

고객유지를 위한 접촉스케줄링시스템의 설계*

이재식

아주대학교 경영대학 e-비즈니스학부 교수
(leejsk@ajou.ac.kr)

조유정

아주대학교 대학원 경영학과
(mispower@empal.com)

.....

고객유지는 갈수록 경쟁이 심화되고 있는 생명보험산업에서 핵심이슈 중에 하나이다. 생명보험사들은 고객을 유지하기 위해서 많은 활동들을 한다. 그 가운데 대표적인 것이 바로 지속적으로 고객과 접촉하는 것이다. 본 연구는 접촉스케줄링시스템(CSS: Contact Scheduling System)의 설계에 대한 것으로 해축된 모집설계사의 고객을 관리해야만 하는 지원설계사의 고객관리를 돕는 것을 목표로 한다. 지원설계사는 모집설계사의 고객관리 경험과 지식을 공유할 수 없다. 이런 지원설계사의 고객접촉을 지원하기 위해서, 본 연구에서는 CSS를 설계한다. CSS 설계는 두 단계로 이루어지고, CART(Classification And Regression Tree)와 SPM(Sequential Pattern Mining)의 데이터 마이닝 기법을 활용한다. 단계 1에서는 CART 기법을 이용하여 고객을 8개의 고객군으로 분류한다. 단계 2에서는 각 분류고객군에 적합한 접촉내용, 접촉간격 그리고 접촉방법 등의 접촉스케줄링 정보를 생성한다. 접촉내용은 스케줄 접촉내용, 이벤트접촉내용 그리고 비즈니스규칙에 의한 접촉내용의 결합으로 결정되는데 스케줄접촉내용은 SPM 모델의 결과를 통해 생성된다. 또한 본 연구에서 설계한 CSS가 실제상황에서 어떻게 작동하는지를 제시함으로써, CSS가 효율적이고 효과적인 고객접촉에 실용적임을 보인다.

.....

논문접수일 : 2005년 7월

게재 확정일 : 2005년 12월

교신저자 : 조유정

1. 서론

국내의 생명보험산업은 경쟁의 새로운 국면을 맞이하고 있다. 생명보험과 손해보험의 구분이 명확하던 과거와는 달리 생명보험과 손해보험의 구분이 없어지고, 방카슈랑스가 도입됨에 따라 산업간 경계 또한 사라지고 있다. 이에 생명보험사들은 더 이상 상품과 서비스에 대해 자신들만의 고유영역을 보호받을 수 없게 되었고, 이에 대한 대안으로 고객관계관리에 더 많은 관심을 기울이게 되었다.

생명보험사에서 고객관계관리를 직접 실천하는 사람은 바로 설계사이다. 잠재고객의 탐색에서부터 상품계약 또는 고객모집 그리고 상품계약이후 관리까지 전 과정을 직접 관할하는 설계사를 모집설계사라고 한다. 이 모집설계사는 직접 고객을 모집하면서 고객과 개별적인 접촉경험을 하고 각 고객에게 적합한 접촉 노하우 즉, 지식을 축적한다. 하지만 이런 모집설계사의 해축이 빈번해지면서 해축된 모집설계사의 관리고객들에 대한 관리 또한 소홀지고 있다. 이는 고객의 만족도 및 로열티를 감소시켜서 고객이탈을 유발하는 원인이 된다.

* 본 연구는 21세기 프론티어 연구개발 사업의 일환으로 추진되고 있는 정보통신부의 유비쿼터스 컴퓨팅 및 네트워크원천 기반기술개발사업의 지원에 의한 것임.

이러한 현상을 방지하기 위해서 생명보험사는 해촉된 모집설계사의 관리고객만을 관리하는 별도의 설계사를 지정하고 있는데 이를 지원설계사라고 한다.

이 지원설계사는 모집설계사가 가진 고객에 대한 경험과 지식에 접근할 수 없기 때문에 관리고객을 접촉하는데 어려움을 겪는다. 실제 지원설계사의 고객접촉현황을 보면 관리대상고객의 53%만이 접촉되고 있음을 확인할 수 있다. 이에 모집설계사의 해촉과 동시에 그들의 경험과 지식이 사라지는 것과 같은 문제를 해결하기 위해서 지원설계사들은 자신들의 접촉노하우를 데이터로 축적하였고, 이 데이터를 접촉관리가 이루어지지 않고 있는 고객에게 활용할 필요성이 대두되었다.

이에 본 연구에서는 지원설계사가 당면한 문제들을 해결하고자 Classification And Regression Tree(CART)와 순차패턴마이닝(SPM: Sequential Pattern Mining) 등의 데이터 마이닝 기법들을 활용하여 접촉스케줄링시스템(CSS: Contact Scheduling System)을 설계한다. CSS는 지원설계사에게 고객접촉에 대하여 무엇을 위하여, 어떻게 그리고 언제 해야 하는지에 대한 정보를 제공한다. 이를 통해 지원설계사는 고객에게 효율적이고 효과적인 접촉을 수행할 수 있게 되고, 이것은 접촉관리 없이 방치되던 고객들과의 관계를 개선시킬 뿐만 아니라 그들의 만족도를 향상시킬 수 있게 된다.

본 연구는 고객의 미래행동 예측에 초점이 맞추어져 있는 것이 아니라 지원설계사의 현재 미접촉 고객 및 미래접촉고객에게 적용할 수 있는 과거접촉패턴을 찾아서 활용하는 것에 초점이 맞추어져 있다.

본 연구는 다음과 같이 구성되어 있다. 제 2절에서는 보험산업에서 데이터 마이닝의 활용에 대한 기존연구를 살펴보고, 제 3절에서는 순차패턴마이

닝에 대한 개념 및 연구분야를 설명한다. 제 4절에서는 본 연구에 사용된 데이터를 설명하고, 제 5절에서는 CSS의 설계과정에 대해서 서술한다. 제 6절에서는 고객 접촉스케줄링시스템의 적용의 예를 설명하고, 끝으로 제 7절에서는 본 연구의 결론과 향후 연구과제를 제시한다.

2. 보험산업에서 데이터 마이닝의 활용에 대한 기존연구

지금까지 보험산업에서 데이터 마이닝의 활용에 관한 연구는 주로 미래행동을 예측하는 것에 초점이 맞추어져 있었다. Ambross 등(1994)은 Logit Model을 통해 지불불능 보험사(Insolvent life-insurer)를 예측하는 연구를 수행하였다. 이들은 또한 지불불능 보험사를 예측하는데 있어서 Best's recommendation, IRIS(Insurance Regulatory Information System) ratios 그리고 재무적인 비율 지표(Financial measures) 등이 얼마나 효율적으로 영향을 미치는가에 대해서도 조사하였다. 그 결과, 지불불능 여부를 판정하는데 있어서 재무적인 비율지표와 IRIS ratios가 Best's recommendation 보다 효율적인 것으로 나타났다. 하지만 이 세 종류의 예측자들을 개별적으로 사용하지 않고 동시에 적용했을 때 더 좋은 성능을 보였다. Sancho 등(2005)은 보험사 가운데서 비생명보험사(Non-life insurance company)의 지불불능 여부를 Genetic Programming 응용모델로 예측했고, 19개의 재무적인 비율지표들을 속성으로 생성 및 사용하였다. 그들은 속성선택(Feature Selection)의 효과를 살펴보기 위해 Genetic Algorithm 기반으로 선별한 3개 속성으로 구성된 데이터세트(3-데이터세트)와 19개의 속성을 전부 사용한 데이터세트(19-데

이터세트를 생성하였고 두 데이터세트의 결과를 비교하였다. 3-데이터세트의 적합도(Fitness Value)는 0.38로써 0.31인 19-데이터세트보다 약간 높았다. 하지만 재현율(Recall) 및 정확도(Precision)는 모두 19-데이터세트가 높았다. 3-데이터세트의 재현율은 66%인 반면 19-데이터세트의 재현율은 93%였고, 3-데이터세트의 정확도는 61%인 반면 19-데이터세트의 정확도는 64%였다. 또한, Genetic Programming 응용모델의 재현율 및 정확도는 모두 SVM(Support Vector Machine) 모델보다 높았다. 이 연구자들은 궁극적으로 Genetic Programming 응용모델을 이용하여 비생명보험사를 분석하기 위한 자동진단시스템(Automatic Diagnostic System)을 개발하였다.

