

연관규칙을 이용한 고객의 구매경향에 관한 연구
(A Study on Customer's Purchase Trend Using Association Rule)

임 영 문, 최 영 두
강릉대학교 산업공학과
Young-Moon Leem, Young-Doo Choi
ymleem@knusun.kangnung.ac.kr

Department of Industrial Engineering, Kangnung National University
123. Chibyun-Dong, Kangnung, Kangwon-Do, 210-702, Korea

Abstract

General definition of data mining is the knowledge discovery or is to extract hidden necessary information from large databases. Its technique can be applied into decision making, prediction, and information analysis through analyzing of relationship and pattern among data. One of the most important work is to find association rules in data mining.

The objective of this paper is to find customer's trend using association rule from analysis of database and the result can be used as fundamental data for CRM(Customer Relationship Management). This paper uses Apriori algorithm and FoodMart data in order to find association rules.

1. 서 론

오늘날의 기업환경에서 효율적으로 경쟁하기 위해서는 기업은 경쟁사들보다 시장 상황을 잘 파악하고 분석할 수 있는 능력이 요구되고 있으며, 기업환경은 정보화 사회를 맞이하여 갈수록 치열해 지고 있다. 최근 들어 의사결정 지원시스템의 필요성이 증가함에 따라서 데이터웨어하우스의 구축이 증가하고, 이를 기반으로한 데이터 마이닝의 적용으로 의사결정의 지원수단으로 사용되고 있다. 그러한 데이터웨어하우스를 기반으로 한 데이터 마이닝을 이용하면 분석, 예측, 규칙, 통계정보등 의사결정에 필요로 되는 양질의 고급 정보를 얻을 수 있다.

본 연구에서는 데이터베이스를 이용한 고객의 물품 구매경향을 분석해 보고자 하는 것이며 이를 위한 알고리즘으로 연관규칙의 Apriori알고리즘을 적용하였다. 이를 통한 결과의 분석으로 FoodMart의 고객 물품 구매경향을 알 수 있으며, 추후 CRM(Customer Relationship Management)의 기초 자료로도 사용되어 질 수 있다고 보여진다.

2. 연관규칙 정의

2.1 연관규칙 탐사(Association Rule Mining)

연관규칙은 빈발 항목집합에 대한 연관규칙을 생성해 내는 작업으로 나눌 수 있다. 전체 항목집합(상품들의 집합)이라 불리는 문자들의 집합을 $I = \{i_1, i_2, \dots, i_m\}$ 라고 주어진다면, D는 트랜잭션(Transaction ; 판매기록 - 판매데이터의 한 레코드)의 집합이고, 각 트랜잭션 T는 $T \subseteq I$ 인 항목들의 집합이라고 정의된다. 한 트랜잭션에서 구입하는 항목들의 양에는 상관하지 않고, 각 항목은 구입여부를 나타내는 이진 변수이다. 각 트랜잭션은 TID라 불리는 식별자와 연계되어 있다. X를 항목들의 집합(Itemset)이라 하고 트랜잭션 T가 X를 포함한다면 $X \subseteq T$ 이라 한다. 연관규칙은 $X \Rightarrow Y$ 의 형식으로 표시되고, $X \subseteq I$, $Y \subseteq I$ 및 $X \cap Y = \emptyset$ 는 공집합이다. 또한 $X \Rightarrow Y$ 신뢰도(Confidence) c를 가지고 있다는 것은 X를 포함하는 D의 트랜잭션들 중 c%가 Y도 포함하고 있음을 의미한다. 또한 $X \Rightarrow Y$ 는 트랜잭션 집합 D내에 지지도(Support) s%를 갖는데 이는 D의 트랜잭션들 중 s%는 XUY를 포함하고 있음을 뜻한다. 최소 지지도(Minimal Support) 이상을 갖는 항목집합을 빈발항목집합(Large Itemset or Frequently Itemset)이라 한다. K개의 항목들로 이루어진 빈발 항목집합을 빈발 k-항목집합이라고 한다. 빈발 k-항목 집합들의 집합을 L_k 라하고 이를 위한 후보 k-항목집합들의 집합을 C_k 라고 한다.

