

이차원 고객충성도 세그먼트 기반의 고객이탈예측 방법론*

김형수

한성대학교 스마트경영공학부
(hskim@hansung.ac.kr)

홍승우

한성대학교 대학원 산업경영공학과
(seung6858@gmail.com)

CRM의 하위 연구 분야로 진행되었던 고객이탈예측은 최근 비즈니스 머신러닝 기술의 발전으로 인해 빅데이터 기반의 퍼포먼스 마케팅 주제로 더욱 그 중요도가 높아지고 있다. 그러나, 기존의 관련 연구는 예측 모형 자체의 성능을 개선시키는 것이 주요 목적이었으며, 전체적인 고객이탈예측 프로세스를 개선하고자 하는 연구는 상대적으로 부족했다. 본 연구는 성공적인 고객이탈관리가 모형 자체의 성능보다는 전체 프로세스의 개선을 통해 더 잘 이루어질 수 있다는 가정하에, 이차원 고객충성도 세그먼트 기반의 고객이탈예측 프로세스 (CCP/2DL: Customer Churn Prediction based on Two-Dimensional Loyalty segmentation)를 제안한다. CCP/2DL은 양방향, 즉 양적 및 질적 로열티 기반의 고객세분화를 시행하고, 고객세그먼트들을 이탈패턴에 따라 2차 그룹핑을 실시한 뒤, 이탈패턴 그룹별 이질적인 이탈예측 모형을 독립적으로 적용하는 일련의 이탈예측 프로세스이다. 제안한 이탈예측 프로세스의 상대적 우수성을 평가하기 위해 기존의 범용이탈예측 프로세스와 클러스터링 기반 이탈예측 프로세스와의 성능 비교를 수행하였다. 글로벌 NGO 단체인 A사의 협력으로 후원자 데이터를 활용한 분석과 검증을 수행했으며, 제안한 CCP/2DL의 성능이 다른 이탈예측 방법론보다 우수한 성능을 보이는 것으로 나타났다. 이러한 이탈예측 프로세스는 이탈예측에도 효과적인 뿐만 아니라, 다양한 고객통찰력을 확보하고, 관련된 다른 퍼포먼스 마케팅 활동을 수행할 수 있는 전략적 기반이 될 수 있다는 점에서 연구의 의의를 찾을 수 있다.

주제어 : 고객이탈예측, CRM, 고객로열티, 고객 빅데이터, CCP/2DL

논문접수일 : 2020년 6월 29일 논문수정일 : 2020년 12월 10일 게재확정일 : 2020년 12월 16일
원고유형 : 학술대회 Fast-track 교신저자 : 김형수

1. 서론

성숙기에 접어든 대부분의 업종들이 치열한 경쟁환경에 노출되면서 고객생애가치의 중요성을 인식하고, 이에 따라 신규고객의 확보보다는 기존고객의 이탈방지가 더욱 중요한 비즈니스 이슈가 되고 있다(Hung and Tsai, 2008). 이는 기

존 고객을 유지하는 것이 새로운 고객을 확보하는 것에 비해 경제적인 측면에서 훨씬 유리하기 때문인데(Nie et al., 2009), 실제로 신규 고객의 획득 비용은 기존 고객의 유지비용 대비 5~6배에 달하는 것으로 알려져 있다(Athanassopoulos, 2000). 또한, 이탈 고객은 기업의 안정적인 매출을 급격하게 감소시키기 때문에 기존 매출을 유

* 본 연구는 한성대학교 교내 학술연구비 지원에 의해 수행되었습니다.

지하기 위해서는 일정 수준의 신규 고객을 지속적으로 유치해야 하는 비효율적인 경영구조에 빠지게 한다(Kim and Shin, 2012). 이와는 반대로 고객이탈을 효과적으로 방지하여 고객 유지율을 개선시키는 기업은 회사의 수익성 증대뿐만 아니라, 고객만족도 향상을 통해 브랜드 이미지 개선에도 긍정적인 효과를 가져오는 것으로 알려져 있다(Hung et al., 2006). 즉, 현재와 같은 무한 경쟁 환경 속에서는 고객의 이탈을 효과적으로 예측하고, 방지하는 것이 재무적, 비재무적 관점에서 기업에게 효과적이다.

그 동안의 고객이탈예측 연구는 이동통신업, 금융업, 유통업, 그리고 게임산업 등 경쟁이 치열하고, 이탈관리가 시급한 업종에서 활발히 진행되어 왔으며, 주로 머신러닝과 인공지능 기술을 활용한 예측모형 개발에 초점이 맞추어져 왔다. 이러한 기존의 이탈예측 연구는 단순히 다양한 모형의 성능을 비교하거나(Kim et al., 2002), 이탈예측에 효과적인 변수 (feature)를 탐색하거나(De Bock and Van den Poel, 2012), 새로운 앙상블 기법을 개발하는 것(Lee and Lee, 2006)과 같이 이탈예측 모형 자체의 성능을 개선하는 것에 집중되었고, 대부분의 연구에서 전체 고객집단을 하나의 그룹으로 간주하고 예측모형을 개발했기 때문에 실질적인 활용도 측면에서 한계를 가진다. 실제 비즈니스상에서의 고객들은 이질적인 거래패턴으로 인해 고객의 행태 특성이 다르고, 이에 따른 이탈율도 다르기 때문에 고객 전체를 하나의 고객집단으로 전제하는 것은 무리가 따르기 때문이다(Kim and Shin, 2012). 이처럼 전체 고객을 대상으로 단일 이탈예측 모형이 적용될 경우, 특정 고객 세그먼트의 고객행태 특성이 달라질 때 변화된 고객특성이 이탈예측 모형에 반영되기 어렵기 때문에 이러한 방법론은

개인별 기여가치가 이질적인 업종에서 효과적인 고객이탈예측 성과를 기대하기 어려워진다.

