

# 대리점 이탈예측모델 개발 - 동적모델(Pattern Model)과 정적모델(Matrix Model)의 예측적중률 비교 -

안봉락\* · 이새봄\* · 노인성\*\* · 서영호\*\*†

\* 경희대학교 대학원 경영학과

\*\* 경희대학교 경영대학

## Development of Prediction Model for Churn Agents -Comparing Prediction Accuracy Between Pattern Model and Matrix Model-

Bong-Rak An\* · Sae-Bom Lee\* · In-Sung Roh\*\* · Yung-Ho Suh\*\*†

\* Department of Management, Graduate School, Kyung Hee University

\*\* School of Management, Kyung Hee University

### ABSTRACT

**Purpose:** The Purpose of this study is to develop a model for predicting agent churn group in the cosmetics industry. We develop two models, pattern model and matrix model, which are compared regarding the prediction accuracy of churn agents. Finally, we try to conclude if there is statistically significant difference between two models by empirical study.

**Methods:** We develop two models using the part of RFM(Recency, Frequency, Monetary) method which is one of customer segmentation method in traditional CRM study. In order to ensure which model can predict churn agents more precisely between two models, we used CRM data of cosmetics company A in China.

**Results:** Pattern model and matrix model have been developed. we find out that there is statistically significant differences between two models regarding the prediction accuracy.

**Conclusion:** Pattern model and matrix model predict churn agents. Although pattern model employed the trend of monetary mount for six months, matrix model that used the amount of sales per month and the duration of the employment is better than pattern model in prediction accuracy.

**Key Words:** CRM, Customer, Churn, RFM, Agent, Pattern, Matrix, Model

● Received 7 May 2014, revised 20 May 2014, accepted 21 May 2014

† Corresponding Author(suhy@khu.ac.kr)

© 2014, The Korean Society for Quality Management

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-Commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

# 1. 서론

본 연구는 한국과 중국의 합작기업인 A화장품 기업을 대상으로 대리점 이탈 현상을 연구하고자 한다. A화장품 회사는 1995년에 설립되어 올해 18년이 되는 화장품 생산업체로 중국시장의 급격한 성장과 함께 빠르게 성장해 왔다. 그러나 최근 다양한 해외 화장품 회사들 및 중국 화장품 회사들과의 경쟁으로 성장이 정체국면에 들어와 있다. 그 원인중의 하나는 경쟁 환경의 악화 이외에도 이로 인한 대리점 이탈률이 증가하고 있다는 것이다.

중국화장품 시장의 경우, 판매채널이 다양하게 존재하는데 그중에서도 외국화장품들은 주로 백화점에서 판매되고 있는 경우가 많으며, 중국기업의 화장품들은 대리점을 통한 직접 판매가 주를 이루고 있다. A기업도 마찬가지로 아래 <그림 1>처럼 두 번째 유형에 속하는 형태의 유통구조를 가지고 있다. <그림 1>의 소매점포(retailer)를 우리나라에서는 대리점이라고 부르지만 중국에서는 경소상(经销商)이라고 부르며 이들을 통해 대부분의 판매가 이루어지고 있다. 따라서 경소상 관리가 매출관리에 있어 매우 중요한 이슈이다. A기업의 경우 경소상 이탈율이 최근 증가하고 있어 경소상 관리에 많은 노력을 경주하고 있다. 경소상 수의 감소는 결국 매출에 영향을 주기 때문에 회사 입장에서 가장 우선적으로 관리해야 할 이슈이다.

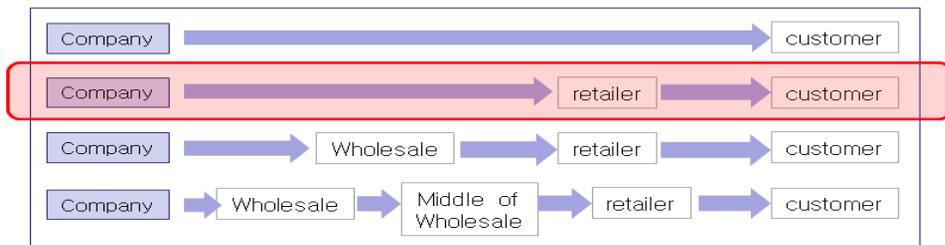


Figure 1. Distribute Courses of Cosmetics in China

이러한 문제를 해결하기 위해 CRM 자료분석을 통한 대리점 이탈예측모델의 개발이 본 연구의 목표이다. 본 연구는 최종소비자와 기업을 연결하는 대리점 중심의 CRM연구로 이탈방지 모델에 앞서 이탈방지의 대상이 되는 이탈예상 경소상을 예측할 수 있는 모델을 개발하고자 한다. 본 연구에서 경소상 이탈예측모델 개발방법으로 동적모델(Pattern Model)과 정적모델(Matrix Model)을 제시하는 것과 어느 모델이 경소상 이탈 예상 적중률이 높은 우수한 모델인지 입증하기 위하여 비율차이검정을 수행하는 것 까지가 본 연구의 내용이라고 할 수 있다. 이를 위하여 본 연구는 2장 이론적 배경 및 선행 연구, 3장 연구방법에 대한 소개, 4장 분석결과와 이탈 예측적중률 차이에 대한 동적모델과 정적모델 우수성 검증, 5장 결론 및 시사점으로 구성되어 있다.

## 2. 이론적 배경 및 선행연구

### 2.1 CRM(Customer Relationship Management)

기존 CRM 연구는 CRM에 대한 일반적인 내용의 설명 혹은 개념에 대한 연구(Abbott, 2001; Parvatiyar and Sheth, 2001), CRM 계획 및 전략에 대한 연구(Chang et al., 2002; Winer, 2001; Payne and Frow, 2005), 마케팅 활용에 대한 연구(Moe and Fader, 2001), 고객세분화 및 타겟팅에 대한 연구(Kim et al., 2006; Ha et al., 2002;

Hwang et al., 2004)등 다양한 분야에서 많이 수행되어 왔다. 그럼에도 불구하고 화장품 산업을 중심으로 한 연구는 찾아보기 힘들다. 대부분의 CRM 연구들은 통신사(Hwang et al., 2004), 전자 거래(Fader et al., 2005), 금융(Peter, 2003) 분야에서 주로 수행되어져 왔다. 지금까지 주요 CRM 연구에서 제시되어온 CRM에 대한 정의와 설명을 표로 정리하면 아래와 같다.

