

# 데이터 마이닝을 이용한 농산물 전자상거래의 온 오프라인 통합시스템

주종문\* · 황승국\*\*

\*경남대학교 산업공학과 박사과정 · \*\*경남대학교 벤처창업학부

## Integrated System of On-Off Line in Agricultural Products Electronic Commerce Based on Data Mining

Jong-Moon Ju\* · Seung-Gook Hwang\*\*

\* Kyungnam University Graduate Schools Department of Industrial Engineering

\*\* Kyungnam University Division of Venture Technology

The Internet, as a commercial tool, presented a new market that connects producers with consumers through the E-commerce. Now, E-commerce spreads over almost all industries through the Internet excluding some. This research indicates the reason why the E-commerce is not activated in agricultural industry, which is less developed than other industries. And it suggests a good example of E-commerce on the agricultural products combining on and off line markets. In addition, data-mining technique is suggested to analyze whole information in system.

**Key Words** : internet, Data Mining, E-commerce

### 1. 서론

인터넷과 전자상거래의 발달은 많은 물품들이 전자상거래를 통하여 거래가 이루어지게 하고 있다. 미국의 센던트 사(Cendant Corporation)는 100만종 이상의 물품과 서비스를 웹(web)상에서 판매하고 있는데, 주로 판매되는 것은 자동차, 전자제품, 카메라, 책, 향수, 꽃, 선물, 컴퓨터 하드웨어와 소프트웨어, 비디오게임 등이다[3]. 센던트 사의 조사를 토대로 볼 때, 거래되는 물품들의 대부분이 가공품 혹은 표준화된 공산품이라는 것을 알 수 있다. 농업을 통해 얻어지는 것은 꽃 외는 전자상거래가 거의 이루어지지 않고 있는 것이다. 이렇듯, 아직까지 농산물의 전자상거래가 활성화되지 않는 이유를 보면 다음과 같다[1].

첫째, 농산물은 신선도의 유지가 필수인데 신선도 유지를 위해 소비자의 주문을 받고 출하를 하면 소비자의

요구에 적절히 대응하기 힘들고, 소비자의 요구에 적절히 대응하기 위해 미리 출하하여 저장을 하고 있다면, 신선도를 잃어버릴 염려가 있다.

둘째, 농산물의 상품 재생산 주기(Product reproduction cycle)는 다른 상품에 비해 매우 길어 공산품과 같이 소비자의 요구에 신속하고, 민첩하게, 대응하기가 힘들다.

셋째, 소비자에게 적절한 신뢰를 줄 수 있는 정보가 부족하다는 것이다. 즉 소비자가 상품을 선택하는데 필요한 정보가 가격과 사진 외에는 특별한 것이 없어서 판단이 힘들다는 것이다. 공산품과 달리 농산물은 맛이나 향기와 같은 감각적인 정보가 필요한데 이러한 요구를 전자상거래 시스템만으로 충족시키기는 힘들다는 것이다.

본 연구에서는 이러한 농산물 전자상거래의 활성화를 가로막는 문제점을 해결하기 위해 전자상거래에서 데이터 마이닝을 기반으로 하여 온라인과 오프라인을 통합

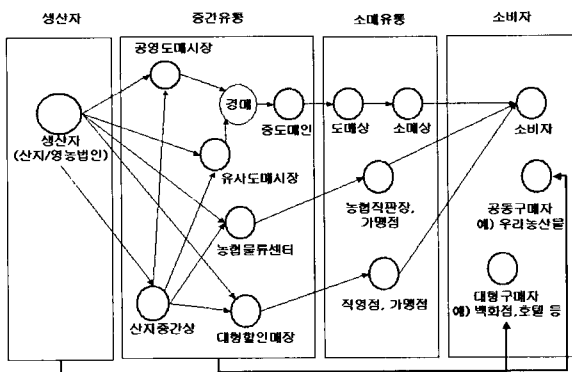
하는 전자상거래 시스템을 제안하였다.

