전자상거래에서 FSM을 이용한 고객구매패턴 분석

Analysis of Customer Purchase Patterns for Electronic Commerce Using FSM

주종문(Jong-Moon Ju)*, 황승국(Seung-Gook Hwang)**

초 록

웹 마이닝은 전자상거래(Electronic Commerce)의 발전과 함께 그 중요성이 대두 되고 있으며, 주로 전자상거래에서 구매자의 구매 경향을 분석하기 위해 중요한 주제로서 연구되고 있는 분야이다.

본 연구에서는 전자상거래에서 구매자의 구매과정을 퍼지환경으로 정의하고 기존의 웹 마이닝 방법론에 퍼지이론을 도입한 새로운 방법론을 제안하였다.

ABSTRACT

The importance of web Mining is highlighted with growth of Electronic Commerce.

Web Mining is the important field of subject for studying customer's purchasing trend in Electronic Commerce.

This research defined customer's purchasing process as Fuzzy environment in Electronic Commerce. And it suggests new methodology that introduces Fuzzy theory based on current Web Mining methodology

키워드 : 퍼지구조모델링법, 웹마이닝, 전자상거래
FSM, Web Mining, Electronic Commerce

* 경남대학교 산업공학과 박사과정(Kyungnam University Graduate Schools Department of Industrial Engineering)

** 경남대학교 테크노경영학부 부교수(Kyungnam University Division of Techno-Management professor)

1. 서 론

Internet과 web기술의 발전으로 인한 전자상거래의 비약적인 발전은 기존의 상거래에 많은 변화를 가져왔다. 특히 기업과 소비자간(B2C)의 전자상거래에서는 하루가 다르게 새로운 가상상점이 생겨나고 있다.

가상상점의 발전은 가상상점을 이용하는 다양한 고객에게 좀 더 나은 서비스를 위해 고객의 구매행태에 대한 깊은 연구를 요구하게 되었다.

이러한 요구는 웹상에 존재하는 디지털화된 자료를 이용하여 필요한 고객의 구매행태 정보를 추출해내고자 하는 Web Mining분야의 비약적인 발전을 가져왔다. 그러나 이제까지의 Web Mining 방법은 고객이 구매한 이력이나 검색한 이력 등과 같은 구매정보와 고객정보를 통계적으로 분석하여 고객의 미래 구매행태를 예측하는 것이었다. 이러한 Web Mining방법은 충분한 구매정보가 있는 경우에는 오차가 적어질 수 있으나 그렇지 않을 경우에는 통계의 특성상 오차가 높아 질 수밖에 없는 문제점을 가지고 있다.

본 연구에서는 이러한 기존 Web Mning방법의 문제를 해결하기 위해 구매정보에서 동일 구매행태를 가질 것으로 예측되는 동일 유형의 고객이 동일 구매행태을 나타내지 않고 최적의 구매행태에 어느 정도 일치하는 구매행태를 나타낸다는 사실을 근거로 하여 이것을 퍼지환경으로 정의하고 Web Mining기법에 퍼지이론을 도입하였다.

즉 Web Mining방법에 퍼지이론을 도입함으로써 최소의 자료만으로 최적의 구매행태에 가장 접근하는 결과를 만들고자 하는 것이 본 연구의 목적이다.

본 연구에서는 특히 웹 사용 마이닝의 패턴 발견기법으로 기존방법과 다원적인 가치가 복합되어 있는 시스템의 구조인식에 적합하다고 알려진 퍼지구조모델링법(FSM)[1]을 적용한 것을 비교하여 Web Mining방법에 퍼지이론의 적용한 실질적인 예를 제시하였다.

