

인터넷 쇼핑물에서의 동적 고객 분류에 관한 연구

임승재, 서의호, 정태수

포항공과대학교 산업공학과

sjlim76@postech.ac.kr, ehsuh@postech.ac.kr, white978@postech.ac.kr

Abstract

고객 분류는 고객관계관리(CRM)의 한 부분으로서 기업에게 이익을 주는 고객의 속성과 구매패턴을 분석함으로써 목표 고객을 설정하는 것을 의미한다. 현재까지 고객 분류에 관한 연구는 특정한 시점에서 고객의 속성과 구매 패턴을 분석함으로써 이루어졌다. 그러나 인터넷 쇼핑물과 같은 동적인 환경에 있어서, 기존의 정적인 분석방법은 시간에 따라 지속적으로 변하는 고객의 행동 변화를 찾아내고 예측하는데 적합하지 않다.

본 논문에서는 Decision Tree, ANOVA 분석, ARIMA모형을 사용하여, 특정한 시점에서의 고객 분류뿐만 아니라 미래 시점에서의 고객 분류를 예측하고 패턴을 분석하는 동적인 고객 분류 방법을 제안한다. 동적인 고객 분류를 통해 인터넷 쇼핑물 기업은 효율적인 마케팅 전략을 작성하여 기업의 이익을 증진시킬 수 있다.

1. Introduction

고객 분류는 고객관계관리(CRM)의 한 부분으로서 기업에게 이익을 주는 고객의 속성과 구매패턴을 분석함으로써 목표 고객을 설정하는 것을 의미한다.(A.Vellido et al., 1999) 고객 분류는 B2C 시장에서 중요한 연구분야가 되고 있으며(Chang, 1998) 특히 인터넷 기업에게 있어서 관심이 증대되고 있다.(O'Connor & O'Keefe, 1997)

고객 분류의 목적은 마케팅의 대상을 선정하고 분류함으로써 고객의 특성을 파악하여 효율적인 마케팅 전략을 도출하는 데 있다. ((P.C.Verhoef et al., 2002) 고객 분류는 여러 가지 Data Mining 기법과 통계기법을 통하여 이루어지는데 대표적인 방법으로는 Cross tabulations, RFM, Linear regression analysis, Cluster analysis Factor analysis, CHAID or CART, Discriminant analysis, LOGit probit analysis, Neural network,

Genetic algorithms 등이 있다. (P.C.Verhoef et al., 2002) RFM 방법을 사용한 Cross tabulation이 고객 분류에 있어서 가장 많이 사용되는 방법론이다.

특히 인터넷 쇼핑물 고객의 분류를 위하여 여러가지 다양한 방법론이 제시되었는데 대표적인 경우로 RFM(S.H.Ha & S.C.Park, 1998), Neural Network (A.Vellido et al, 1999, 김시환 외, 2000)와 Fussy c means를(M.Ozer, 2001) 사용한 연구가 있다. 기존 연구에서 사용하였던 방법은 특정시점에서 이루어진 정적인 방법이었다. 시간이 지남에 따라 고객의 정적인 지식을 통한 분류는 한계를 나타내었다.(S.H.Ha, 2002) 또한 기존의 고객 분류 방법은 기업의 마케팅 전략과의 연결이 미흡하였다. 시간에 따른 고객의 변화가 심한 E-commerce 상황에서 고객 분류를 마케팅 전략으로 연결하기 위해서는 고객의 변화에 능동적으로 대응할 수 있는 동적인 고객 분류 방법이 필요하다. 정적인 고객 분류는 고객 분류에 있어서 다음과 같은 약점이 있다. 첫째, 고객 분류의 시기를 결정할 수 없다. 둘째, 고객의 등급 변화를 살펴볼 수 없다. 셋째, 시간에 따르는 고객의 구입패턴 변화를 찾아낼 수 없다.(S.H.Ha et al., 2002)

본 논문에서는 정적인 고객 분류의 문제점을 극복하여 시간에 따르는 동적인 고객 분류 방법을 제안하고자 한다. 본 논문의 목적은 미래 시점에서의 고객 분류를 예측하고 그 원인을 규명하여 마케팅 전략에 도움을 주는 데 있다.