**Table 1.** Various Definitions of CRM

| Authors                    | Description  |
|----------------------------|--|
| Vavra(1992)                | “CRM only as seeking customer retention by using a variety of after marketing tactics that lead to customer bonding or staying in touch with the customer after a sale is made.”   |
| Shani and Chalasani (1992) | “An integrated effort to identify, maintain, and build up a network with individual consumers and to continuously strengthen the network for the mutual benefit of both sides, through interactive, individualized and value-added contacts over a long period of time”  |
| Berry(1995)                | in somewhat broader terms, also has a strategic viewpoint concerned with CRM. He has stressed that attracting new customers should be viewed only as an intermediate step in the marketing process and that developing closer relationship with these customers and turning them into loyal ones should be equally important aspects of marketing. Thus, he proposed that relationship marketing be seen as “attracting, maintaining, and – in multi-service organizations – enhancing customer relationships” |
| Kincaid(2003)              | “CRM comprises three major functional areas: Marketing, Sale, Services and Support. These three components may be seen as the life cycle of a customer relationship that moves from marketing, to sales, to service and support. Indeed, IT and IS are the other crucial components in supporting and maintaining these three functional areas as well as the whole CRM process”   |

## 2.2 고객이탈관리(Churn Management)

고객이탈을 관리하는 방법으로는 어느 고객이 이탈할 가능성이 있는지를 예측한 후 대상 고객에게 인센티브나 다양한 유인책을 통해 이탈을 감소시키는 방안을 들 수 있다. 기업은 이러한 접근방식을 통해 이탈할 가능성이 높은 고객에게 노력과 자원을 집중할 수 있고, 이탈가능성이 낮은 고객에게 인센티브를 제공하는 등의 낭비되는 비용을 절감할 수 있다(Scott et al., 2006). 이탈고객은 자발적(Voluntary) 이탈과, 비자발적 (Non-Voluntary) 이탈의 두 가지 그룹으로 나눌 수 있다. 이탈을 하는 이유에 대해서는 연구자들에 따라 의견이 다양하다. 많은 연구자들이 자발적 이탈에 관하여 인구통계학적(Demographical) 데이터를 통해 이탈을 예측하였다. Wei and Chiu (2002)는 인구통계학적 데이터를 통해 계약(Contract)과 소명 의식(calling) 이라는 두 가지 유의한 요소를 측정하였다. 그러나 인구통계학적 데이터는 다소 정보가 제한적이라서 최근에는 많은 연구들이 다양한 정보기술들을 사용하여 이탈 고객을 예측하고 있다(Hung et al., 2006; John et al., 2008). 이탈예측을 위하여 많은 연구들이 고객의 정보 데이터 혹은 거래 데이터를 이용하여 예측하고자 하였는데 데이터를 통한 CRM연구는 분석적(Analytical) CRM 프레임워크에 포함된다.

Kim et al.(2006)은 무선이동통신 회사에 대해 사례연구를 수행하였고, LTV 모델을 제시하여 고객 가치를 분석하고 고객을 세분화하기 위해 의사결정나무(Decision Tree) 모형을 이용하였다. John et al.(2008)은 선형회귀분석(Linear Regression)과 회귀나무(Regression Tree) 그리고 인공신경망(Neural Network)과 같은 기술적 접근법을 사용하여 트레이닝 데이터셋을 테스트하였다. Hung et al.(2006)은 이동통신에서의 이탈예측에 대한 연구를 제안

하였는데 데이터마이닝 기법을 통하여 이탈을 예측할 수 있는 우수한 모델을 찾고자 하였다. 그들은 의사결정나무(Decision Tree)와 인공신경망기법이 가장 이탈을 예측하기에 좋은 기법이라고 주장하였다.

### 2.3 RFM 방법론

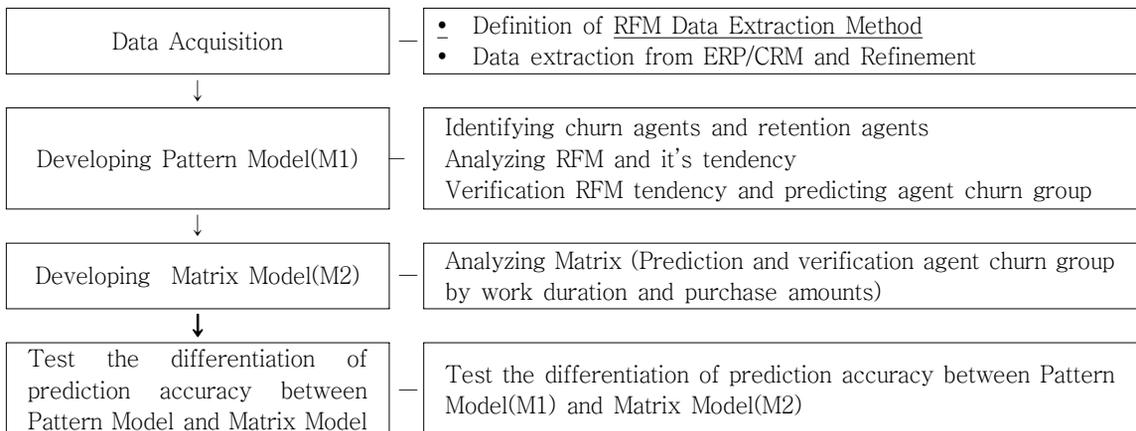
고객세분화 기법 중 데이터마이닝 외에 기본적으로 많이 사용되는 것이 RFM 방법론이다. RFM의 기본 개념은 Bult and Wansbeek(1995)에서 처음 소개되었으며, RFM 방법은 고객의 행위를 분석하기 위한 마케팅 기법으로 얼마나 최근에 구매했는가(Recency), 얼마나 자주 구매했는가(Frequency), 얼마나 많이 구매했는가(Monetary)라는 세 가지 지표들을 통해 고객을 구분한다.

최근 연구되어지는 RFM에 대한 연구들은 기본 데이터를 RFM 변수로 변환한 후 데이터마이닝 기법 혹은 알고리즘을 통하여 고객세분화 연구를 진행하고 있다(Liu and Shih, 2005; Kaymak, 2001; Hsieh, 2004). Liu and Shih(2005)는 RFM 변수들의 상대적인 가중치를 결정하기 위하여 분석계층 프로세스(AHP)를 적용하였다. Kaymak(2001)은 RFM 변수들을 사용하여 Fuzzy 클러스터링 기법을 통해 특성화된 고객을 세분화하고자 하였다. Hsieh(2004)는 통합된 데이터마이닝 기법과 RFM기반의 행동 점수화 모델을 제시하였는데 신용카드 이용 고객을 대상으로 인공신경망을 이용한 자기조직지도(self-organizing map)를 통하여 수익성이 높은 고객 그룹을 예측하고자 하였다. 또한, 많은 기존 연구들이 RFM의 기본 세 가지 변수 외에도 다른 변수를 추가하여 연구하고 있다. Hosseini et al.(2010)은 B2B를 대상으로 RFM 변수 외에 제품활동기간(Period of product activity)을 추가하여 고객의 제품 충성도를 구별하고자 하였다. Yeh et al.(2009)은 다이렉트 마케팅을 위하여 고객군을 세분화하고자 하였는데 데이터베이스로 추출된 RFM 변수에서 첫 구매이후 시간(Time since first purchase)과 이탈 가능성(churn probability)을 변수로 추가하여 연구하였다. 주로 RFM 방법론은 고객가치측정도구나 고객세분화 도구로 사용되고 있으며 본 연구도 RFM을 고객세분화를 위한 데이터 분석 도구로 사용하고자 한다.