2. 퍼지환경

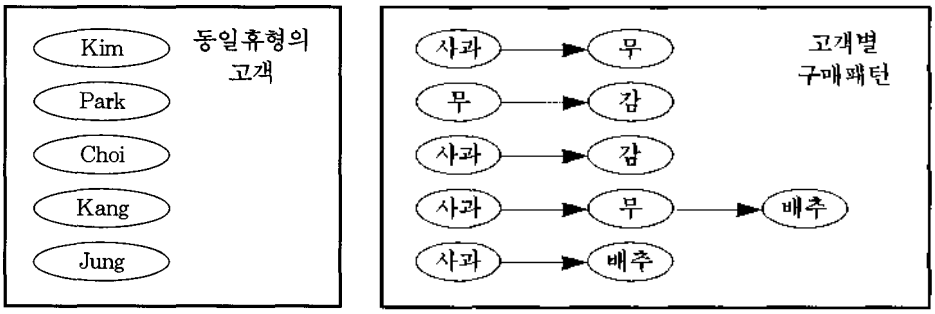
2.1 Web Mining상의 퍼지환경

본 연구에서 퍼지환경에 대한 정의는 동일 유형의 고객이 나타낼 수 있는 특정한 유형의 구매패턴에 각각의 고객들이 완전히 일치하는 구매패턴을 나타내는 것이 아니라 다양한 일치의 정도를 나타낸다는 사실을 바탕으로 하였다.

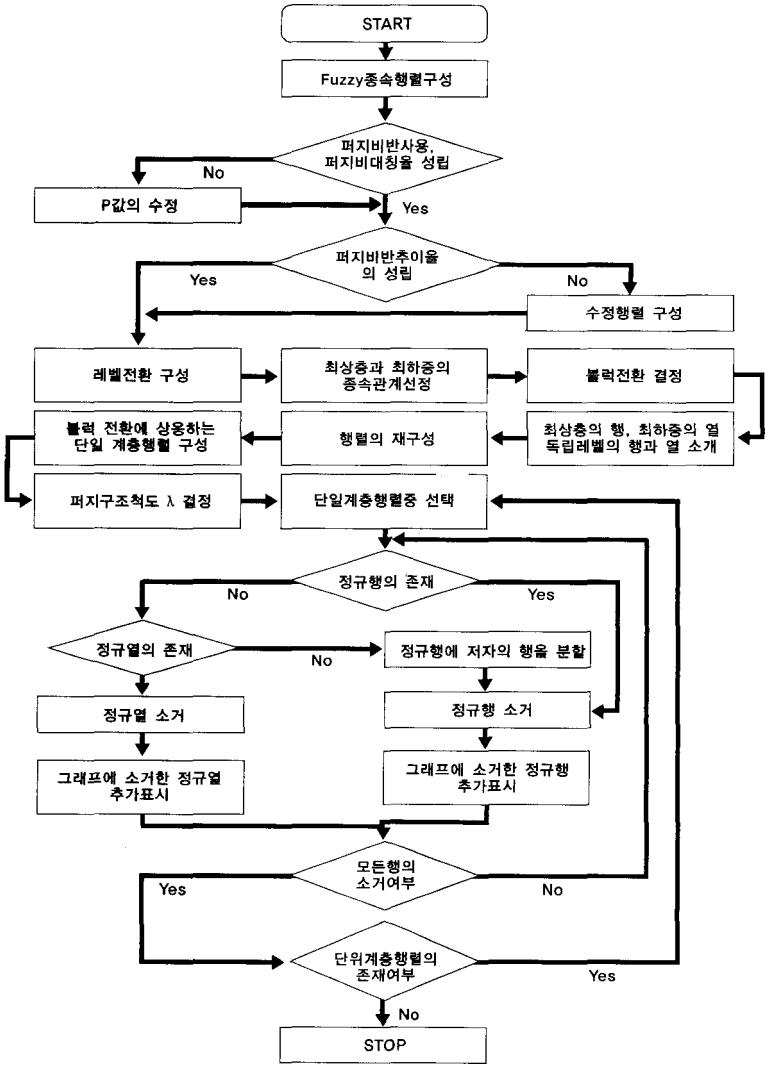
〈그림 1〉에서 보는 바와 같이 Kim, Park, Choi, Kang, Jung이라는 5명의 고객은 동일유형의 고객이다.

이러한 동일유형의 고객은 공통된 형태의 구매패턴을 가지고 있다. 그러나 5명이 고객이 나타내는 구매패턴은 그들이 가지고 있는 공통의 구매패턴과 어느 정도 일치는 하지만 완전히 동일하지 않은 것이 현실이다.

