

대화식 데이터 마이닝 기법을 활용한 자동차 보험사의 인입 콜량 예측 사례

백 응
애니카자동차손해사정서비스주
(w.baek@samsungfire.com)

김남규
국민대학교 경영정보학부 조교수
(ngkim@kookmin.ac.kr)

최근 고객들의 비대면 점점 서비스 이용도가 높아짐에 따라, 비대면 채널은 다양한 데이터의 분석을 통해 고객 만족도를 향상시킬 수 있는 유용한 창구로 인식되고 있다. 이러한 비대면 채널의 대표적 영역으로 콜센터를 들 수 있으며, 콜센터 운영에서 고객 만족도에 가장 큰 영향을 미치는 요소는 상담 인력의 규모인 것으로 알려져 있다. 즉, 일정 수준 이상의 고객 만족도를 유지하기 위해서는 충분한 상담 인력을 확보하는 것이 관건이지만, 불필요하게 많은 인력을 확보하는 것은 인건비 측면에서 비용의 낭비를 초래할 수 있다. 따라서 부족하지도 않고 넘치지도 않을 정도의 적정 인력을 산출하는 능력은 콜센터 운영의 핵심 경쟁력으로 인식되고 있으며, 최근 콜센터에서는 적정 인력의 규모를 예측하기 위해 WFM(Work Force Management) 업무 전담 부서를 설치하고 콜량을 정확하게 예측하기 위한 노력을 기울이고 있다.

콜량 예측을 위해 현업에서 주로 사용되는 방법은 담당자의 직관에 의존하는 방법으로, 일정기간의 콜량 평균을 담당자가 주관적 판단에 의해 보정함으로써 이루어진다. 하지만 이러한 방식은 담당자의 주관적 성향에 크게 좌우된다는 한계를 갖고 있어서, 최근에는 다양한 예측 모형을 시스템화한 WFMS(Workforce Management System) 패키지가 널리 활용되고 있다. 하지만 이 시스템은 초기 도입 시 매우 고가의 구축비용이 발생하며, 신규 요인 발굴 시 이를 즉각적으로 시스템에 반영하기 어렵다는 한계점을 갖고 있다. 이를 극복하기 위해 본 연구에서는 데이터 마이닝의 대화식 의사결정나무 기법을 이용함으로써, 객관적이면서도 업무 배경 지식을 충분히 활용할 수 있는 예측 모형을 수립하고자 한다. 또한, 본 연구에서 수립한 모형의 정확성 평가를 위해, 국내 최대 규모의 한 자동차 보험사 콜센터의 4년 8개월 간의 실 데이터를 사용한 실험을 수행하고 그 결과를 제시하였다. 실험에서는 기존의 WFMS와 본 연구에서 제안하는 두 가지 모형인 대화식 의사결정나무 기반의 예측 모형, 일반 의사결정나무 기반의 예측 모형의 세 가지 모형에 대해, 다양한 오차 허용범위 하에서의 사고콜 및 고장콜에 대한 예측 적중률을 평가하였다.

논문접수일 : 2000년 07월 06일 논문수정일 : 2000년 07월 10일 게재확정일 : 2000년 07월 21일 교신저자 : 김남규

1. 서론

최근 고객 데이터 분석을 통한 고객 만족도 향상 사례가 증가함에 따라, 데이터 관리 및 분석은 점차 CRM(Customer Relationship Management)의 핵심 요소로 자리 잡고 있다. 특히 고객들의 비대면 점점 서비스에 대한 이용도가 점차 높아짐에