Artis 등(1999)은 스페인 자동차보험시장(Spanish Automobile Insurance Market)에서 고객의 보험사기(Fraud) 행동을 예측하는 연구를 수행하였다. 보험사기는 보험고객의 행동 중에 하나로서, 스페인 자동차보험시장에서는 그 발생률이 15%~60%로 비교적 높게 나타났기 때문에 그에 따른 손실을 축소하기 위해 보험사기에 대한 연구가 필요했다. 이에 Artis 등(1999)은 고객의 보험사기 행동을 예측하는데 MNL(MultiNomial Logit model)과 NMNL(Nasted MultiNomial Logit model)의 두 가지 접근법을 사용하였고, 예측모델은 고객의 보험사고(Claim)를 합법적인 것(Legitimate), 개인의 이익을 위한 보험사기(Fraud for Personal Profit) 그리고 제3자를 위한 보험사기(Fraud for a Third Party Benefit)로 예측하였다. 예측에 활용된 데이터는 장소, 현장조사경찰의 판단 등의 사고특성 데이터, 이전사고경력 등의 운전자이력 데이터 그리고 자동차제조년월 등과 같은 자동차관련 데이터이다. 예측성능면에서 MNL이 우월한 결과를 보였다. Bently (2000) 역시 금

용분야에서 거대한 손실을 야기하는 보험사기 문제를 연구하였다. 그는 Genetic Algorithm을 사용해서 종합주택보험에 대한 보험사고(Home Insurance Claims)의 보험사기 여부를 감지 및 예측하는 Evolutionary-Fuzzy System을 구축하였고, 이 시스템의 최고 예측률은 62%였다.

이재식과 이진천(2000)은 예측성능향상을 위한 결합모델을 구축하는 그들의 연구에서 자동차보험고객의 이탈을 예측문제를 다루었다. 이 결합모델은 기본모델, 판별모델 그리고 지원모델 등으로 구성된다. 자동차보험가입고객의 이탈여부를 예측하기 위해서 먼저 인공신경망, 의사결정나무 그리고 사례기반추론 기법을 이용하여 단일모델을 만든 후 예측성능이 가장 좋은 것을 기본모델로 선택한다. 그런 다음 기본모델의 학습데이터에 대한 예측적중여부를 새로운 목표속성으로 설정하여 판별모델을 C5.0으로 구축한다. 그리고 기본모델이 잘못 예측한 데이터만을 선별해서 이 데이터들만으로 새로운 예측모델을 구축하는데 이를 지원모델이라고 한다. 지원모델은 인공신경망 기법으로 설계한다. 이러한 모델설계에는 성별, 연령, 보험가입경력, 전계약사, 차량연식 등의 변수들을 사용하였다. 이렇게 설계한 결합모델의 고객이탈예측적중률은 71.8%로 인공신경망(64.5%), 의사결정나무(63.2%) 그리고 사례기반추론(65.6%)의 단일모델들보다 우수한 성능을 보였다. 위의 보험산업에서 데이터 마이닝 활용에 대한 기존연구들을 <표 2.1>에 간략하게 비교해 놓았다. 기존의 연구들이 미래행동을 예측하는 것에 초점이 맞추어져 있었던 것과는 달리 본 연구는 고객유지에 활용되었던 과거의 접촉데이터로부터 활용가능한 패턴을 도출하여 현재 및 미래고객에게 적용시키는 것에 초점이 맞추어져 있다.

<표 2.1> 보험산업에서의 데이터 마이닝에 대한 기존연구 비교

	예측			설명	기법
	지불불능	사기	이탈	유지	
Ambrose 등 (1994)	√				Logit Model
Artis 등 (1999)		√			MultiNomial Logit model, Nasted MultiNomial Logit model
Bently (2000)		√			Fuzzy Logic Rules
이재식과 이진천 (2000)			√		Case-Based Reasoning, Decision Tree, Neural Network
Sancho 등 (2005)	√				Genetic Programming
본 연구				√	CART, Sequential Pattern Mining

3. 순차패턴마이닝

3.1 연관규칙마이닝(Association Rule Mining)

데이터 마이닝은 이전에 알려지지 않았으면서 잠재적인 유용성을 지닌 정보 및 지식을 데이터베이스로부터 추출하는 것이다(Dimitris 등, 2003). 데이터 마이닝에는 우리가 예측하고자 하는 것이 무엇인지를 사전에 아는 경우에 사용하는 지도(Directed) 데이터 마이닝과 이와 달리 데이터 속에서 유용한 패턴을 발견하고자 할 때 사용하는 비지도(Undirected) 데이터 마이닝이 있는데, 비지도 데이터 마이닝의 대표적인 예가 연관규칙이다(Berry & Linoff, 1997). 연관규칙은 Agrawal 등 (1993)에 의해서 처음 소개되었고, 다음과 같이 설명할 수 있다.

판매데이터베이스가 주어졌다고 가정하자. 항목집합(Itemset) X 와 Y 가 주어지면, 연관규칙은 $X \rightarrow Y$ 의 형태로 나타낼 수 있다. 이 규칙은 만약 고객이 X 를 구매했다면 Y 를 구매할 가능성도 높다는 것을 의미한다. 데이터베이스로부터 연관

규칙을 찾기 위해서는 먼저 항목집합 X 와 Y 의 Support를 계산하여야 한다. 항목집합 X 와 Y 의 Support는 전체 거래량 중에서 X 와 Y 를 포함하는 거래량이 어느 정도인가를 측정하는 것이다. 전체 거래량이 N 일 때, 식 (3.1)에 따라 계산된 Support가 사용자가 명시한 Minimum Support보다 크거나 같은 경우에 해당 항목집합이 자주 발생한다고 정의한다.

$$Support = \frac{n(X \cap Y)}{N} \quad \text{식 (3.1)}$$

항목집합 X 와 Y 가 자주 발생한다면 식 (3.2)에 의해 Confidence를 계산할 수 있는데, 이 Confidence는 X 가 발생한 거래 중에서 Y 가 포함된 거래의 정도를 측정한다. Support와 마찬가지로 Confidence도 사용자가 명시한 Minimum Confidence보다 크거나 같으면, 규칙 $X \rightarrow Y$ 는 유효한 것으로 간주한다(Chen 등, 2002).

$$Confidence = \frac{Support}{P(X)} \quad \text{식 (3.2)}$$

또한, 규칙의 평가지표로 Improvement가 사용되는데 이것은 처음단계에서 결과를 가정한 것보다 규칙의 예측력이 얼마나 더 향상되었는지 여부를 측정하는 것이다. 즉, X 가 발생했을 때, 그 거래에 Y 가 포함된 경우와 Y 가 임의로 거래된 경우의 비율을 보는 것이다. 수식으로 표현하면 다음의 식 (3.3)과 같다. Improvement 값이 1보다 크면, 결과는 무작위적(Random Choice)으로 예측한 것보다 더 좋은 예측력을 가진다고 해석할 수 있다(Berry & Linoff, 1997). 의미 있는 규칙을 찾아내기 위해서는 Improvement 값이 1보다 큰지 작은지를 확인하면 된다.

$$Improvement = \frac{Confidence}{P(Y)} \quad \text{식 (3.3)}$$

이 연관규칙에 관한 연구는 거래에 대해서 구매한 항목들 간에 순서가 없다는 것을 가정한다. 이 가정이 합리적인 것처럼 보이나, 순서가 매우 중요하게 고려되어야 하는 상황에 봉착하기도 한다. 이에 Agrawal & Srikant(1995)는 구매한 항목들도 중요하지만 고객의 행동을 이해하는 데는 거래시간(Transaction Time)도 매우 중요하다고 주장했다.