본 논문은 다음과 같이 구성되었다. 2장에서는 동적인 고객분류에 연관된 RFM, Decision Tree, ARIMA 방법론의 기본적인 개념에 대하여 설명한다. 3장에서는 2장에서 소개된 개념을 통해 과거 및 현재시점과 미래시점으로 나누어서 동적 고객분류의 방법을 제시한다. 마지막으로 4장에서는 본 논문의 요약과 공헌도 그리고 앞으로의 연구방향을 제시한다.

2. Related Work

본 단원에서는 동적 고객 분류를 위해 사용되는 ARIMA 모형, Decision Tree, ANOVA에 관하여 기본적인 개념을 소개한다.

2.1 ARIMA 모형

ARIMA(Autoregressive Integrated Moving Average) 모형은 시계열 자료의 예측을 위해서 Box and Jenkins에 의해서 만들어지는 방법론이다.(Box & Jenkins, 1970) ARIMA 모형은 일반적으로 Model Identification, Parameter Estimation, Diagnostic Checking 세 단계를 반복적으로 수행한다.

ARIMA 모형은 과거의 자료와 Random Error를 고려하여 미래의 값을 예측한다. ARIMA 모형을 수식으로 표현하면 다음과 같다.

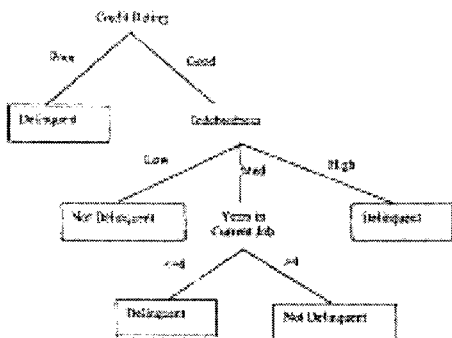
$$y_t = \theta_0 + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \dots - \theta_q \epsilon_{t-q}$$

여기서 y_t 와 ϵ_t 는 시점 t 에서의 actual value 와 random error를 가르킨다. ϕ_i ($i=1,2,\dots,p$) and θ_j ($j=0,1,2,\dots,q$)는 ARIMA 모형의 파라미터들이다. p 와 q 는 정수값을 가지며 모형의 차수를 결정한다.(G.P.Zhang, 2003)

2.2 Decision Tree

Data Mining의 여러 기법 중에서 Decision Tree는 일반적으로 예측력보다는 설명력이 강한 것으로 알려져 있다. Decision Tree는 Target을 예측하기 위해서 Tree의 형식으로 표현한다.(R.O.Duda, 2001)

Decision Tree의 일반적인 형식은 [그림1]와 같다. 각각의 노드는 Decision을 위한 규칙을 의미한다. 마지막 노드는 Decision Tree를 통해 얻어진 분류를 나타낸다. 높은 곳의 있는 노드는 아래쪽의 노드보다 높은 결정력을 갖는다.



[그림1] Decision Tree(Bose et al., 2001)

2.3 ANOVA

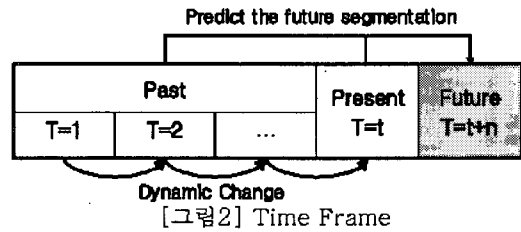
ANOVA(Analysis of Variance)는 가설 검증에 있어서 많이 사용되는 도구이다. 본 논문에서는 특정한 요소가 고객 분류에 영향을 미친다는 가설을 검증하는데 사용한다. 즉, 고객 분류에 영향을 준 요소들을 밝혀내는 도구로 사용한다.

일반적으로 One Factor ANOVA와 two Factor ANOVA를 사용한다. 본 연구에서도 두 가지 요소 사이의 교호작용까지만을 고려한다.