이렇게 일치의 정도로만 타나나는 고객들의 구매패턴에서 공통된 형태의 구매패턴을 찾아내기 위해 본 연구에서는 Web Mining에 퍼지환경을 정의하였다.



〈그림 1〉 Web Mining에서 Fuzzy환경의 정의



〈그림 2〉 Flowchart of FSM Method[2]

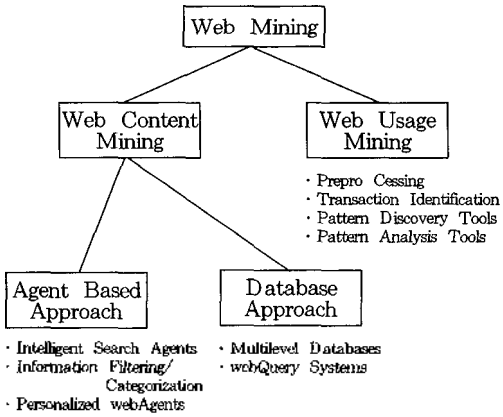
2.2 FSM(Fuzzy Structural Modeling)

동일유형의 고객들이 가지는 구매패턴을 퍼지환경으로 정의하고 이들 동일 유형의 고객들이 공통으로 가지는 공통구매패턴을 찾기 위해 본 연구에서는 퍼지구조모델링(FSM)법을 사용하였다.

퍼지구조모델링(FSM)법은 Tazaki등이 다원적인 가치가 복합되어 있는 시스템의 구조인식을 위해 제안한 방법[1]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. FSM방법의 전체적인 흐름을 보면 다음의 〈그림 2〉와 같다.

3. Web Mining

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



〈그림 3〉 Web Mining의 분류[3]

〈그림 3〉에서 살펴볼 수 있듯이 웹 마이닝은 어떻게 하면 웹 상에 존재하는 데이터를 효율적으로 분석하여 고객의 요구에 맞는 자료를 제공하는가를 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[3].

2.1 Web Content Mining

웹 내용 마이닝은 야후, 알타비스타, 라이코스와 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 데이터마이닝 기술을 확장한 접근 방식이다.

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

1) Intelligent Search Agent

관계 있는 정보를 검색하고 분야의 특성과

사용자의 프로파일을 이용하여 검색한 정보를 구조화하거나 해석하는 에이전트와 익숙하지 않은 정보 제공처의 구조를 학습하고 접촉하는 에이전트로 구분할 수 있다.

첫번째에 해당하는 에이전트로는 Harvest[4], FAQ-Finder[5], Information Mainfold[6], OCCAM[7], ParaSite[8] 등이 있으며, 두번째에 해당하는 에이전트로는 ShopBot[9], ILA(Internet Learning Agent)[10] 등이 있다.

2) Information Filtering/ Categorization Agent

많은 웹 에이전트는 다양한 정보 검색 기법[11]을 이용하여 자료를 자동으로 검색하고, 검색한 정보를 자동적으로 필터링하고 분류하기 위해 하이퍼텍스트 웹 문서의 특성을 이용한다[12],[13],[14],[15],[16].

HyPurshit[17]는 링크 구조안에 포함되어 있는 semantic information을 이용한다. BO(Bookmark Organizer)[18]는 웹 문서의 수집을 구조화하기 위해 구조화된 클러스터링기법과 사용자의 상호작용을 같이 이용한다.

3) Personalized Web Agent

이 분류에 속하는 에이전트는 사용자의 좋아하는 경향을 학습하고, 이 경향성에 기초하여 웹에서 관련 정보를 검색한다. 이런 에이전트들의 예로는 PAINT, Syskill & Webert, GroupLens, Firefly 등이 있다.

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

1) 다중레벨 데이터 베이스

이 접근 방식은 데이터 베이스의 가장 낮은 레벨은 하이퍼텍스트 문서와 같은 다양한 웹 저장소안에 저장되어 있는 반 구조화된 정보를 포함한다는 생각에 기초한다. 더 높은 레벨에서 메타 데이터 또는 일반화된 데이터는 더 낮은 레벨로부터 추출되고, 관계형 데이터베이스 또는 객체지향 데이터베이스와 같은 구조화된 집합 안에 구성된다. 예를 들어, Han.et.Al.[19] 은 각 계층이 더 낮은 계층에 일반화와 변환작용을 수행하여 얻어지는 다중 계층 데이터 베이스를 사용하였다.