연관규칙을 발견하는 과정은 다음 두 단계로 구성된다. 첫째는 빈발항목 집합들을 찾는 다. 즉 주어진 최소 지지도 이상의 트랜잭션들의 지지를 갖는 항목 집합들의 모든 집합을 찾는 것이며, 둘째는 데이터 베이스에 대한 연관규칙을 찾기 위해 빈발항목집합을 사용한다. 발견된 연관규칙은 지지도 및 신뢰도를 가지는 수치 값으로 정량화 된다.[1,3]

2.2 Apriori Algorithm

Apriori 알고리즘은 각각의 반복시행에서 빈발 항목 집합들에 대한 후보 집합을 생성하고 각 후보 항목 집합의 빈도수를 계산한 후 주어진 최소 지지도 이상의 빈발 항목 집합들을 결정하게 된다. 최초의 반복시행에서 Apriori는 각 항목에 대한 빈도수를 계산하기 위해 모든 트랜잭션을 읽는다. 트랜잭션 데이터 베이스가 [표1]과 같이 주어졌을 때 데이터 베이스 D를 검색하여 1-항목집합들로 구성된 빈발 1-항목집합들 L_1 이 결정된다. 빈발 2-항목집합들의 집합 L_2 을 찾을 때 빈발 항목집합들의 모든 부분 집합들이 최소 지지도 이상을 가져야 한다는 사실로부터 Apriori는 항목 집합들의 후보 집합인 C_2 를 생성하기 위해 $L_1 * L_1$ 을 사용한다. 여기서 *는 접속(concatenation)연산이다. 다음으로 데이터베이스

[표 1] 트랜잭션 데이터베이스 D의 예

TID	Items
100	ACD
200	BCE
300	ABCE
400	BE

D내의 4개의 트랜잭션들을 읽고 C₂의 후보항목집합의 지지도를 세어본다. 빈발 2-항목집합들의 집합 L₂는 C₂의 후보 2-항목집합의 지지도를 기반으로 결정된다.

후보 3-항목집합들의 집합 C₃는 다음과 같이 L₂로부터 생성된다. {BC}, {BE}와 같이 동일한 첫 번째 항목을 가지고 있는 두개의 빈발 2-항목집합들을 찾는다. 그 다음 Apriori는 이들의 두 번째 항목들로 구성된 {CE}가 빈발 2-항목집합들을 이루는지 여부를 검사한다. {CE} 자체가 빈발항목 집합이고 {BCE}의 모든 부분집합들이 빈발항목집합이므로 {BCE}는 후보 3-항목집합이 된다. 이와 같은 절차를 통한 후보 생성방법의 제시가 Apriori알고리즘이 기여한 가장 중요한 부분으로 평가받고 있다. L₂로부터 그 밖의 다른 후보 3-항목집합이 나오지 않는다. 이때 Apriori는 모든 트랜잭션들을 읽어 빈발 3-항목집합들인 L₃을 찾는다. L₃로부터 후보 4-항목집합들이 나오지 않으므로 Apriori는 빈발 항목집합들을 찾는 작업을 끝낸다. [표 2]은 전체적인 수행과정을 보여준다.[4,5,6]