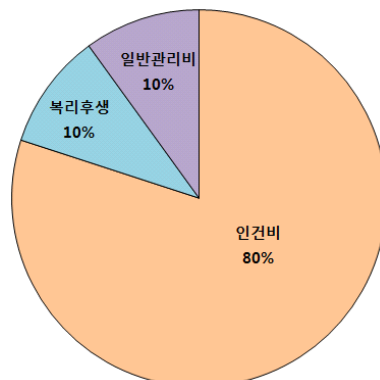
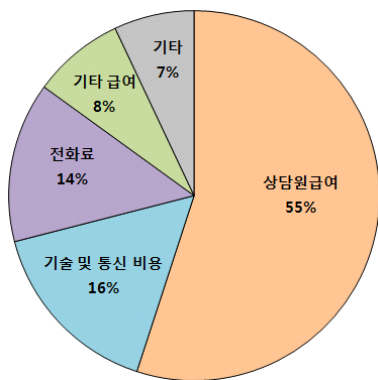
따라(KMAC, 2009), 비대면 채널은 데이터의 관리 및 분석을 통해 최소한의 비용으로 고객 만족도를 높일 수 있는 가장 효율적인 창구로 인식되고 있다. 비대면 채널의 대표적인 영역으로는 기업과 고객 간에 다양한 정보통신 수단을 통한 직접적인 접촉이 이루어지는 콜센터(오봉연, 김용석, 2007)를 들 수 있는데, 콜센터는 콜의 요청 주체에 따라

인바운드(Inbound) 콜센터와 아웃바운드(Outbound) 콜센터로 구분된다(송현수, 2002). 인바운드 콜센터는 CS(Customer Satisfaction) 콜센터 혹은 고객센터 등으로 불리며 고객의 민원, 문의, 업무처리를 주로 수행한다. 한편 아웃바운드 콜센터는 TM센터, 마케팅 센터 등으로 불리며 광고 혹은 판촉활동을 담당한다. 일반적으로 콜센터라 함은 인바운드 형태의 콜센터를 말하고, 실제로 고객만족에 더욱 직접적인 영향을 미치는 것 역시 이 유형의 콜센터이다.

콜센터에서 고객의 콜에 대한 응대는 실시간에 이루어지는 것을 가정하므로, 충분한 상담원을 확보하지 못하면 이는 고객의 대기시간을 증가시켜서 고객 만족도를 저하시키는 부작용을 가져오게 된다. 따라서 일정 수준 이상의 고객 만족도를 유지하기 위해서는 충분한 상담 인력을 확보하는 것이 관건이지만, 필요 이상의 많은 인력을 확보하는 것은 오히려 인건비의 낭비를 초래할 수 있다. 특히 상담원의 채용, 교육훈련, 관리에 소요되는 비용 및 급여, 복리후생, 4대보험 등의 인건비로 지출되는 비용이 콜센터 전체 운용비용 중 60~70% 이상(신상기, 2005)을 차지하고 있기 때문에, 부족하지도 않고 넘

치지도 않을 정도의 적정 인력 규모를 산출하는 것은 콜센터 운영의 CSF(Critical Success Factor)로 인식되고 있다<그림 1>.