2) 웹 질의 시스템

많은 웹 기반 질의 시스템들과 언어들은 SQL과 같은 데이터 베이스 표준 질의 언어, 웹 문서에 대한 구조화된 정보 또는 심지어 웹 검색에 사용되는 질의를 위해 자연어 처리를 이용한다. W3QL[20]은 하이퍼텍스트 문서의 구조에 기초한 구조화 질어어들과 정보 추출기법에 기초한 내용 질의어를 같이 이용한다. 재구조화를 위한 로직 기반 질의 언어인 WebLog[21]는 웹 정보 소스로부터 정보를 추출한다.

2.2 Web Usage Mining

웹 사용 마이닝은 웹 로그에서 불필요한 정보를 제거하고 필요한 정보만을 추출하는 전처리(Preprocessing)와 다양한 로긴 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[3].

1) Clustering analysis

비슷한 특징들을 가진 아이템들을 모아서 하나의 그룹으로 모으는 것을 가능케 한다.

2) Classification rule Dicovery

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

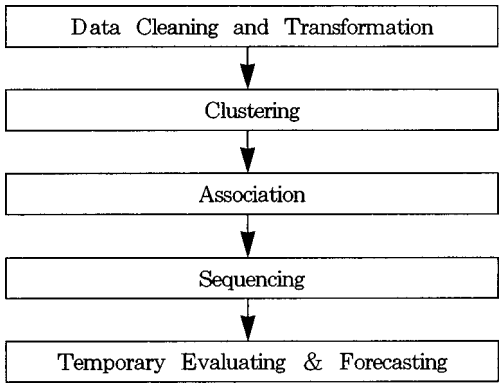
일반적으로 여러 아이템들의 집합으로 구성된 트랜잭션들이 저장된 데이터베이스에 적용된다. 문제는 한 트랜잭션안에 어떤 아이템들이 존재한다는 사실을 근거로, 다른 아이템들이 존재할 것이라는 것을 추정할 수 있는 아이템들 간의 모든 연관성(association)과 상호관계(correlation)을 찾는 것이다.

4) Sequence Pattern discovery technique

아이템들의 한 집합의 존재가 시간이 찍힌 순서화된 트랜잭션의 집합안에서 다른 아이템들에 의해 뒤따라진다는 상호관계 트랜잭션 팬턴들을 찾는 것이다.

4. 퍼지환경하의 Web Mining 방법

본 연구의 퍼지환경하의 Web Mining은 전자상거래 가상상점을 이용하는 고객들이 다양한 상품중 특정한 상품을 선택을 하여 구매할 때 그 선택의 과정에서 상품에 대한 선호의 정도가 있다는 현실적인 환경을 고려하여 고객의 유형별 구매패턴을 분석하는 방법을 말하는 것이다. 그 분석의 과정은 다음의 〈그림 4〉와 같다.



〈그림 4〉 Framwork of Analysis

4.1 Data Cleaning and Transformation

분석을 수행 하기전에 웹 로그에서 에러가 발생한 페이지 데이터 및 이미지 관련 데이터

와 같은 불필요한 데이터를 선택하여 제거하는 사전처리(Preprocessing)가 선행되어야 한다. 이것은 Data Cleaning과는 별개로 구분하기도 하나 본 연구에서는 하나로 보기로한다.

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

이렇게 작성된 사용기록은 분석을 위해 변환이 된다.

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

$$A_{ij} = \frac{(J_t)}{(i_t)+(J_t)} \quad (1)$$

i : 상품 을 구매한 고객의 수
j : 상품 을 구매한 고객의 수
t : 분석대상이 되는 특정한 기간

즉 분석대상이 되는 특정한 기간동안 상품j를 구매한 고객의 수에 상품 i와 j를 구매한 고객의 총수에 나누어 i와 j의 일대비교 행렬 개념을 도입한 퍼지 종속행렬을 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 구매고객들이 비슷한 계층이므로 하나의 그룹군으로 파악하여 공동 마케팅, DM발송과 같은 이벤트를 공동으로 수행할 수 있다.