## 3. 연구방법

### 3.1 연구절차

Table 2. Research Process



연구절차는 먼저 A회사의 DB에서 데이터를 추출하고 이를 통하여 양 모델을 개발한 후 그들의 성과를 비교하는 과정을 거친다. 본 연구는 이탈예상대리점을 예측하는데 있어 예측적중률이 높은 모델 개발을 위하여 시간의 흐름에 따른 고객군의 변화패턴을 예측할 수 있는 동적인 모델(M1)과 특정 시점에서의 대리점 상태를 나타내는 정적인 모델(M2)을 동시에 개발하였다. 또한 두 모델간의 비교를 통하여 어떤 모델이 더 높은 예측적중률을 보이는가를 확인하고 모델 간 이탈예측률 차이가 통계적으로 유의한지를 분석하였다.

### 3.2 연구가설 및 검증방법

귀무가설(H0): 동적모델(M1)과 정적모델(M2)의 이탈예측률은 차이가 없을 것이다(H0: M1=M2).

연구가설(H1): 동적모델(M1)과 정적모델(M2)의 이탈예측률은 차이가 있을 것이다(H0: M1≠M2).

비율차이 검증(Z)은 아래의 집단 간 비율차이 검증수식을 이용하여 검증한다.

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (\hat{p}_1 - \hat{p}_2)}{\sqrt{\frac{\hat{p}\hat{q}}{n_1} + \frac{\hat{p}\hat{q}}{n_2}}}$$

$\hat{p}_1$  = 비율추정치로서 표본1의 비율값

$\hat{p}_2$  = 비율추정치로서 표본2의 비율값

$p_1$  = 모집단1의 비율값

$p_2$  = 모집단2의 비율값

$\hat{p}$  =  $\frac{(x_1 + x_2)}{(n_1 + n_2)}$  ( $x_1$ 과  $x_2$ 는 각 표본에서 특정 속성을 갖는 구성원의 수)

$\hat{q}$  =  $1 - p$

$\sqrt{\frac{\hat{p}\hat{q}}{n_1} + \frac{\hat{p}\hat{q}}{n_2}}$  =  $(\hat{p}_1 - \hat{p}_2)$ 의 표준오차

### 3.3 연구방법

#### 3.3.1. 데이터 추출

앞서 이론적 배경에서 설명한 것처럼 RFM 방법은 고객의 행위를 분석하기 위한 마케팅 기법으로 얼마나 최근에 구매했는가(Recency), 얼마나 자주 구매했는가(Frequency), 얼마나 많이 구매했는가(Monetary)라는 세 가지 변수들을 통해 고객을 선별한다. 그러나 대리점을 기반으로 사업을 영위하고 있는 A기업으로부터 추출할 수 있는 데이터는 최종소비자가 아닌 대리점 구매 데이터이고, 본사의 판촉활동(프로모션) 혹은 대리점이 확보하고 있는 재고량에 의해 대리점의 구매 최신성(Recency)과 빈도(Frequency)가 최종소비자의 구매 의도와는 다르게 나타날 수 있다. 본 연구에서는 이러한 대리점 구매의 특수성을 고려하여 RFM변수 중 구매금액(Monetary)변수만을 사용하며 최신성과 빈도는 고려하지 않기로 하였다.

A기업의 ERP/CRM 시스템에서 2010년 1월~2013년 9월(약45개월) 사이의 A기업 대리점들에 대한 월평균 구매

금액 데이터를 추출하였다.

Monetary(M): 2010년 1월~2013년 9월(약4년) 사이의 A기업 대리점들에 대한 월평균 구매금액

### 3.3.2. 동적(Pattern) 모델 개발

동적모델은 다음과 같은 단계별 접근방법으로 개발하였다.

- Step 1 : 2013년 3월을 기준으로 이탈 및 유지 그룹 구분
- Step 2 : 6개월간의 월별 매출액 데이터(Monetary Data) 추출
- Step 3 : 6개월간의 대리점별 매출 위상차이 분석
- Step 4 : 이탈대리점들의 패턴 파악 및 이탈예측률 검증

동적(pattern)모델은 대리점별 2010년 1월~ 2013년 9월 사이의 월평균 금액을 추출하여 데이터 세트(Data set)를 만들고 기준월인 2013년 3월을 기준으로 Training set(TRS)과 Test set(TS1, TS2)을 구분하였다.

- 1st set(Training set): 2013년 3월 이전 이탈대리점
- 2nd set(Test set 1): 2013년 4~9월까지 6개월 이내 이탈대리점
- 3rd set(Test set 2): 2013년 4월 이후 6개월 이상 유지대리점