3.2 순차패턴마이닝(Sequential Pattern Mining)

바코드 기술의 진보로 인해 대형소매점에서 장바구니데이터(Market Basket Data)라 불리는 판매데이터를 수집하고 저장하는 것이 가능해졌다. 판매데이터는 거래날짜와 구매한 항목들로 구성되는데, 고객이 신용카드로 구매했거나 자사고객카드를 가졌을 경우 고객ID도 포함된다. 이러한 데이터를 기반으로 Agrawal & Srikant(1995)가

SPM을 아래와 같이 소개하였다.

고객거래데이터베이스 D 는 각 거래에 대해 고객ID, 거래시간 그리고 구매한 항목의 속성들로 구성된다. 각 항목들은 구입했는지 아닌지 여부에 따라 이진값을 가진다. 항목집합은 항목들의 모임을 말하고, 순서(Sequence)는 항목집합의 리스트를 말한다. 항목집합 i 는 $\{i_1, i_2, \dots, i_m\}$ 으로 표현되고, 이때 i_j 는 항목을 말한다. 또한 순서 s 는 $\langle s_1, s_2, \dots, s_n \rangle$ 으로 표현되고, s_j 는 순서화된 항목집합을 일컫는다.

순서 $\langle a_1, a_2, \dots, a_n \rangle$ 와 순서 $\langle b_1, b_2, \dots, b_m \rangle$ 이 있을 경우에, $a_1 \subseteq b_{i_1}, a_2 \subseteq b_{i_2}, \dots, a_n \subseteq b_{i_n}$ ($i_1 < i_2 < \dots < i_n$)이 성립하면 순서 $\langle a_1, a_2, \dots, a_n \rangle$ 는 순서 $\langle b_1, b_2, \dots, b_m \rangle$ 에 포함된다. 예를 들면, 순서 $\langle (3)(45)(8) \rangle$ 은 순서 $\langle (7)(38)(9)(456)(8) \rangle$ 에 포함되는데, $(3) \subseteq (38)$, $(45) \subseteq (456)$, $(8) \subseteq (8)$ 이 성립하기 때문에 그러하다. 하지만 순서 $\langle (3)(5) \rangle$ 는 $\langle (35) \rangle$ 에 포함되지 않는다. 이렇게 순서 s_1 이 다른 순서 s_2 에 포함되지 않은 경우를 Maximal이라 한다.

각 거래가 항목집합에 상응하고 시간 순으로 정렬되어 있으면 고객의 모든 거래는 위에서 말한 순서와 함께 보여질 수 있는데, 이런 순서를 고객순서라고 부른다. 공식적으로 고객의 거래는 시간에 따라 T_1, T_2, \dots, T_n 으로 정렬된다. T_i 의 항목집합은 ($itemset(T_i)$)로 표현할 수 있다. 즉, 순서 $\langle itemset(T_1), itemset(T_2), \dots, itemset(T_n) \rangle$ 이 고객순서가 된다. 순서 s 가 고객순서에 포함되어질 때, 고객은 이 순서를 Support한다. 이 때, 순서에 대한 Support는 해당 순서를 지지하는 고객에 대한 전체고객의 비율로 정의할 수 있다.

D 가 주어지면, SPM에 대한 문제는 모든 순서

들 가운데서 사용자가 정의한 Minimum Support 를 만족하는 Maximal 순서를 찾는 것으로 귀결된다. 이 Maximal 순서가 바로 순차패턴인 것이다.

예를 들면, <표 3.1>처럼 고객ID와 거래시간으로 정렬된 데이터베이스는 <표 3.2>와 같이 고객 순서화한 데이터베이스로 표현할 수 있다 (Agrawal & Srikant, 1995).

<표 3.1> 고객ID와 거래시간으로 정렬된 데이터베이스

고객ID	거래시간	구매 항목
1	2004년 6월 25일	30
1	2004년 6월 30일	90
2	2004년 6월 10일	10, 20
2	2004년 6월 15일	30
2	2004년 6월 20일	40, 60, 70
3	2004년 6월 25일	30, 50, 70
4	2004년 6월 25일	30
4	2004년 6월 30일	40, 70
4	2004년 7월 25일	90
5	2004년 7월 12일	90

<표 3.2> 고객순서화한 데이터베이스

고객ID	고객순서
1	<(30) (90)>
2	<(10 20) (30) (40 60 70)>
3	<(30 50 70)>
4	<(30) (40 70) (90)>
5	<(90)>

여기서, Minimum Support가 25%라는 것은 적어도 둘이상의 고객에게 지지를 받아야 한다는 것

을 의미한다. <표 3.2>를 보면, 주어진 Support 제약조건을 충족하면서 Maximal인 순서 즉, 순차패턴은 <(30) (90)>과 <(30) (40 70)> 뿐이다. 순차패턴 <(30) (90)>은 고객ID 1과 4의 지지를 받는다. 고객ID 4의 경우엔 항목 (30)과 (90) 사이에 항목 (40 70)이 존재하지만 궁극적으로는 순차패턴 <(30) (90)>을 지지한다. 왜냐하면, 우리가 찾는 패턴이 반드시 인접해야만 하는 것은 아니기 때문이다. 순차패턴 <(30) (40 70)>은 고객ID 2와 4의 지지를 받는데, 고객ID 2의 경우 (40 70)이 (40 60 70)의 부분집합이기에 순차패턴 <(30) (40 70)>을 지지하는 것으로 간주한다. <(10 20) (30)>은 고객ID 2에 의해서 지지를 받지만 Minimum Support를 충족하지는 못하기 때문에 유효한 순차패턴이 아니다. <(30)>, <(40)>, <(70)>, <(90)>, <(30) (40)>, <(30) (70)> 그리고 <(40) (70)> 또한 Minimum Support를 충족하지만 Maximal이 아니기 때문에 유효한 순차패턴이 될 수 없다 (Agrawal & Srikant, 1995).

SPM에 관한 연구는 크게 세범주로 분류할 수 있다(Chen 등, 1996). 첫번째는 유사한 패턴을 찾는 것이다. 한 패턴에 대해 다른 패턴이 사용자가 정의한 임계치에 대해서 유효한 거리(Distance) 또는 상관관계(Correlation)를 가진다면 두 패턴을 유사한 것으로 간주한다(Agrawal 등, 1993; Agrawal & Srikant, 1995; Faloutsos 등, 1994; Li 등, 1996). 두번째는 주기적인 패턴을 찾아내는 것이다(Han 등, 1999; Ozden 등, 1998). 마지막 세번째는 자주 발생하는 즉, 사용자가 정의한 임계치를 넘는 패턴을 찾아내는 것이다(Chen 등, 1998; Mannila 등, 1997; Tsechansky 등, 1999). 이 가운데 최근 가장 활발하게 연구되는 것이 세번째이고, 웹 서핑, 실행계획(Plan Execution) 그리고 쇼핑순서 분석 등에 널리 활용된다.

4. 연구에 사용된 데이터

4.1 고객접촉관리 데이터 개요

본 연구에서는 고객데이터, 계약데이터 그리고 접촉데이터가 사용되었다. 고객데이터는 각 고객의 이름, 직업, 최종학력, 주소, 연락처 등과 같은 개인적인 특성들을 담고 있다. 계약데이터는 각 계약별로 계약번호, 고객번호, 계약상태, 납입방법, 납입보험료, 보험 상품 종류 등의 계약관련특성들을 담고 있다. 마지막으로 접촉데이터는 지원설계사들에 의해 생성되고 관리되는 것으로써 총 8개월간 접촉에 대해서 기록한 것이다. 이것은 접촉시간, 접촉방법, 접촉내용 그리고 메모 등으로 구성되어 있다. 본 연구는 이 데이터를 바탕으로 지원설계사들의 체계적인 고객접촉관리를 돕기 위해 CSS를 설계한다. CSS는 지원설계사의 전체 관리대상고객 약 120만명 중 접촉관리가 원활한 57% 고객의 접촉데이터를 바탕으로 접촉내용, 접촉방법 그리고 접촉간격으로 구성된 접촉스케줄링 정보를 생성한다.