[표 2] 후보 항목집합들과 빈발 항목집합들의 생성

	C ₁		L ₁																														
	<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr><th>Itemset</th><th>Sup</th></tr> </thead> <tbody> <tr><td>{A}</td><td>2</td></tr> <tr><td>{B}</td><td>3</td></tr> <tr><td>{C}</td><td>3</td></tr> <tr><td>{D}</td><td>1</td></tr> <tr><td>{E}</td><td>3</td></tr> </tbody> </table>	Itemset	Sup	{A}	2	{B}	3	{C}	3	{D}	1	{E}	3	SCAN D ==>	<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr><th>Itemset</th><th>Sup</th></tr> </thead> <tbody> <tr><td>{A}</td><td>2</td></tr> <tr><td>{B}</td><td>3</td></tr> <tr><td>{C}</td><td>3</td></tr> <tr><td>{E}</td><td>3</td></tr> </tbody> </table>	Itemset	Sup	{A}	2	{B}	3	{C}	3	{E}	3								
Itemset	Sup																																
{A}	2																																
{B}	3																																
{C}	3																																
{D}	1																																
{E}	3																																
Itemset	Sup																																
{A}	2																																
{B}	3																																
{C}	3																																
{E}	3																																
C ₂		SCAN D ==>																															
<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr><th>Itemset</th></tr> </thead> <tbody> <tr><td>{AB}</td></tr> <tr><td>{AC}</td></tr> <tr><td>{AE}</td></tr> <tr><td>{BC}</td></tr> <tr><td>{BE}</td></tr> <tr><td>{CE}</td></tr> </tbody> </table>	Itemset	{AB}	{AC}	{AE}	{BC}	{BE}	{CE}	<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr><th>Itemset</th><th>Sup</th></tr> </thead> <tbody> <tr><td>{AB}</td><td>1</td></tr> <tr><td>{AC}</td><td>2</td></tr> <tr><td>{AE}</td><td>1</td></tr> <tr><td>{BC}</td><td>2</td></tr> <tr><td>{BE}</td><td>3</td></tr> <tr><td>{CE}</td><td>2</td></tr> </tbody> </table>	Itemset	Sup	{AB}	1	{AC}	2	{AE}	1	{BC}	2	{BE}	3	{CE}	2	<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr><th>Itemset</th><th>Sup</th></tr> </thead> <tbody> <tr><td>{AC}</td><td>2</td></tr> <tr><td>{BC}</td><td>2</td></tr> <tr><td>{BE}</td><td>3</td></tr> <tr><td>{CE}</td><td>2</td></tr> </tbody> </table>	Itemset	Sup	{AC}	2	{BC}	2	{BE}	3	{CE}	2
Itemset																																	
{AB}																																	
{AC}																																	
{AE}																																	
{BC}																																	
{BE}																																	
{CE}																																	
Itemset	Sup																																
{AB}	1																																
{AC}	2																																
{AE}	1																																
{BC}	2																																
{BE}	3																																
{CE}	2																																
Itemset	Sup																																
{AC}	2																																
{BC}	2																																
{BE}	3																																
{CE}	2																																
C ₃		SCAN D ==>																															
<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr><th>Itemset</th></tr> </thead> <tbody> <tr><td>{BCE}</td></tr> </tbody> </table>	Itemset	{BCE}	<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr><th>Itemset</th><th>Sup</th></tr> </thead> <tbody> <tr><td>{BCE}</td><td>2</td></tr> </tbody> </table>	Itemset	Sup	{BCE}	2	<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr><th>Itemset</th><th>Sup</th></tr> </thead> <tbody> <tr><td>{BCE}</td><td>2</td></tr> </tbody> </table>	Itemset	Sup	{BCE}	2																					
Itemset																																	
{BCE}																																	
Itemset	Sup																																
{BCE}	2																																
Itemset	Sup																																
{BCE}	2																																

3. 실험 방법

본 연구를 위한 시스템 구성을 위해 Apriori알고리즘을 MS-SQL Server 7.0, Visual Basic 6.0과 ADO(ActiveX Data Object)로 구현하였으며, FoodMart[2]의 데이터로부터 특성을 파악한 뒤, 필요 항목을 SQL 질의를 통하여, Txt파일이나, MDB파일로 트랜잭션 파일을 생성한다.

프로그램의 개략적인 구성은 크게 3가지 단계로 구성할 수 있다. 판매 일자와, 고객에 따른 판매 상품이 순차적으로 저장되어 있는것을 트랜잭션 데이터 형태로 변환하기 위한 작업이 전처리 단계인 첫 번째 단계([그림1])이고, 두 번째 단계로는 연관규칙의 Apriori 알고리즘을 적용하기 위한 단계로 빈발항목과 빈발 후보 항목에 대한 질의를 구성하여 결과를 출력하는 부분이며, 마지막 단계로는 알고리즘 적용 후 생성된 연관규칙을 가지

고, 분석하는 단계로 구분할 수 있다.

