

유전자 알고리즘을 활용한 개인화된 상품추천시스템 개발

A personalized recommender system using genetic algorithms

김병국, 김경재

동국대학교 정보관리학과

ubf@dongguk.edu, kikim@dongguk.edu

Abstract

규칙기반의 상품추천시스템은 많은 인터넷 쇼핑몰에서 활용되고 있지만 규칙을 추출할 수 있는 마케팅 전문가 확보와 방대한 양의 고객 데이터 처리의 어려움으로 유용한 규칙을 찾는 것이 매우 어렵다. 본 연구에서는 이러한 규칙기반 상품추천시스템의 단점을 보완할 수 있는 방법으로 전역 최적화 기법의 하나인 유전자 알고리즘을 활용하여 고객정보를 토대로 추천 규칙을 도출할 수 있는 방안을 제시한다. 또한 본 연구에서 제안한 유전자 알고리즘에 기반한 추천 규칙들이 장착된 웹 기반의 개인화된 상품추천시스템의 프로토타입을 개발하고 이에 대한 실제 사용자들의 이용 만족도를 확인함으로써 본 연구에서 제안한 방법론의 유용성을 확인하고자 한다.

1. 서론

전자상거래 활성화와 함께 수많은 사이버 쇼핑몰 업체가 등장함으로써 각 업체들은 다른 업체에 비해 경쟁 우위를 가질 수 있는 마케팅 전략이 필요하게 되었다. 또한 고객은 상품 정보의 과다로 인해 효과적으로 상품을 선택할 수 없게 되는 상품 과부화 현상을 맞이하게 되었다. 위 문제를 극복할 수 있는 방안으로 최근 일대일 마케팅(Peppers et al., 1993) 전략이 주목을 받고 있으며, 일대일 마케팅을 위한 정보기술 수단의 하나로서 추천시스템(Recommender Systems)을 사용하고 있다.

추천시스템은 인터넷 상거래 업체에게는 고객에게 개인화 서비스를 제공하여 고객의 웹사이트 충성도를 높일 수 있는 경쟁력으로 작용하며, 사용자에게는 정보 과다 현상을 해결함으로써 상품 선택의 탐색 비용을 감소시켜, 양측 모두에게 의미 있는 가치를 지니게 된다.

추천시스템 구축을 위해서 사용되는 방법론은 다양하다. 협업 필터링(Collaborative Filtering) 방법론(Resnick et al., 1994)은 지금까지 가장 성공적인 추천 기술로 알려져 있으며

내용기반(Content-based) 방법론과 함께 추천시스템 연구 방향의 두 축을 이룬다. 이 두 가지 방법론 이외에 규칙기반(Rule-based) 방법론 및 혼성(Hybrid) 방법론에 관한 연구가 있다.

본 연구에서는 규칙기반 상품추천시스템의 단점 즉, 규칙 추출의 어려움을 보완할 수 있는 방법으로 전역 최적화 기법의 하나인 유전자 알고리즘(Genetic Algorithms)을 활용하여 고객정보를 토대로 추천 규칙을 도출할 수 있는 방안을 제시한다. 규칙 추출을 위해서는 일반적으로 의사결정나무, 연관성 규칙 등이 활용되고 있으나, 유전자 알고리즘은 주어진 문제 영역에서 전역 검색을 통해 최적해를 찾아 나가는 데이터 마이닝 기법으로서, 다양한 고객 변수 데이터를 활용하여 추천 규칙을 생성하는데 매우 효과적으로 사용될 수 있다. 또한 일련의 추천 규칙들을 모델 기반(model-based) 방식을 이용하여 추천 규칙과 사용자 정보와의 매칭이 온라인 상에서 실시간으로 이루어지게 함으로써 즉각적인 상품 추천이 가능한 시스템을 제안한다.