3. Methodology

본 단원에서는 시간에 따른 동적 고객 분류 방법과 절차에 관하여 설명한다. 시간에 따른 고객 분류를 위해서 먼저 Time Frame을 정의하고 시점에 따른 동적 고객 분류 방법을 제시한다.

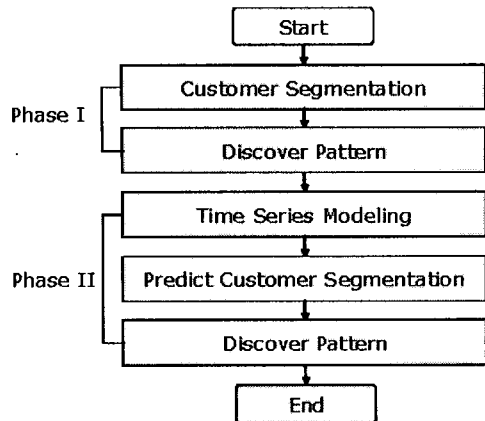
3.1 Time Frame 정의



시간에 따른 고객 분류를 설명하기 위해서는 먼저, 고객 분류 시점을 정의해야 한다. 고객 분류의 시점은 [그림2]와 같이 일반적으로 과거, 현재, 미래의 세가지 측면으로 나뉜다. 이 단원에서 설명하는 동적 고객 분류에서 사용되는 시간은 T 로 표시하며 이는 월별 시간 단위를 나타낸다. 동적 고객 분류 방법을 처음 시작한 시점은 $T=1$ 으로 나타내며, 현재 시점은 $T=t$ 로 나타낸다. 현재부터 n 시간단위 후의 미래는 $T=t+n$ 으로 나타낸다. 각 시간단위의 고객 분류는 다음 시간단위에 영향을 미치게 되며, 미래 시점의 예측을 위해서 과거와 현재의 고객 분류 자료가 사용된다.

3.2 동적 고객 분류를 위한 절차

동적 고객 분류방법의 절차는 [그림3]과 같이 과거와 현재시점의 고객 분류 자료를 시계열 자료로 변환한 후 미래의 시점의 고객 분류를 예측한다. 동적 분류를 위한 절차는 크게 두 가지 단계로 이루어진다.



[그림3] Procedure for Dynamic Segmentation

첫째 단계는 과거와 현재 시점에서 이루어지는 고객 분류 방법으로서 분류시점에서의 과거 고객 자료를 통하여 고객을 분류하고 패턴을 분석한다. 이 단계에서 이루어지는 고객 분류는 기존의 정적인 고객 분류 방법과 유사하지만 과거 시점에서의 고객 분류를 시간에 따라 정리하고 반영한다는 점에서 기존의 고객 분류와 차이가 있다. 이 단계에서 초점을 맞추고 있는 부분은 시간의 흐름에 따른 고객 분류 자료의 수집과 단기적인 마케팅 전략에 도움을 줄 수 있는 월별 고객 분류와 패턴 분석이다.

둘째 단계는 미래의 고객 분류를 예측하는 부분이다. 첫째 단계에서 얻어진 시간에 따른 고객 분류 자료는 시계열 자료로 변화되는 단계를 거쳐 미래의 고객 분류를 예측하고 패턴을 분석한다. 이 단계를 통해서 기업은 미래 고객의 위치를 파악함으로써 현재의 고객의 상황만을 고려하는 것이 아니라 미래의 고객 상황까지 고려한 마케팅 전략을 세울 수 있다. 동적 고객 분류는 과거와 현재 그리고 미래의 유기적인 연결을 통하여 이루어진다.

각 시점에서 적용되는 고객 분류의 목적과 방법은 서로 다르기 때문에 본 연구에서는 고객 분류의 시점을 크게 과거 및 현재 시점과 미래 시점으로 분류하였다. 다음 장에서는 동적 고객 분류를 위하여 두 시점에 따른 부적인 고객 분류 절차와 방법에 대하여 설명하게 된다.