따라서, 고객의 행태특성이 이질적인 업종에서 효과적인 고객이탈 예측을 수행하기 위해서는 충성도와 같은 고객 분류 기준에 따라 고객을 세분화하고, 이를 기반으로 적합한 이탈예측 모형이 개별적으로 운영되는 것이 바람직하다. 물론 일부 연구에서는 클러스터링 기법으로 고객을 세분화하여 개별 고객그룹에 대한 이탈예측 모형을 적용한 연구는 존재한다(Oh et al., 2018). 이러한 이탈예측 프로세스는 전체 고객집단을 대상으로 한 단일 이탈예측 모형에 비해 우수한 예측결과를 가져올 수 있지만, 클러스터링은 투입변수들을 바탕으로 거리를 계산하는 기계적이고, 탐색적인 그룹핑 기법이기 때문에 로열티와 같은 기업의 전략적 의도가 제대로 반영되지 못한다는 점에서 여전히 개선의 여지가 남아있다.

이에 본 연구에서는 비즈니스 머신러닝의 성과가 알고리즘이나 모형 자체의 우수성보다는 전체 머신러닝 프로세스의 우수성에 의해 더 큰 영향을 받는다는 가정을 기반으로 이차원 고객 충성도 세그먼트 기반의 고객이탈예측 프로세스(CCP/2DL: Customer Churn Prediction based on Two-Dimensional Loyalty segmentation)를 제안하고자 한다.

2. 문헌연구

2.1. 머신러닝 기반 CRM 연구

머신러닝은 비즈니스 분야 중 CRM과 같은 고객 데이터 분석 주제에 많이 활용되어 왔으며(Kim et al., 2019), 특히, CRM 분야 중에서도 아

<Table 1> Machine Learning-Based CRM Studies

Category	Research		Key Finding
Customer Segmentation	Single model	Hong and Suh 2004	K-means applied to the customer's psychological statistical data and used discriminant analysis and artificial neural network to predict the customer segmentation.
		Chang 2005	Prove that it is more meaningful to manage only selected customer groups through pre-segmentation.
	Hybrid model	Oh et al., 2018	Suggests how to predict churn based on online role-playing game (MMORPG) user types
Churn Prediction	Single model	Y. Xie and Lee, 2008	Prediction of Customer churn in Unbalanced Data Using Linear Discrete Boosting (LD-Boosting)
		Kawale et al., 2009	Propose a churn prediction model to investigate the social influence of game players and their personal participation in games.
	Hybrid model	Tsai et al., 2009	Development of a new model combining clustering techniques and classifier ensembles for churn prediction.
		Xie et al., 2009	Proposal of Balanced Random Forest (IBRF) Method for Predicting Customer churn
Recommendation Service	Single model	Wen and Zhou, 2012	Proposal of Dynamic Item Cluster-Based Collaborative Filtering Recommendation System.
		Pham et al., 2011	Proposal of clustering approach based on the social information of users to derive the recommendations.
	Hybrid model	Cho and Bang, 2011	Propose a method for the new customer recommendation by using a combined measure based on three well-used centrality measures to identify the customers.

래 <Table 1>에서 정리한 고객세분화, 고객이탈 예측, 그리고 개인화 추천 분야에서의 머신러닝 기반 CRM 연구가 다른 분야들에 비해 활발히 이루어져 왔다. CRM 활동의 출발점이라고 할 수 있는 고객세분화는 동질적인 고객의 특성을 중심으로 고객을 그룹핑하는 활동으로서 고객그룹별 차별화된 CRM 활동의 기반을 제공한다(Chang and Kim, 2018). 고객세분화를 위해 사용된 머신러닝 모형은 의사결정나무와 같은 지도 학습모형이나 SOM(Self-Organizing Map) 또는 K-means 모형과 같은 비지도학습모형이 주로 사용되었다(Chang, 2005). 최근 머신러닝 기반 고

객세분화 연구의 주요 특징 중 하나는 단순히 의미 있는 고객 세그먼트를 도출하는 것에서 끝나는 것이 아니라, 고객이탈예측과 같은 연관된 다른 CRM 연구목적에 위해 사전적인 고객세분화가 수행되고 있다는 점이다(Lee et al., 2001; Tsai et al., 2009; Xie et al., 2009; Kim and Ahn, 2018; Oh et al., 2018).

고객이탈예측 역시 머신러닝 기반의 주요 CRM 연구 주제 중 하나이다(Kim and Shin, 2012). 경쟁이 치열한 현대 경영환경 하에서 고객의 이탈이 증가함에 따라 효과적인 이탈 예측은 CRM뿐만 아니라, 전사적인 경영전략 차원에

서도 매우 중요한 연구 주제로 인식됨에 따라 (Athanasopoulos, 2000), 고객이탈을 성공적으로 예측하기 위한 새로운 모형개발 연구가 많이 이루어져 왔다(Kim and Shin, 2012; Oh et al., 2018). 과거에는 고객의 이탈을 예측하기 위해 의사결정나무, 로지스틱회귀, 인공신경망과 같은 단일 알고리즘을 사용해 모형을 학습하고, 성능을 비교 평가하려는 연구가 주를 이루었다면 (Hung et al., 2006), 최근에는 다양한 알고리즘을 결합한 앙상블 모형이나 단계적으로 이중 모형을 연결한 하이브리드 모형을 개발하려는 시도가 많아졌다(Oh et al., 2018; Tsai and Lu 2009).

한편, 개인화 추천 역시 이탈예측과 더불어 가장 활발하게 진행되어 온 머신러닝 기반 CRM 연구 주제 중 하나이다(Park and Kim 2012; Wen and Zhou, 2012). 과거에는 전형적인 비지도 학습모형인 연관성 분석이나 개별 상품에 대한 구매확률 추정을 통한 개인화 추천 연구가 주를 이루었다면(Cho et al., 2008; Park and Kim 2012), 최근에는 아마존이나 넷플릭스 등의 추천서비스에 적용된 협업필터링 기법을 활용한 개인화 추천 연구가 늘어나고 있는 추세이다. 협업필터링에 의한 개인화 추천 연구도 이탈예측 연구와 마찬가지로 예측 성능 자체를 높이기 위한 모형개발 연구(Pham et al., 2011; Wen and Zhou, 2012)가 주를 이루었으나, 최근에는 협업 필터링에 내재된 한계점을 극복하기 위해 다양한 보조적인 처리기술을 결합한 하이브리드 협업 필터링 방법론에 대한 연구가 활발해지고 있다(Cho and Bang., 2011; Joe and Nam 2017).