본 연구의 구성은 먼저 현재 농산물 유통현황과 농산물 전자상거래 현황에 대해 알아보고 제안 전자상거래 시스템의 기반이 되는 데이터 마이닝에는 어떤 방법론이 있는지 알아본다. 그리고 이를 바탕으로 하여 앞서 제기한 문제점을 해결하기 위한 농산물 전자상거래에서 온라인과 오프라인의 통합한 시스템을 제안하고, 제안된 시스템을 적용하기 위한 데이터의 수집과 가공에 대해 연구하였다.

## 2. 농산물의 유통현황

### 2.1 국내 농산물의 유통구조

[그림1]에서 보면 생산자의 소비자 판매를 위한 다양한 유통경로가 개발되지 못하여 농민들의 출하처 선택의 폭이 제한적이고 산지에서 수집상들에 의해 50%이상이 출하되고 있다. 또한, 소비자에게 상품이 전달되기 전에



[그림1] 국내 농산물의 유통구조[2]

여러 도, 소매상을 거치기 때문에, 복잡한 유통단계와 중간 상인들의 과도한 이윤 추구로 인해 생산자와 소비자 모두가 피해를 입게 된다[2].

이러한 문제점을 해결하기 위한 가장 적절한 대안으로 제시되고 있는 것이 전자상거래를 통한 생산자와 소비자의 직거래라고 할 수 있다.

### 2.2 농산물 전자상거래 현황

한국을 비롯한 세계 여러 나라에서 운영되고 있는 전자상거래 시스템을 운영현황은 [표1]과 같다.

[표1]에서 보면 대부분의 전자상거래 사이트에서 소비자에게 가격과 통계정보를 제공하고 있지만 출하와 상품

정보에 대해서 제공하지 못하고 있다.

출하정보는 생산자가 생산한 물품의 량과 생산시기, 생산지 등을 포함하는 정보로 농산물의 경우 소비자의 구매결정에 매우 중요한 정보이다.

또한 농산물의 효능이나 식용방법과 같은 다양한 상품 정보 역시 소비자의 신뢰를 이끌고 구매의욕을 높일 수 있는 매우 귀중한 정보라고 할 수 있다.

이러한 정보의 제공은 생산자의 일방적인 정보제공만으로 이루어질 수 없다.

생산자 역시 소비자의 요구에 적절히 대응하기 위해 소비자의 선호물품이나 구매시기, 구매량과 같은 다양한 정보가 필요하고 이것은 앞서 말한 정보를 제공하는 원천이 되기 때문이다.

이렇게 소비자와 생산자의 요구정보를 적절히 제공해 주기 위해 데이터마이닝기법이 필요한 것이다.

## 3. 데이터 마이닝

### 3.1 데이터 마이닝의 활용

Michael(1997)에 의하면 데이터 마이닝은 “자동적 (automatic)이거나 반자동적(semiautomatic)인 방법들을 이용하여 많은 양의 데이터로부터 의미 있는 패턴 (pattern)이나 규칙(rule)을 찾아내는 것[4]”이라고 하였다. 또한, Peter(1998)는 “많은 데이터베이스(database)로부터 지금까지 알려지지 않고(unknown), 유용한(valid),

[표1]농산물 전자상거래 사이트[1]

사이트	가격 정보	통계 정보	출하 정보	상품 정보
www.rda.go.kr	○			○
www.garak.co.kr	○	○	○	
www.maf.go.kr	○	○		
www.afmc.co.kr	○			○
www.kati.net	○	○		○
www.nacf.co.kr	○			
www.affis.net	○		○	
www.ams.usda.gov	○			
www.agribiz.com	○			
www.ibcweb.co.jp	○			
www.honjo-osaka.or.jp	○		○	
www.tge.or.jp	○			
www.maff.go.jp		○		
www.fruitonline.com	○	○		
ats-sea.agr.ca		○		
www.agriculture.com	○	○		
www.dainet.de		○		