본 연구에서는 Clustering을 위한 방법으로 일반적인 Web Mining에 사용되는 SOM(Self-Organization Map)을 이용하였다.

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 퍼지종속행렬을 바탕으로 퍼지구조모델링법을 이용하여 구매상품의 상호종속성을 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 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. 기존 Web Mining방법과의 비교

5.1 기존 Web Mining방법

5.1.1 연관성발견

연관성 발견은 MBA분석을 통해서 이루어진다. MBA를 이용한 분석을 위해서는 먼저 판매정보와 고객정보로부터 〈표 1〉과 같이 소비자별 구매품목을 추출한다. 추출된 정보를 이용하여 〈표 2〉와 같은 각 상품별 Co-occurrence Matrix를 구하여야 한다.

〈표 1〉 소비자별 판매정보

고객	판매정보
Kim	(사과, 무)
Park	(무, 감)
Choi	(사과, 감)
Kang	(사과, 무, 배추)
Jung	(사과, 배추)

〈표 2〉 연관성분석 : 각 상품별 Co-occurrence Matrix

	사과	무	감	배추
사과	0.80	0.40	0.20	0.40
무	0.40	0.60	0.20	0.20
감	0.20	0.20	0.40	0
배추	0.40	0.20	0	0.40

Co-occurrence Matrix를 이용하여 규칙을 발견하기 위해 confidence와 improvement라는 두가지 값을 계산하게 된다. 규칙은 조합(condition과 result)두 부분으로 이루어진다.

If condition, the then result.

Confidence는 다음식(2)에 의해 구해진다.

$$Confidence = \frac{p(condition \text{ and } result)}{p(condition)}$$

그러나 Confidence에는 문제점이 존재한다. 그것은 때에 따라서 임의로 예측하는것보다 결과가 올바르지 못한 경우가 있다는 것이다.

이러한 점을 보완하기 위해 improvement를 이용한다.

Improvement를 구하는 식은 식(3)과 같다.

$$Confidence = \frac{p(condition \text{ and } result)}{p(condition)p(result)}$$

Improvement가 1보다 클 경우에 Rule이 유효하다고 할 수 있다. 만약 improvement가 1보다 작다면 그 값은 효용성이 없다고 할 수 있다.

예를 들어 사과를 사면 무를 살 확률이 〈표 3〉의 confidence Matrix값에 의해 0.50이다. 즉 사과를 사는 사람 중에 50%는 무를 산다는 것이다. 이 값은 〈표 2〉에서 다른 것에 상관없이 무를 살 확률인 0.60보다 낮게 나온다. 즉 이규칙은 효용성이 없다.

〈표 4〉의 improvement값을 보면 0.83으로 1보다 낮아 효용성이 없다고 할 수 있다.

다른 예로 사과를 사면 배추를 살 확률이 〈표 3〉의 confidence값이 0.50이고 〈표 2〉에서

〈표 3〉 연관성분석 : Confidence Matrix

	사과	무	감	배추
사과	1	0.50	0.25	0.50
무	0.67	1	0.33	0.33
감	0.50	0.50	1	0
배추	1	0.50	0	1

〈표 4〉 연관성분석 : Improvement Matrix

	사과	무	감	배추
사과	·	0.83	0.625	1.25
무	0.83	·	0.83	0.83
감	0.625	0.83	·	0
배추	1.25	0.83	0	·

다른것에 상관없이 배추를 살 확률인 0.40보다 높게 나온다. 즉 이규칙은 효용성이 있다. 이것은 〈표 4〉의 improvement값을 보면 1.25로 1보다 높아 효용성이 있다고 할 수 있다.

이런 방법으로 연관규칙을 발견하여 서로 연관성이 높은 상품끼리 묶어서 소비자들에게 제공함으로써 소비자들의 구매의욕을 높이고, 판매자들은 판매효과도 높일 수 있게 된다. 또한 2차원 분석 뿐만 아니라, 3차원 분석, 4차원 분석 등을 함으로써, 보다 복합적인 연관 규칙을 발견할 수 있다.