2.2. 고객이탈예측

고객이탈의 예측과 방지는 로열티 경영에 있

어서 늘 핵심적인 이슈로 연구되어 왔다(Kim and Shin, 2012). 로열티 (loyalty) 즉, 충성도는 고객들의 구매행태에 직접 관련된 행동적 로열티 (behavioral loyalty)와 구매 외적인 행태와 심리적 애착에 관련된 태도적 로열티 (attitudinal loyalty)로 구분할 수 있으며, 이 두 가지 로열티의 이차원적 배치에 따라 고객들의 이탈위험도 다르게 나타난다(Kim and Kim, 2015). 즉, 행동적 로열티나 태도적 로열티가 각각 높아질수록 고객들의 이탈위험도 역시 낮아지고, 두 가지 로열티가 모두 높은 경우 고객이탈율의 확연한 감소를 보인다. 따라서, 두 가지 특성의 로열티를 모두 고려하여 동질적인 그룹별 이탈관리가 이루어진다면 보다 효율적인 이탈관리가 가능하다 (Kim and Kim, 2015).

한편, 유통업종과 같은 비계약업종은 계약업종이나 구독 비즈니스처럼 명시적인 계약관계가 존재하지 않는 자율거래 업종이기 때문에 고객 이탈이라는 명확한 조작적 정의가 존재하지 않으며, 이는 이탈이라는 명확한 목표변수가 존재하지 않으므로 원칙적으로 지도학습 머신러닝 기법으로 해결하기 어렵다(Kim and Shin, 2012). 이 경우 고객의 구매패턴을 모델링하여 자신의 구매패턴에서 일정 수준 벗어난 지점을 이탈이라고 간주하는 상대적 이탈예측 모형을 개발해야 한다(Kim and Shin, 2012; Oh et al., 2018).

고객이탈예측 연구는 연구의 목적 관점에서 고객이탈을 정의하는 모형을 제시하거나 (Kim and Shin, 2012), 이탈을 예측하는 모형을 개발하거나(Nath and Behara, 2003; Kraljević and Gotovac, 2010), 고객이탈에 대한 인과관계를 연구하거나(Han and Maeng, 2004), 고객이탈 방지를 위한 전략을 제시하는 연구(Kim et al., 2002) 등으로 구분할 수 있다. 또한, 연구방법론 관점

에서 고객이탈 관련 연구는 가설검증을 통한 실증연구보다는 머신러닝 기법이나 수리적 모델을 활용한 고객이탈 예측 연구가 주를 이루고 있다 (Xie and Li, 2008; Kawale et al., 2009). 이러한 머신러닝 기반의 고객이탈예측 연구는 대부분 이탈예측의 정확도를 높이고자 하는 공통적인 목적을 가지고 있으나, 크게 두 가지의 연구 방향으로 구분된다. 첫 번째 연구 방향은 다양한 머신러닝 알고리즘을 결합한 하이브리드 모델 (또는 앙상블 모델)을 활용하여 고성능의 이탈예측 모형을 구축하려는 것이고(Oh et al., 2018), 두 번째 연구 방향은 분석에 사용되는 데이터의 불균형 문제를 개선하거나, 예측성능에 영향을 미치는 아웃라이어(Outlier)를 효과적으로 제거하는 등 모형 개발 이외의 프로세스적인 차원의 개선을 이루려고 하는 연구이다(Tsai and Lu 2009; Xie et al., 2009).

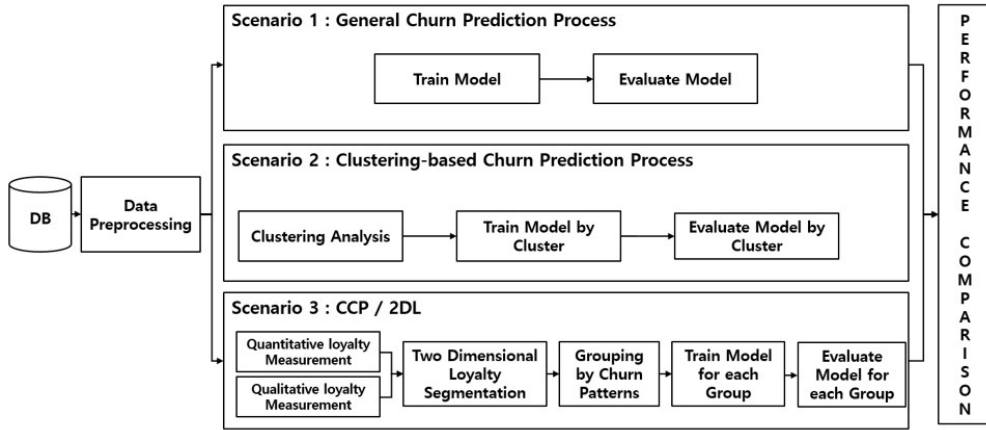
즉, 머신러닝 기반의 고객이탈예측 연구들은 대부분 예측모형의 개발이나 데이터 전처리와 같은 기술적이고, 지엽적인 문제에 집중되어 온 것이 사실이다. 이러한 기술적인 연구 역시 점차 인공지능화 되는 고객경영의 발전에 반드시 필요한 노력이지만, Python이나 R과 같은 오픈소스를 통해 우수한 머신러닝 기법들을 누구나 사용할 수 있는 상황에서 기술 중심적인 연구의 성과는 자칫 분석 패키지나 함수의 기능에 의존적일 수 있다는 한계가 존재한다. 따라서, 기술적인 모형개발과 더불어 모형을 개발하는 전반적인 프로세스를 전략적인 관점에서 개선하려는 노력 역시 필요한 시점이다.