2. 관련 연구

2.1 추천시스템 주요 추천 기법

협업필터링(Collaborative Filtering) 추천시스템은 가장 성공적인 추천 방식 가운데 하나이며, 사용자 사이의 연관성을 기반으로, 선호도 또는 구매 패턴이 유사한 고객군을 분류하고, 동일 고객군 내의 다른 사람들이 선호하는 상품을 추천하는 방법(Funakoshi et al., 2001)이다. 내용기반(Content-based) 추천시스템은 상품이 지닌 속성을 파악하고 고객의 구매 이력을 바탕으로 고객이 구매한 상품의 속성들과 유사한 상품을 추천하는 방법으로서 제품 정보를 중심으로 추천하는 방법(Balabanovic, 1997)이다. 규칙기반(Rule-based) 추천시스템은 사용자의 프로필 데이터, 구매 데이터, 웹로그 데이터 등에 근거하여 조건문 형식의 규칙을 이용한 개인화된 추천을 제공하는 방법이다.

2.2 유전자 알고리즘

유전자 알고리즘은 데이터 마이닝의 대표적인 기법 중의 하나라고 할 수 있다. 유전자 알고리즘은 자연계에 존재하는 어떤 개체집단이 적자생존과 자연도태를 통해 다음 세대의 집단을 형성해나가는 과정을 수학적 알고리즘 형태로 구성한 것이다. 즉, 유전자 알고리즘은 최적화 알고리즘의 하나로서 여러 가지 가능한 해집합 가운데 가장 적당한 최적의 해를 찾아주는 알고리즘이라고 할 수 있다.

유전자 알고리즘은 우선 임의의 값을 가진 초기화된 개체집단을 생성한다. 즉 전체 탐색공간 내에서 임의의 N개의 점들을 선택하여 개체집단(Population)을 형성하게 된다. 이렇게 생성된 개체집단이 문제에 얼마나 적합한지를 적합도 함수(Fitness Function)를 이용하여 평가한다. 다음 세대의 개체 생성은 평가된 개체집단을 확률적으로 선택(Selection)하여 교배(Crossover)와 돌연변이(Mutation)를 통해 이루어진다. 다음 세대의 개체집단은 다시 평가를 하여 종결조건에 도달할 때까지 즉, 목표로 한 적합도 수준에 도달하거나 미리 정해진 최대 진화 수를 넘을 때까지 반복적으로 다음 세대를 생성한다(조희연 외, 2003).

3. 제안 추천시스템

본 연구에서 제안하는 추천시스템은 두 개의 추천 모듈을 가지고 있다. 첫 번째 모듈은 유전자 알고리즘을 활용하여 생성된 마케팅 규칙들이 장착되었고, 두 번째 모듈은 의사결정나무를 활용하여 생성된 마케팅 규칙들이 장착되어 있다.

각 모듈은 규칙을 생성하기 위해서 고객 데이터와 거래 데이터를 활용한다. 유전자 알고리즘과 의사결정나무 분석을 통해 도출된 상품 추천 규칙은 모델 기반의 추천 엔진에 통합되어 저장된다. 사용자 정보와 각 모듈 내의 규칙은 순서대로 매칭되어 N개의 개인화된 맞춤 상품을 추천하게 된다. 그림 1.은 제안 추천시스템의 모형을 보여준다.

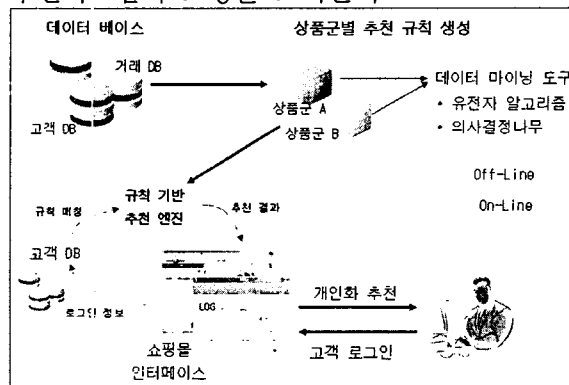


그림 1. 제안 추천시스템 모형

4. 데이터 분석

4.1 분석 데이터 셋

본 연구에서 사용된 데이터 셋은 인터넷 리서치를 통해 수집되었으며 마케팅 규칙을 생성하기 위한 데이터로 사용되었다.