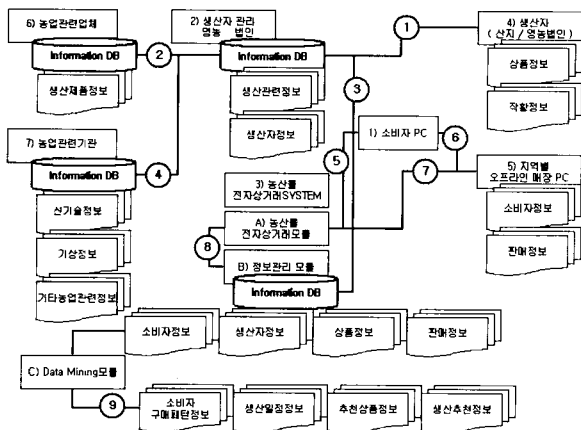
그리고 활용 가능한(actionable) 정보를 추출(extracting) 하는 과정[5]"이라고 정의하였다. 즉, 데이터 마이닝은 많은 데이터 내에서 의미 있는 경향(trend)이나 패턴(pattern)을 찾아내는 것이라고 할 수 있다. 이러한 데이터 마이닝은 여러 가지 방법이 존재한다.

Cooley(1997)는 군집분석(clustering analysis), 분류규칙발견(classification rule discovery), 연관규칙발견(association rule discovery), 연속패턴발견(sequence pattern discovery), 시각화(visualization)방법의 5가지로 분류하였고[6], Michael(1997)은 분류(classification), 판단(estimation), 예측(prediction), 유사그룹화(affinity grouping) 혹은 연관규칙(association rule), 군집(clustering), 설명(description)과 시각화의 6가지로 분류하였다[4]. 또한 Peter(1998)는 예측모델(Predictive Modeling), 데이터베이스 세분화(database segmentation), 연결분석(link analysis), 편향성탐지(deviation detection)방법의 4가지로 분류하였다[5].

본 연구에서는 이러한 다양한 데이터 마이닝방법중 연결분석과 예측을 사용한다.

#### 4. 시스템의 구성

본 연구에서 제안하는 농산물 전자상거래에서 온라인과 오프라인을 통합하는 방법의 구성은 [그림2]와 같이 제안되어 진다.



[그림2] 온라인과 오프라인을 통합하는 농산물 전자상거래 시스템구조

[그림2]에 나타난 시스템의 구성원을 보면 다음 [표2]와 같다.

특히 농산물 전자상거래 시스템은 온라인과 오프라인의 유기적인 통합을 위해 3가지 모듈로 구성하였다. 시스템구성간의 관계를 살펴보면 [그림2]에서 표시된 번

[표2] 시스템의 구성원

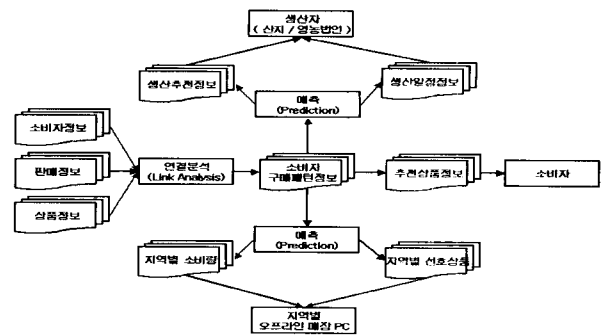
구성원	담당업무
1) 소비자	직접적인 소비의 주체
2) 생산자관리 영농법인	농산물의 공급과 품질관리
3) 농산물 전자상거래 시스템	농산물의 온라인과 오프라인 거래 중개와 정보제공
4) 생산자	농산물의 생산
5) 지역별 오프라인 매장	소비자들의 요구에 직접적으로 대 응
6) 농업관련 업체	농산물 생산과 판매의 지원
7) 농업관련 기관	농산물 생산과 판매의 지원

[표3] 농산물 전자상거래 시스템의 구성

구성모듈	역할
A) 농산물 전자상거래 모듈	온라인과 오프라인의 모든 거래를 관리
B)정보관리 모듈	전자상거래 모듈의 거래정보와 생산 자, 농업관련업체나 기관등 에서 제 공되는 다양한 정보를 관리
C)데이터 마이닝 모듈	정보관리 모듈의 정보를 데이터 마이 닝을 이용하여 소비자 구매패턴정보 나 생산일정정보를 생성

호에 따라 다음 [표4]와 같이 정의할 수 있다.

