

퍼지환경의 Web Mining방법에 관한 연구

A Study on Web Mining Method under A Fuzzy Environment

주종문, 황승국

(주)원스텝소프트, 경남대학교 벤처창업학부

Jong-Moon Ju, Seung-Gook Hwang

OneStepSoft Co. Ltd, Division of Venture Technology Kyungnam University

E-mail : cto@onestepsoft.com

요 약

web Mining은 Network의 발전과 함께 그 중요성이 대두되고 있는 전자상거래(EC)에서 디지털화된 자료의 분석을 통하여 가상상점을 이용하는 고객의 이용경로, 검색 및 구매경로, 상품에 대한 고객의 검색 및 구매 경향을 정확히 파악하기 위해 중요한 주제로서 연구되고 있는 분야이다.

본 연구에서는 전자상거래(EC) 시스템의 디지털화된 자료를 분석할 때 기존의 통계적인 방법론에서 벗어나 고객의 주관적인 검색 및 구매의 의사결정 과정분석에서 퍼지이론을 도입하여 새로운 Web Mining 방법론을 제안하였다.

1. 서론

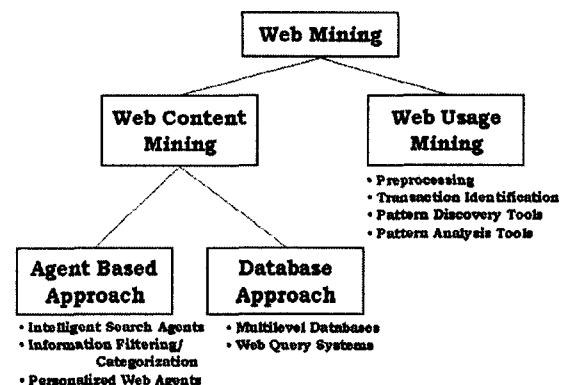
Internet과 web기술의 발전은 기존의 상거래에 많은 변화를 가져와 전자상거래(EC)의 비약적인 발전을 가져왔다. 특히 기업과 소비자간(B2C)의 전자상거래에서는 하루가 다르게 새로운 가상상점이 생겨나고 있다.

이러한 가상상점의 발전은 가상상점을 이용하는 다양한 고객에게 좀 더 나은 서비스를 위해 웹상에 존재하는 디지털화 된 자료를 이용하여 고객에게 필요한 정보를 추출해내고자 하는 Web Mining분야의 비약적인 발전을 가져왔다.

본 연구에서는 이러한 Web상의 디지털화 된 자료의 분석에 사용되는 Web Mining기법으로 가상상점의 이용하는 고객의 주관적인 검색 및 구매의 의사결정 과정상에 애매모호성을 가정하고 퍼지이론을 도입하고자 한다. 특히 웹 사용 마이닝의 패턴발견기법에 가장 적합한 방법으로 알려진 퍼지구조모델링법[1]을 적용하여 웹마이닝 방법론의 실질적인 예를 제시하였다.

2. Web Mining

웹 마이닝은 대체적으로 웹상에서 유용한 정보를 발견하고 분석하는 것으로서 정의할 수 있다.



[그림1] Web Mining의 분류[2]

[그림1]에서 살펴볼 수 있듯이 웹 마이닝은 어떻게 하면 웹 상에 존재하는 데이터를 효율적으로 분석하여 고객의 요구에 맞는 자료를 제공하는가

를 연구하는 웹 내용 마이닝(Web Content Mining)분야와 웹 서버 로그(Web server logs), 조회로그(refrral logs), 등록 데이터(Registration data) 등을 이용하여 사용자 경로 분석(Path analysis), 사용자 이용 패턴 분석(pattern analysis)를 중심으로 하는 웹 사용 마이닝(Web Usage Mining)분야로 나눌 수 있다[2].

2.1 Web Content Mining

웹 내용 마이닝은 야후, 알타비스타, 라이코스 와 같은 전통적인 검색 엔진들이 제공하지 못하는 구조화된 정보를 제공할 수 있도록 하기 위해 두 가지 접근 방식을 취하고 있다.