5.1.2 순차 패턴 발견

순차 패턴 발견은 정적인 연관성 분석과 달리 시간개념이 포함된 동적 분석이다. 그렇지만, 분석기법은 연관성 분석엣 사용한 confidence와 improvement를 사용한다.

〈표 1〉을 이용하여 예를 들어보자. 단, 여기서는 시간의 개념이 있으므로 〈표 1〉에 적혀진 순서대로 상품을 구입하였다고 가정하자. 〈표 5〉는 〈표 1〉에 시간을 고려하여 co-occurrence matrix를 작성한 결과이다.

〈표 5〉는 〈표 1〉과 달리 대각행렬을 중심으로 서로 대칭인 행렬이 아니다. 이러한 이유는 시간 순서가 존재하기 때문에, (사과, 무)와 (무, 사과)는 서로 다른 패턴이기 때문이다.

5.1.3 유사 시간 순차 발견

유사 시간 순차 발견은 연관성 분석이나 순차 패턴 발견과는 다른 방법으로 진행된다. 유사 시간 순차 발견은 특정 시간 동안의 패턴을 발견하는 것이 중심이 된다.

이러한 유사 패턴 발견은 신경회로망을 이용하면 쉽게 파악이 가능하다. Rigoll(1996)과 Kandel(1997) 등은 신경회로망을 이용하여 패턴 인식에 대한 연구를 수행하였고, 신경회로망을 이용한 패턴 인식에 관한 연구는 많이 이루어졌고, 아직도 많이 이루어 지고 있다.

서로 유사한 패턴을 찾음으로써 앞으로의 소비량 예측을 쉽게 할 수 있게 된다. 소비량 예측이 보다 정확해지면, 소비자의 요구가 있기 전에 미리 출하함으로써 농산물의 신선도가 떨어지는 것을 방지할 수 있는 등 농산물 전자상거래의 문제점을 적절히 해결할 수 있다.

5.2 Fuzzy Web Mining방법

본 연구에서 제안한 Fuzzy Web Mining의 소비자 구매패턴정보를 추출하기 위한 연결 분석은 다음과 같은 두 단계로 이루어진다.

〈표 5〉 순차 패턴 발견 : Co-occurrence Matrix

	사과	무	감	배추
사과	·	0.40	0.20	0.40
무	0	·	0.20	0.20
감	0	0	·	0
배추	0	0	0	·

[1단계] 연관성 및 순차 패턴 발견
: 각각의 소비자가 어떤 상품을 동시에 구매하는지와 소비자들이 임의의 상품을 구매한 후 어떤 상품을 구매하는가 분석
[2단계] 유사 시간 순차 발견
: 어떠한 특정 기간 동안에 어떠한 상품들이 서로 유사한 패턴을 가지고 거래가 이루어졌는지와 특정한 시간 차이(lead or leg time)를 두고 서로 유사한 패턴이 있는가를 분석

5.2.1 연관성 분석 및 순차패턴의 발견

본 연구에서 제안한 Fuzzy Web분석을 위해서 먼저 판매정보와 고객정보로부터 기존 Web Mining방법에서와 같이 〈표 1〉과 같은 소비자별 구매품목을 추출한다.

그러나 이 판매정보는 분석대상이 되는 시간 내에서라는 제한이 하나 더해진다. 그리고 이렇게 추출된 판매정보와 식(1)을 이용하여 〈표 6〉과 같은 각 상품별 Fuzzy subordination Matrix를 구하여야 한다.