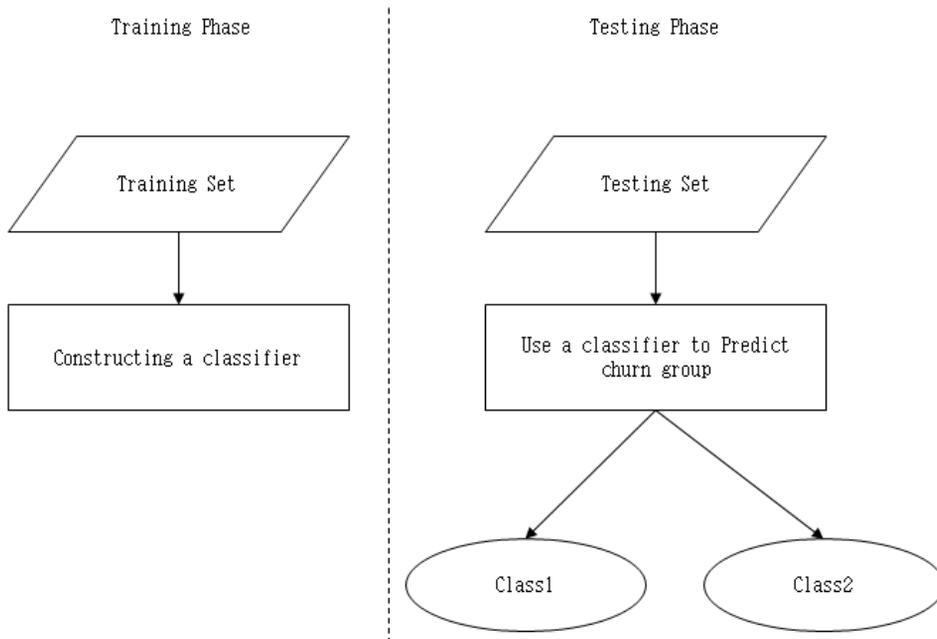


Figure 2. Hold-out Validation Framework

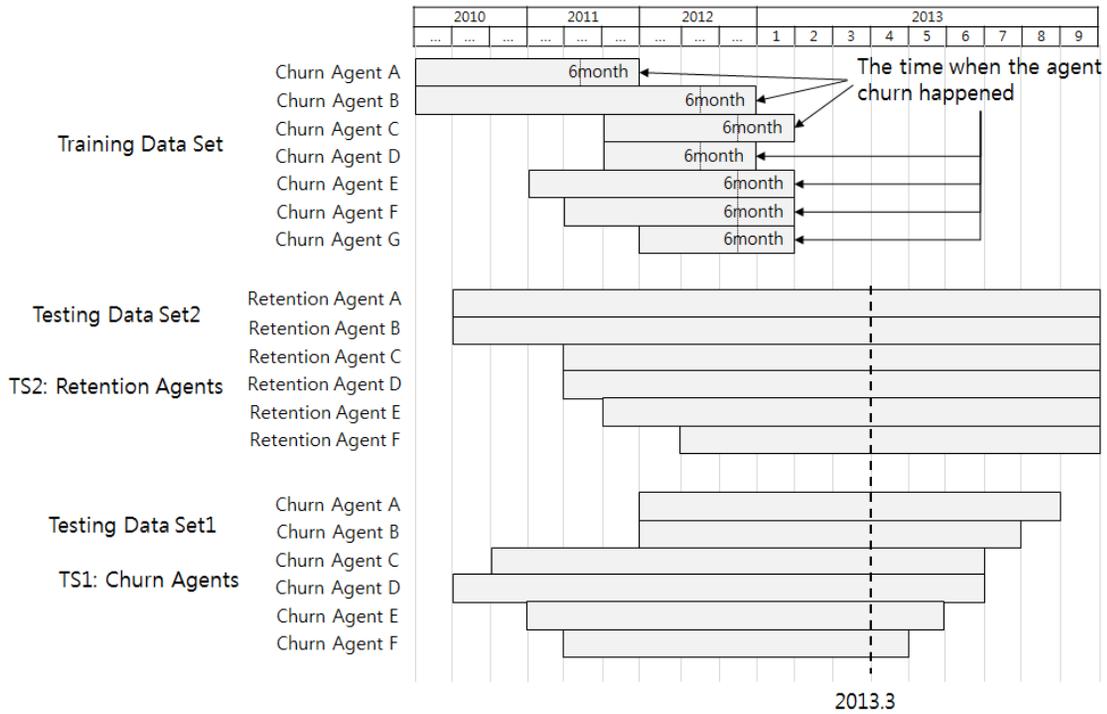


Figure 3. Training Data Set(before April 2013), Test Data Set1 and 2(Since April 2013)

두 번째 단계로 6개월간의 구매금액(Monetary) 데이터를 추출하여 트레이닝 데이터 세트(Training Data Set)와 테스트 데이터 세트1, 테스트 세트2(Test set 1, Test set 2)를 만든다.

- 트레이닝 데이터 세트: 이탈일자를 기준으로 각 대리점의 이탈 전 6개월간 월평균 구매금액 추출
- 테스트 데이터 세트1: 2013년 3월을 기준으로 지난 6개월(2012년 10월~2013년 3월) 월평균 구매금액 추출
- 테스트 데이터 세트2: 2013년 3월을 기준으로 지난 6개월(2012년 10월~2013년 3월) 월평균 구매금액 추출

Table 3. Examples of Extracting Monetary Data for 6 months

| <Training set> |            |          |             | <Test set> |            |         |             |
|----------------|------------|----------|-------------|------------|------------|---------|-------------|
| ID             | Start_DATE | End_DATE | Test_Period | ID         | Start_DATE | M_DATE  | Test_Period |
| 2E0307         | 2010-10    | 2011-03  | 6           | 0H2619     | 2012-10    | 2013-03 | 6           |
| 0b8289         | 2010-11    | 2011-04  | 6           | 0H2909     | 2012-10    | 2013-03 | 6           |
| 0B9095         | 2011-04    | 2011-09  | 6           | 0H3343     | 2012-10    | 2013-03 | 6           |
| 0b9307         | 2012-02    | 2012-07  | 6           | 0H3926     | 2012-10    | 2013-03 | 6           |
| 0B9406         | 2011-04    | 2011-09  | 6           | 0h4684     | 2012-10    | 2013-03 | 6           |
| 0B9447         | 2012-04    | 2012-09  | 6           | 0I0448     | 2012-10    | 2013-03 | 6           |
| 0B9524         | 2011-08    | 2012-01  | 6           | 0I8881     | 2012-10    | 2013-03 | 6           |
| 0B9591         | 2010-12    | 2011-05  | 6           | 0J2802     | 2012-10    | 2013-03 | 6           |

세 번째 단계로 트레이닝 세트에 대하여 6개월간의 대리점별 매출 위상차이를 분석한다. 금월에 전월보다 구매금액(Monetary)이 증가하면 +(양)로, 감소하거나 매출이 0이면 -(음)로 표시하여 6개월 간 총 5개 위상차의 패턴을 분석한다.

| ID     | Start_Date | End_Date | W_Period | 2012.09 | 2012.10 | 2012.11 | 2012.12 | 2013.1 | 2013.2 | Total  |
|--------|------------|----------|----------|---------|---------|---------|---------|--------|--------|--------|
| OK3182 | 2010-01    | 2013-02  | 38       | 12,690  | 15,011  | 20,000  | 15,000  | 18,000 | 0      | 80,701 |

+   +   -   +   -

Figure 4. Converting Monetary data to negative or positive signal to Develop Dynamic Prediction Model

마지막 단계로 이탈대리점 패턴을 기준으로 예측가능대리점들을 파악한다.

Table 4. Analyzing Pattern of Churn Agents to Develop Dynamic Prediction Model

| Rull Name    | Pattern | TR6              |            |                        | TS1_6            |            |                        | TS2_6            |            |                        |
|--------------|---------|------------------|------------|------------------------|------------------|------------|------------------------|------------------|------------|------------------------|
|              |         | Number of Agents | Percentage | Accumulated Percentage | Number of Agents | Percentage | Accumulated Percentage | Number of Agents | Percentage | Accumulated Percentage |
| LP1VOT06_R1  | -----   | 643              | 14.4%      | 14.4%                  | 120              | 14.17%     | 14.2%                  | 84               | 2.64%      | 2.64%                  |
| LP1VOT06_R2  | ----+   | 696              | 15.6%      | 30.1%                  | 92               | 10.86%     | 25.0%                  | 85               | 2.67%      | 5.3%                   |
| LP1VOT06_R3  | ---+-   | 591              | 13.3%      | 43.3%                  | 49               | 5.79%      | 30.8%                  | 112              | 3.52%      | 8.8%                   |
| LP1VOT06_R4  | --+-    | 398              | 8.9%       | 52.3%                  | 27               | 3.19%      | 34.0%                  | 56               | 1.76%      | 10.6%                  |
| LP1VOT06_R5  | -+-+    | 357              | 8.0%       | 60.3%                  | 32               | 3.78%      | 37.8%                  | 112              | 3.52%      | 14.1%                  |
| LP1VOT06_R6  | +++     | 466              | 10.5%      | 70.8%                  | 55               | 6.49%      | 44.3%                  | 127              | 3.99%      | 18.1%                  |
| LP1VOT06_R7  | ---+    | 333              | 7.5%       | 78.2%                  | 21               | 2.48%      | 46.8%                  | 68               | 2.14%      | 20.2%                  |
| LP1VOT06_R8  | -+---   | 184              | 4.1%       | 82.4%                  | 9                | 1.06%      | 47.8%                  | 35               | 1.10%      | 21.3%                  |
| LP1VOT06_R9  | -++-    | 67               | 1.5%       | 83.9%                  | 24               | 2.83%      | 50.6%                  | 101              | 3.17%      | 24.5%                  |
| LP1VOT06_R10 | -+++    | 110              | 2.5%       | 86.3%                  | 46               | 5.43%      | 56.1%                  | 158              | 4.97%      | 29.5%                  |
|              |         |                  |            |                        | ⋮                |            |                        |                  |            |                        |
| LP1VOT06_R23 | +----   | 18               | 0.4%       | 98.2%                  | 19               | 2.24%      | 90.1%                  | 141              | 4.43%      | 79.5%                  |
| LP1VOT06_R24 | ++++    | 9                | 0.2%       | 98.4%                  | 5                | 0.59%      | 90.7%                  | 80               | 2.51%      | 82.0%                  |
| LP1VOT06_R25 | +---    | 11               | 0.2%       | 98.6%                  | 7                | 0.83%      | 91.5%                  | 52               | 1.63%      | 83.6%                  |
| LP1VOT06_R26 | +++-    | 10               | 0.2%       | 98.9%                  | 12               | 1.42%      | 92.9%                  | 92               | 2.89%      | 86.5%                  |
| LP1VOT06_R27 | +--+    | 22               | 0.5%       | 99.4%                  | 17               | 2.01%      | 94.9%                  | 131              | 4.12%      | 90.7%                  |
| LP1VOT06_R28 | ++++    | 11               | 0.2%       | 99.6%                  | 14               | 1.65%      | 96.6%                  | 89               | 2.80%      | 93.5%                  |
| LP1VOT06_R29 | ++++    | 1                | 0.0%       | 99.6%                  | 5                | 0.59%      | 97.2%                  | 53               | 1.67%      | 95.1%                  |
| LP1VOT06_R30 | +---    | 8                | 0.2%       | 99.8%                  | 17               | 2.01%      | 99.2%                  | 92               | 2.89%      | 98.0%                  |
| LP1VOT06_R31 | ++++    | 7                | 0.2%       | 100.0%                 | 6                | 0.71%      | 99.9%                  | 43               | 1.35%      | 99.4%                  |
| LP1VOT06_R32 | ++++    | 2                | 0.04%      | 100.0%                 | 1                | 0.12%      | 100.0%                 | 21               | 0.66%      | 100.0%                 |
| Total        |         | 4453             | 100.0%     |                        | 847              | 100.00%    |                        | 3182             | 100.00%    |                        |