맨 처음 모든 데이터를 구입일자, 고객번호, 상품번호 순으로 질의하여 MS-SQL에 임포트 한 후, MS-SQL에서 1-항목집합의 지지도 수를 카운트 한 뒤, 질의 결과를 MDB파일이나, Txt파일로 트랜잭션을 저장한다. 그 후 Apriori의 알고리즘을 적용하여 2-항목집합서부터 k-항목집합까지 구하기 위한 루프를 적용한다. 그 후 도출된 연관규칙에 대하여, 관련상품과의 관계를 파악, 분석한다.

레코드수, 구입일자, 고객번호, 상품번호	T10, 상품1, 상품02, 상품03, 상품04, 상품05, 상품06, 상품07, 상품08, 상품09, 상품10, 상품11, 상품12, 상품13, 상품14, 상품15
1, 732, 534, 12	1, 12, 166, 212, 322, 405, 921, 1287
2, 732, 534, 166	2, 59, 115, 119, 225, 309, 907
3, 732, 534, 212	3, 50, 61, 127, 267, 416, 576, 1409
4, 732, 534, 322	4, 46, 112, 153, 246, 249, 561, 665, 1162, 1434
5, 732, 534, 405	5, 23, 1063, 1267
6, 732, 534, 921	6, 61, 1215
7, 732, 534, 1287	7, 342, 403, 509, 532, 1091
8, 732, 1013, 59	8, 4, 320, 391, 663, 1006, 1339
9, 732, 1013, 115	9, 491, 803, 1056, 1273, 1339, 1479
10, 732, 1013, 119	10, 877, 1081, 1176, 1210, 1226, 1297, 1326
11, 732, 1013, 225	11, 16, 89, 332, 682, 965, 1071, 1464
12, 732, 1013, 309	12, 56, 402, 432, 502, 1039, 1339, 1372
13, 732, 1013, 607	13, 366, 382, 575, 596, 652, 662, 685, 941, 945, 1045, 1534
14, 732, 1325, 90	14, 11, 377, 443, 556, 1232, 1366, 1432
15, 732, 1325, 61	15, 445, 1181, 1348, 1368
16, 732, 1325, 127	16, 180, 246, 906, 1028, 1035, 1189, 1372
17, 732, 1325, 267	17, 201, 328, 434, 438, 1170, 1253, 1509
18, 732, 1325, 416	18, 182, 599, 838, 1058, 1148
19, 732, 1325, 576	19, 75, 165, 231, 756, 1174, 1357, 1434
20, 732, 1325, 1409	20, 371, 594
21, 732, 1425, 46	21, 139, 582, 811, 855, 1188
22, 732, 1425, 112	22, 404, 465, 571, 581, 814, 910, 1226, 1447
23, 732, 1425, 169	23, 126, 163, 236, 1032
24, 732, 1425, 248	24, 375, 612, 645, 662, 761, 984, 1456
25, 732, 1425, 249	25, 909, 932, 1392
26, 732, 1425, 561	26, 249, 682, 735, 892, 898, 1341
27, 732, 1425, 665	27, 90, 228, 461, 683, 693, 1000, 1070, 1066, 1269, 1543
28, 732, 1425, 1162	28, 449, 455, 1225, 1470
29, 732, 1425, 1434	29, 252, 330, 371, 521, 773, 778, 1029, 1046, 1376, 1378, 1406, 1416, 1427, 1513
30, 732, 1550, 23	30, 248, 365, 436, 558, 602, 668, 1076, 1065, 1107, 1239, 1528
31, 732, 1550, 1063	31, 130, 240, 323, 1235, 1228
32, 732, 1550, 1287	32, 14, 20, 204, 424, 703
33, 732, 1641, 61	33, 406, 485, 541, 676, 1015, 1161, 1317, 1332, 1502
34, 732, 1641, 1215	34, 657, 680
35, 732, 2229, 342	35, 5, 409, 1170
36, 732, 2229, 403	36, 121, 697, 1076
37, 732, 2229, 509	37, 198, 248, 265, 751, 757, 907, 1200, 1473
38, 732, 2229, 592	38, 361, 1197
39, 732, 2229, 1091	39, 45, 931, 1374
40, 732, 2439, 4	40, 142, 176, 192, 292, 319
41, 732, 2439, 320	41, 69, 319
42, 732, 2439, 391	42, 6, 27, 73, 352, 362, 626, 782, 878, 982, 1508, 1527
43, 732, 2439, 663	43, 188, 435, 1142, 1144
44, 732, 2439, 1006	44, 299, 549, 1048
45, 732, 2439, 1339	45, 11, 71, 94, 350, 727, 833, 837, 1139, 1522, 1532
46, 732, 2481, 491	46, 681, 791
47, 732, 2481, 803	47, 435, 645, 945, 1095
48, 732, 2481, 1058	48, 58, 201, 652, 945, 1095, 1500, 1543
49, 732, 2481, 1287	49, 38, 201, 652, 945, 1095, 1500, 1543