본 연구에서는 제안되어진 온, 오프라인을 통합하는 전자상거래 시스템에서 가장 핵심적인 역할을 하는 데이터 마이닝 모듈의 분석구조는 [그림3]과 같다. 데이터 마이닝 분석구조는 2단계로 나누어서 판매, 소비자, 상품정보를 연결분석을 통하여 소비자의 구매패턴정보를 만들어낸다.



[그림3] 데이터 마이닝 분석구조

[표4] 시스템 구성간의 관계

	내용
①	생산자 역시 자신이 생산하는 농산물의 생산량이나 품질과 같은 상품정보와 작황정보를 생산자 관리 영농법인 DB로 전송
②	농업관련업체에 있는 정보데이터베이스에서 농업관련업체에서 생산하는 생산정보를 생산자관리 영농법인 DB로 전송
③	생산자와 농업관련업체, 농업관련기관의 정보가 모인 생산자 관리 영농법인의 DB는 농산물 전자상거래 시스템의 DB와 실시간으로 연동됨
④	농업관련기관에 있는 정보데이터베이스에서 농산물 생산에 필요한 신기술정보나 기상정보와 같은 농업관련정보를 생산자관리 영농법인 DB로 전송
⑤	소비자는 집안의 PC나 기타 다양한 온라인망을 통하여 농산물 전자상거래 시스템에 회원으로 가입
⑥	지역별 오프라인 매장을 통하여 회원으로 가입, 농산물을 주문
⑦	소비자의 주문은 농수축산물 전자상거래 모듈(A)에 의해 종합되어지고 실시간으로 지역별 매장으로 배송정보가 전달되어짐
⑧	회원가입을 통하여 축적되어진 소비자정보와 소비자의 주문과 배송정보를 통하여 축적되어진 판매정보는 정보관리 모듈(B)의 DB에 실시간으로 입력
⑨	축적되어진 정보관리모듈(B)의 DB는 데이터마이닝 모듈(C)을 통하여 다양한 방법으로 분석되어져 농산물 전자상거래 시스템(A)을 통하여 소비자와 생산자에게 다양한 방법을 제공

소비자의 구매패턴정보라는 것은 시간과 지역 기타 소비자정보에 따른 선호상품과 소비량 등을 포함하고 있어서 생산자는 이를 바탕으로 예측을 통하여 다음해 생산일정과 생산품목을 결정할 수 있다. 또한 오프라인의 지역별 매장은 예측을 통하여 다음해 구비량과 품목을 결정할 수 있다.

또한 새로운 신규소비자들은 예측을 통하여 자신의 요구에 맞는 제품을 추천 받을 수 있는 것이다.

## 5. 시스템 운영

### 5.1 소비자 구매패턴정보

고객정보와 판매정보, 상품정보를 이용하여 각각 고객

의 구매 형태를 분석하여 앞으로의 소비자 구매패턴을 발견하여 효과적인 고객 관리가 가능해지고, 고객들의 소비량과 선호상품을 예측할 수 있다. 또한 이러한 소비자 구매패턴 정보에 맞게 생산자의 출하시기를 조절하여 고객들의 요구에 유연하게 대응하게 할 수 있다. 소비자 구매패턴정보를 추출하기 위한 연결분석은 다음과 같은 세가지 단계로 이루어진다.