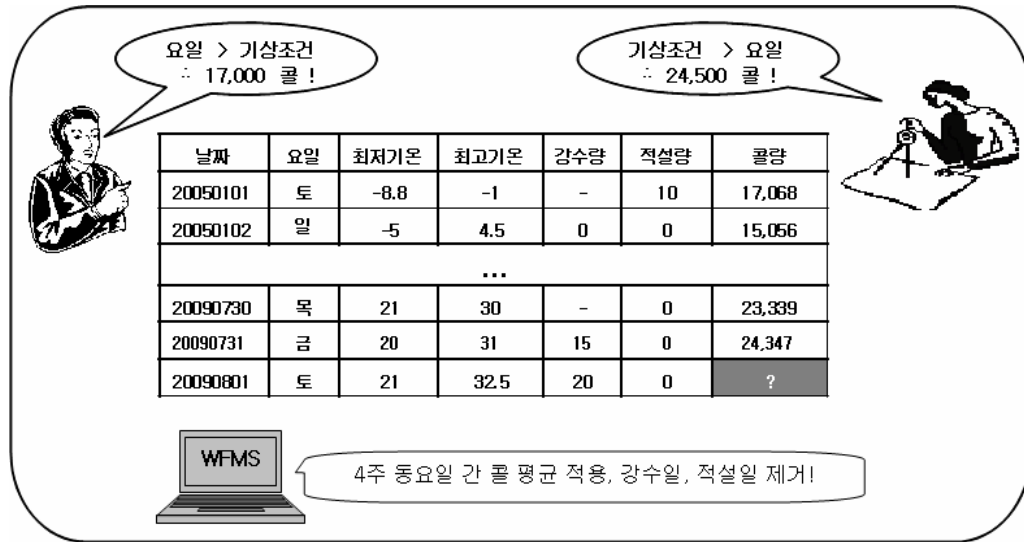
콜센터에서 현실적으로 필요한 적정 인력은 RSF(Rostered Staff Factor)(김윤배 외, 2003)를 고려하여 산출되며, 이 값은 업무일 당일에 발생하는 콜을 일정한 대기 지연시간 내에 모두 응대하기 위해 필요한 최소 인력을 의미한다. 최근 콜센터에서는 이러한 적정 운영 인력의 규모를 예측하기 위해 WFM(Work Force Management) 업무를 전담하는 부서를 별도로 운영함으로써 효율적인 인력 관리로도 모하고 있다. 일반적인 콜센터의 업무 부하는 콜량, 통화시간, 후처리시간 등의 요인들로 산출이 가능하며, 통화시간과 후처리시간을 합하여 상담처리시간이라 한다. 상담 처리시간은 상담 프로세스 및 화법 교육에 따라 조절이 가능한 항목임에 반해, 콜량의 경우 외부적인 요인, 즉 고객이 콜센터로 요청한 건수에 따라 결정되는 통제 불가능 항목이다. 따라서 업무부하의 체계적 관리를 위해서는 콜량의 정확한 예측이 반드시 필요하며, 정확한 콜량 예측을 통해 적정 필요인원을 산출하는 능력은 WFM 부서의 핵심 역량으로 인식되고 있다.



(a) 일반적인 콜센터의 평균 구성

(b) S사 콜센터의 구성 예

<그림 1> 콜센터 운용비 구성 비교



<그림 2> 예측자의 상이한 주관에 기인한 예측치의 불일치

콜량 예측을 위해 현업에서 실제로 가장 널리 적용되는 방법은 담당자의 직관에 의한 방법으로, 오랜 업무 경험을 가진 예측 담당자가 콜량에 영향을 주는 변수를 식별하고 콜량 패턴을 발견할 수 있는 충분한 능력을 갖고 있다는 가정 하에 수행된다. 하지만 이러한 방식은 사람의 주관적인 판단과 직관에 의존하므로 담당자의 역량 및 경험에 따라 예측 결과가 상이하게 나타난다는 한계를 갖고 있다. 즉 담당자들 간에도 콜량에 영향을 주는 변수 및 그 우선순위에 대한 견해 차이가 존재하므로, 동일한 데이터에 대해서도 담당자들이 서로 다른 예측치를 산출하는 경우 <그림 2>가 비일비재하게 발생한다. 또한 이 방법은 결과 도출 과정에 대한 논리적 설명이 어렵기 때문에, 예측 모형을 일반화하여 공유할 수 없다는 본질적인 한계를 갖고 있다.

담당자의 직관에 근거한 예측 방식의 한계를 극복하기 위해, 최근 콜센터 업계에서는 WFMS(Workforce Management System) 패키지의 활용도가 높아지고 있다. WFMS는 WFM 부서의 전반적 업무 수행