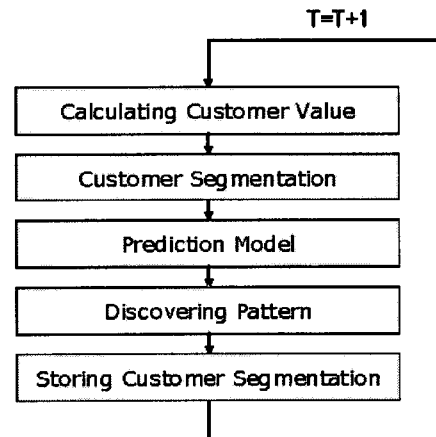
3.3 과거와 현재시점에서의 고객분류(Phase I)

과거와 현재 시점에서의 고객 분류의 목적은 크게 두 지로 나눌 수 있다. 첫째는 현재 상태의 고객의 가치를 파악하여 단기적인 마케팅 전략을 세우는 것이고 둘째는 미래 시점

의 고객 분류를 예측을 위한 자료로 사용된다.

이 시점에서 고객 분류는 [그림4]와 같이 진행된다. [그림4]에서 제시된 절차를 간단히 설명하면 다음과 같다.

먼저 고객 가치를 산정하기 위해 RFM 분류법에 의해서 고객의 점수를 매긴다. 고객의 점수를 통해 고객별로 등급을 나누어 고객 분류를 실시한다. 여기서 나온 고객 분류는 일부 고객을 통해서 얻은 분류이므로 마케팅 전략을 세우기에 부적합하다. 따라서 고객의 분류에 영향을 미치는 고객 특성을 얻기 위해서 등급을 Target 변수로 정하고 Training 절차를 거친다. 여기서 사용되는 방법은 여러 가지 Data Mining 기법 중에서 분석이 용이한 Decision Tree를 사용한다. Training 절차를 걸친 Prediction Model은 새로운 고객들의 분류와 기존 고객의 패턴을 분석하는 도구로 사용된다. Decision Tree는 어느 정도의 패턴을 예측하는데 유용하지만 구체적으로 어떤 요소가 고객 분류에 영향을 미치는가에 대한 답을 주기에는 미흡하다. Decision Tree와 함께 ANOVA 분석을 사용함으로써 패턴 분석을 시행한다. 이러한 절차는 월별로 반복적으로 시행되며 월별 고객 분류 자료는 저장되어 미래 시점의 고객 분류를 예측하는 데 사용된다.



[그림4] 과거와 현재시점에서의 고객분류절차

3.3.1 Calculating Customer Value

Customer Value를 계산하기 위해서 사용되는 방법은 RFM 분석법이다. 기업에게 있어 사용되는 RFM을 표현하는 속성을 정의하면 다음과 같다. 첫째, R은 Recency를 약어로서 얼마나 최근에 상품을 구입했는가를 나타내며 둘째, F는 Frequency의 약어로서 상품을 구매하는 빈도를 나타내며 셋째, M은 총 구입금액을 나타낸다. 고객의 가치를 산정하는 방법

으로 이 세가지 변수를 가장 많이 사용하고 예측력이 높은 것으로 알려져 있다. 각각의 변수에 점수를 부여하는 것은 데이터의 성격과 상황에 따라 다르다. RFM 변수를 써서 고객의 가치는 다음과 같이 계산된다.

$$\text{Customer Value} = R * F * M \text{ -----(1)}$$

3.3.2 Customer Segmentation

계산된 Customer Value를 통하여 고객을 세분화 한다. 고객의 등급을 나누기 위해서 기준을 선정하여야 한다. 고객을 분류하는 기준으로는 Loyalty에 초점을 맞추어 Loyal, Habitual, Variety Seeker, Switcher로 구분하기도 하고(SIMON KONX, 1998), Privacy와 Interest에 따라 Complacent Customer, Efficiency Experts, Agitated Activists로 구분하기도 한다.

본 논문에서는 위의 세그멘테이션 방법과 더불어 기업에서 기존에 사용하는 분류가 있다면 그것을 사용하기로 한다. 이 단계에서는 Customer Value를 통해 각 고객의 점수와 등급을 부여하게 된다.