[1단계] 연관성분석(Associations) : 각각의 소비자가 어떤 상품을 동시에 구매하는지를 분석

[2단계] 순차 패턴 발견(Sequential Pattern Discovery) : 소비자들이 임의의 상품을 구매한 후 어떤 상품을 구매하는가 분석

[3단계] 유사시간 순차 발견(Similar Time Pattern Discovery) : 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

#### 5.1.1 연관성발견

본 연구에서 연관성 발견은 MBA(Market Basket Analysis)분석[4]을 통해서 이루어진다. MBA를 이용한 분석을 위해서는 먼저 판매정보와 고객정보로부터 [표5]와 같이 소비자별 구매품목을 추출한다. 추출된 정보를 이용하여 [표6]와 같은 각 상품별 Co-occurrence Matrix를 구하여야 한다.

[표5] 소비자별 판매정보

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

[표6] 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 다음과 같은 식으로 구한다.

$$confidence = \frac{p(\text{condition and result})}{p(\text{condition})} \quad (1)$$

[표7] 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

그러나 Confidence에는 문제점이 존재한다. 그것은 때에 따라서 임의로 예측하는것보다 결과가 올바르게 맞지 않는 경우가 있다는 것이다. 이러한 점을 보완하기 위해 improvement를 이용한다. Improvement를 구하는 식은 다음과 같다.

$$Improvement = \frac{p(\text{condition and result})}{p(\text{condition})p(\text{result})} \quad (2)$$

[표8] 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	.

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 [표7]의 confidence Matrix값에 의해 0.50이다. 즉 사과를 사는 사람 중에 50%는 무를 산다는 것이다. 이 값은 [표6]에서 다른 것에 상관없이 무를 살 확률인 0.60보다 낮게 나온다. 즉 이 규칙은 효용성이 없다. [표8]의 improvement 값을 보면 0.83으로 1보다 낮아 효용성이 없다고 할 수 있다.

다른 예로 사과를 사면 배추를 살 확률이 [표7]의 confidence값이 0.50이고 [표6]에서 다른 것에 상관없이 배추를 살 확률인 0.40보다 높게 나온다. 즉 이 규칙은 효용성이 있다. 이것은 [표8]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 분석뿐만 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 분석에서 사용한 confidence와 improvement를 사용한다.

[표5]을 이용하여 예를 들어보자. 단, 여기서는 시간의 개념이 있으므로 [표5]에 적혀진 순서대로 상품을 구입하였다고 가정하자. [표9]는 [표5]에 시간을 고려하여 co-occurrence matrix를 작성한 결과이다.

[표9]는 [표5]과 달리 대각행렬을 중심으로 서로 대칭인 행렬이 아니다. 이러한 이유는 시간 순서가 존재하기 때문에, (사과, 무)와 (무, 사과)는 서로 다른 패턴이기 때문이다.

[표9] 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	.

### 5.1.3 유사시간 순차 발견

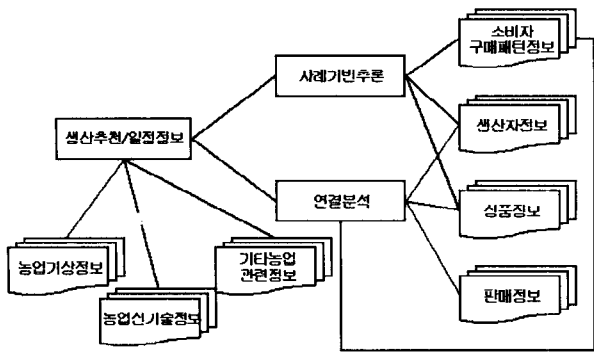
유사시간 순차 발견은 연관성 분석이나 순차 패턴 발견과는 다른 방법으로 진행된다. 유사시간 순차 발견은 특정 시간 동안의 패턴을 발견하는 것이 중심이 된다.

이러한 유사 패턴 발견은 신경회로망을 이용하면 쉽게 파악이 가능하다. Rigoll[7]과 Kandel[8] 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 생산추천 및 일정정보

농산물의 추천 및 일정정보를 제공하는 시스템은 [그림 4]과 같은 구조를 가지고 있다.