지원설계사들의 역할은 모집설계사의 해촉으로 인해 동반이탈 가능성이 있는 고객들을 접촉관리를 통해 유지시키는 것이다. 본 연구의 분석대상은 유지고객으로 정하였고, 이는 지원설계사의 전체 관리대상고객 중 접촉기록이 존재하고 계약이후 추가계약이나 실효계약 또는 해지계약 없이 계약관계를 유지하고 있는 고객들을 말한다. 여기서 실효고객은 보험료를 두달 이상 연체하여 보험계약의 효력이 상실된 고객을 말하고, 해지고객은 자발적으로 보험계약을 중단한 고객을 일컫는다. 본 연구에서 사용한 데이터는 지원설계사의 접촉기록이 2회 이상인 유지고객 352,520명의 계약데이터와 고객데이터 그리고 접촉데이터 약 170만 건이

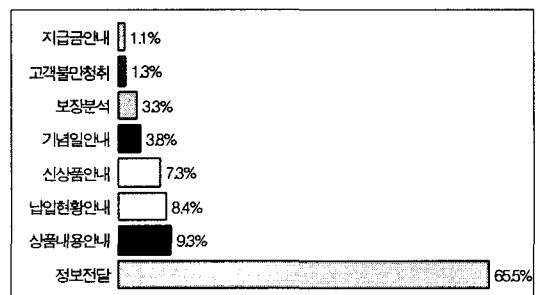
다. 본 연구에서는 장기적인 접촉관리를 목적으로 하기 때문에 접촉기록이 1회만 남아있는 고객의 데이터는 제외하였다.

4.2 데이터 정제

4.2.1 접촉내용

접촉내용은 접촉데이터의 속성 중에 하나로서, “무엇을 위해” 고객을 접촉하였는가를 기록한 것이다. 접촉내용의 세부항목으로는 정보전달, 상품내용안내, 납입현황안내, 신상품안내, 기념일안내, 보장분석, 고객불만청취 그리고 지급금안내가 있고 [그림 4.1]에 세부항목의 분포가 나타나있다.

[그림 4.1]에 제시된 접촉내용의 세부항목에 대한 분포를 보면 특정 세부항목 즉, 정보전달이 전체 분포를 지배하고 있음을 알 수 있다. 이 현상은 접촉내용을 결정하는 모델을 구축할 때 특정 세부항목만 산출하는 결과를 가져오므로 발생원인에 대한 분석과 그에 따른 조치가 필요하다.



[그림 4.1] 접촉내용 세부항목의 분포

특정 세부항목 지배현상에 대한 원인은 두가지로 정리할 수 있다. 첫번째는 입력상의 오류이고 두번째는 세부항목의 부적절한 구성이다.

첫번째 원인은 지원설계사들이 접촉내용의 세

부항목에 대한 이해를 잘못하여 부적절한 값을 입력한 것을 말한다. 지원설계사의 실제 접촉내용과 데이터에 기록된 접촉내용이 서로 다른 것으로 이는 “메모” 속성을 통해서 확인할 수 있다. 이 원인에 따른 문제를 해결하기 위해 입력오류에 대한 수정을 했다. 메모 속성과 접촉내용의 세부항목이 일치하지 않는 경우 메모 속성을 기준으로 접촉내용의 세부항목을 수정하였다. 무작위로 추출한 샘플에서 각 접촉내용의 세부항목에 적합한 키워드들을 추출하여 그 키워드들을 토대로 접촉내용 세부항목을 수정하였다. <표 4.1>에는 접촉데이터에 기록된 접촉내용과 메모 그리고 키워드기반으로 수정된 접촉내용이 차례로 나타나있다. 메모 칼럼에 굵게 표시된 글자는 키워드이다.

두번째 원인을 해결하기 위해 고객의 반응에 따라 접촉내용의 세부항목을 재분류하였다. 고객이 특정한 반응을 보였을 경우에만 수행해야하는 접촉내용 세부항목을 이벤트접촉내용의 세부항목으로 간주하고, 그렇지 않은 경우를 스케줄접촉내용의 세부항목으로 간주하였다. 본 연구의 분석 및 모델링대상은 스케줄접촉내용으로 한정한다.

스케줄접촉내용의 세부항목은 납입현황안내, 보장분석, 상품내용안내, 신상품안내 그리고 정보전달 등이 있다. 납입현황안내는 고객에게 보험계약일, 월납입보험료 그리고 납부이력 등과 같은 보험료 납입상태를 설명하는 것이다. 보장분석은 고객이 가지고 있는 보험상품의 전체 보장내용을 중

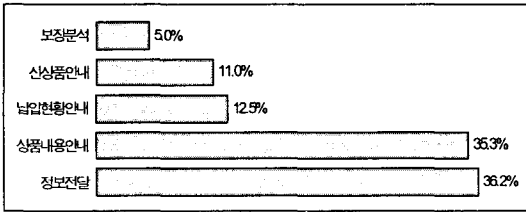
합적으로 분석하여 안내하는 것이고, 상품내용안내는 이미 가입되어 있는 보험계약의 중요 내용들 즉, 보험의 보장내용이나 보장기간 등을 재설명해주는 것을 말한다. 신상품안내는 고객이 가입한 상품 이후에 출시된 상품이나 고객이 가입하지 않은 상품을 소개하는 것이고, 정보전달은 보험계약과 무관하지만 고객이 원하거나 필요로 하는 정보를 제공하는 것을 말한다. 고객의 취미나 건강관리에 필요한 정보들을 제공하는 것이 정보전달의 대표적인 예이다.

이벤트접촉내용은 고객불만청취, 기념일안내 그리고 지금급안내와 같은 세부항목들로 구성된다. 고객불만청취는 콜센터를 통해 고객의 민원이 접수되는 경우에 고객에게 직접 해당사항에 대한 내용을 청취 및 접수하는 것을 말한다. 기념일안내 는 고객의 각종 기념일에 축하 메시지를 전달하는 것을 말하는데 고객생일, 가족생일, 결혼기념일 그리고 계약 1주년 등이 그 예이다. 지금급안내는 계약기간 내에 발생하는 분할지급보험금이 있을 경우에 이 사실을 고객에게 알리는 것을 말한다.

이처럼 접촉내용의 특정 세부항목이 전체분포를 지배하게 된 원인을 찾고 그에 대한 조치를 취한 결과 [그림 4.2]와 같은 분포를 얻을 수 있었다. [그림 4.2]는 재구성된 스케줄접촉내용 세부항목의 분포로서 [그림 4.1]에서 보았던 정보전달의 지배현상이 완화되었음을 확인할 수 있다.

<표 4.1> 접촉내용 세부항목을 수정한 예

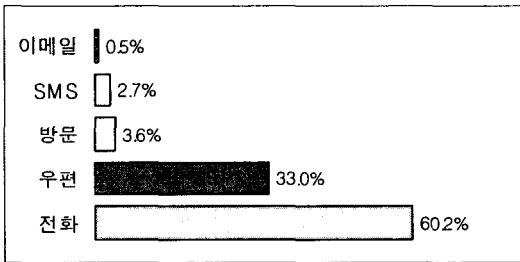
접촉내용	메모	수정된 접촉내용
정보전달	XXX 상품설명차 방문.	신상품안내
정보전달	증권 전달하러 가서 계약한 상품에 대해 설명해 줌.	상품내용안내
정보전달	종신보험권유 및 오늘 저녁때 방문하기로 함	신상품안내



[그림 4.2] 스케줄접촉내용 세부항목의 분포

4.2.2 접촉방법

접촉방법은 고객을 “어떻게 접촉했는가.”를 기록한 것이다. 접촉방법의 세부항목은 이메일, SMS(모바일 폰을 이용한 문자서비스), 방문, 우편 그리고 전화 등이 있고, [그림 4.3]에 이에 대한 분포가 나타나있다.