3. 실험 및 평가

3.1. 모형의 개요

본 연구는 문헌연구 파트에서 기술한 기존 이탈예측 연구의 한계를 극복하고자 이차원 고객충성도 세그먼트 기반의 고객이탈예측 프로세스 (CCP/2DL: Customer Churn Prediction based on Two-Dimensional Loyalty segmentation)를 제안하고자 한다. CCP/2DL 프로세스는 고객의 양적, 질적 로열티를 모델링하여 2차원 고객세분화를 수행한 후 도출된 여러 고객 세그먼트를 고객이탈율에 따라 소수의 그룹으로 다시 그룹핑하고, 각 그룹별 최적의 이탈예측 모형을 개별적으로 적용하여 전체 고객에 대한 이탈예측 최적화를 이루고자 하는 방법론이다. 제안 모형은 일반적인 범용이탈예측 프로세스, 클러스터링 기반 이탈예측 프로세스와 함께 성능 비교를 통해 제안 모형의 타당성을 입증하고자 하였으며 각각은 Scenario 1, 2, 3으로 구분하여 진행하였다. 3가지 이탈예측 프로세스에 사용된 머신러닝 알고리즘은 의사결정나무, 로지스틱 회귀, 랜덤포레스트, 그라디언트 부스팅, 에다부스트, 스택킹 총 6가지이다.

<Figure 1>은 이러한 연구의 전반적인 흐름을 보여주고 있다. Scenario 1은 전체 고객을 하나의 집단으로 간주하고, 학습 데이터 세트와 테스트 데이터 세트로 구분하여 모형을 학습하고, 평가하는 범용이탈예측 프로세스이다(Hung et al., 2006). 이 시나리오 상에서 최적의 모형은 앞서 기술한 6가지 머신러닝 알고리즘을 모두 적용하여 가장 우수한 성능을 보이는 모형을 선택한다. Scenario 2는 의미 있는 고객행태 변수들을 기반으로 탐색적 군집분석을 먼저 시행하여 적절한



〈Figure 1〉 Research Model

군집 수로 고객을 세분화하고, 각 군집 별로 6개의 예측 알고리즘을 적용하여 최적의 모형을 선택하는 클러스터링 기반 이탈예측 프로세스이다 (Oh et al., 2018). 끝으로 Scenario 3는 앞서 설명한 CCP/2DL 프로세스를 의미하며, 다른 예측 프로세스와 마찬가지로 6가지 예측 알고리즘을 적용하여 각 이탈패턴 그룹별 최적의 모형을 선택하였다. 이탈예측 성능을 평가하기 위해 정확도 (Accuracy), 정밀도(Precision), 재현율(Recall), 특이도(Specificity)와 더불어 정밀도와 재현율의 조화평균으로 계산하는 F1-score를 함께 검토하였다.

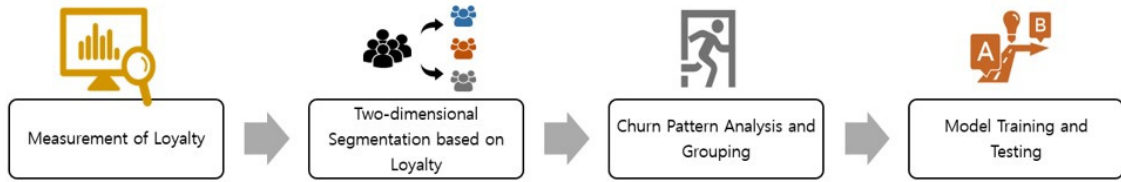
3.2. 데이터 및 실험설계

본 연구를 위해 글로벌 NGO 단체 중 하나인 A사의 지원으로 2019년 3월 31일 시점의 472,007명의 전체 후원자와 2017년 4월 ~ 2019년 3월 기간 동안에 발생한 69,068명의 해지자 데이터를 개인 식별정보를 제거하고 사용하였다. 고객 데이터에 포함된 200여개의 기본 변수 및 파

생 변수 중에서 이탈예측에 유의미하다고 판단되는 98개의 변수를 선별하였고, 전체 데이터의 70%를 모형 학습에, 나머지 30%는 모형 검증에 사용하였다. 데이터전처리 과정을 통해 범주형 데이터 변환과 이상치 데이터 처리를 시행하였고, 데이터 불균형으로 인한 모형의 성능 감소 문제를 해결하기 위해 오버샘플링 기법을 적용하였다. 본 연구에서 사용한 A사 고객 데이터의 전반적인 구성은 <Table 2>와 같다.

〈Table 2〉 Customer Data

	Label	Number	Rate	Total
Gender	Male	294,795	54.48%	541,075
	Female	200,248	37.01%	
	Uncertain	46,032	8.51%	
Age	20s	91,499	16.91%	541,075
	30s	98,686	18.24%	
	40s	158,465	29.29%	
	50s	118,428	21.89%	
	60s	53,348	9.86%	
	70s +	20,649	3.82%	
Churn	Not Churned	472,007	87.2%	541,075
	Churned	69,068	12.8%	



<Figure 2> CCP/2DL Processes

<Table 3> Measurements of Customer Loyalty

Quantitative Loyalty	Longevity (months), Total Donation during recent 1 year (won)
Qualitative Loyalty	Donation at one time(Y/N), No. of Inbound Interaction (#), No. of Outbound Interaction (#), Intimacy of Information Receiving (score), etc.

한편, 클러스터링 기반 이탈예측 프로세스에서 군집분석을 위해서는 나이나 후원금액, 총 후원기간과 같은 기본적인 거래 이력정보 외에도 커뮤니케이션 채널 수, 인바운드 커뮤니케이션 횟수, 아웃바운드 커뮤니케이션 횟수 등과 같은 마케팅 행태 변수도 추가하여 군집분석에 사용하였다. 우선 계층적 군집분석을 통해 최적의 군집 수를 3개로 결정한 뒤 전체 후원자에 대한 K-평균 군집화 알고리즘을 실시하였다. 도출된 3개의 고객 세그먼트별로 의사결정나무, 로지스틱 회귀, 랜덤포레스트, 그래디언트 부스팅, 에다부스트, 스택킹 등 6가지 모형을 모두 적용하여 가장 우수한 성능을 보이는 모형을 개별적으로 선택하였다. 본 연구의 제안 모형인 Scenario 3의 CCP/2DL은 <Figure 2>와 같이 4단계의 프로세스로 구현되었다.

