

시차 연관성 규칙과 RFM 기법에 기반한 고객관계관리

김은나* (고려대 정보통계학과 석사과정 kenanana@korea.ac.kr)

하덕주 (한국신용정보 CB기획마케팅실)

최종후 (고려대 정보통계학과 교수)



차례

- I. 연구배경 및 목적
 - II. 분석자료
 - III. 시차 연관성 규칙
 - IV. RFM 기법
 - V. RFM 기법에 기반한 시차 연관성 규칙
 - VI. 토의 및 향후 연구과제
- 참고문헌

I . 연구배경 및 목적

최근 기업의 **마케팅 전략수립**은 구매이력 데이터에 기반한 과학적 접근법 요구



고객확보와 유지를 위한 마케팅 활동 및 고객가치 분석



고객관계관리(CRM) 구축



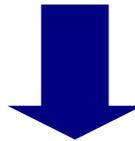
고객구매이력데이터



시차 연관성 규칙

(Sequential Association Rule)

- 시차 연관성 규칙은 시차 패턴 마이닝(Sequential Pattern Mining)
- 상품 간 교차판매(Cross-Sell)관계 탐색
- RFM(Recency-Frequency-Monetary) 기법을 이용하여 고객세분화 시도
- 세분화된 고객군 별로 특징적으로 발견되는 시차 연관성 규칙 탐색



합리적 고객관계관리 전략수립

II. 분석자료

■ 각 기업의 제품 구매 내역(일부)

거래코드 (Transaction)	품목 (Item)
0001	AIM
0001	ITS
0001	Informix
0001	Rational
0001	Storage
0002	AIM
0002	DB2
0002	ITS
0002	Lotus
0002	iSeries
0002	Rational
0003	ITS
0004	pSeries
0004	xSeries
0005	AIM
0005	ITS
0005	Storage
:	:

■ 자료설명

2003년에서 2005년까지 I社 제품에 대한 기업들의 구매이력데이터

■ 구매된 기업정보에 관한 데이터

기업의 고유번호, 종업원 수, 지역 등...

■ 구매된 제품에 관한 데이터

구매가 이루어진 최근 분기, 매출액, 제품의 구매 빈도 등...

■ 이용된 자료는 3,886건이며, 총 12개 제품 중 회사당 평균 제품 구매 빈도는 3번



■ 분석에 사용된 12개 제품 설명

S/W	Information Management : 데이터관리 및 통합
	Lotus : 포털 / 협업 / 기업메시징
	Rational : 소프트웨어 개발
	Tivoli : 보안 / 시스템 / 스토리지 관리
	WebSphere : 애플리케이션 서버 및 비즈니스 통합
	DB2 : 데이터베이스 관리
	ITS : 시스템 환경 최적화 서비스
	AIM : e-비즈니스를 위한 인프라 소프트웨어
H/W	pSeries : 유닉스 서버
	iSeries : 통합 서버
	xSeries : 인텔기반 서버
	Storage : 저장매체

Ⅲ. 시차 연관성 규칙

시차 연관성 규칙

- 순서에 따라 발생하는 행위 및 연속적인 사건으로부터 나타나는 일련의 반복적인 사례를 관찰하기 위한 데이터마이닝 기법 (Jiawei Han and Micheline Kamber, 2006)



- 지지도(Support):

두개의 항목이 동시에 일어날 확률 $\Pr(A \cap B)$

$$'A \rightarrow B' \text{의 지지도} = \frac{\text{(항목 A와 B를 동시에 포함하는 거래의 수)}}{\text{전체 거래의 수}}$$

- 신뢰도(Confidence):

조건부 확률 $\Pr(B|A)$

$$'A \rightarrow B' \text{의 신뢰도} = \frac{\text{(항목 A와 B를 동시에 포함하는 거래의 수)}}{\text{항목 A를 포함하는 거래의 수}}$$

(Berry, M.J.A & Linoff G, 1997; 최종후 외, 2006)

지지도(%)	신뢰도(%)	거래빈도	시차 연관성 규칙 ¹⁾
23.91	71.89	931	ITS → ITS → ITS
15.79	75.37	615	AIM → ITS → ITS
13.59	76.56	529	Storage → ITS → ITS
12.33	76.92	480	ITS & AIM → ITS → ITS