[그림 1] 전처리과정 후 생성된 데이터와 트랜잭션 데이터

4. 데이터 분석

FoodMart는 MS-SQL 7.0 OLAP[2] 예제 데이터이고, 본 연구에서 사용하는 데이터는 상품 판매 실적을 기록한 데이터로, 장바구니 분석을 위한 데이터로는 충분하다고 생각된다. 다만, FoodMart데이터가 인위적으로 만들어 졌는지, 실제 판매 데이터에 대한 기록인지에 대한 내용은 파악되지 않았다. FoodMart의 데이터 테이블 구조는 총 15개의 테이블로, 상품테이블, 고객테이블, 재고테이블, 판매테이블, 상품분류항목 테이블, 체인점 테이블 등으로 이루어져 있으며, 체인점의 소유형태에 따라 6개 범주로 구성되어 있다. [표3]는 1997년과 1998년 데이터의 속성에 대한 비교를 보여준다.

FoodMart의 상품 판매 데이터는 sales_fact_1997과 sales_fact_1998 테이블이 있는데, 각각 상품판매 기록으로는 86837건과, 164558건이 있으며, 1997년 판매기록의 경우 20522개의 트랜잭션 데이터를 얻을 수 있었으며, 평균 4.838개의 상품구매와, 최대 17개의 상품

[표 3] FoodMart 데이터 연도별 비교

	1997년	1998년
레코드수	86837건	164558건
트랜잭션수	20522건	34015건
평균 상품구매수	4.838개	4.2288개
최대 상품구매수	17개	28개

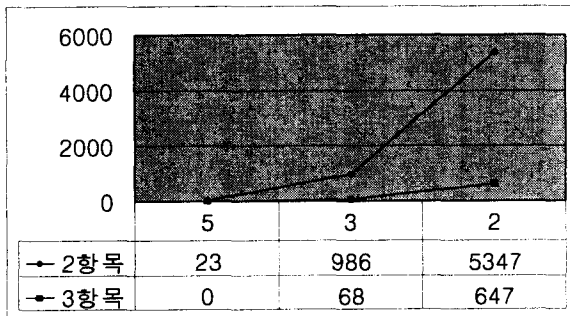
을 구매 한 것으로 나타났다. 1998년의 데이터를 가지고 트랜잭션 데이터를 생성해 본 결과 34015건의 트랜잭션이 발생되었으며, 이는 한 고객이 평균 4.2288개의 상품을 구매하였고, 최대 28의 상품 구매 한 것으로 나타났다. 등록된 고객은 10281명이며, 상품은 1560개에 110개의 범주로 분류된다.

데이터베이스 내에서 연관규칙을 찾기 위해서 판매기록 테이블과, 고객테이블, 상품테이블 및 상품 분류 테이블을 가지고 전처리 과정을 거치면, [그림1]의 왼쪽 프레임과 같은 데이터를 얻을 수 있으며, 알고리즘 적용에 필요한 트랜잭션 데이터 형태로 만들면 [그림1]의 오른쪽 프레임과 같은 예제 데이터 베이스 [표1]의 형태를 얻을 수 있다. 여기서 생성된 총 54537건의 트랜잭션에 대해 연관규칙을 적용한 뒤 생성된 연관규칙에 따라, 1997년과, 1998년의 데이터에 적용, 연도별 트랜잭션에서 생성된 연관규칙을 가지고 총 트랜잭션에 적용 비교 분석하였다.