[그림 4.3] 접촉방법 세부항목의 분포

접촉방법 세부항목의 분포 역시 특정 세부항목에 편향되어 있음을 알 수 있다. 이러한 현상이 발생한 것은 데이터 수집시점이 이메일과 SMS를 고객접촉에 활용하기 시작하는 과도기였기 때문인 것으로 사료된다. 현재 전화와 우편이 전체분포의 93.2%를 차지하기 때문에 이를 바탕으로 접촉방법을 결정하는 모델을 구축하면 그 결과 또한 전화와 우편만 산출된다. 따라서 모델의 결과로 특정 접촉방법만이 산출되는 것을 방지하고, 이메일과 SMS 등의 활용도를 높이기 위해서 접촉방법을 결정하는 모델을 구축하는 대신 영역지식(Domain Knowledge)에 기반을 둔 접촉방법 리스트를 생성하였다. 이 접촉방법 리스트가 <표 4.2>에 제시되어 있고, 주어진 접촉내용에 대해 사용할 수 없는 접촉방법을 X로 표시했다.

5. CSS의 설계

5.1 CSS 개요

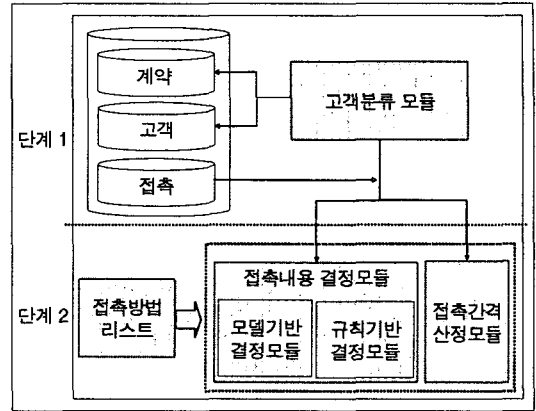
CSS는 고객들을 특성에 따라서 고객군으로 분류하고 분류고객군별로 접촉스케줄링 정보인 접

<표 4.2> 접촉내용에 대한 접촉방법 리스트

접촉내용	접촉내용의 종류	접촉방법				
		전화	우편	방문	SMS	이메일
남입현황안내	스케줄접촉	○	○	○	○	○
보장분석	스케줄접촉	○	X	○	X	X
상품내용안내	스케줄접촉	○	X	○	X	X
신상품안내	스케줄접촉	○	○	○	X	○
정보전달	스케줄접촉	○	○	○	○	○
고객불만청취	이벤트접촉	○	X	○	X	X
기념일안내	이벤트접촉	○	○	○	○	○
지급금안내	이벤트접촉	○	○	○	○	○

축내용, 접촉간격 그리고 접촉방법을 산출한다. 이러한 역할을 하는 CSS의 설계개념도가 [그림 5.1]에 제시되어 있다. [그림 5.1]을 보면, CSS는 세계의 모듈로 구성되어 있다. 고객분류 모듈, 접촉내용 결정모듈 그리고 접촉간격 산정모듈이 그것들이다. 여기서 접촉내용 결정모듈은 모델기반 결정모듈과 규칙기반 결정모듈의 하위모듈로 이루어져 있다.

고객분류 모듈은 단계 1에서 CART 알고리즘을 이용하여 유사한 특성을 가진 고객들을 고객군으로 분류한다. 접촉내용 결정모듈은 단계 2에서 접촉스케줄링 정보의 구성요소 중 접촉내용을 결정한다. 이 접촉내용은 모델기반 결정모듈과 규칙기반 결정모듈의 결합에 의해서 완성된다. 모델기반 결정모듈은 스케줄접촉내용을 결정하고 규칙기반 결정모듈은 이벤트접촉내용과 비즈니스규칙에 의한 접촉내용을 결정한다. 접촉간격 산정모듈은 결정된 접촉내용과 산정된 접촉간격에 <표 4.2>의 접촉방법 리스트를 결합함으로써 최종적으로 고객의 접촉스케줄링 정보가 완성된다.



[그림 5.1] CSS의 설계도

5.2 고객분류 모듈

고객분류 모듈은 계약데이터와 고객데이터를 기반으로 해서 인구통계학적 및 개인적인 특성과 계약관련 특성이 유사한 고객들을 고객군으로 분류한다. 주어진 데이터를 기반으로 고객분류 모듈 구축에 필요한 속성들을 생성하고 선별하였는데 <표 5.1>에 선별한 속성들이 나타나있다.

<표 5.1> 고객분류를 위해 선별한 속성들

속성명	설명	역할
상품보유여부	해당 상품을 보유했으면 1의 값을 부여, 그렇지 않으면 0의 값을 부여	입력
LTV집단번호	LTV값을 기준으로 5개 집단으로 구분하여 값을 부여	입력
성별	보험가입 고객의 남과 여를 구분하여 값을 부여	입력
직업	고객의 직업을 사무직, 서비스직, 제조업, 주부, 기타로 그룹화하여 그룹별로 값을 부여	입력
연령	보험가입고객의 나이	입력
기존계약건수	고객의 보험계약건수의 합	입력
평균합계보험료	총합계보험료를 총계약건수로 나눈 값	입력
접촉간격	각 접촉간의 일수 차이	목표

최초와 최종접촉 간격이 90일 이상인 고객들을 대상으로 총 8개의 분류고객군을 생성하였다. CART 모델에 의한 분류고객군 생성결과가 [그림 5.2]에 나타나있다. [그림 5.2]를 보면, 합계보험료, 직업, 연령 그리고 종신상품보유여부의 속성들이 고객분류에 영향을 미쳤음을 확인할 수 있다. 이 결과는 IF-THEN 규칙의 형태로 저장되어 고객의 분류고객군을 판정할 때 사용된다.

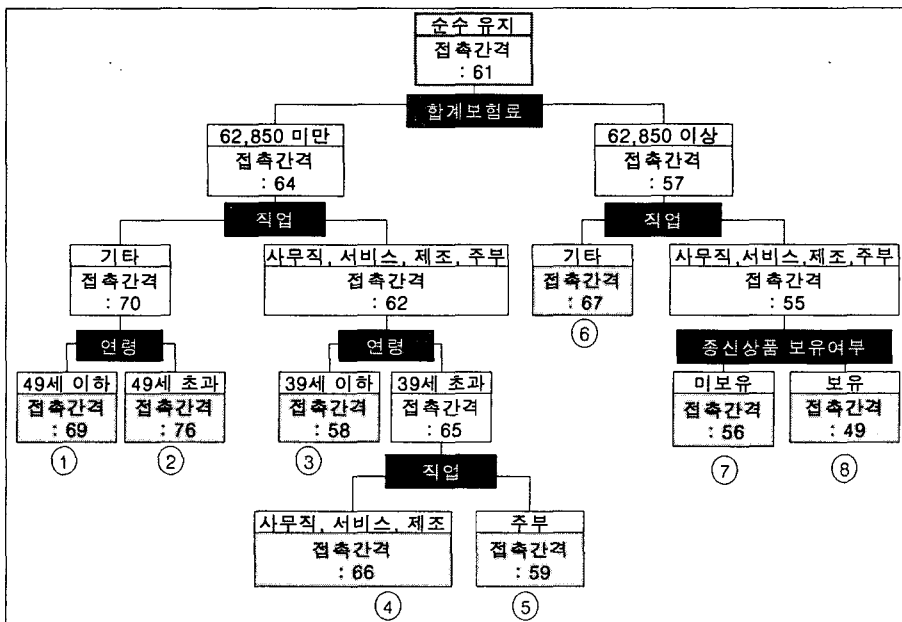
5.3 접촉내용 결정모듈

접촉내용 산정모듈은 두개의 하위모듈로 구성된다. 하나는 모델기반 결정모듈이고 다른 하나는 규칙기반 결정모듈이다. 모델기반 결정모듈에 의해 스케줄접촉내용이 결정되고 규칙기반 결정모듈에 의해서 이벤트접촉내용과 비즈니스규칙에 의한 접촉내용이 결정된다. 접촉내용은 스케줄접

촉내용에 이벤트접촉내용 및 비즈니스규칙에 의한 접촉내용을 결합하여 완성한다.