〈표 6〉 Fuzzy subordination Matrix

	사과	무	감	배추
사과	1	3/7	1/3	1/3
무	4/7	1	2/5	2/5
감	2/3	3/5	1	1/2
배추	2/3	3/5	1/2	1

〈표 7〉 Classification of Fuzzy Subordination Matrix($p=0.55$, $\lambda=-0.5$)

	Item	영향주고	영향받고
Top level	사과		감
Middle level	감		
low level	배추	사과, 감	
independence level	무		

〈표 8〉 Classification of Fuzzy Subordination Matrix($p=0.55$, $\lambda=0.5$)

	Item	영향주고	영향받고
Top level	사과		감, 배추
Middle level	감		
low level	배추		
independence level	무		

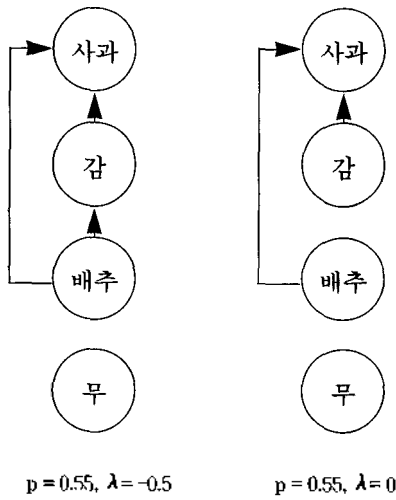
이렇게 만들어진 〈표 6〉의 Fuzzy subordination Matrix를 FSM방법에 의해 분석해보면 〈표 7〉 과 〈표 8〉와 같은 결과들을 얻을 수 있다.

여기서 p는 역치(threshold)로 자극에 대해 반응하는 분계점으로 미리 주어진 반개구간 (0,1]의 실수로 주어지게 된다.

λ 는 퍼지구조측도로 $-1 < \lambda < \infty$ 로 주어지게 된다.

본 연구에서는 역치 p의 경우 0.55를 기점으로 하여 그것보다 작은 경우는 FSM을 실행하기 위한 기본조건중 하나인 퍼지반추이율을 만족하지 못하였다.

구조파라메터 의 경우에는 - 부분과 + 부분으로 나누어 분석해보면 절대값이 같은 경우에 -인 경우가 +인 경우보다 구조가 많이 복잡한 것으로 알려져 있으나 대부분 동일하게 나타났다.



〈그림 5〉 FSM으로 도식화된 소비자 구매패턴

항목분류에 해당하는 Top Level, Middle Level, Low Level항목에 대해서 영향을 주고, 영향을 받는 항목으로 나누어진 것을 도식화해보면 〈그림 5〉와 같다.

이것은 하나의 상품 구매가 다음의 어떤 상품 구매에 영향을 미치는지를 나타낼 수 있다. 즉 위의 〈그림 5〉에서 보는 바와 같이 소비자 구매패턴은 일반적으로 배추를 구매한 고객이나 감을 구매한 고객은 사과를 구매하는 패턴을 나타내고 있었다.

이러한 결과는 기존의 Web Mining방법으로 분석한 결과인 사과를 구매한 고객이 배추를 구매한다는 패턴과 일치하였다. 또한 기존 Web Mining방법에서는 순차패턴분석을 따로 해야했던 것을 퍼지구조모델링법을 이용한 경우 그 순서가 명확히 나타난 것을 알 수 있다.

6. 결 론

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 중요한 연구분야중 하나로 대두되고 있는 퍼지 구조모델링법(FSM)을 활용하여 Web Mining에 퍼지이론을 접목시켜보았다. 즉 가상상점에서 자동화된 상품관리를 위해 사용되고 있는 Web Mining에서 기존의 분석방법에서 벗어나 동일유형의 구매패턴을 가진 고객들도 차이를 나타내는 상품에 대한 선호의

정도를 퍼지환경으로 정의하고 퍼지이론을 도입한 것이다.

기존 Web Mining방법과 Fuzzy Web Mining 방법의 비교에서 본바와 같이 동일한 결과를 나타냄은 물론 좀 더 세부적인 분석의 결과를 나타내었다.

또한 기존의 방법이 가지는 2단계를 1단계로 축소하여 분석할 수 있었다.

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.

참 고 문 헌

[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] 박영화, "퍼지환경하에서 다기준 의사결정법에 의한 중소기업의 경쟁력 평가", 동아대학교 박사학위논문, pp61-62, 1996.

[3] 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.

[4] T.Sullivan.Reading reader reaction : A proposal for for inferential analysis of web server log files. In Proc.3rd Conf.Human Factors & the Web, Denver, Coloradeo, June 1997.