제품 ITS 구매에 관련된 시차 연관성 규칙 분석 결과



1) 지지도(Support)를 기준으로 선별한 우수한 4개의 시차 연관성 규칙 결과

IV. RFM 기법

■ 고객세분화 방법

<p>단일 기준에 의한 고객세분화</p>	<ul style="list-style-type: none"> - 상품에 의한 세분화 - 구매액에 의한 세분화
<p>多 기준 스코어에 의한 고객세분화</p>	<ul style="list-style-type: none"> - RFM에 의한 세분화 - LTV에 의한 세분화 - 통계모형에 기반한 세분화



(Don Peppers and Martha Rogers, 1999)

RFM에 의한 세분화

RFM 모형

고객세분화를 위한 점수화는 최근 구매시기(Recency), 구매빈도(Frequency), 구매금액(Monetary)에 대한 고객의 수익기여도를 나타내는 세 가지 지표들의 결합으로 수행

$$\text{RFM} = A \times \text{Recency} + B \times \text{Frequency} + C \times \text{Monetary}$$

A, B, C : 각각의 R값, F값, M값에 대한 가중치

가중치



(하덕주, 2006)

가중치 선정

- 파레토 법칙(Pareto's rule)의 상위 20% 고객의 평균값을 기준으로 실제 점수화를 하였을 때의 평균값과 비교하여 가장 유사한 수준에서 결정하는 방법(이강태, 2002)
- 다중회귀분석을 통해 추정하는 방법(김승아, 1998)
- RFM의 가중치를 각각 50%, 35%, 15%로 산정하는 방법 (R.S. Hodson, 1988)
- RFM의 세 가지 요인에 고객이 구입한 상품의 종류, 고객의 지불방법, 주거지 등을 고려하여 평가하는 방법 (Bob Stone and Ron Jacobs, 2001)

가중치는 상품이나 시장환경에 따라 많은 차이를 보임



가장 예측력 있는 가중치를 알아내는 것이 핵심 문제임

(류귀열 외, 2006)

파레토 법칙

- 기업의 수익창출에 가장 영향력을 끼친다고 생각되는 총 구매금액을 가지고 파레토²⁾ 분석 실시

구매액	빈도	%	누적%
26,516,190	1	0.0	0.3
26,516,190	1	0.0	0.3
:	:	:	:
422,000	1	0.0	20.0
420,000	3	0.1	20.0
418,881	1	0.0	20.1
:	:	:	:
합계	3886		100.0

상위 20% 절사값



**상위 20% 고객의 수는 3886개의 기업 중 779개 기업
총 구매금액은 전체의 90.2% 차지**

2) 파레토 법칙(Pareto rule) : 상위 20%에 속하는 고객들이 전체 매출의 80%를 차지한다는 이론

RFM값 설정

■ R값 설정

최근 구매시기 R값		빈도	퍼센트	누적퍼센트
1, 2, 3분기	1	203	5.2	5.2
4, 5, 6분기	2	349	9.0	14.2
7, 8분기	3	408	10.5	24.7
9, 10분기	4	894	23.0	47.7
11분기	5	551	14.2	61.9
12분기	6	1,481	38.1	100.0
합 계		3,886	100.0	

■ F값 설정

총 구매빈도 F값		빈도	퍼센트	누적퍼센트
1회	1	1,344	34.6	34.6
2회	2	821	21.1	55.7
3회	3	502	12.9	68.6
4회	4	347	8.9	77.6
5~7회	5	642	16.5	94.1
8회 이상	6	230	5.9	100.0
합 계		3,886	100.0	