모델기반 결정모듈은 접촉데이터를 사용하여 분류고객군별로 SPM 모델을 구축하고, 구축한 모델결과를 기반으로 납입현황안내, 보장분석, 상품내용안내, 신상품안내 그리고 정보전달의 배합을 결정한다. 이렇게 결정된 것이 스케줄접촉내용이고, 스케줄접촉내용에 기반을 둔 접촉을 스케줄접촉이라 한다. 동일한 분류고객군에 속한 고객들의 스케줄접촉내용은 모두 동일하다. <표 5.2>는 SPM 모델이 산출한 결과의 일부분이다.

<표 5.2>의 첫번째 칼럼인 NITEMS는 항목집합의 길이를 말하는 것으로 순차패턴에 포함된 총 항목의 수를 나타내고, COUNT는 해당 순차패턴을 포함하고 있는 사례의 수를 말한다. SUPPORT 칼럼과 CONF 칼럼은 Support와 Confidence의 값을 차례로 나타내고, 이어서 보이는 SET1, SET2,



<그림 5.2> 고객분류 결과

SET3, ... 는 각 순차패턴에 포함된 항목 즉, 접촉 내용의 세부항목들을 순서대로 보여주고 있다. 이렇게 산출된 분석결과를 통해서 각 분류고객군별로 대표순차패턴을 선정한다.

<표 5.2> SPM 모델 결과의 예

ITEMS	COUNT	SUPPORT	CONF	SET1	SET2	SET3	SET4	SET5
2	21,750	71.7	77.9	정보전달	정보전달			
2	8,788	28.0	68.1	상품내용	정보전달			
2	7,228	24.0	63.0	납입현황	정보전달			
2	7,280	24.0	67.4	신상품	정보전달			
2	1,611	6.0	6.5	정보전달	정보전달 & 상품내용			
3	4,984	16.4	62.3	신상품	정보전달	정보전달		
3	4,513	14.8	57.2	납입현황	정보전달	정보전달		
3	4,049	13.4	61.5	정보전달	상품내용	정보전달		
3	3,520	11.6	18.2	정보전달	정보전달	상품내용		
3	3,437	11.3	62.0	정보전달	신상품	정보전달		
3	3,413	11.3	56.3	정보전달	납입현황	정보전달		
3	3,344	11.0	64.1	상품내용	상품내용	정보전달		
3	3,157	10.4	14.5	정보전달	정보전달	신상품		
3	3,147	10.4	14.5	정보전달	정보전달	납입현황		
4	3,350	11.2	58.5	상품내용	정보전달	정보전달	정보전달	
4	2,870	9.8	58.6	신상품	정보전달	정보전달	정보전달	

대표순차패턴을 선정하는 과정은 두 단계로 이루어진다. 먼저 항목집합의 길이를 선정하는데, 이는 각 분류고객군의 평균접촉회수를 따른다. 분류고객군 ③의 평균접촉회수가 4라면 항목집합의 길이 또한 4가 되는 것이다. 그런 다음, 해당 항목집합의 길이 내에서 Support와 Confidence 값을 비교하여 최상위값을 가진 순차패턴을 분류고객군의 대표패턴으로 선정한다. 이렇게 선정된 분류고객군별 순차패턴이 [그림 5.3]에 나타나있다.

규칙기반 결정모듈은 두 종류로 구성된다. 하나는 4.2.1절에서 분류된 이벤트접촉내용이고 다른 하나는 비즈니스규칙에 의한 접촉내용이다. 이벤트접촉내용은 앞서서 살펴본 것처럼 과거데이터에 기반을 두어서 파생된 것이지만 비즈니스규칙은 과거데이터가 아니라 영역지식에 기반을 두어서 도출된 것이다. 이벤트접촉내용에 기반을 둔 접촉을 이벤트접촉이라 하고 비즈니스규칙에 기반을 둔 접촉을 비즈니스규칙접촉이라고 한다.

비즈니스 관점에서 고객의 특정 상황과 행동이 고객유지에 큰 영향을 미친다고 판단될 때는 부정적인 영향을 최소화하기 위해 긴급조치가 필요한데 이러한 항목과 조치를 명시한 것이 바로 비즈니스규칙이다. 항목에 대한 조치는 스케줄접촉내용 세부항목들의 조합으로 구성되고, 이것은 CSS가 작동할 때 스케줄접촉내용보다 우선순위를 가진다. '보험료 2회 이상 연속 연체' 이벤트 발생은 실효로 이어지기 때문에, 비즈니스규칙에 의해 즉시 납입현황안내 접촉을 수행하고 21일 후에 상품내용안내 접촉을 수행하게 된다. 이것은 실효고객에 대한 고객이탈현상을 줄이기 위함이다. 이와 같은 비즈니스규칙의 예가 <표 5.3>에 나타나있다. 이 규칙기반 결정모듈은 고객의 상황과 행동에 따라서 다르게 결정되기 때문에 같은 분류고객군에 속하는 고객들에 대해서 서로 다른 접촉스케줄링 정보가 생성되게 된다.

<표 5.3> 비즈니스규칙의 예

비즈니스 규칙 항목	내용 및 간격	접촉내용 1	접촉내용 2	접촉내용 3
보험료 2회 이상 연속 연체 고객	활동	납입현황안내	상품내용안내	
	간격	21일		
납입연료 3개월 전, 만기 3개월 전	활동	정보전달+ 납입현황안내	상품내용안내	보정분석
	간격	35일		21일
보험료 지급고객 (분할 보험금 포함)	활동	고객불만처리	상품내용안내	보정분석
	간격	14일		30일
콜센터 문의 고객 (불만 사항)	활동	고객불만처리	정보전달	상품내용안내
	간격	15일		28일

5.4 접촉간격 산정모듈

접촉간격 산정모듈은 각 접촉내용 세부항목간 접촉일수의 평균거리를 계산하여 그 값을 접촉간격으로 결정한다. 분류고객군별로 해당 접촉내용의 세부항목을 순차적으로 포함하고 있는 사례들만을 추출하여 식 (5.1)에 따라 접촉간격을 계산한다.

$$\text{접촉간격} = \frac{\sum_{k=1}^n D_{ab}}{n} \quad \text{식 (5.1)}$$

n 은 해당 접촉내용의 세부항목을 포함하고 있는 고객수를 의미하고, D 는 두 접촉내용 세부항목 간 거리 즉, 두 접촉내용 세부항목간 접촉일수의 차이를 말한다. a 는 첫번째 접촉내용의 세부항목을 의미하고 b 는 두번째 접촉내용의 세부항목을 의미한다.

분류고객군별 접촉간격 산정결과가 [그림 5.3]에 스케줄접촉내용과 함께 나타나있다. 분류고객군 ①의 예를 보면, 분류고객군 ①의 스케줄접촉내용은 납입현황안내, 상품내용안내 그리고 정보전달이다. 접촉내용1의 납입현황안내와 접촉내용2의 상품내용안내에 대한 접촉간격은 40일이고, 접촉내용2의 상품내용안내와 접촉내용3의 정보전달에 대한 접촉간격은 36일이다.