### 3.3.3. 정적(Matrix)모델 개발

정적모델은 경소상 이탈을 예측하기 위하여 BCG Matrix의 포지셔닝 의미를 활용하여 근무기간과 월평균 매출액에 따른 대리점 분석으로 이탈예상대리점을 예측한다. 예를 들어 근무기간이 길고 월평균매출액이 크다면 <그림 7>에서와 같이 다른 영역에 위치한 경소상들보다 상대적으로 높은 가치가 있는 경소상이라고 할 수 있다(Cash Cow). 근무기간이 짧지만 월평균매출액이 크다면 다른 영역에 위치한 경소상들보다 상대적으로 성장률이 높은 경소상이라고 할 수 있다(Star). 근무기간이 길지만 월평균매출액이 작다면 그러한 이유가 무엇인지 해당 경소상의 여러 가지 상황을 면밀히 살펴볼 필요가 있는 경소상이라고 할 수 있다(Question Mark). 마지막으로 근무기간도 짧고 월평균매출액도 작다면 다른 영역에 위치한 경소상들보다 이탈가능성이 높은 경소상이라고 볼 수 있다. 본 연구에서는 경소상 이탈 예측 정적(Matrix)모델을 개발하고 실제 이탈한 경소상 데이터를 활용하여 위와 같은 매트릭스 관계가 성립하는지를 실증분석한다.

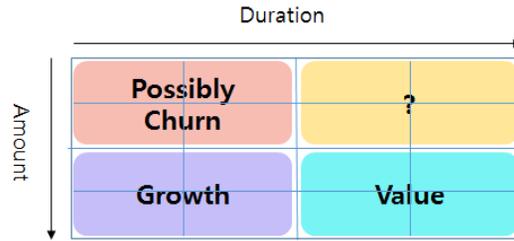


Figure 5. A Matrix for Static Prediction Model

## 4. 분석결과

### 4.1 동적모델 이탈예측률 분석결과

패턴(Pattern)을 활용하여 2013년 3월말 기준, 향후 6개월의 예측률을 분석하는 동적모델(Pattern-6개월)에서 트레이닝 데이터 세트에는 4,453개의 데이터가 사용되었으며, 테스트 데이터 세트1은 847개, 테스트 데이터 세트2는 3,182개의 데이터가 사용되었다. 이를 표로 나타내면 아래 <표3>과 같다.

Table 5. Number of Agents for Pattern Model separated by Data set Type

| Data set              | Description                   | Number of Agents |      |
|-----------------------|-------------------------------|------------------|------|
| Training Set_6 months | agent churn group in the past | 4453             |      |
| Test Set1_6 months    | Test1_agent churn group       | 847              | 4029 |
| Test Set 2_6 months   | Test 2_agent retention group  | 3182             |      |

각각의 데이터 세트별 패턴에 대한 결과는 아래 <표4>와 같다. 패턴은 이탈예상이 높은 순서대로 정렬하여 각각에 룰(Rule) 명칭을 부여하였다. 룰 명칭의 M은 Monetary라는 의미이며 P06은 6개월을 의미한다.

Table 6. Churn Agents' Patterns for Dynamic Model

| rule name | pattern | TR_6             |       |                    | TS1_6            |       |                    | TS2_6            |       |                    |
|-----------|---------|------------------|-------|--------------------|------------------|-------|--------------------|------------------|-------|--------------------|
|           |         | number of agents | ratio | accu-mulated ratio | number of agents | ratio | accu-mulated ratio | number of agents | ratio | accu-mulated ratio |
| MP06_R1   | -----   | 643              | 14.4% | 14.4%              | 120              | 14.1% | 14.2%              | 84               | 2.6%  | 2.6%               |
| MP06_R2   | -----+  | 696              | 15.6% | 30.1%              | 92               | 10.2% | 25.0%              | 85               | 2.6%  | 5.3%               |
| MP06_R3   | ----+-  | 591              | 13.3% | 43.3%              | 49               | 5.8%  | 30.8%              | 112              | 3.5%  | 8.8%               |
| MP06_R4   | ----++  | 398              | 8.9%  | 52.3%              | 27               | 3.2%  | 34.0%              | 56               | 1.7%  | 10.6%              |
| MP06_R5   | ---+--  | 357              | 8.0%  | 60.3%              | 32               | 3.7%  | 37.8%              | 112              | 3.5%  | 14.1%              |
| MP06_R6   | --+--+  | 466              | 10.5% | 70.8%              | 55               | 6.49% | 44.3%              | 127              | 3.99% | 18.1%              |
| MP06_R7   | --++-   | 333              | 7.5%  | 78.2%              | 21               | 2.48% | 46.8%              | 68               | 2.14% | 20.2%              |
| MP06_R8   | ---+++  | 184              | 4.1%  | 82.4%              | 9                | 1.06% | 47.8%              | 35               | 1.10% | 21.3%              |
| MP06_R9   | --+---  | 67               | 1.5%  | 83.9%              | 24               | 2.83% | 50.6%              | 101              | 3.17% | 24.5%              |
| MP06R10   | --+++   | 110              | 2.5%  | 86.3%              | 46               | 5.43% | 56.1%              | 158              | 4.97% | 29.5%              |