첫 번째 접근 방식은 에이전트-기반 접근 방식(Agent-based Approach)으로서 인공지능 웹 에이전트와 같은 더 지능적인 툴을 개발하여 구조화된 정보를 제공하고자 하는 접근 방식이다. 또 다른 접근 방식은 데이터베이스 접근방식(Database Approach)로서 웹 상에서 사용 가능한 정보의 구조를 더 높게 하기 위하여 데이터 마이닝 기술을 확장한 접근 방식이다.

에이전트 기반 접근 방식은 대체적으로 다음과 같이 3가지로 분류할 수 있다.

- 1) Intelligent Search Agent
- 2) Information Filtering/Categorization Agent
- 3) Personalized Web Agent

데이터베이스 접근 방식은 다중레벨 데이터베이스(Multilevel Database)와 웹 질의 시스템(Web Query System)으로 분류할 수 있다.

2.2 Web Usage Mining

웹 사용 마이닝은 웹 로그에서 불필요한 정보를 제거하고 필요한 정보만을 추출하는 전처리(Preprocessing)와 다양한 로그인 정보에서 한 사용자의 트랜잭션을 규정하는 트랜잭션 검증(Transaction Identification)을 수행한다. 이 과정이 끝나면, 로그 데이터 상에 나타나는 사용자의 검색 또는 구매 패턴을 발견하는 패턴 발견(Pattern Discovery)을 수행하게 되며, 발견된 패턴을 사용목적에 맞게 분석하는 패턴 분석(Pattern Analysis)를 수행한다.

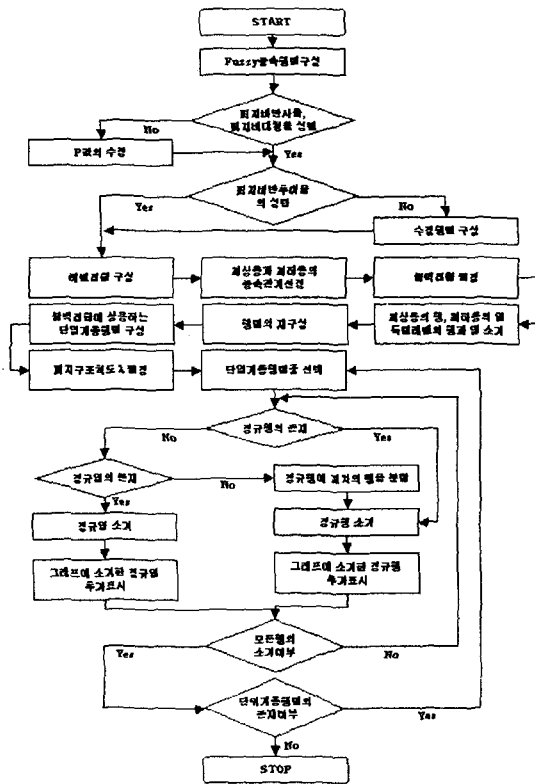
- 패턴 발견을 위한 기법으로는 다음과 같은 것이 있다[2].

- 1) Clustering analysis
비슷한 특징들을 가진 아이템들을 모아서 하나의 그룹으로 모으는 것을 가능케 한다.
- 2) Classification rule Discovery
공통된 특성(Attribute)에 의한 특정한 그룹에 속하는 아이템들의 프로파일을 찾을 수 있게 한다. 이 프로파일은 데이터베이스에 추가된 새로운 데이터 아이템들을 분류하기 위해 사용될 수 있다.
- 3) Association rule discovery technique
일반적으로 각 트랜잭션이 여러 아이템들의 집합으로 구성된 트랜잭션들이 저장된 데이터베이스에 적용된다. 문제는 한 트랜잭션안에 어떤 아이템들이 존재한다는 사실을 근거로, 다른 아이템들이 존재할 것이라는 것을 추정할 수 있는 아이템들 간의 모든 연관성(association)과 상호관계(correlation)을 찾는 것이다.
- 4) Sequence Pattern discovery technique
아이템들의 한 집합의 존재가 시간이 찍힌 순서화된 트랜잭션의 집합안에서 다른 아이템들에 의해 뒤따라진다는 상호관계 트랜잭션 패턴들을 찾는 것이다.