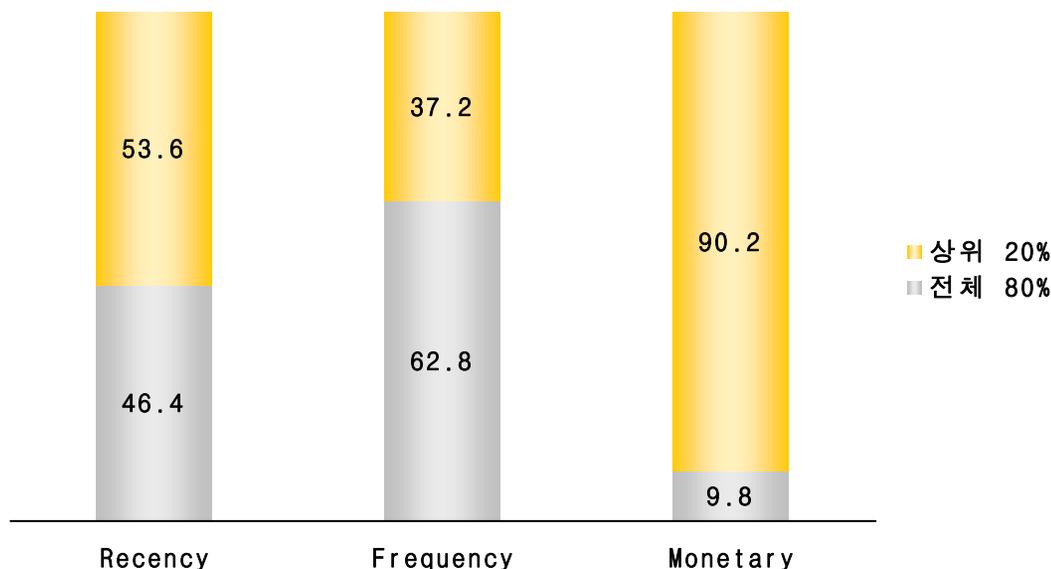
■ M값 설정

총 구매금액 M값		빈도	퍼센트	누적퍼센트
~ \$10,000	1	824	21.2	21.2
\$10,000 ~ \$30,000	2	583	15.0	36.2
\$30,000 ~ \$70,000	3	554	14.3	50.5
\$70,000 ~ \$150,000	4	536	13.8	64.3
\$150,000 ~ \$500,000	5	706	18.2	82.4
\$500,000 ~	6	683	17.6	100.0
합 계		3,886	100.0	

RFM값 설정

- 고객의 평균 RFM 점수 = $0.3 \times \text{Recency} + 0.2 \times \text{Frequency} + 0.5 \times \text{Monetary}$
- RFM 점수 = (고객의 평균 RFM 점수 $\times 100$)/6

I社 제품의 총 구매금액에 의한 상위 20% 고객에 대해 R값, F값, M값의 비율에 따라 가중치를 구함



기업의 특성을 고려한 적절한 가중치 부여가 핵심 과제

(강창완 외, 2004)

RFM 점수를 통한 세분화 그룹의 절사점

RFM 점수	빈도	퍼센트	누적퍼센트
42	77	2.0	17.7
:	:	:	:
50	120	3.1	23.7
52	21	0.5	23.8
:	:	:	:
60	90	2.3	49.2
62	85	2.2	49.4
:	:	:	:
70	92	2.4	64.3
72	86	2.2	64.5
:	:	:	:
80	84	2.2	79.5
82	13	0.3	79.6
:	:	:	:
90	28	0.7	84.9
92	42	1.1	86.0
:	:	:	:
98	14	0.4	97.1
100	112	2.9	100

우수그룹 절사점
상위 20% 고객

최우수그룹 절사점

V. RFM 기법에 기반한 시차 연관성 규칙

시차 연관성 규칙

RFM 점수	빈도	퍼센트	누적퍼센트
42	77	2.0	17.7
:	:	:	:
80	84	2.2	79.5
82	13	0.3	79.6
:	:	:	:
90	28	0.7	84.9
92	42	1.1	86.0
:	:	:	:
98	14	0.4	97.1
100	112	2.9	100

우수그룹 절사점
상위 20% 고객

최우수그룹 절사점

RFM 점수 상위 20%에 속하는 최우수그룹과 우수그룹에 대해 시차 연관성 분석 시도



구매 가능성이 높은 최우수그룹과 우수그룹에 집중적 마케팅 실시

시차 연관성 규칙

최우수그룹의 제품간의 시차 연관성 규칙

지지도(%)	신뢰도(%)	거래빈도	시차 연관성 규칙 ³⁾
64.72	81.89	321	ITS → ITS → ITS
59.48	75.26	295	ITS → ITS → pSeries
58.87	85.88	292	Storage → ITS → ITS
56.05	82.74	278	AIM → ITS → ITS
54.84	80.00	272	Storage → ITS → pSeries
53.43	67.60	265	ITS → ITS → xSeries
50.20	74.11	249	AIM → ITS → pSeries
48.59	80.33	241	AIM → AIM → ITS

