithm for optimal clustering in RN", *Information Sciences*, Vol. 146, 2002, pp. 221-237.
- [15] Berry, J.A. and Linoff, G., *Data Mining Techniques : For Marketing, Sales, and Customer Relationship Management*, Second Edition, John Wiley & Sons, 2004.
- [16] Cho, Y.H. and Kim, J.K., "Application of Web usage mining and product taxonomy to collaborative recommendations in e-commerce", *Expert Systems with Applications*, Vol. 26, 2004, pp. 233-246.
- [17] Cho, Y.H., Kim, J.K., and Kim, S.H., "A personalized recommender system based on Web usage mining and decision tree induction", *Expert Systems with Applications*, Vol. 23, 2002, pp. 329-342.
- [18] Funakoshi, K. and Ohguro, T., "A Content-based collaborative recommender system with detailed use of evaluations", *Proceedings of the 4th International Conference on Knowledge-Based Intelligent Engineering Systems & Allied Technologies*, 2000, pp. 253-256.
- [19] Gehrt, K.C. and Shim, S., "A shopping orientation segmentation of French consumers: Implications for catalog marketing", *Journal*

- of Interactive Marketing*, Vol. 12, No. 4, 1998, pp. 34-46.
- [20] Goldberg, D.E., *Genetic algorithms in search, optimization, and machine learning*, Reading, MA, Addison-Wesley, 1989.
- [21] Goldberg, D., Nichols, D., Oki, B.M., and Terry, D., "Using collaborative filtering to weave an information tapestry", *Communications of the ACM*, Vol. 35, No. 12, 1992, pp. 61-70.
- [22] Ignizio, J.P. and Soltys, R., "Simultaneous design and training of ontogenic neural network classifiers", *Computers & Operations Research*, Vol. 23, No. 6, 1996, pp. 535-546.
- [23] Kim, J.K., Cho, Y.H., Kim, W.J., Kim, J.R., and Suh, J.H., "A personalized recommendation procedure for Internet shopping support", *Electronic Commerce Research and Applications*, Vol. 1, 2002, pp. 301-313.
- [24] Kim, K.-S. and Han, I., "The cluster-indexing method for case-based reasoning using self-organizing maps and learning vector quantization for bond rating cases", *Expert Systems with Applications*, Vol. 21, No. 3, 2001, pp. 147-156.
- [25] Kohonen, T., *Self-organizing maps*, Berlin : Springer, 1995.
- [26] Konstan, J., Miller, B., Maltz, D., Herlocker, J., Gordon, L., and Riedl, J., "GroupLens : Applying collaborative filtering to usenet news", *Communication of the ACM*, Vol. 40, 1997, pp. 77-87.
- [27] Kuo, R.J., Ho, L.M., and Hu, C.M., "Integration of self-organizing feature map and K-means algorithm for market segmentation", *Computers & Operations Research*, Vol. 29, 2002a, pp. 1475-1493.
- [28] Kuo, R.J., Ho, L.M., and Hu, C.M., "Cluster analysis in industrial market segmentation through artificial neural network", *Computers & Industrial Engineering*, Vol. 42, 2002b, pp. 391-399.
- [29] Kuo, R.J., Liao, J.L., and Tu, C., "Integration of ART2 neural network and genetic K-means algorithm for analyzing Web browsing paths in electronic commerce", *Decision Support Systems*, Vol. 40, 2005, pp. 355-374.
- [30] Laflamme, L. and Eilert-Petersson, E., "School-injury patterns : A tool for safety planning at the school and community levels", *Accident Analysis & Prevention*, Vol. 30, No. 2, 1998, pp. 277-283.
- [31] Maniezzo, V., "Genetic evolution of the topology and weight distribution of neural networks", *IEEE Transactions on Neural Networks*, Vol. 5, No. 1, 1994, pp. 39-53.
- [32] Michaud, Pierre., "Clustering techniques", *Future Generation Computer Systems*, Vol. 13, No. 2-3, 1997, 135-147.
- [33] Murthy, C.A. and Chowdhury, N., "In search of optimal clusters using genetic algorithms", *Pattern Recognition Letters*, Vol. 17, 1996, pp. 825-832.
- [34] Pazzani, M.J., "A framework for collaborative, content-based and demographic filtering", *Artificial Intelligence Review*, Vol. 13, No. 5-6, 1999, pp. 393-408.
- [35] Pena, J.M., Lozano, J.A., and Larranaga, P., "An empirical comparison of four initialization methods for the K-Means algorithm," *Pattern Recognition Letters*, Vol. 20, 1999, pp. 1027-1040.
- [36] Resnick, P., Iacovou, N., Suchak, M., and Bergstrom, P., "GroupLens : An open architecture for collaborative filtering of netnews", *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, 1994, pp.175-186.
- [37] Roh, T.H., Oh, K.J., and Han, I., "The collaborative filtering recommendation based on

- SOM cluster-indexing CBR”, *Expert Systems with Applications*, Vol. 25, No. 3, 2003, pp. 413-423.
- [38] Sarwar, B., Karypis, G., Konstan, J., and Riedl, J., “Application of dimensionality reduction in recommender systems : A case study”, Proceedings of the WebKDD Workshop at the ACM SIGKDD, 2000.
- [39] Selim, S.Z. and Ismail, M.A., “K-means-type algorithms : a generalized convergence theorem and characterization of local optimality”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 6, No. 1, 1984, pp. 81-87.
- [40] Sexton, R.S., Dorsey, R.E., and Johnson, J.D., “Toward global optimization of neural networks : A comparison of the genetic algorithm and backpropagation”, *Decision Support Systems*, Vol. 22, 1998, pp. 171-185.
- [41] Sexton, R.S., Dorsey, R.E., and Johnson, J.D., “Optimization of neural networks : A comparative analysis of the genetic algorithm and simulated annealing”, *European Journal of Operational Research*, Vol. 114, 1999, pp. 589-601.
- [42] Shin, H.W. and Sohn, S.Y., “Segmentation of stock trading customers according to potential value”, *Expert Systems with Applications*, Vol. 27, No. 1, 2004, pp. 27-33.
- [43] Smith, W.R., “Product differentiation and market segmentation as an alternative marketing strategy”, *Journal of Marketing*, Vol. 21, 1956, pp. 3-8.
- [44] Tsai, C.-Y. and Chiu, C.-C., “A purchase-based segmentation methodology”, *Expert Systems with Applications*, Vol. 27, 2004, pp. 265-276.
- [45] Vellido, A., Lisboa, P.J.G., and Meehan, K., “Segmentation of the on-line shopping market using neural networks”, *Expert Systems with Applications*, Vol. 17, No. 4, 1999, pp. 303-314.

저자소개



김 경 재

저자는 KAIST에서 경영정보 시스템을 전공으로 박사학위를 취득하였으며, 현재 동국대학교 경영대학 정보관리학과 조교수로 재직하고 있다.

주요 관심분야는 데이터 마이닝, 고객관계관리, 지식경영, 전사적자원관리 등이다.



안 현 철

저자는 KAIST에서 산업경영 학사 및 경영공학석사를 취득하였으며, 현재 한국과학기술원 테크노경영대학원 경영공학 박사과정에 재학 중이다.

주요 관심분야는 인공지능/데이터마이닝을 활용한 재무예측, 고객관계관리, m-CRM 등이다.