3. FSM(Fuzzy Structural Modeling)

웹 사용 마이닝의 패턴발견기법에 적용하기 위한 퍼지구조모델링법은 Tazaki등이 다원적인 가치가 복합되어 있는 시스템의 구조인식을 위해 제안한 방법[*]으로 시스템의 구조화라는 것은 어떤 대상 시스템을 구성하고 생각할 수 있는 요인을 적당한 방법으로 추출정리하고, 어떤 문맥상의 관계에 대해 추출된 요인을 계층화하고, 계층간 혹은 계층에 속한 요인간의 종속관계를 결정하고, 그것을 그래프로 나타내는 것이다.

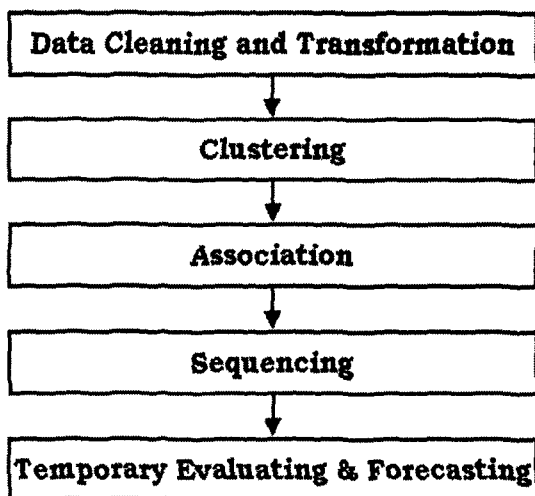
그 흐름을보면 다음의 [그림2]와 같다.



[그림2] Flowchart of FSM Method[3]

4. 퍼지환경하의 Web Mining 방법

본 연구의 퍼지환경하의 Web Mining은 전자상거래 가상상점을 이용하는 고객들이 다양한 상품 중 특정한 상품을 선택을 하여 구매할 때 그 선택의 과정에서 상품에 대한 선호의 정도가 있다는 현실적인 환경을 고려하여 고객의 유형별 구매패턴을 분석하는 방법을 말하는 것이다.



[그림 3] Framework of Analysis

4.1 Data Cleaning and Transformation

분석을 수행 하기전에 웹 로그에서 에러가 발생한 페이지 데이터 및 이미지 관련 데이터와 같은 불필요한 데이터를 선택하여 제거하는 사전처리(Preprocessing)가 선행되어야 한다. 이것은 Data Cleaning과는 별개로 구분하기도 하나 본 연구에서는 하나로 보기로 한다.

이렇게 정제된 웹 로그 데이터와 데이터 베이스 자료를 이용하여 분석대상이 되는 사용자의 사용 기록(Access History)를 작성하게 된다.

이렇게 작성된 사용기록은 분석을 위해 변환이 된다.

다음 (1)식에 의해 일대비교 행렬 개념을 도입한 퍼지중속행렬을 구성하게 된다.

$$A_{ij} = \frac{(j_i)}{(i_i) + (j_i)} \tag{1}$$

i : 상품 i 을 구매한 고객의 수

j : 상품 j 을 구매한 고객의 수

t : 분석대상이 되는 특정한 기간

즉 분석대상이 되는 특정한 기간동안 상품 j 를 구매한 고객의 수에 상품 i 와 j 를 구매한 고객의 총수에 나누어 i 와 j 의 일대비교 행렬 개념을 도입한 퍼지 중속행렬을 구할 수 있는 것이다.

4.2 Clustering

Clustering은 상품과 고객간의 연관성을 추출하는 기능을 수행한다.

이러한 Clustering을 수행한 결과에서 각각의 Cluster는 고객 대 구매상품의 연관성을 나타낸다. 즉, 고객들이 상품을 구매할 때 비슷한 구매 성향을 가진 고객들이 구매하는 상품들이 하나의 Cluster로 구분되게 된다. 그러므로 같은 Cluster로 분류되는 상품군들은 구매 고객들이 비슷한 계층이므로 하나의 그룹군으로 파악하여 공동 마케팅, DM발송과 같은 이벤트를 공동으로 수행할 수 있다.