을 지원하기 위한 예측, 스케줄링, 트래킹, 레포팅 등의 모듈을 포함하고 있다. WFMS의 예측 모듈은 담당자의 직관에 의한 예측 과정을 시스템화한 것으로, 과거 일정 기간의 동요일 중 특수한 이벤트가 있었던 날을 제외한 콜량의 평균을 기준으로 하고, 이에 담당자가 경험에 의한 보정을 수행하는 방식으로 예측이 이루어진다(MPC, 2003). WFMS은 예측 결과를 신속하게 산출할 수 있을 뿐 아니라, 관련 교육을 통해 통계적 지식이나 업무지식이 미비한 담당자라 하더라도 쉽게 사용이 가능하다는 장점을 갖는다. 하지만 이 방법은 기본적으로 담당자의 직관적 판단에 기초한 콜예측 방법 중 Best Practice를 시스템화한 것이므로 예측 결과에 대한 설명이 거의 불가능하다는 한계를 갖고 있다. 더욱이 시스템 도입 초기에 매우 고가의 구축비용이 발생하며, 콜량에 영향을 주는 신규 요인 발굴시 이를 시스템에 즉시 반영하기가 매우 어렵다는 현실적인 한계점을 갖고 있다. 이와 같이 콜센터 업계에서는 상담 인력 운용의 최적화를 위한 인입 콜량 예측의 필요성이

꾸준히 제기되어 왔음에도 불구하고, 실제 현업에서는 매우 제한된 형태의 단순한 예측 방법만이 사용되고 있으며 관련 연구의 실무 적용 시도 역시 활발히 이루어지지 않고 있다. 본 연구에서는 이러한 현실을 극복하기 위해 설명력이 뛰어날 뿐 아니라 실무 적용 가능성이 높은 콜량 예측 모델을 개발하고자 한다.

본 연구에서는 데이터마이닝의 Classification-/Prediction 기법을 이용함으로써 설명 가능하고 일반화 가능한 체계적인 콜 예측모형을 도출하고자 한다. 또한 순수하게 과거 데이터에만 근거하여 도출된 모형과 업무 배경지식을 일부 반영하여 도출된 모형간의 성능 차이를 비교하고자 한다. 즉 실제 콜량 예측 업무에 있어서는 콜량에 예측을 미치는 변수들의 우선순위가 상황에 따라 다르게 적용되는데, 이러한 현상을 데이터마이닝을 통한 예측 모형의 수립 과정에 적용하고자 한다. 즉 연구의 목적은 단순하게 모형을 개발하는 것에 국한되지 않으며, 제안하는 모형과 현재 가장 널리 사용되고 있는 상용화된 기법간의 예측 적중률을 비교하는 것을 포함한다. 연구 결과의 실무 활용 가능성을 평가하기 위해, 시뮬레이션을 통해 생성된 데이터가 아닌 국내 최대 수준의 자동차 보험사가 운영하고 있는 콜센터의 실 데이터를 사용하여 모형의 성능을 평가하고자 한다.

본 논문의 이후 구성은 다음과 같다. 다음 장인 제 2장에서는 콜센터 업무 일반, 콜 예측 그리고 데이터마이닝 기법을 이용한 예측에 대한 기존의 연구를 간략하게 소개한다. 제 3장에서는 본 연구의 범위를 규정하고, 업무 배경지식을 수용한 데이터마이닝 기반 콜 예측 모형을 제안한다. 제안하는 두 가지 모형과 현업에서 활발하게 사용되고 있는 WFMS에 대한 예측력 평가는 제 4장의 사례 연구에서 수행되며, 마지막 장인 제 5장에서는 본 연구의

기여와 한계, 그리고 향후 연구 방향을 제시한다.

2. 관련 연구

2.1 콜센터 운영 관련 연구

최근 기업들은 고객의 요구가 다양해지고 기대 수준이 높아짐에 따라 고객 개개인의 기호와 성향에 맞는 서비스를 제공하기 위해 고객과의 관계 형성을 중요시하게 되었으며, IT기술을 활용한 강력한 고객관계관리 능력이 기업의 핵심 경쟁력이라는 인식(김재전 외, 2004)을 공유하게 되었다. 특히 최근에는 고객들의 비대면 접점 서비스에 대한 이용도가 높아짐에 따라(KMAC, 2009), 대표적인 비대면 채널의 하나인 콜센터 운영을 통한 고객만족도 향상에 대한 연구가 활발하게 이루어지고 있다. CMF(Callcenter Manager Forum)의 연구(박종태, 2007)에 의하면 콜센터의 성과는 기업의 매출과 밀접하게 연결되어 있으며, 콜센터의 역할이 증가할수록 콜센터 성과 관리의 필요성이 더욱 증대됨을 알 수 있다.