[그림4] 생산추천 및 일정정보 시스템 구조

연결분석에 의해 분석된 구매패턴정보를 바탕으로 생산자 정보의 생산자의 이력이나 상품정보의 상품별 이력을 사례기반추론을 통하여 생산추천 농산물을 추천하고 이 추천농산물을 구매패턴 정보를 바탕으로 다양한 농업관련 정보를 참고하여 생산일정을 제시하도록 하는 것이다.

## 6. 결론

현재 전자상거래업계에서는 기존 Web 위주의 전자상거래 시스템에서 발생하는 문제점들을 해결하기 위한 연구가 많이 이루어지고 있다.

본 연구는 아직까지 활성화되지 않은 농산물 전자상거래에서 Web 위주의 전자상거래로 발생할 수 있는 문제점들을 확인해보고 이 문제점을 해결하기 위한 방법으로서 온라인과 오프라인을 통합한 전자상거래 시스템을 제안하고, 데이터마이닝을 이용하여 온라인과 오프라인의 유기적 연결을 위한 정보분석시스템을 제안하였다.

본 연구에서 제안한 시스템을 요약하면, 온라인과 오프라인의 판매정보, 고객정보를 통합하여 데이터 마이닝의 연결분석을 통하여 소비자 구매패턴을 분석함으로써 소비자의 선호상품과 소비량을 예측할 수 있게 하였다. 또한 소비자의 선호상품과 소비량 예측을 바탕으로 생산자에게 생산품목을 추천해주고, 생산일정을 제시하여 줄 수 있도록 하였다.

소비자의 선호상품과 소비량을 예측하고 그에 따라 생산품목을 결정하고 생산일정을 결정함으로써 소비자는 요구에 맞는 상품을 신선함을 유지한 상태로 구매할 수 있고 생산자는 소비자의 요구에 적절하게 대응하는 것이 가능하게 되는 것이다.

또한 생산이 결정된 상품과 그 생산일정을 소비자에게 제시함으로써 소비자의 구매의사결정에 광범위한 정보를 제공할 수 있다.

본 연구에서 제안한 농산물 전자상거래의 온 오프라인 통합시스템의 궁극적인 발전 방향은 물류(Logistics)와 전사적자원관리(ERP) 개념의 도입을 통하여 현재 시스템 구성원의 유기적 결합을 더욱 공고히 하는 것이 필요하다.

## 참고문헌

1. 김태현, "데이터마이닝을 이용한 농산물전자상거래에서의 공급자 및 고객 관계관리에 관한 연구", 석사학위논문, KAIST, 2001.
2. 이용효, 류광택, 김민수, 전현철, 윤기봉, 정보화를 통한 농산물유통업무 개선방안, 한국전산원, 1998.
3. Secretariat for Electronic Commerce, U.S. Department of Commerce., *The Emerging Digital Economy*, .1999.
4. Michael J. A. Berry, Gordon Lindoff, "Data Mining Techniques For Marketing, sales, and Customer Support", John Wiley & Sons, Inc., 1997.
5. Peter C., Pablo H., Rolf S., Jaap V., Alessandro Z., "Discovering Data Mining From Concept To Implementation", Prentice Hall PTR, 1998.
6. R. Cooley, B. Mobasher, and J. Srivastava, "Web Mining : Information and Pattern Discovery on the World Wide Web", *Proc. of the 9th IEEE International Conference*, pp. 558-567, 1997.
7. Rigoll G., "Mutual information neural networks for dynamic pattern recognition tasks", *Proceedings of the IEEE international Symposium on Industrial Electronics*, pp. 80-85, 1996
8. Kandel A., Yan-Qing Zhang, BunkeH., "A genetic fuzzy neural network for pattern recognition", *Proceedings of the Sixth IEEE International Conference on Fuzzy Systems*, pp.75-78,1997

\* 본 연구는 2002학년도 경남대학교 학술논문게재연구비 지원으로 이루어졌음