본 연구에서는 Clustering을 위한 방법으로 SOM(Self-Organization Map)을 이용한다.

SOM은 신경회로망(Neural Network)의 일종으로 Kohonen에 의해 일반화되었기 때문에 Kohonen net이라고도 한다.

4.3 Association

Association은 상품과 상품간의 일정시간의 정적인 연관성을 분석하게 된다.

본 연구에서는 기존 Web Mining방법에서 많이 사용되는 상관계수(Correlation)를 사용하지 않고 Data Cleaning & Transformation와 같이 일대비교행렬의 개념을 도입한 퍼지중속행렬을 바탕으로 퍼지구조모델링법을 이용하여 구매상품의 상호중속성을 도식화하여 보여 주도록 하였다.

이러한 Association의 결과는 가상상점의 관계성이 있는 상품간의 재고량을 조절하거나, 하나의 상품을 구매한 동일 유형의 고객에게 관계성이 있는 상품을 추천하는 등 재고관리와 상품관리에 다양하게 사용될 수 있다.

4.4 Sequencing

Sequencing은 Association에 연속적인 시간의 선후개념을 도입하여 동적인 연관성을 분석하게 된다.

본 연구에서는 Association의 시간 t 를 선후관계에 따라 계속적으로 연결하여 하나의 흐름을 분석할 수 있도록 하였다.

이것은 Association에서 나타난 상품구매의 기본적인 선후관계에 동적인 시간의 개념이 더해져 구매까지 걸리는 시간을 추출할 수 있다.

4.5 Temporary Evaluating & Forecasting

Temporary Evaluating & Forecasting은 Clustering과 Association과 Sequencing의 결과를 바탕으로 하여 동일한 유형의 고객이 어떤 구매 패턴으로 갈것인지를 예측하게 된다.

이러한 구매 패턴에 대한 예측이 중요한 이유는 전자상거래 이용자의 증가와 함께 그 중요성이 증가 되고 있는 상품관리와 재고관리를 위해 가장 필요한 정보이기 때문이다.

5. 결론

오늘날 전자상거래는 상거래에서 하나의 큰 흐름으로 자리를 잡아가고 있다. 특히 기업과 소비자간 전자상거래(B2C)는 가상상점의 발전과 함께 이미 대중화 되어 우리의 생활의 일부분이 되었다.

본 연구에서는 이러한 전자상거래(EC)의 중요

한 연구분야중 하나로 대두되고 있는 가상상점에서 자동화된 상품관리를 위해 사용되고 있는 Web Mining에서 기존의 분석방법에서 벗어나 고객의 구매의사결정시 상품에 대한 선호의 정도를 표현하고, 그 각각의 상품을 선택하는 의식구조까지를 분석하기 위해 퍼지구조모델링법을 도입하여 Web Mining을 위한 새로운 방법론을 제안하였다.

향후 연구방향은 현재 Association와 Sequencing에만 적용된 퍼지환경에 대한 적용을 확대하여 Clustering과 Temporary Evaluating & Forecasting까지 적용시켜 Web Mining전반에 퍼지환경을 고려한 방법론을 만들어야 할 것이다. 또한 이러한 방법론을 적용하여 기존의 재고관리시스템과 연동하여 상품관리가 가능한 시스템의 구현도 필요하다.

6. 참고문헌

- [1] E.Tazaki and M.Amagasa, 'Structural Modeling in a Class of System Using Fuzzy Sets Theory', Fuzzy Sets and Systems, Vol.2, No.1, 1979.
- [2] R.Cooley, B.Mobasher, and J.Srivastava, 'Web Mining : Information and Pattern Discovery on the World Wide Web', In Proc.of the 9 , IEEE International Conference, pp558-567, 1997.
- [3] 박영화, "퍼지환경하에서 다기준 의사결정법에 의한 중소기업의 경쟁력 평가", 동아대학교 박사학위논문, pp61-62, 1996.