콜센터 업무의 성과 달성 수준을 측정하기 위한 많은 방법들이 제시되었으며, 그 가운데 콜센터 전문 연구기관인 CCDQ(Center for Customer Driven Quality)에서 제시한 성과지표(Anton et al., 2001)가 업계 표준처럼 사용되고 있다. CCDQ에서는 인입 콜센터의 핵심 성과지표로 최초 콜 처리율, 스케줄 준수율, 평균 대기시간, 평균 포기율, 그리고 평균 응대속도를 제시했는데, 이러한 요인들은 실제 현업에서도 가장 중요한 핵심 지표로 활용되는 예가 많다. 콜센터 업무의 성과 측정은 이러한 핵심 성과지표의 목표수준을 정하고, 각각의 지표에 대한 현재 달성도를 산출하는 방식으로 수행되는 것이 일반적이다.

콜센터 업무는 상담원과 고객과의 대면적 상호 작용(서창적, 이세영, 2007)을 중심으로 수행되며, 이러한 특성으로 인해 콜센터는 CRM의 전초기지로 인식되고 있다. 더불어 최근 고객만족도의 중요성이 부각되면서 고객 만족도가 콜센터 성과 측정에도 포함되는 경향을 보인다. 즉, 외부기관의 콜센터 평가 과정에서 한국산업의 서비스 품질지수(KSQI), 한국산업의 고객 만족지수(KCSI) 등이 결과에 포함되는 경우가 많을 뿐 아니라, 내부적으로도 정기적, 비정기적 모니터링을 통해 고객 만족도를 측정하고 있다. 콜센터에서의 고객만족도는 물리적 측면과 주관적 측면의 두 가지 측면(오봉연, 김용석, 2007)으로 구성된다. 고객이 정확한 정보를 제공받으며 상담원이 필요한 모든 정보를 조회할 수 있어야 한다는 것이 서비스 품질의 물리적 측면이라면, 고객이 가능한 한 빠른 시간 이내에 상담을 시작해야 해야 한다는 것은 품질의 주관적인 측면에 해당된다. 이처럼 물리적 측면 뿐 아니라 주관적 측면에서도 고객 만족도를 향상시키기 위해서는 각 상담원에게 지나치게 많은 부하가 걸리는 것을 방지해 줄 수 있는 스케줄링이 필요하며(김윤배 외, 2003), 최적의 스케줄 수립을 위해서는 응대에 필요한 적정 인원을 산출하기 위해 콜량을 정확히 예측하는 것이 반드시 필요하다.

2.2 콜량 예측 관련연구

콜센터의 운영 프로세스(Kungle and Maluchik, 1998)는 예측, 피크타임의 최대 전화 수용량 계산, 필요 인력 계산, 그리고 스케줄 작성으로 구성된다. 즉 인입 콜량에 근거하여 적정 인력 산출 및 인력의 배치가 단계적으로 이루어지기 때문에, 콜량의 정확한 예측은 최소 비용의 고객만족도 향상(Noiman, 2006)의 성과를 좌우하게 된다. 콜센터 운영비용은