|          |       |       |       |       |     |       |       |       |       |       |
|----------|-------|-------|-------|-------|-----|-------|-------|-------|-------|-------|
| MP06_R11 | --++- | 113   | 2.5%  | 88.9% | 51  | 6.02% | 62.1% | 166   | 5.22% | 34.7% |
| MP06_R12 | --+++ | 79    | 1.8%  | 90.7% | 29  | 3.42% | 65.5% | 82    | 2.58% | 37.3% |
| MP06_R13 | --+-- | 40    | 0.9%  | 91.6% | 30  | 3.54% | 69.1% | 86    | 2.70% | 40.0% |
| MP06_R14 | --+++ | 63    | 1.4%  | 93.0% | 20  | 2.36% | 71.4% | 157   | 4.93% | 44.9% |
| MP06_R15 | -++++ | 39    | 0.9%  | 93.8% | 15  | 1.77% | 73.2% | 64    | 2.01% | 46.9% |
| MP06_R16 | -++++ | 29    | 0.7%  | 94.5% | 6   | 0.71% | 73.9% | 20    | 0.63% | 47.5% |
| MP06_R17 | +---- | 24    | 0.5%  | 95.0% | 14  | 1.65% | 75.6% | 82    | 2.58% | 50.1% |
| MP06_R18 | +---- | 18    | 0.4%  | 95.4% | 19  | 2.24% | 77.8% | 116   | 3.65% | 53.8% |
| MP06_R19 | +---+ | 35    | 0.8%  | 96.2% | 22  | 2.60% | 80.4% | 160   | 5.03% | 58.8% |
| MP06_R20 | +---+ | 12    | 0.3%  | 96.5% | 19  | 2.24% | 82.6% | 131   | 4.12% | 62.9% |
| MP06_R21 | +--+  | 25    | 0.6%  | 97.1% | 13  | 1.53% | 84.2% | 152   | 4.78% | 67.7% |
| MP06_R22 | +--+  | 32    | 0.7%  | 97.8% | 31  | 3.66% | 87.8% | 234   | 7.35% | 75.0% |
| MP06_R23 | +--+  | 18    | 0.4%  | 98.2% | 19  | 2.24% | 90.1% | 141   | 4.43% | 79.5% |
| MP06_R24 | +---- | 9     | 0.2%  | 98.4% | 5   | 0.59% | 90.7% | 80    | 2.51% | 82.0% |
| MP06_R25 | ++--- | 11    | 0.2%  | 98.6% | 7   | 0.83% | 91.5% | 52    | 1.63% | 83.6% |
| MP06_R26 | ++--- | 10    | 0.2%  | 98.9% | 12  | 1.42% | 92.9% | 92    | 2.89% | 86.5% |
| MP06_R27 | ++--  | 22    | 0.5%  | 99.3% | 17  | 2.01% | 94.9% | 131   | 4.12% | 90.6% |
| MP06_R28 | ++--  | 11    | 0.2%  | 99.6% | 14  | 1.65% | 96.6% | 89    | 2.80% | 93.4% |
| MP06_R29 | +++-- | 1     | 0.02% | 99.6% | 5   | 0.59% |       | 53    | 1.67% | 95.1% |
| MP06_R30 | +++-- | 8     | 0.2%  | 99.8% | 17  | 2.01% | 2.0%  | 92    | 2.89% | 98.0% |
| MP06_R31 | ++++- | 7     | 0.2%  | 100%  | 6   | 0.71% | 2.7%  | 43    | 1.35% | 99.3% |
| MP06_R32 | +++++ | 2     | 0.04% | 100%  | 1   | 0.12% | 2.8%  | 21    | 0.66% | 100%  |
| Total    |       | 4,453 | 100%  |       | 847 | 100%  |       | 3,182 | 100%  |       |

전체 대상 경소상중 20%를 이탈 예상으로 추출할 경우 2013년 3월말 예측 6개월 동적모델에서 Rule 5까지의 패턴을 적용하면 19.1%의 추출률로 769개의 대리점을 이탈예상대리점으로 예측하게 된다. 이 중 실제로 320명이 이탈하므로써 41.6%의 좋은 이탈 예측적중률을 보였다. 이탈 예상 경소상을 예측할 때 추출률 20%를 넘을 경우 추출시 뿐만 아니라 분석시에도 시간과 비용이 대폭 늘어나 모델의 효율성이 감소하므로 적정 추출률을 20%에 근접하도록 설정하였고 이때 이탈 예측적중률은 Rule5(MP06\_R5)까지를 적용한 모델이 더 높은 것으로 나타났다(Rule5까지: 41.6% > Rule6까지: 39.4%). 따라서 Rule5(MP06\_R5)까지를 동적모델 솔루션으로 선택하였다.

Table 7. Prediction Results of Pattern Model

| Prediction at March/2013(6month)   |   | MP06_R2        | MP06_R3        | MP06_R4        | MP06_R5        | MP06_R6 |
|--|---|----------------|----------------|----------------|----------------|---------|
| number of working agents over six months (TS)  |   | 4,029          |                |                |                |         |
| number of churn agents from April to September (TS1)<br>(churn rate: TS1/TS)                       | 847(6 month churn rate: 21%)                  |                |                |                |                |         |
|  | number of retention agents at September (TS2) | 3,182          |                |                |                |         |
| number of predicted churn agents (A)<br>(acquisition rate: C=A/TS)                                 | 381<br>(9.5%)                                 | 542<br>(13.5%) | 625<br>(15.5%) | 769<br>(19.1%) | 951<br>(23.6%) |         |
| number of actual churn agents among the predicted churn agents (B)<br>(prediction accuracy: D=B/A) | 212<br>(55.6%)                                | 261<br>(48.2%) | 288<br>(46.1%) | 320<br>(41.6%) | 375<br>(39.4%) |         |
| number of actual churn agents among the predicted churn agents (B)<br>(churn accuracy: E=B/TS1)    | 212<br>(25%)                                  | 261<br>(30.8%) | 288<br>(34%)   | 320<br>(37.8%) | 375<br>(44.3%) |         |

### 4.2 정적모델(Matrix) 이탈예측률 분석결과

매트릭스를 기반으로 한 정적모델은 경소상을 근무기간과 월평균매출액을 2×2매트릭스로 맵핑하여 그룹별 경소상 수(2013.3월 기준)와 이탈자 수(2013.4월~9월 이탈)를 입력한다. 상대적으로 근무기간이 짧고(PQ\_1~PQ\_2: 26개월 이하) 월평균매출액이 적은 그룹(AQ\_1~AQ\_2: 16,500위안 이하)을 이탈예상그룹으로 예측하였을 때 이탈예상 그룹(PQ\_1~PQ\_2 & AQ\_1~AQ\_2)내 경소상 수 1,427명 중 실제 이탈 경소상 667명을 예측하여 46.7% 이탈예측적중률을 나타내었다.