6. CSS의 적용

6.1 CSS 적용 개요

지원설계사가 해축된 모집설계사의 관리고객을 접촉할 때 CSS의 작동과정은 다음과 같다. [그림 6.1]의 적용 흐름도를 참고하면, 접촉스케줄링 정보를 생성해야하는 고객이 주어졌을 때 먼저 고객 분류 모듈에 의해 만들어진 고객분류 규칙에 따라서 고객의 분류고객군을 판정한다. 판정한 분류고객군에 따라서 스케줄접촉내용과 접촉간격이 자동적으로 결정된다. 이벤트접촉내용과 비즈니스규칙에 의한 접촉내용이 해당되지 않는 경우에는 ①을 따라 스케줄접촉내용에만 기반을 두어서 접촉스케줄링 정보가 생성된다. 반면에 고객의 개인적 특성에 따라서 이벤트접촉내용과 비즈니스규칙에 의한 접촉내용을 고려해야 할 경우 ②를 따라 접촉스케줄링이 진행된다. 이벤트접촉내용과 비즈니스

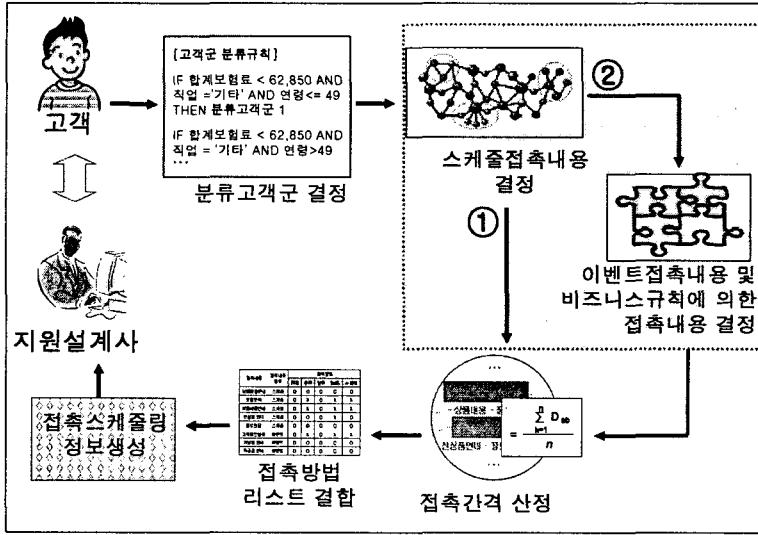
* 접촉간격 단위: 일

분류 고객군	접촉내용1	접촉간격	접촉내용2	접촉간격	접촉내용3
①	납입현황	40	상품내용	36	정보전달
②	납입현황	36	상품내용	36	정보전달
⑥	납입현황	40	상품내용	40	정보전달

분류 고객군	접촉내용1	접촉간격	접촉내용2	접촉간격	접촉내용3	접촉간격	접촉내용4
③	신상품	40	상품내용	44	정보전달	32	정보전달
④	신상품	40	상품내용	40	정보전달	36	정보전달
⑤	신상품	48	정보전달	36	상품내용	32	정보전달
⑦	납입현황	28	상품내용	32	정보전달	32	정보전달

분류 고객군	접촉내용1	접촉간격	접촉내용2	접촉간격	접촉내용3	접촉간격	접촉내용4	접촉간격	접촉내용5
⑧	납입현황	36	정보전달	28	상품내용	20	정보전달	24	정보전달

[그림 5.3] 분류고객군별 접촉내용과 접촉간격



[그림 6.1] CSS의 적용 흐름도

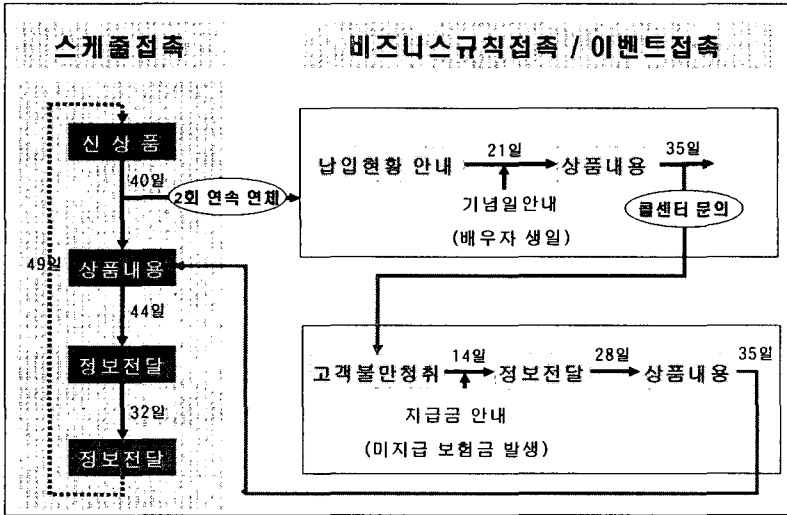
스규칙에 의한 접촉내용은 스케줄접촉내용에 실시간으로 결합한다. 이렇게 결정된 접촉내용과 접촉간격에 이미 생성된 접촉방법 리스트가 결합하여 접촉스케줄링 정보가 완성된다. 이 접촉스케줄링 정보는 지원설계사에게 제공되고 지원설계사는 이 정보를 바탕으로 효율적이고 효과적인 고객 접촉관리를 수행하게 된다.

6.2 CSS 적용의 예

본 절에서는 CSS가 어떻게 작동하는지에 대한 예를 설명한다. [그림 6.2]는 세 종류의 접촉내용 즉, 스케줄접촉내용, 이벤트접촉내용 그리고 비즈니스규칙에 의한 접촉내용이 동시에 발생한 경우 CSS가 어떻게 작동하는지를 보여준다.

[그림 6.2]를 보면, 먼저 고객에 대해서 스케줄 접촉내용에 기반을 둔 스케줄접촉이 주어진다. 아무 이벤트가 발생하지 않는다면 지원설계사는 주

어진 스케줄접촉에 따라 고객을 접촉한다. 첫번째 접촉은 신상품안내이고 40일 후에 상품내용안내 접촉을 한다. 그런데, 다음 접촉까지 대기시간인 40일이 지나기 전에 '2회 이상 연속 연체' 이벤트가 발생한다면 스케줄접촉에 앞서서 비즈니스규칙접촉을 수행한다. 즉, 납입현황안내 접촉을 수행하고 21일 후에 상품내용안내 접촉을 한다. 납입현황안내 접촉과 상품내용안내 접촉 사이에 배우자 생일 이벤트가 발생하면 바로 기념일안내 이벤트 접촉을 수행한 다음 상품내용안내 접촉을 한다. 이렇게 상품내용안내 접촉이 끝나면 비즈니스규칙 접촉이 완료되고 다시 두번째 스케줄접촉인 상품내용안내로 환원하여야 한다. 하지만 환원하기 전에 '콜센터 문의' 이벤트가 발생한다면 다시 비즈니스규칙접촉을 우선적으로 수행한다. 즉, 고객불만처리, 정보전달 그리고 상품내용안내 접촉을 각각 주어진 접촉간격에 맞춰 수행해야 한다. 하지만 그 중간에 또다른 이벤트가 발생하면 해당 이벤트



[그림 6.2] CSS 작동의 예

접촉을 먼저 수행한다. 비즈니스규칙접촉이 최종적으로 완료되면 다시 두번째 스케줄접촉으로 환원한다. 이 때 접촉간격은 사전에 산정된 결과를 그대로 적용하고, 접촉방법은 <표 4.2>에서 제시한 접촉방법 리스트를 따른다.

7. 결론 및 향후연구과제

본 연구에 사용한 데이터에 의하면 지원설계사의 전체 접촉관리 대상고객 중 57%만이 접촉관리가 되고 있다. 지원설계사들을 인터뷰를 통하여 나머지 43%의 고객의 접촉관리가 되고 있지 않은 데는 특별한 이유가 없음을 알 수 있었다. 전체 관리대상고객의 유지율을 증가시키기 위해서는 43%의 고객도 꾸준히 접촉관리를 수행할 필요가 있다. 하지만 해축된 모집설계사의 경험과 지식에 접근할 수 없는 지원설계사들은 그들의 고객을 접촉하는데 어려움을 겪을 수밖에 없다. 본 연구에서 설

계한 CSS는 지원설계사들이 보다 효율적이고 효과적으로 고객접촉을 수행할 수 있도록 한다. 이것은 모집설계사의 해축으로 담당설계사를 잃은 고객들의 만족도 및 로열티가 저하되는 것을 방지하여서 궁극적으로 고객을 유지하는데 기여한다.