[5] K. Hammond, R.Burke, C. Martin, and S. Lytinen. Faq-finder : A case-based approach to knowledge navigation. In working Notes of the AAAI spring Symposium : Information Gathering from Heterogeneous, Distributed Environments.AAAI Presss, 1995.

[6] T. Kirk, A. Y. Levy, Y.Sagiv, and D.Srivastava. The information manifold. In Working Notes of the AAAI spring Symposium : Information Gathering form Heterogeneous. Distributed Environments, AAAI Press, 1995

[7] C. Kwork and D. Weld. Planning to gather information. In Proc. 14th National Conference on AI, 1996

[8] E. Spertus. Parasite : Mining structural information on the web. In Proc. Of 6th International World Wide Web Conference, 1997.

[9] R. B. Doorenbos, O.Etzioni, and D. S. Weld. A scalable comparison shopping agent for the world wide web. Technical Report 96-01-03, University of Washington, Dept. of Computer Science and Engineering, 1996.

[10] M. Perkowitz and O.Etzioni. Category translation : learning to understand

information on the internet. In Proc. Of the 15th International Joint Conference on AI, pages 930-936, Montral, Canada, 1995.

[11] W. B. Frakes and R. Baeza-Yates. Information Retrieval Data Structures and Algorithms. Prentice Hall, Englewood Cliffs, NJ, 1992

[12] A. Z. Broder, S.C. Glassman, M.S. Manasse, and G Zweig. Systactic clustering of the web. In Proc. Of 6th International World Wide Web Conference, 1997.

[13] C. Chang and C.Hsu. Customizable multi-engine search tool with clustering. In Proc. Of 6th International World Wide Web Conference, 1997

[14] Y. S. Maarek and I.Z. Ben Shaul. Automatically organizing bookmarks per content. In Proc. Of 5th International World Wide Web Confernece, 1996.

[15] R. Weiss, B. Velez, M. A. Sheldon, C. Namprempre, P. Szilagyi, A. Duda,and D. K.Gifford. Hypursuit : a hierarchical network search engine that exploits content-link hypertext clustering. In Jypertext'96 : The Seventh ACM Conference on Hypertext, 1996

[16] M. R. Wulfekuhler and W. F. Punch. Finding salient features for personal web page categorization. In Proc. Of 6th International World Wide Web Conference, 1997.

[17] R. Weiss, B. Velez, M. A. Sheldon, C. Namprempre, P. Szilagyi, A. Duda,and D. K.Gifford. Hypursuit : a hierarchical network search engine that exploits content-link hypertext clustering. In Jypertext'96 : The Seventh ACM Conference on Hypertext, 1996

[18] Y. S. Maarek and I.Z. Ben Shaul. Automatically organizing bookmarks per content. In Proc. Of 5th International World Wide Web Confernece, 1996

[19] O. R. Zaiane and J. Han. Resource and Knowledge discovery in global information systems : A preliminary design and experiment, In Proc. of the First Int'l Conference on Knowledge Discovery and Data Mining, pages 331-336, Montreal, Quebec, 1995.

[20] D.Konopnicki and O. Shmueli. W3ql : A query system for the world wide web. In Proc. Of the 21th VLDB Conference, pages 54-65, Zurich, 1995.

[21] L.Lakshmanan, F. Sadri, and I.N. Subramanian. A declarative language for querying and restructuring the web. In Proc.6th International Workshop on Research Issues in data Engineering : Interoperability of Nontraditional Database Systems (RIDE-NDS'96), 1996.

저 자 소 개

주종문 (E-mail : econsulting@dgu.ac.kr)
1996. 경남대학교 산업공학과(학사)
1998. 경남대학교 산업공학과(석사)
1998 ~ 현재 경남대학교 산업공학과 박사과정
관심 분야 : Intelligent e-business System, CRM

황승국
동아대학교 산업공학과(학사)
동아대학교 산업공학과(석사)
일본오오사까부립대학교 경영공학 박사
현재 경남대학교 산업공학과 교수
관심 분야 : Fuzzy, TQM