**Table 8.** Prediction Results of Matrix Model

| Employment Duration<br>(as of March, 2013)<br>Aver. Monthly sales |                       | PQ_1<br>(15↓) | PQ_2<br>(16~26) | PQ_3<br>(27~33) | PQ_4<br>(34↑) | Total        |
|---|-----------------------|---------------|-----------------|-----------------|---------------|--------------|
| AQ_1<br>(10,000↓)   | 2013.03 Retention     | 409           | 420             | 383             | 332           | 1,544        |
|   | 2013.04 ~ 09 Churning | 253<br>62%    | 170<br>40%      | 131<br>34%      | 87<br>26%     | 640<br>41%   |
| AQ_2<br>(10,000~16,500)   | 2013.03 Retention     | 311           | 287             | 360             | 294           | 1,252        |
|   | 2013.04 ~ 09 Churning | 156<br>50%    | 8<br>31%        | 66<br>18%       | 68<br>23%     | 378<br>30%   |
| AQ_3<br>(16,500~27,000)   | 2013.03 Retention     | 297           | 316             | 361             | 364           | 1,338        |
|   | 2013.04 ~ 09 Churning | 95<br>32%     | 53<br>17%       | 37<br>10%       | 43<br>12%     | 228<br>17%   |
| AQ_4<br>(27,000 ↑)  | 2013.03 Retention     | 408           | 400             | 302             | 258           | 1,368        |
|   | 2013.04 ~ 09 Churning | 33<br>8%      | 26<br>7%        | 21<br>7%        | 23<br>9%      | 103<br>8%    |
| Total   | 2013.03 Retention     | 1,426         | 1,423           | 1,406           | 1,248         | 5,502        |
|   | 2013.04 ~ 09 Churning | 536<br>38%    | 337<br>24%      | 255<br>18%      | 221<br>18%    | 1,350<br>25% |

### 4.3 동적모델과 정적모델 간의 이탈예측적중률 차이 검정

4.1과 4.2를 통하여 정적모델의 이탈예측적중률이 46.7%로 동적모델의 41.6% 보다 높게 나타남을 확인하였다. 아래의 수식을 통하여 비율차이 검정(Z)결과 정적모델과 동적모델의 이탈예측적중률에는 유의한 차이가 있음을 통계적으로 검증하였다( $P=0.021 < 0.05$ ). 따라서 연구가설 즉, “H1: 동적모델(M1)과 정적모델(M2)의 이탈예측률은 차이가 있을 것이다(H1:  $M1 \neq M2$ )”가 채택되며 귀무가설 “H0: 동적모델(M1)과 정적모델(M2)의 이탈예측률은 차이가 없을 것이다(H0:  $M1 = M2$ )”은 기각된다. 다시 말하면, 정적모델의 이탈예측적중률이 더 높다고 할 수 있다.

정적모델(M1)과 동적모델(M2) 차이 검정

$$Z_{obs} = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\hat{p}\hat{q}}{n_1} + \frac{\hat{p}\hat{q}}{n_2}}} = \frac{(0.46 - 0.42) - 0}{\sqrt{\frac{(0.45)(0.55)}{1427} + \frac{(0.45)(0.55)}{769}}} = 2.305$$

$$\hat{p}_1 = \frac{320}{769} = 0.42$$

$$\hat{p}_2 = \frac{667}{1427} = 0.46$$

$$\hat{p} = \frac{(320 + 667)}{(769 + 1427)} = 0.45 \rightarrow \hat{q} = 0.55$$

$$Z_{crit} = Z_{\alpha} = Z_{0.01} = 2.305$$

$$P = 0.021$$

### 4.3 동적모델과 정적모델 간의 이탈예측적중률 차이 비교

위에서 수식을 통하여 검정(비율의 차이: 정적모델 Vs. 동적모델)한 바와 같이 정적모델은 통계적으로 동적모델보다 뛰어난 이탈예측적중률을 보여준다는 것을 증명하였다. 동적모델은 6개월간의 패턴을 분석하여 추출률 20%에서 높은 예측율을 보이는 룰(MP06\_R1~MP06\_R5)을 탑재한 모델로 이탈예측적중율은 41.6%며, 정적모델은 근무기간(Duration)과 월평균매출액(Amount)을 동시에 고려하여 월평균매출액이 16,500위안 이하이면서 근무기간이 26개월 이하인 경소상 1,427명을 이탈 예상경소상으로 예측하였고, 이들 중에서 실제 이탈한 경소상들은 667명으로 이탈예측적중률 46.7%를 보였다. 이탈예측적중률이 동적모델보다 정적모델에서 높게 나타나는 차이는 통계적으로도 유의한 차이가 있음을 검정하였으며 정적모델이 동적모델보다 통계적으로 이탈경소상 예측이 우수하다고 할 수 있다.

## 5. 결론 및 시사점

본연구의 결론 및 시사점은 아래와 같이 정리할 수 있다.

첫째, 본 연구는 이탈예상대리점을 예측할 수 있는 두 가지 모델을 개발하였다. 하나는 매트릭스를 기반으로 특정 시점에서의 상태를 나타내는 정적 모델이며 다른 하나는 시간의 흐름에 따른 고객(대리점)군의 변화패턴을 예측할 수 있는 동적 모델이다. 정적모델(46%)과 동적모델(42%) 모두 40% 이상의 이탈예측적중율을 보였다. 실제 A기업의 6개월 동안의 이탈예측적중율은 약 18%(2013년 3월 기준 3.2%, 6개월 누적시 약 18%)에 불과하며 본 연구에서 개발된 두 모델 모두 이보다 높은 예측율을 보였다.

둘째, 개발된 정적모델과 동적모델에 대하여 비율차이 검정(Z)을 수행하여 두 모델 간의 이탈예측적중률 차이가 통계적으로 유의한 지를 확인하였다. 그 결과 유의확률 0.05미만(p<0.05)으로 정적모델과 동적모델의 이탈예측적중

률에는 통계적으로 의미 있는 차이가 있음을 확인하였다. 즉, 이탈예측적중률이 더 높게 나온 정적모델이 동적모델보다 우수하다고 할 수 있으며 이러한 점은 보다 정확한 이탈예측모델을 선택해야하는 기업이나 조직들에게 시사하는 바가 있다.

셋째, 본 연구는 최종소비자가 아닌 대리점 단위의 CRM연구라는 점에서 기존연구와 차별성이 있으며, 특히 화장품 산업처럼 대리점의 역할이 큰 산업 내 기업들은 본 연구에서 개발된 대리점 이탈예측모델의 상대적으로 높은 이탈예측적중율을 활용하여 효과적인 대리점 이탈 관리가 가능하다.

넷째, 본 연구에서는 대리점이탈을 예측하는데 있어 정적인 개념과 함께 동적인 개념을 제시하였다. 즉, 6개월 동안의 총 액수가 아닌 월별로 주문 매출 액수가 늘어나거나, 줄어드는 6개월 동안의 추세를 패턴화하여 이탈예측에 반영하였다는 점에서 기존연구와 차별성이 있다. 기존 연구들도 고객 행동 패턴을 모델에 반영하려는 시도가 있었으나, 본 연구처럼 위상 차이 분석을 통한 추세변화를 모델에 반영하지 못하였다.