3.3.3 Prediction Model

고객 분류의 패턴과 원인을 규명하고 새로운 고객의 분류를 위해 Prediction Model은 유용하다. Prediction Model을 위한 방법론은 Data Mining의 여러가지 기법 중에 Decision Tree를 사용한다. Decision Tree는 일반적으로 예측을 측면에서는 다른 기법에 비해 떨어지지만 사용자에게 원인을 설명하는 측면에서 유용하다.

3.3.4 Discovering Pattern

고객 분류에 영향을 미치는 변수를 구분하기 위하여 ANOVA 분석을 실시한다. 고객 분류를 Target 값으로 설정하고 고객 속성을 변수로 지정하여 얼마나 영향을 미치는가를 결정한다. Prediction Model을 구축함으로써 얻어진 패턴에 관하여 ANOVA 분석을 실시함으로써 보다 정확한 원인을 분석한다.

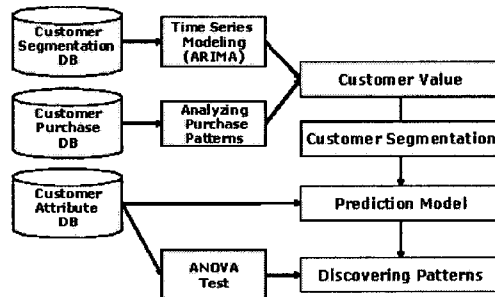
3.3.5 Storing Customer Segmentation

위의 과정은 월 단위로 반복적으로 시행되며 각 달의 Customer Segmentation 자료는 미래 시점의 고객 분류를 예측하는데 유용하게 쓰이게 된다.

3.4 미래 시점에서의 고객 분류 예측(Phase II)

본 단원에서는 과거 및 현재 시점의 축적된

고객 분류 자료를 토대로 미래 시점의 고객 분류를 예측하고 패턴을 분석한다. 미래 시점에서의 고객 분류를 위한 방법은 과거 고객 분류 자료를 시계열 자료로 변화하여 ARIMA 모형을 통하여 예측을 하게 된다. 시계열 분석을 통하여 얻어진 고객 분류 결과는 미래의 어느 시점에서든 예측이 가능하지만 일정한 시간이 지나면 동일한 결과를 나타내거나 특정한 패턴을 나타낼 것으로 기대한다. 미래 시점에 대한 문제는 Case Study를 통해서 증명하게 된다. 본 단원에서는 특정 미래 시점에서의 고객 분류 예측 방법을 제시하도록 한다.



[그림5] 미래시점에서의 고객분류절차

미래 시점에서의 고객 분류 예측의 절차를 간단히 요약하면 [그림5]과 같다. Customer Value를 예측하기 위해서 과거의 고객 분류 자료와 구입 패턴 자료가 필요하다. 고객 분류 자료는 ARIMA 모형을 통해서 시계열 자료로 변환된다. 또한 고객 구입패턴이 고객 가치를 예측하는데 중요한 요소이기 때문에 과거의 자료를 분석하여 반영한다. 고객 가치가 예측되면 고객 분류 작업이 이루어지고 새로운 고객을 위해 Prediction Model을 작성한다. 마지막으로 Prediction Model과 ANOVA 분석을 통해서 고객 분류에 영향을 주는 요소를 찾아낸다. 위 절차를 부분별로 정리하면 다음과 같다.

3.4.1 Customer Value 예측

Customer Value를 예측하기 위해서 본 논문에서는 과거의 고객 분류자료와 구매 패턴을 고려하였다. 우선 고객의 분류자료는 시계열 자료 변환을 통하여 ARIMA 모형을 통하여 예측을 하게 된다. ARIMA 모형은 시계열 분석 중에서도 Autocorrelation과 Moving Average를 통합하여 만든 모형이다. ARIMA 모형은 Box-Jenkins 방법론이라고도 하며 ACF, PACF를 결정하여 미래 시점의 값을 예

측하는 시스템이다. 시계열 자료에 적합한 값은 자료에 따라 다르므로 구체적인 절차는 생략한다.