과거 보험산업에서 데이터 마이닝의 활용에 대한 기존연구가 미래행동을 예측에 초점이 맞추어져 있던 것과 달리 본 연구는 고객의 유용하고 사실적인 순차패턴을 찾아서 이것을 유사한 특성을 가진 고객에게 적용하는 방법에 초점을 두었다. 이렇게 하는데 CART와 SPM을 활용함으로써 고객접촉을 지원하는 시스템 설계에 데이터 마이닝 기법을 활용할 수 있다는 가능성을 제시하였다.

본 연구에 사용한 접촉데이터는 데이터생성의 초기단계에 해당되는 것이었기 때문에 최적의 CSS를 구축하는데 한계점이 있었다. CSS는 지원설계사의 유지고객에 대한 접촉기록이 모두 바람직하다는 가정하에 의미가 있다. 이 가정을 일반화하고 접촉의 성공여부를 객관적으로 측정하기 위

해서는 각 접촉마다 고객의 만족도를 수집할 필요가 있다. 이와 같은 방법으로 각 접촉의 유효성을 논리적으로 뒷받침할 수 있고 CSS의 성과도 측정할 수 있을 것이다. 또한 접촉내용 세부항목의 분포에서 나타난 문제점 즉, 지원설계사들의 입력오류와 세부항목의 부적절한 구성을 보완하기 위해 본 연구에서는 수작업이 이루어졌으나 텍스트 마이닝 기법 등을 활용함으로써 CSS 구축의 효율성도 도모할 수 있을 것이다. 접촉방법의 결정도 현재 데이터로는 접촉방법 리스트를 생성하는 것이 최선이었으나 이메일과 SMS 등의 활용도가 안정된 후에는 과학적인 모델의 구축을 꾀할 수 있을 것으로 사료된다.

참고문헌

- [1] 이재식, 이진천, "입력자료 판별에 의한 데이터 마이닝의 성능개선", 한국지능정보시스템학회 2000 춘계정기학술대회지, 2000, pp. 293-303.
- [2] Agrawal R. and R. Srikant, "Mining Sequential Patterns", *Proceedings of the 11th International Conference on Data Engineering*, Taipei, Taiwan, 1995, pp. 3-14.
- [3] Agrawal R., C. Faloutsos and A. Swami, "Efficient similarity search in sequence databases", *Conference on Foundations of Data Organization and Algorithms*, Chicago, Illinois, Springer, Berlin, 1993, pp. 69-84.
- [4] Agrawal R., K. Lin, H.S. Sawhney and K. Shim, "Fast similarity search in the presence of noise, scaling, and translation in time-series databases", *Proceedings of the 21st International Conference on Very Large Data Bases*, Zurich, Switzerland, Morgan Kaufmann, Los Altos, CA, 1995, pp. 490-501.
- [5] Ambrose J.M. and A.M. Carroll, "Using Best's ratings in life insurer insolvency prediction", *Journal of Risk and Insurance*, Vol. 61, 1994, pp. 317-327.
- [6] Artis M., Mercedes A. and Montserrat G., "Modeling different types of automobile insurance fraud behaviour in the Spanish market", *Insurance: Mathematics and Economics*, Vol. 24, 1999, pp. 67-81.
- [7] Bentley P.J., "Evolutionary, my dear Watson, Investigating Committee-based Evolution of Fuzzy Rules for the Detection of Suspicious Insurance Claims", *Proceedings of the Second Genetic and Evolutionary Computation Conference*, 2000, pp. 8-12.
- [8] Berry M J. A. and Gordon S. Linoff, *Data Mining Technique*, John Wiley & Sons, Inc., 1997.
- [9] Chen M.S., J. Han and P.S. Yu, "Data Mining: an overview from a database perspective", *IEEE Trans. Knowledge Data Eng.*, Vol. 8(6), 1996, pp. 866-883.
- [10] Chen M.S., J.S. Park and P.S. Yu, "Efficient data mining for path traversal patterns", *IEEE Trans. Knowledge Data Eng.*, Vol. 10(2), 1998, pp. 209-221.
- [11] Chen Y., S. Chen and P. Hsu, "Mining hybrid sequential patterns and sequential rules", *Information Systems*, Vol. 27, 2002, pp. 345-362.
- [12] Dimitris K., B. Karakostas and E. Papathanassiou, "The potential of virtual communities in the insurance industry in the UK and Greece", *International Journal of Information Management*, Vol. 23, 2003, pp.41-53.

- [13] Faloutsos C., M. Ranganathan, and Y. Manolopoulos, "Fast subsequence matching in time-series databases", *Proceedings of the 1994 ACM SIGMOD International Conference on Management of Data*, Minneapolis, Minnesota, ACM Press, New York, 1994, pp. 419-429.
- [14] Han J., G. Dong and Y. Yin, "Efficient mining of partial periodic patterns in time series databases", *Proceedings of the 15th International Conference on Data Engineering, Sydney, Australia, 1999*, pp. 106-115.
- [15] Li C., P.S. Yu, and V. Castelli, "Hierarchyscan: a hierarchical similarity search algorithm for databases of long sequences", *Proceedings of the 12th International Conference on Data Engineering*, New Orleans, Louisiana, IEEE Computer Society, 1996, pp.546-553.
- [16] Mannila H., H. Toivonen and A.I. Verkamo, "Discovery of frequent episodes in event sequences", *Data Mining Knowledge Discovery*, Vol. 1(3), 1997, pp. 259-289.
- [17] Ozden B., S. Ramaswamy and A. Silberschatz, "Cyclic association rules", *Proceedings of the 14th International Conference on Data Engineering*, Orlando, Florida, IEEE Computer Society, 1998, pp. 412-421.
- [18] Sancho S.S., J. Fernandez, M.J. Segovia and C. Bousoño, "Genetic Programming for the prediction of insolvency in non-life insurance companies", *Computers and Operations Research*, Vol. 32, 2005, pp. 749-765.
- [19] Tsechansky M.S., N. Pliskin, G. Rabinowitz and A. Porath, "Mining relational patterns from multiple relational tables", *Decision Supports Systems*, Vol. 27, 1999, pp. 177-195.

Abstract

Design of Contact Scheduling System(CSS) for Customer Retention

Jae Sik Lee* · You Jung Cho**

Customer retention is one of the major issues in life insurance industry, in which competition is increasingly fierce. There are many things for the life insurers to do many things to retain the customers. One of those things is to make sure to keep in touch with all customers. When an insurance-planner resigned, his/her customers must be taken care of by some planner-assistants. This article outlines the design of Contact Scheduling System (CSS) that supports planner-assistants for contacting the customers. Planner-assistants are unable to share the resigned insurance-planner's experience and knowledge regarding the customer relationship management. The CSS developed by employing both Classification And Regression Tree (CART) technique and Sequential Pattern Mining (SPM) technique has a two-stage process. In the first stage, it segments the customers into eight groups by CART model. Then it generates contact scheduling information consisting of contact-purpose, contact-interval and contact-channel, according to the segment's typical contact pattern. Contact-purpose is derived by schedule-driven, event-driven, or business-rule-driven. Schedule-driven contact is determined by SPM model. In the operation of CSS in a realistic situation, it shows a practicality in supporting planner-assistants to keep in touch with the customers efficiently and effectively.

Key words : Customer Relationship Management, Sequential Pattern Mining, Scheduling, Segmentation, Decision Tree

* e-Business, School of Business Administration, Ajou University

** Business Administration, Graduate School, Ajou University