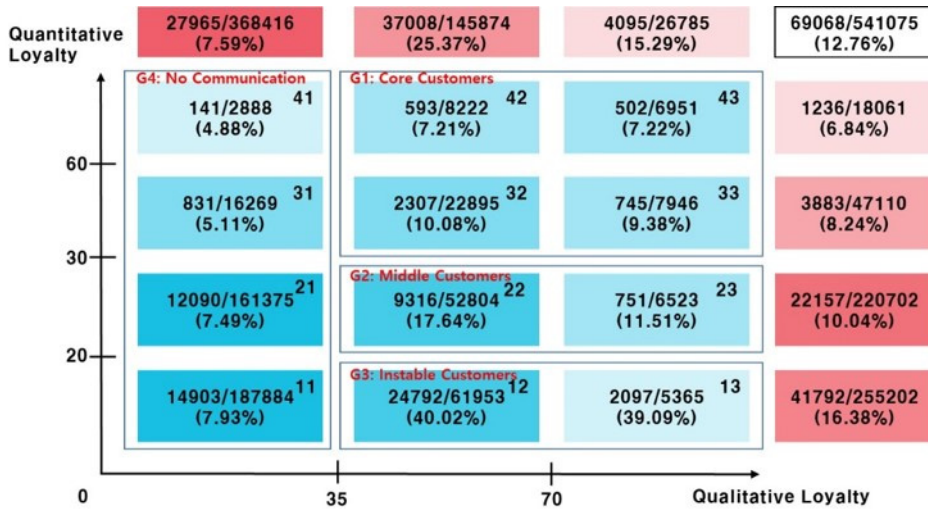
Step 1: 로열티의 측정

본 연구에서는 고객 로열티를 양적 로열티와 질적 로열티로 구분하고, A사와의 협의를 통해 적절한 후보 변수를 선정하여 두 가지 차원의 로

열티를 측정하였다. 양적 로열티와 질적 로열티의 측정변수는 <Table 3>와 같이 2개 이상의 고객행태 변수를 가중 결합하여 사용했으며, 고객별 양적 로열티 점수를 종속변수로, 질적 로열티 측정변수들을 독립변수로 하는 다중 회귀분석을 시행하고, 각 독립변수의 회귀계수를 바탕으로 가중치로 도출하여 결국 후원자 별 하나의 질적 로열티 값을 생성하였다.

Step 2: 로열티 기반 이차원 세분화

양적/질적 로열티 변수 각각의 분할 기준점을 설정하기 위해 계층적 군집분석과 K-평균 군집화 알고리즘을 사용하였다. 구체적인 방법을 살펴보면 5,000 명을 무작위 추출하여 계층적 군집분석을 수행해 각 로열티 차원의 최적 군집 수를 파악하고, 전체 후원자에 K-평균 군집화 알고리즘을 시행하여 군집 간의 분할 기준점을 도출하였다. 분석 결과, 양적 로열티와 질적 로열티에 대해 각각 4개와 3개의 세그먼트로 분할되어 총 12개의 고객 세그먼트가 생성되었다.



〈Figure 3〉 Two-dimensional Segmentation and Churn Pattern Grouping

Churn Score	Churn Alerts		Customer		Churned		Churn Ratio in Alert Level
			#	%	#	%	
95~	●	Very Dangerous	4,481	2.77%	4,406	21.72%	98.33%
80~95	●	Dangerous	9,727	6.02%	7,957	39.23%	81.80%
50~80	●	Cautious	18,123	11.21%	4,236	20.88%	23.37%
20~50	●	Need Attention	51,601	31.92%	2,922	14.41%	5.66%
~20	●	Safe	77,748	48.09%	763	3.76%	0.98%
Total			161,680	100%	20,284	100%	12.55%

〈Figure 4〉 Validation Test Result for the Selected Model

Step 3: 이탈패턴 분석 및 그룹화

총 12개의 로열티 세그먼트에 대해 이탈패턴으로 재그룹화를 하기 위해 각 세그먼트의 고객 이탈율을 계산하고, 12개의 세그먼트 중 유사한 이탈율을 보이는 세그먼트들을 하나의 그룹으로 병합시켜 총 4개의 고객그룹으로 분류하였다.

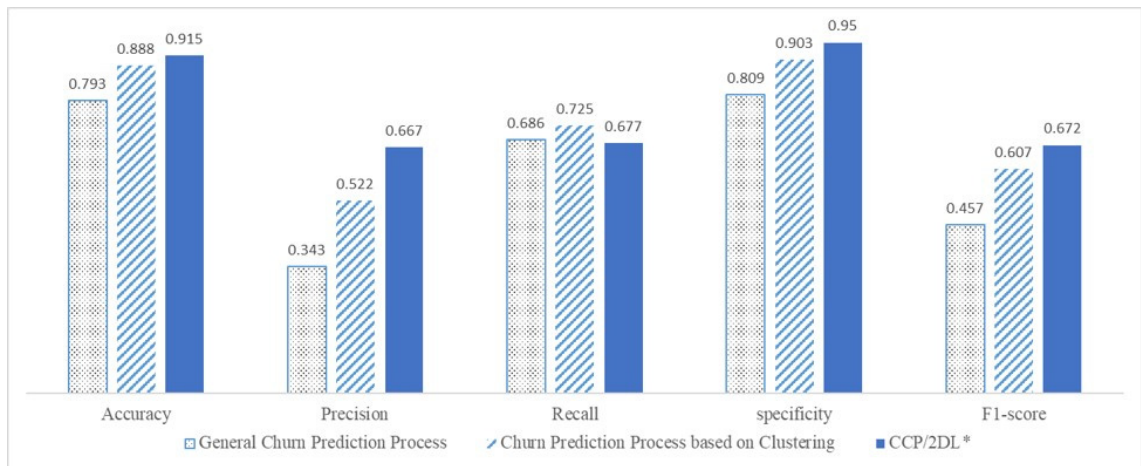
Step 4: 모형 학습 및 평가

이탈패턴 그룹화를 통해 산출된 4개의 그룹에

대해 각각 이탈 예측모형을 개발하고, 평가하였다. 모든 그룹에 대해 상기 기술한 6개의 예측모형을 모두 적용하였고, 각 그룹별로 가장 우수한 성능을 보이는 모형을 선택하였다. 최종적으로 선정된 이탈예측 모형에 의해 구간별 이탈위험 등급을 설정하고, 실제 이탈율을 재평가한 결과 이탈위험등급이 높을수록 실제 이탈율도 높아져 이탈의 조기경보 시그널로 상당한 효과가 있다는 것을 알 수 있다.