다섯째, 본 연구는 RFM의 기본 변수 외에 근무기간(duration)이라는 새로운 변수를 삽입하여 대리점의 특성을 반영하였다는 점에서 기존 연구와의 차별성을 가지고 있다.

본 연구의 한계로는 RFM변수들 중에서 구매액(Monetary)변수만을 사용하였다는 점을 들 수 있다. 대리점의 속성(일정수준이상으로 재고를 관리)과 본사 프로모션(관촉활동)은 A화장품 경소상들의 최근성(Recency)과 빈도(Frequency) 변수를 원래 의미와 다르게 변질 시킬 수 있다는 점과 두 변수간의 상관관계도 높아 다중공선성(Multicollinearity)문제가 발생할 가능성도 높으므로 본 연구에는 최근성과 빈도를 포함하지 않았다.

향후의 연구로는 첫째, RFM변수들 중 Recency와 Frequency를 추가로 포함하여 다차원적인 예측모델을 개발하는 방안을 검토할 수 있으며 둘째, 정적인 연구와 동적인 연구를 통합하는 통합모델 개발에 관한 연구가 필요하다고 할 수 있다. 셋째, 비즈니스적 관점에서 대리점을 기반으로 사업을 영위하는 조직이나 기업에서는 본 연구의 이탈예측모델을 통해 이탈고객을 사전에 파악하여 이탈방지전략을 수립할 수 있다. 또한 본 연구결과를 바탕으로 다양한 산업분야에 맞춤형된 고객이탈예측모형 개발도 가능하다. 다만 유의할 점으로, 본 연구가 대리점을 기반으로 하고 있는 중국 화장품 기업의 데이터를 이용하여 분석하였기 때문에 국내 개인 고객을 대상으로 사업을 영위하고 있는 기업들이 적용하는 것은 신중할 필요가 있다. 마지막으로 본 연구를 실제 산업 현장에서 응용하는 방법으로 이탈예측모델뿐만 아니라 성공한 대리점들을 분석하여 대리점 성공예측모델 또는 이를 발전시킨 대리점 성공전략을 수립하여 효율적이고 효과적으로 대리점을 성공시킬 수 있는 경로를 충분히 예측해 볼 수 있다.

## REFERENCES

- Abbott, J. 2001. "Data data everywhere and not a byte of use?." *Qualitative Market Research* 4(3):182-192.
- Adrian, Payne, and Pennie, Frow. 2005. "A Strategic Framework for Customer Relationship Management." *Journal of Marketing* 69(4):167-176.
- Berry, L. L. 1995. "Relationship Marketing of services—Growing Interest, Emerging Perspectives." *Journal of the Academy of Marketing Science*:236-245.
- Bult, J. R., and Wansbeek, T. 1995. "Optimal Selection for Direct Mail." *Market Sci* 14(4):378-394.
- Chang, J., Yen, D. C., Young, D., and Ku, C. Y. 2002. "Critical issues in CRM adoption and implementation." *International Journal of Services Technology and Management* 3(3):311-324.
- Fader, P. S., Hardie, B. G. S., and Lee, K. L. 2005. "Counting your customers, The easy way: An alternative to the Pareto/NBD model." *Marketing Science* 24(2):275-284.
- Shin, Geoncheol, Ku, Yeongae, and Choi, sunghwan. "The Study on the Influence of Consumers' Value that Affects Brand Charisma, Brand Attitude and Purchase Intention With Relevance to Masstige Brand." *Journal of*

- Korean Society for Quality Management 39(4).
- Ha, S. H., Bae, S. M., and Park, S. C. 2002. "Customer's time-variant purchase behavior and corresponding marketing strategies: an online retailer's case." *Computers & Industrial Engineering* 43(4):801–820.
- Hsieh, N. C. 2004. "An integrated data mining and behavioral scoring model for analyzing bank customer." *Expert Systems with Applications* 27:623–633.
- Hung, S. Y., Yen, D. C., and Wang, H. Y. 2006. "Applying data mining to telecom churn management." *Expert Systems with Applications* 31:515–524.
- Hwang, Hyunseok, Jung, Taesoo, and Suh, Euiho. 2004. "An LTV model and customer segmentation based on customer value: A case study on the wireless telecommunication industry." *Expert Systems with Applications* 26:181–188.
- Kim, Hyun-Mo et al. 2012. "Variables Affecting End-User Satisfaction in Application Market." *Journal of Korean Society for Quality Management* 40(2).
- John, H., Ashutosh, T., Rajkumar, R., and Dymitr, R. 2008. "Churn Prediction: Does Technology Matter?." *World Academy of Science, Engineering and Technology* 16.
- Kaymak, U. 2001. "Fuzzy target selection using RFM variables." *Proceedings of the Joint 9th IFSA World Congress and 20th NAFIPS International Conference*, 1038–1043.
- Kim, Gye-Soo, and Xu, Da-Peng, 2013. "Excellence for Organization Quality Management Innovation Challenge-Comparison of Korea and China." *Journal of Korean Society for Quality Management* 41(4).
- Kim, Su-Yeon, Jung, Tae-Soo, Suh, Eui-Ho, and Hwang, Hyun-Seok. 2006. "Customer segmentation and strategy development based on customer lifetime value: A case study." *Expert Systems with Applications* 31:101–107.
- Kincaid, J. W. 2003. "Customer Relationship Management: Getting it Right!." New Jersey: Prentice-Hall PTR.
- Liu, D-R., and Shih, Y-Y. 2005. "Hybrid approaches to product recommendation based on customer lifetime value and purchase preferences." *The Journal of Systems and Software* 77(2):181–191.
- Moe, Wendy w., and Peter, S. Fader. 2001. "Uncovering Patterns in Cybershopping." *California Management Review* 43(4):106–117.
- Paravtiyar, A., and Sheth, Jagdish N. 2001, "Customer Relationship Management: Emerging Practice, Process, and Discipline." *Journal of Economic and Social Research* 3(2):1–34.
- Scott, A. Neslin, Sunil, Gupta, Wagner, Kamakura, Junxiang, Lu, and Charlotte, H. 2006. "Defection Detection: Measuring and Understanding the Predictive Accuracy of Customer Churn Models." *Journal of Marketing Research* 43(2):204–211.
- Shani, D., and Chalasani, S. 1992. "Exploiting Niches Using Relationship Marketing." *Journal of Consumer Marketing* 9(3):33–42.
- Hosseini, S. M. S. et al. 2010. "Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty." *Expert Systems with Applications* 37:5259–5264.
- Vavra, Terry G. 1992. "How to Keep Customers for life through relationship marketing." Homewood: Busienss one Irwin.
- Verhoef, Peter C. 2003. "Understanding the Effect of Customer Relationship Management Efforts on Customer Retention and Customer Share Development." *Journal of Marketing* 67(4):30–45.
- Wei, C. P., and Chiu, I. T. 2002. "Turning telecommunications call details to churn prediction: A data mining approach." *Expert Systems with Applications* 23:103–112.
- Winer, R. S. 2001. "A framework for customer relationship management." *California Management Review* 43(4):89–105.
- Yeh, C., Yang, K., and Ting, T. 2009. "Knowledge discovery on RFM model using Bernoulli sequence." *Expert Systems with Applications* 36:5866–5871.