3) 시차 연관성 규칙 발견에 포함될 최대 항목 수는 3이고, 시차에 대한 제품의 구매비율 기준은 5%임지지도를 기준으로 선별한 우수한 8개의 시차 연관성 규칙 결과임

- 우수그룹의 제품간의 시차 연관성 규칙

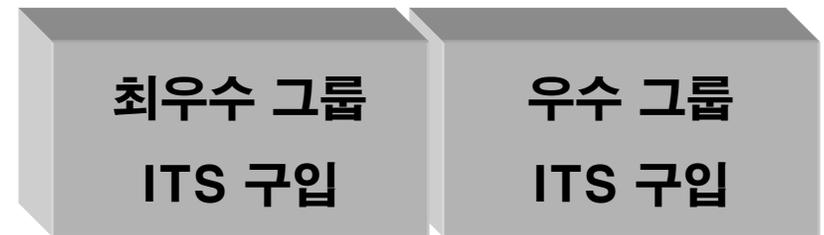
지지도(%)	신뢰도(%)	거래빈도	시차 연관성 규칙
34.08	70.98	137	ITS → ITS → ITS
29.60	73.46	119	AIM → ITS → ITS
23.13	71.54	93	Storage → ITS → ITS
22.14	74.17	89	ITS & AIM → ITS → ITS
20.90	67.74	84	AIM → AIM → ITS
20.40	70.69	82	ITS → AIM → ITS
18.41	76.29	74	AIM → ITS & AIM → ITS
16.92	78.16	68	Storage → Storage → ITS

VI. 토의 및 향후 연구과제

Cross-Selling Up-Selling



ex) 시차 연관성 규칙에 기반한
DM 전략 ...



토의 및 향후 연구과제

- 고객관계관리 전략수립을 위하여 세분화 고객 중 최우수 그룹과 우수 그룹에 대한 특성을 바탕으로 효율적인 전략을 통해 고객관계관리 전략 수행
- 세분화된 고객들에 대한 차별화 된 마케팅 전략을 수행 할 수 있음
- 향후 신규고객 확보에 있어서 유사한 특성을 지닌 고객들에 대해 선별적이고 집중적인 고객확보전략 수행
- 고객의 충성도를 높이고 개별 고객의 평생가치를 극대화하기 위해 무엇보다 정확하고 사실적인 고객정보의 획득이 중요
- 정확한 고객예측과 수익창출을 위한 고객세분화 기법들에 대한 연구가 이루어져야 할 것임
- 가장 예측력 있는 가중치 선정에 대한 연구가 이루어져야 함



- [1] 강창완, 이소영, 최승배, 김규곤 (2004). 고객 세분화를 위한 최적 RFM 모형 구축에 관한 연구, *Journal of the Korean Data Analysis Society*, Vol. 6, No. 6, pp. 1829-1840.
- [2] 강현철, 한상태, 최종후 외 (2006). 『고객관계관리(CRM)를 위한 데이터마이닝 방법론』, 서울: 자유아카데미.
- [3] 김승아 (1998). 『DB 마케팅 분석법』, 서울: 경영과 컴퓨터.
- [4] 류귀열, 문영수 (2006). 연관분석을 이용한 데이터마이닝 기법에 관한 사례연구, *Journal of the Korean Data Analysis Society*, Vol. 8, No. 3, pp. 1021-1033.
- [5] 이강태 (2002). *eCRM 환경에서 LTV 극대화를 위한 고객세분화 기업에 관한 연구*, 전주대학교 석사학위논문.
- [6] 하덕주 (2006). *연관성 규칙과 RFM 기법에 기반한 고객관계관리*, 고려대학교 석사학위논문.
- [7] Berry, M.J.A & Linoff G (1997). *Data Mining Techniques For Marketing, Sales, and Customer Support*, New York : John Wiley & Son.
- [8] Bob Stone and Ron Jacobs (2001). *Successful Direct Marketing Methods*, 7rd Edition, New York : McGraw-Hill.
- [9] Don Peppers and Martha Rogers (1999). *Enterprise One to One*, 1rd Edition, New York : Currency Doubleday.
- [10] Richard S. Hodgson (1988). *Direct Mail and Mail Order Handbook*, 3rd Edition, Probus Publishing.
- [11] Jiaewi Han and Micheline Kamber (2006). *Data Mining: Concepts and Techniques*, 2rd Edition, San Francisco : Morgan Kaufmann.