<Table 4> Research Result

Scenario	Accuracy	Precision	Recall	Specificity	F1-score
1. General Churn Prediction Process	0.793	0.343	0.686	0.809	0.457
2. Clustering-based Churn Prediction Process	0.888	0.522	0.725	0.903	0.607
3. CCP / 2DL*	0.915	0.667	0.677	0.950	0.672



<Figure 5> Performances between Predictive Processes

3.3. 경쟁모형 비교

범용이탈예측 프로세스의 최종 선택 모형은 랜덤 포레스트였으며, 클러스터링 기반 이탈예측 프로세스는 2개의 군집에서 랜덤 포레스트가, 나머지 한 개의 군집에서 에다부스트가 선택되었다. 이차원 고객충성도 세그먼트 기반 이탈예측 프로세스의 경우 3개의 고객그룹에서 랜덤 포레스트가, 나머지 한 개 군집에서 그래디언트 부스팅 모형이 선택되었다. 각 시나리오 별 이탈예측 성과를 평가한 결과 CCP/2DL, 클러스터링 기반 이탈예측, 그리고 범용이탈예측 프로세스의 순으로 예측 성과가 뛰어난 것으로 나타났다. CCP/2DL 모형은 5가지 모형평가지표 중 재현율 (Recall)을 제외한 4가지 지표에서 다른 이탈예측

방법론에 비해 우수한 성능을 보였으며, 정밀도와 F1-score의 경우 비교적 큰 차이를 보였다. 이러한 분석결과는 <Table 4>와 <Figure 5>에서 제시하고 있다.

4. 결론

4.1. 연구요약

본 연구에서는 효과적인 고객이탈 예측을 위해 고객충성도 기반의 이차원 세분화와 이탈패턴 분석을 활용하는 새로운 이탈예측 프로세스 모델을 제안하였다. CCP/2DL이라고 명명한 이 프로세스는 고객의 양적 로열티와 질적 로열티

로 기준으로 세분화한 12개의 고객 세그먼트를 도출하고, 유사한 이탈패턴에 따라 다시 4개의 고객그룹을 생성하여 그룹 별 최적의 예측 모형을 적용한다. 제안 모형은 기존의 범용이탈예측 프로세스 및 클러스터링 기반 이탈예측 프로세스와 비교한 결과 상대적으로 우수한 예측성과를 보이는 것으로 나타났다.

4.2. 연구의 의미

4.2.1. 연구의 학술적 의미

본 연구의 학술적 의미로는 첫째, 전체 이탈예측 프로세스의 개선이 예측 모형의 개별적인 최적화보다 전체 예측 성능에 더 큰 영향을 미친다는 것을 검증했다는 점이다. 기술적인 관점에서 고객그룹에 적용한 예측 알고리즘과 모형의 학습방식이 동일하기 때문에 본 연구의 실험설계는 프로세스의 차별화를 관찰하고자 하는 방향으로 구성되었다고 볼 수 있다. 둘째, 하나의 전체 고객그룹에 대한 이탈예측 보다는 세분화된 고객그룹에 대한 개별적인 이탈예측이 더 유효하다는 기존 연구의 결과를 재확인하였고, 더 나아가 단순히 기술적인 통계분석에 의한 그룹화 보다는 고객로열티와 이탈패턴과 같은 전략적인 기준에 의해 분류된 고객세분화가 이탈예측에 있어 더 효과적이라는 것을 제시하였다는 점이다. 셋째, 고객 충성도 관련 연구에서 개념적으로 중요시 되어왔던 양적 로열티와 질적 로열티의 개념을 정량적인 측정변수로 모델링하고, 이를 고객세분화에 직접 적용했다는 점이다.

4.2.2. 연구의 실무적 의미

본 연구의 실무적 의미로는 첫째, 본 연구의

결과를 사용하여 기업에서는 보다 효율, 효과적인 고객이탈관리를 수행할 수 있다는 점이다. 실제로 본 연구에 참여한 NGO A사는 본 연구결과를 파이썬 모듈로 개발하여 현업에서 후원자 이탈방지에 직접 사용하고 있다. 둘째, 본 연구에서 CCP/2DL과 같은 세분화된 이탈예측 시스템이 운영될 경우 여타 모형에 비해 모형의 유지보수와 전체 최적화에 더 유리하다는 점이다. 전체 고객에 대해 단일 이탈예측모형이 적용될 경우 특정 고객 세그먼트의 행태변화로 인해 전체 이탈예측 모형이 변경되어야 하므로 모형 변경에 따른 리스크가 매우 크다. 셋째, 이탈예측 연구가 상대적으로 일천했던 NGO 업종에 과학적인 CRM 기법을 적용했다는 점이다. NGO의 시장이 점점 커지지만, 경쟁 NGO가 급증하는 현 시점에서 NGO 역시 후원자를 ‘고객 (Customer)’으로 인식해야 할 것이며, 이러한 관점에서 NGO에 대한 CRM 전략은 향후 더 중요해질 것으로 판단한다.