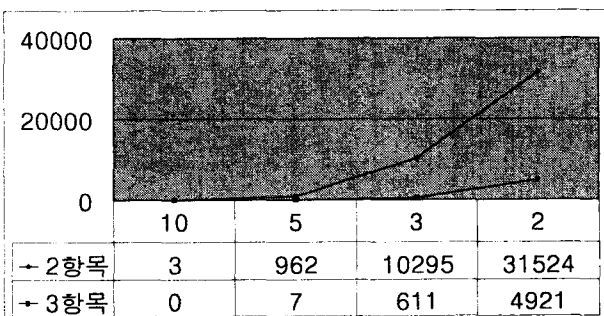
5. 결과분석

연관규칙은 고객의 물건 구매행위에 대한 경향이나 패턴을 찾는 것으로, 전체 구매데이터에서 고객이 A라는 물건을 구매한 뒤에 B라는 물건을 살 확률을 구하는 것으로, 여러 항목집합과, 판매되는 물건들의 특성 및 그러한 물건을 구매한 고객들의 특성 등을 분석하여, 판매 경향이나, 마케팅 전략 수립에 많은 도움을 줄 것이다.

FoodMart의 데이터를 분석해 본 결과 1997년은 최소 지지도가 5이상(0.025%)인 경우와, 1998년은 10이상(0.03%)인 경우에 관련 상품은 없는 것으로 나타났으며, 최소 지지도가 5인 경우의 1997년 빈발 2항목은 23, 3항목은 0개며, 1998년의 경우 빈발 2항목은 962개, 빈발 3항목은 7건으로 나타났다. [그림2]와 [그림3]에서 이런 차이를 보이는 이유는 동일한 상품(1560개)에 대한 상품판매 건수 즉 트랜잭션 데이터가 다르기 때문에 빈발 항목의 건수도 차이가 난 것으로 분석된다. 1997년의 경우 최소 지지도가 워낙 낮기 때문에 이 규칙에 의한 실적용에 있어서 우연에 의한 일치에 대한 생각도 고려해야 한다.

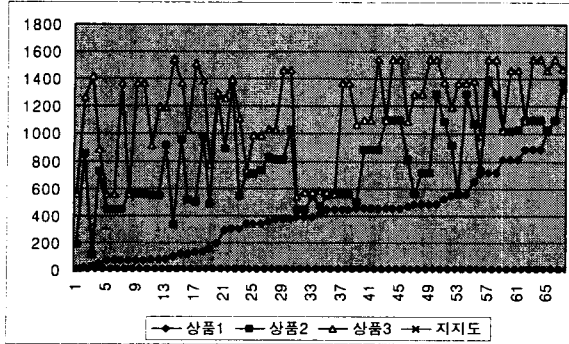


[그림 2] 1997년 지지도별 항목집합 개수

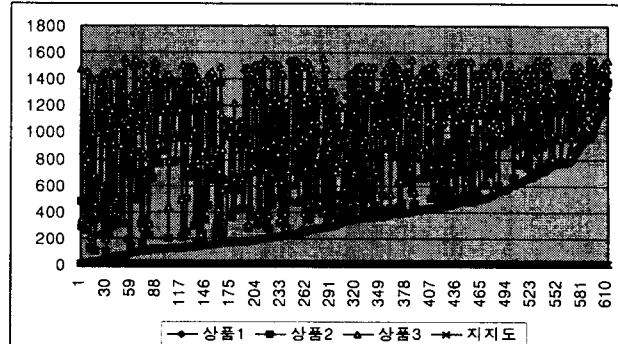


[그림 3] 1998년 지지도별 항목집합 개수

[그림4]와 [그림5]는 1997년, 1998년 지지도 3인 경우의 상품데이터를 나타내고 있는 것으로, 상품1을 기준으로 상품2와 상품3에서는 특정상품에 대해 군을 형성하고 있는것을 볼 수 있다.