인터넷 리서치는 2004년 6월 10일부터 15일까지 총 6일간 대학생 및 일반인을 대상으로 실시되었으며, 설문 참여자는 개인 신상정보, 맞춤 서비스를 위한 부가 정보, 인터넷 쇼핑물 이용 현황에 관한 정보를 입력한 후, 8개 장르의 영화 DVD 상품 36개 가운데 구매를 원하는 상품 5개를 선택하게 된다. 이를 통해 총 134개의 고객 신상에 관한 데이터 및 615건의 거래 데이터를 수집하였다.

4.2 유전자 알고리즘 분석

유전자 알고리즘을 수행하기 위해서는 탐색공간 내의 다양한 변수 집합(표 1.)을 선형 스트링(염색체)에 사상(mapping)하는 과정이 필요하다. 유전자 알고리즘이 효율적인 탐색을 하도록 하기 위해서는 탐색 목적에 적합한 효과적인 사상방법을 찾아야만 한다. 본 연구에서 유전자 알고리즘을 수행하는 목적은 고객 데이터를 활용하여 특정 상품을 구매하는 고객을 예측할 수 있는 추천 규칙을 찾는 것이다. 즉, 표 1.에 열거된 16개의 변수와 그 속성의 조합으로 모집단 염색체를 구성하고 유전자 알고리즘 고유의 연산과정 적합도평가, 복제연산, 교배연산, 돌연변이연산을 거쳐 추천 규칙을 생성하게 된다. 유전자 분석을 위해 사용된 분석 툴은 Evolver 4.0.2이며 모집단의 크기는 50, 교차율은 0.5, 변이율은 0.06으로 설정하였다.

기본 개인 정보 입력 내용		
변수명	내용	선택 범위
Gender	성별	2
Age	연령	7
Phone	이동전화	7
Mail	이메일	9
Addr	주소	28
개인화 서비스를 위한 맞춤 정보 입력 내용		
Edu	교육수준	8
Occ	직업	20
Income	월소득	7
House 1	주거형태	10
House 2	주거소유유무	9
Mar	결혼유무	2
Car	자동차보유유무	2
Pref	선호장르	8
쇼핑물 이용 현황 정보 입력 내용		
itime	구매빈도	5
item	주구매/검색품목	7
Pay	주결제수단	3

표 1. 분석 데이터 셋의 변수

4.2.1 염색체 구성

염색체 유전자는 총 149개로 구성된다. 최종적으로 선택된 16개 변수들의 특성 135개를 나타내기 위한 유전자가 132개이고, 각 변수를

조건들을 아래와 같다.

- Analysis Tool : SAS E-Miner 4.0
- Splitting criterion : Gini Index
- Max.Number of branches from a node : 2
- Maximum depth of tree : 6

5. 유용성 평가

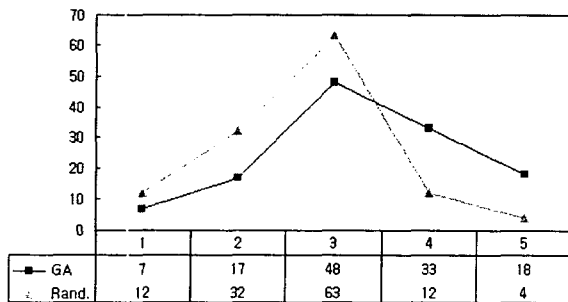
5.1 프로토타입 구축

본 연구에서 제안한 추천 알고리즘의 실제적 유용성을 평가할 수 있는 웹(Web)기반의 프로토타입 시스템을 구축하였다. 본 프로토타입에는 유전자 기반 추천 알고리즘과 의사결정나무 기반 추천 알고리즘 즉, 연구에서 제안된 상품 추천을 위한 마케팅 규칙들이 모델 기반 형태로 장착되어 있다. 프로토타입 시스템 개발 환경을 요약하면 아래와 같다.