4.3. 연구의 한계점과 향후 연구방향

상기 기술한 연구의 의미와 더불어 본 연구는 몇 가지 연구의 한계점이 존재한다. 우선 본 연구에서 제시한 양적, 질적 로열티 변수는 신뢰성과 타당성 분석이 이루어지지 않았다. 물론 두 가지 로열티 변수의 생성이 가설검증을 위한 개념화나 척도 개발이 아니라, 고객세분화에 사용하기 위한 목적이므로 세분화 변수의 적합성 기준인 동질성(Homogeneity), 이질성(Heterogeneity), 크기(Substantial)을 충분히 고려하여 세분화 분류변수를 위한 기술적 타당성은 확보되었다고 볼 수 있다(Chang and Kim 2018). 그러나, 측정변수와의 연관 개발을 위한 전형적인 통계 검증 절차를 거

치지 않았으므로 본 연구에서 사용한 두 가지 로열티 변수는 이론적으로 검증된 개념이라고 할 수 없다. 따라서, 향후에는 양적 로열티와 질적 로열티에 대한 이론적 접근과 더불어 신뢰할 수 있는 측정변수로서의 활용을 위해 별개의 신뢰성 및 타당성 검증 연구가 수행되는 것이 바람직하다. 둘째, 본 연구에 적용된 업종은 NGO로써 계약 업종이기 때문에 유통업과 같은 비계약 업종에는 본 연구의 결과를 그대로 활용하기 어렵다는 점이다. 비계약 업종의 경우 계약의 해지와 달리 명시적인 이탈의 개념이 존재하지 않기 때문에 상이한 고객 구매패턴에 따라 상대적인 이탈기준을 결정하고, 각 이탈수준에 대한 위험도를 측정하는 예측모델을 함수적으로 구현해야 한다. 따라서, 본 연구의 CCP/2DL 프로세스를 적용하되, 본 연구에서처럼 지도학습 머신러닝 모형이 아니라 거래패턴 기반의 이탈예측 모형을 적용하는 연구(Kim and Shin, 2012)가 이루어지는 것이 필요하다.

참고문헌(References)

- Athanassopoulos, A. D., "Customer satisfaction cues to support market segmentation and explain switching behavior," *Journal of business research*, Vol.47, No.3(2000), 191-207.
- Chang, M. S. and H. J. Kim, "A Customer Segmentation Scheme Base on Big Data in a Bank," *Journal of Digital Contents Society*, Vol.19, No.1(2018), 85-91.
- Chang, N. S., "Improving the effectiveness of customer classification models: a pre-segmentation approach," *Information Systems Review*, Vol.7, No.2(2005), 23-40.
- Cho, Y. and J. Bang, "Applying centrality analysis to solve the cold-start and sparsity problems in collaborative filtering," *Journal of Intelligence and Information Systems*, Vol.17, No.3(2011), 99-114.
- Cho, Y. S., M. H. Huh, and K. H. Ryu, "Implementation of Personalized Recommendation System using RFM method in Mobile Internet Environment," *Journal of The Korea Society of Computer and Information*, Vol.13, No.2(2008), 41-50.
- De Bock, K. W. and D. Van den Poel, "Reconciling performance and interpretability in customer churn prediction using ensemble learning based on generalized additive models," *Expert Systems with Applications*, Vol.39, No.8(2012), 6816-6826.
- Han, S. L. and A. Maeng, "Effects of Switching Barrier of Mobile Telecommunication Service on Customer Retention And Churn-out," *Proceedings of the Korea Distribution Association Conference*, (2004), 191-197.
- Hong, T. and B. Suh, "Data Mining for Personalization Model Using Customer Belief under the Internet Banking Environment," *The Journal of Internet Electronic Commerce Research*, Vol.4, No.2(2004), 101-115.
- Hung, C. and C. F. Tsai, "Market segmentation based on hierarchical self-organizing map for markets of multimedia on demand," *Expert systems with applications*, Vol.34, No.1(2008), 780-787.
- Hung, S. Y., D. C. Yen, and H. Y. Wang, "Applying data mining to telecom churn management," *Expert Systems with Applications*, Vol.31, No.3(2006), 515-524.
- Joe, D. Y and K. Nam, "SKU recommender system for retail stores that carry identical

- brands using collaborative filtering and hybrid filtering,” *Journal of Intelligence and Information Systems*, Vol.23, No.4(2017), 77-110.
- Kawale, J., A. Pal, and J. Srivastava, “Churn prediction in MMORPGs: A social influence based approach,” *2009 International Conference on Computational Science and Engineering*, Vol.4, (2009), 423-428
- Kim, C. N., N. S. Chang, and J. W. Kim, “A Study on the Analysis of Comparison of Churn Prediction Models in Mobile Telecommunication Services,” *Asia Pacific Journal of Information Systems*, Vol.12, No.1 (2002), 139-158.
- Kim, D. H. and K. K. Ahn, “A Study on the Customer Segmentation Using Machine Learning,” *The Society of Convergence Knowledge Transactions*, Vol.6, No.2(2018), 115-120.
- Kim, H. S. and D. Y. Shin, “Developing a customer defection model based on relative defection criteria for non-contractual businesses,” *Korean Journal of Marketing*, Vol.27, No.3 (2012), 117-144.
- Kim, H. S., Y. Bak, and J. Lee, “A Personalized Recommendation System Using Machine Learning for Performing Arts Genre,” *Information Systems Review*, Vol.21, No.4(2019), 31-45.
- Kim, H.S. and Y.G. Kim, “*CRM Strategy: Principles and Applications*,” Young Publication, 2015.
- Kraljević, G. and S. Gotovac, “Modeling data mining applications for prediction of prepaid churn in telecommunication services,” *Automatika*, Vol.51, No.3(2010), 275-283.
- Lee, J. S. and J. C. Lee, “Customer Churn Prediction of Automobile Insurance by Multiple Models,” *Journal of Intelligence and Information Systems*, Vol.12, No.2(2006), 167-183.
- Lee, K. C., S. J. Kwon, and K. S. Shin, “Analysis of Defection Customer Using Customer Segmentation on Bank,” *Journal of Intelligence and Information Systems*, Vol.7, No.1(2001), 177-196.
- Nath, S. V. and R. S. Behara, “Customer churn analysis in the wireless industry: A data mining approach,” *Proceedings-annual meeting of the decision sciences institute*, (2003), Vol. 561, 505-510.
- Nie, G., G. Wang, P. Zhang, Y. Tian, and Y. Shi, “Finding the hidden pattern of credit card holder’s churn: A case of china,” *Proceedings of Springer International Conference on Computational Science*, (2009), 561-569.
- Oh, S., E. Lee, J. Woo, and H. K. Kim, “Constructing and Evaluating a Churn Prediction Model using Classification of User Types in MMORPG,” *KIISE Transactions on Computing Practices*, Vol.24, No.5(2018), 220-226.
- Park, S. H. and H. S. Kim, “Design of a Diversity-Based Recommender System for Providing Anti-Churning Rules,” *Korea Intelligent Information System Society*, Vol.11, No.3 (2012), 101-112.
- Pham, M. C., Y. Cao, R. Klamka, and M. Jarke, “A clustering approach for collaborative filtering recommendation using social network analysis,” *J. UCS*, Vol.17, No.4(2011), 583-604.
- Tsai, C. F. and Y. H. Lu, “Customer churn prediction by hybrid neural networks,” *Expert Systems with Applications*, Vol.36, No.10