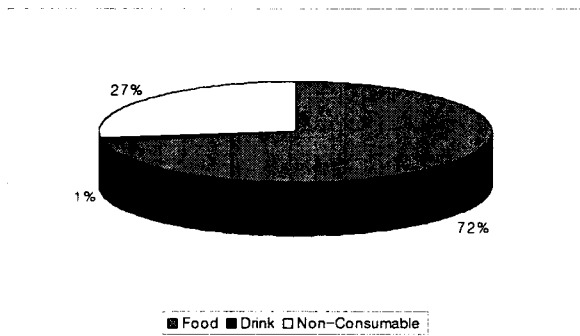


[그림 4] 1997년 3항목집합(지지도 3)

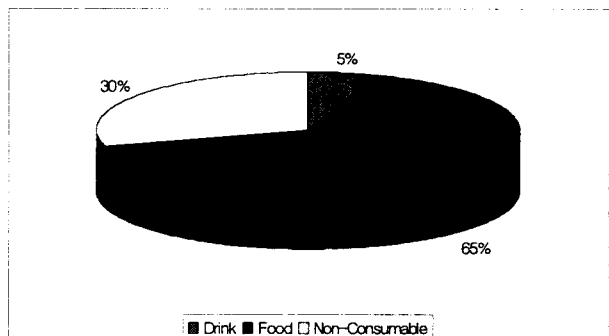


[그림 5] 1998년 3항목집합(지지도 3)

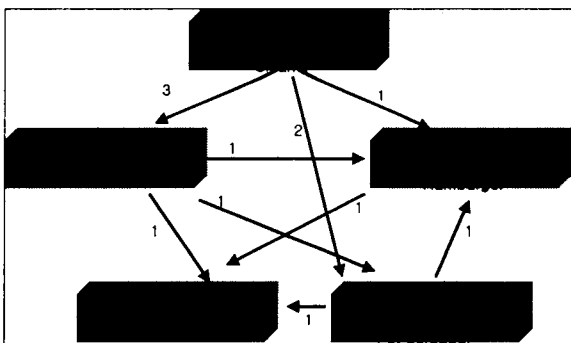
[그림6]과 [그림7]은 1997년과 1998년 지지도3의 결과를 분석한 결과로 1997년은 식품류가 67건중 48건으로 72%를 차지하고 있으며, 음료의 경우 1건으로 1%를 차지하고 있고, 1998년은 식품류 611건중 396건으로 65%를 차지하고 있으며, 음료의 경우 33건으로 5%를 차지하고 있다. 이를 보았을 때 1997의 소비패턴이 음료의 경우 4%가 늘었고, 식품류는 7%가 준 것으로 나타났다. 3항목 모두 한 군으로 모여 있으며, 다른 군과의 관련은 없는 것으로 나타났다.



[그림 6] 1997의 제품군에 의한 비율(지지도3)



[그림 7] 1998의 제품군에 의한 비율(지지도3)



[그림 8] 상품코드 454의 분기현황(6건)

[표 4] 연도별 지지도 3인경우의 항목리스트(일부)

1997년			1998년		
1항목	2항목	3항목	1항목	2항목	3항목
7	190	556	742	925	1291
104	337	1546	468	940	1084
718	1294	1544	116	942	1441
51	727	883	480	946	1096
360	830	1036	115	947	1147
1324	1374	1474	115	947	1363
:	:	:	:	:	:

마지막으로 1997년, 1998년 결과에서 상품별 빈도수가 높은 항목을 예를들어 살펴보면 [그림8]과 같이 나타난다. 즉 레드윙사의 클리너를 구입하면, 모나크사의 쌀을 사는 경우와 하이 퀄리티사의 수세미를 사는 사람과, 젠틸사의 햄버거를 사는 경우로 나타나며, 각 상품마다 분기를 숫자로 표시 해 두었다.

6. 결론 및 향후과제

본 연구에서 취급한 FoodMart 데이터를 분석 해 본 결과, 우선 판매량에 있어서 자료와 상품종류가 많을수록, 지지도가 낮을수록 알고리즘을 적용하는데 많은 시간이 소요되었으며, 각 연관규칙에 의한 지지도와 신뢰도를 바탕으로 판매전략에 대한 자료를 제시할 수 있는데 결론을 요약하면 다음과 같다.