- URL: <http://210.94.165.237/default.htm>
- Operating System: Microsoft Windows XP Professional
- Language: Microsoft Active Server Pages / Javascript
- Database Management System: Microsoft Access 2003

5.2 유용성 평가 및 결과

제안 추천시스템의 유용성 평가를 위한 온라인 설문은 2004년 8월 18일부터 22일까지 총 5일간 진행되었으며 134명의 설문자가 참여하였다. 설문 참여자는 상품 추천을 위해 필요한 정보를 입력한 후, 본 연구에서 제안한 유전자 알고리즘에 기반한 추천 규칙에 의해 선정된 상품과 무작위로 선정되어 추천되는 상품에 대하여 각각의 유용성 정도를 5점 척도로 평가하게 된다.



- GA: 유전자 알고리즘 기반 추천 시스템에 의한 상품 추천
- Rand.: 무작위 상품 추천
- 단위: 가로축-평가자 명수, 세로축-유용성에 대한 5점 척도 평가

표 3. 추천기법간 유용성의 정도 평가 분포도

두 추천 기법 간 유용성에 관한 통계적 유의성 검정을 위해 대응표본 T검정을 실시하였으며, 유의 수준 5% 내에서 두 기법 간의 평균에 관한 차이가 통계적으로 유의한 것으로 나타났다.

대분	평균	N	표준편차	평균의 표준오차
GA	3.3089	123	1.0646	9.599E-02
RAND	2.7073	123	0.8938	8.059E-02

		대용차				
평균	표준편차	평균의 표준오차	t	자유도	유의확률 (양측)	
.6016	1.4243	.1284	4.685	122	.000	

표 4. 추천기법간 유용성 차이에 대한 통계적 검증

6. 결론

본 논문에서는 규칙기반 추천시스템의 문제로 대두되는 효과적인 추천 규칙의 어려움을 해결하고자 최적화 기법 데이터 마이닝 기법 가운데 하나인 유전자 알고리즘을 활용하여 추천을 위한 마케팅 규칙을 생성하고 모델 기반(Model-based) 추천시스템을 구축하였다. 유전자 알고리즘은 특정 문제 영역에서 최적화(near optimization)된 해를 찾는 데 매우 유용하다는 장점을 가지고 있다. 또한 모델 기반 추천시스템은 간단한 정보 입력 후 간단한 연산을 통해 실시간 추천이 가능하다는 점에서 유용하다.

따라서 본 논문은 규칙 기반 추천시스템에서 유전자 알고리즘을 활용하여 규칙 추출의 어려움을 최적화된 방법으로 해결하고 온라인 상점의 특성에 맞도록 실시간 추천이 가능한 시스템을 구축하고자 했다는 데에 그 의의가 있으며, 나아가 유전자 알고리즘에 의해 생성된 추천 규칙의 실제적 유용성을 온라인 설문을 통해 평가해 봄으로써 그 실제적인 유용성 면에서도 의미가 있는 것으로 나타났다.

참고 문헌

1. 조희연, 김영민, "유전자 알고리즘을 이용한 주식투자 수익률 향상에 관한 연구", 한국정보시스템학회지 제 12권 제 2호, 2003.
2. Balabanovic, M. and Shohm, Y., "Fab : Content-Based, Collaborative Recommendation", Communications of the ACM, March 1997.
3. Funakoshi, K. and Ohguro, T., "A Content-Based Collaborative Recommender System with Detailed Use of Evaluations", Proc. 4th International Conference on Knowledge-Based Intelligent Engineering Systems & Allied Technologies, 2000.
4. Resnick, P., Iacovou, N., Suchak, M., and Bergstrom, P., "GroupLens: An Open Architecture for Collaborative Filtering of Netnews", Proceedings of the ACM 1994 Conference on Computer Supported Cooperative Work, pp.175-186, 1994.
5. Peppers, D. and Martha R., "The One to One Future : Building Relationships One Customer at a Time", Currency and Doubleday, New York, 1993.