- (2009), 12547-12553.
- Wen, J. and W. Zhou, "An improved item-based collaborative filtering algorithm based on clustering method," *Journal of Computational Information Systems*, Vol.8, No.2(2012), 571-578.
- Xie, Y. and X. Li, "Churn prediction with linear discriminant boosting algorithm," *Proceedings of IEEE International Conference on Machine Learning and Cybernetics*, (2008), 228-233.
- Xie, Y., X. Li, E. W. T. Ngai, and W. Ying, "Customer churn prediction using improved balanced random forests," *Expert Systems with Applications*, Vol.36, No.3(2009), 5445-5449.

Abstract

A Methodology of Customer Churn Prediction based on Two-Dimensional Loyalty Segmentation

Hyung Su Kim* · Seung Woo Hong**

Most industries have recently become aware of the importance of customer lifetime value as they are exposed to a competitive environment. As a result, preventing customers from churn is becoming a more important business issue than securing new customers. This is because maintaining churn customers is far more economical than securing new customers, and in fact, the acquisition cost of new customers is known to be five to six times higher than the maintenance cost of churn customers. Also, Companies that effectively prevent customer churn and improve customer retention rates are known to have a positive effect on not only increasing the company's profitability but also improving its brand image by improving customer satisfaction.

Predicting customer churn, which had been conducted as a sub-research area for CRM, has recently become more important as a big data-based performance marketing theme due to the development of business machine learning technology. Until now, research on customer churn prediction has been carried out actively in such sectors as the mobile telecommunication industry, the financial industry, the distribution industry, and the game industry, which are highly competitive and urgent to manage churn. In addition, These churn prediction studies were focused on improving the performance of the churn prediction model itself, such as simply comparing the performance of various models, exploring features that are effective in forecasting departures, or developing new ensemble techniques, and were limited in terms of practical utilization because most studies considered the entire customer group as a group and developed a predictive model. As such, the main purpose of the existing related research was to improve the performance of the predictive model itself, and there was a relatively lack of research to improve the overall customer churn prediction process. In fact, customers in the business have different behavior characteristics due to

* Corresponding Author: Hyung Su Kim

Professor: Division of Smart Management Engineering, Hansung University
116, Samseongyo-ro 16-gil, Seongbuk-gu, Seoul, Republic of Korea
Tel: +82-2-760-8035, E-mail: hskim@hansung.ac.kr

** Division of Industrial Management Engineering, Hansung University

heterogeneous transaction patterns, and the resulting churn rate is different, so it is unreasonable to assume the entire customer as a single customer group. Therefore, it is desirable to segment customers according to customer classification criteria, such as loyalty, and to operate an appropriate churn prediction model individually, in order to carry out effective customer churn predictions in heterogeneous industries. Of course, in some studies, there are studies in which customers are subdivided using clustering techniques and applied a churn prediction model for individual customer groups. Although this process of predicting churn can produce better predictions than a single predict model for the entire customer population, there is still room for improvement in that clustering is a mechanical, exploratory grouping technique that calculates distances based on inputs and does not reflect the strategic intent of an entity such as loyalties.

This study proposes a segment-based customer departure prediction process (CCP/2DL: Customer Churn Prediction based on Two-Dimensional Loyalty segmentation) based on two-dimensional customer loyalty, assuming that successful customer churn management can be better done through improvements in the overall process than through the performance of the model itself. CCP/2DL is a series of churn prediction processes that segment two-way, quantitative and qualitative loyalty-based customer, conduct secondary grouping of customer segments according to churn patterns, and then independently apply heterogeneous churn prediction models for each churn pattern group. Performance comparisons were performed with the most commonly applied the General churn prediction process and the Clustering-based churn prediction process to assess the relative excellence of the proposed churn prediction process. The General churn prediction process used in this study refers to the process of predicting a single group of customers simply intended to be predicted as a machine learning model, using the most commonly used churn predicting method. And the Clustering-based churn prediction process is a method of first using clustering techniques to segment customers and implement a churn prediction model for each individual group.

In cooperation with a global NGO, the proposed CCP/2DL performance showed better performance than other methodologies for predicting churn. This churn prediction process is not only effective in predicting churn, but can also be a strategic basis for obtaining a variety of customer observations and carrying out other related performance marketing activities.

Key Words : Customer Churn Predictive Model, Customer Relationship Management, Customer Loyalty, Customer Big data, CCP/2DL

Received : June 29, 2020 Revised : December 10, 2020 Accepted : December 16, 2020

Corresponding Author : Hyung Su Kim

저 자 소개



김형수

현재 한성대학교 스마트경영공학부에서 교수로 재직 중이며, 대학 내에서 CRM 디지털 마케팅 부전공 프로그램의 책임교수를 맡고 있다. 또한, 빅 데이터 및 CRM 산학협력 벤처기업인 고객경영기술원의 원장과 최고연구개발책임자를 겸임하고 있다. 한국과학기술원에서 CRM으로 경영공학 박사학위를 받았으며, 한국CRM협회 회장과 한국CRM학회의 부회장을 역임한 바 있다. 주요 관심분야는 빅 데이터 분석, 고객행태예측, 멤버십 프로그램, 비즈니스 머신러닝 모델 개발 등이다.



홍승우

한성대학교 산업경영공학과 석사과정 중이며 동 대학 영문학과에서 문학사 학위와 CRM·디지털마케팅 연계전공에서 CRM 학사 학위를 취득하였다. SK, 네오팜, 월드비전 등 빅 데이터 및 CRM 프로젝트를 수행한 바 있으며, 주요 관심분야는 머신러닝, CRM, 빅데이터, 고객이탈예측, 로열티 프로그램 등이다.