3.4.2 Customer Segmentation

Customer Value에 의해서 예측된 값은 과거 및 현재시점에서의 동일한 방법으로 고객 분류를 실시한다. 고객 분류가 완성되면 과거의 고객 자료의 변화를 비교를 실시하고 고객 등급의 변화를 기록한다.

3.4.3 Prediction Model

고객 분류 변화를 Target 값으로 정하고 고객의 속성을 통하여 Decision Tree를 통하여 Prediction Model을 만든다. Prediction Model을 통하여 고객의 미래 가치를 추측한다.

3.4.4 Discovering Pattern

Prediction Model을 통해서 얻어진 결과값은 구체적인 통계분석이 없기 때문에 ANOVA 분석을 통하여 영향을 미치는 원인을 규명한다. 이를 통하여 미래 시점의 고객 분류의 원인을 예측한다.

4. Conclusion

본 논문에서는 ARIMA 모형과 Decision Tree와 ANOVA를 이용한 동적 고객 분류방법을 제안하였다. 동적 고객 분류는 과거 및 현재시점과 미래 시점으로 나누어서 이루어지며 각각에 대한 방법을 제안하였다.

고객 분류의 시점을 특정시점에 국한하지 않고 시간에 따라 동적으로 분류하는 것은 마케팅을 위한 중요한 지표가 된다. 특히 인터넷 쇼핑물 기업이 대상이기 때문에 수시로 변하는 고객의 변화를 반영하여 고객 분류를 하는 것은 의미 있는 일이다.

그러나 본 방법론은 사례연구를 통하여 검증되지 않았으며 실제 상황과 다를 수 있다. 사례연구를 통하여 현실 상황을 고려하여 수정을 한 후 타당성 검증을 받아야 한다.

앞으로의 연구과제는 인터넷 쇼핑물 기업의 고객의 구입자료와 속성자료를 통하여 본 방법론을 적용하는 것이다.

또한 시간에 따른 고객 분류 예측을 위하여 ARIMA 모형을 사용하였는데 다른 시계열 예측기법과의 비교 분석이 필요하다.

Reference

- [1]김시환, 권영식, 데이터 마이닝을 이용한 인터넷 쇼핑물 고객 세분화에 관한 연구, 대한산업공학회 2000 추계학술대회
- [2]A.Vellido, P.J.G. Lieboa, K.Meehan, Segmentation of the on line shipping market using neural networks, Expert systems with Application, 17, pp. 303-314, 1999
- [3]Bose, R.K.Mahapatra, Business data mining-a machine learning perspective, Information & Management, 39, 2001
- [4]G.C.O'Connor, &B.O'Keefe, Viewing the Web as a marketplace; the case of small companies, Decision Support Systems,21(3),pp171-183, 1997
- [5]G.E.P.Box, G.Jenkins, Time series Analysis, Forecasting and Control, Holden-Day, San Francisco, CA, 1970
- [6]G.P.Zhang, time series forecasting using a hybrid ARIMA and neural network model, Neurocomputing, 50, pp.159-175, 2003
- [7]M. Ozer, User segmentation of online music services using fuzzy clustering, omega, 29, pp.193-206, 2001
- [8]P.C.Verhoef, P.N.Spring, J.C.Hoekstra, P.S.H.Leefflang, The commercial use of segmentation and predictive modeling techniques for database marketing in the Netherlands, Decision Support System, 34, pp.471-481, 2002
- [9]R.O.Duda, P.E.Hart, D.G.Stork, Pattern Classification, John Willy & Sons, 2001
- [10]S.Chang, Internet Segmentation:state of the art marketing applications, Journal of Segmentation in Marketing,2(1),pp.19-34,1998
- [11]S.Knox, Loyalty-Based Segmentation and the Customer Development Process, European Management, 16(6), pp.729-737, 1998
- [12]S.H.Ha, & S.C.Park, Application of data mining tools to Hotel data mart on the Intranet for database marketing, Expert Systems with Application, 15(4), pp.1-31, 1998
- [13]S.H.Ha, S.M.Bae, S.C.Park, Customer's time variant purchase behavior and corresponding marketing strategies: an online retailer's case. Computer & Industrial Engineering, 43, pp.801-820, 2002