연관규칙 적용에 있어서 낮은 지지도(0.03% - 0.005%)를 나타냈으며, 전체 지지도가 낮은 이유는 상품판매 건수 즉 트랜잭션에 비하여 빈도수가 낮기 때문이며, 최소지지도는 1997년의 경우 5(0.025%)이며, 1998년의 경우 10(0.03%)으로 나타났으며, 항목집합 리스트는 2항목과 3항목에만 나타나고 있으며, 상품분류 구분 내에 포함된 상품끼리 관련성이 있는 것으로 나타났으며 세부적인 항목리스트와 구체적인 예는 결과의 일부인 [표4]와 [그림8]을 참조할 수 있다.

데이터를 분석한 결과, 통계적 기본자료도 중요하지만, 대규모 데이터베이스에서 데이터간의 연관성에 대해 조사해 보는 것에도 큰 의의가 있었으며, 데이터 마이닝의 연관규칙도 확률의 개념에서 나온 것으로 어디까지나 의사결정의 보조수단으로 사용되어야 하며, 결과를 여러 각도에서 이해, 해석하는 것이 필요하다고 생각된다.

본 연구에 미비한 점도 많았지만, 트랜잭션 데이터를 구성하는데 있어서 효율적인 방법을 적용하는데 어려운 점이 많았고, 데이터를 분석해 가는 과정에 여러 가지 기법을 적용할 수도 있음을 깨닫게 되었다. 데이터 마케팅은 어디까지나 의사결정 보조 수단으로 사용되어야 하며, 의사 결정자는 분석된 결과를 잘 이해하고 분석할 수 있어야만 효과의 극대화를 얻을 수 있다. 또한 효율적 데이터베이스 마케팅[9]을 위하여 여러 가지 알고리즘의 혼합적 분석기법[3]과, 선행 연구된 알고리즘에 대한 정확도 및 수행시간에 대한 비교분석, 최적 알고리즘의 선정, 연관규칙을 위한 웹 마이닝에 관한 연구[7,8] 등이 필요하다고 보여진다.

7. 감사의 글

이 논문은 2000년도 두뇌한국 21사업에 의하여 지원되었음.

8. 참고문헌

- [1] Jong Soo Park, Philip S. Yu, Ming-Syan Chen, "Mining Association Rules with Adjustable Accuracy," CIKM 97 Las Vegas Nevada USA, pp151-160, 1997.
- [2] Microsoft SQL Server OLAP Services 7.0 - OLAP Manager Tutorial.
- [3] Mohammed Javeed Zaki, Srinivasan Parthasarthy, Mitsunori Ogihara, and Wei Li,

"New Algorithms for Fast Discovery of Association Rules", The University of Rochester Computer Science Department Rochester, New York 14627, Technical Report 651 July 1997.

- [4] R. Agrawal, and R.Srikant, "Fast algorithms for mining association rules", In Proceedings of the 20th VLDB Congerence, Santiago, Chille, Sept., 1994.
- [5] R. Agrawal, T. Imielinski, and A. Ssami, "Mining association rules in large databases", In Proceeding of ACM SIGMOD Conference on Management of Data, Washington D.C., pp 207-216, May, 1993.
- [6] R. Srikant and R. Agrawal, "Mining Generalized Association Rules", In Proceedings of the 21st VLDB Congerence, Zurich, Swizerland, 1995.
- [7] 김남호, 이동하, 이제현, 이진영, " 연관규칙 탐사를 이용한 웹 사용자 패턴분석 기법, " 정보처리학회 97추계 학술발표논문집, 제4권 2호, pp 425-430, 1997년 10월.
- [8] 남도원, 이동하, 서동렬 이진영, " 웹 로그에서의 사용자 접근 패턴 분석(Analysis of User Access Patterns from Web Log File) , " 정보과학회 인간과 컴퓨터 상호작용 연구회 회보 7권 1호, HCI' 98 학술대회 발표논문집, 1998년 2월. pp. 160-165.
- [9] 남도원, 이동하, 이진영, "관계형 데이터베이스에서의 연관 규칙탐사,", 정보과학회 학술발표대회 1997, pages 217-222, April 26, 1997.