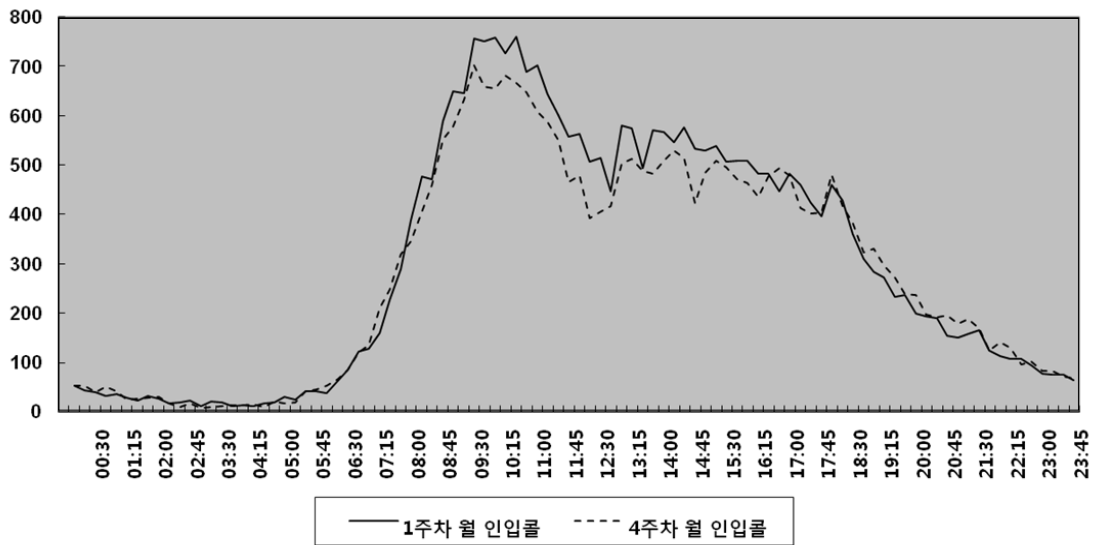
투입되는 인력 규모에 따라 크게 좌우되는데, 필요 서비스 수준보다 적은 인력을 배치하는 언더스태핑(Under Staffing)은 고객의 비용소요를, 제공할 서비스 수준보다 많은 인력을 배치하는 오버스태핑(Overstaffing)은 콜센터의 비용소요를 초래하게 된다(Andrews and Cunningham, 1995). 따라서 언더스태핑 및 오버스태핑으로 인한 비용 낭비를 줄이기 위해, 콜량을 정확하게 예측하기 위한 다양한 시도가 콜센터 업계에서 많은 관심과 지원을 받으며 수행되고 있다.

콜 예측은 수집된 과거 데이터에 대해 계절성과 일별 시간흐름에 따른 트렌드를 분석하고, 이러한 패턴에 따라 연간, 월간, 주간 및 일별 콜을 예측하는 과정(Dasari, 2008)을 통해 수행된다. 이러한 패턴은 <그림 3>에서 보는 바와 같이 일정한 주기를 나타내게 되며, 동요일 콜량의 경우 하루 중 대부분의 시간대에 걸쳐 유사한 패턴을 보이는 것으로 나타났다<그림 4>. 이처럼 주기성을 갖는 데이터에 대해서는 시계열 분석을 통해 예측의 정확도를 높이기 위한 연구가 활발하게 진행되어왔다. 그 예로는 전처리를 통한 시계열 내의 이상치 제거, 시계열 모형화 및 예측, 시계열 예측치의 조정 등을 통해 최적 예측치를 도출하기 위한 시도(지원철, 2002)가 있으며, 데이터의 시간적 흐름에 따라 다차원 회귀 분석을 시도한 연구(Shen and Huang, 2006)도 있다. 또한 ARIMA 모델 기반의 시계열 예측 모델을 제시한 연구(Andrews and Cunningham, 1995)와 K-최근린분류(K-Nearest Neighbour) 알고리즘을 사용하여 데이터 조건이 가장 유사한 과거 데이터를 찾아내는 연구(Bhulai et al., 2005)도 콜 예측의 정확성을 높이기 위한 대표적 연구로 들 수 있다.

본 연구에서는 데이터마이닝의 예측(Classification or Prediction) 기법을 활용하여 콜 예측 모형을 개발하고자 한다. 데이터마이닝의 예측 기법은 목표



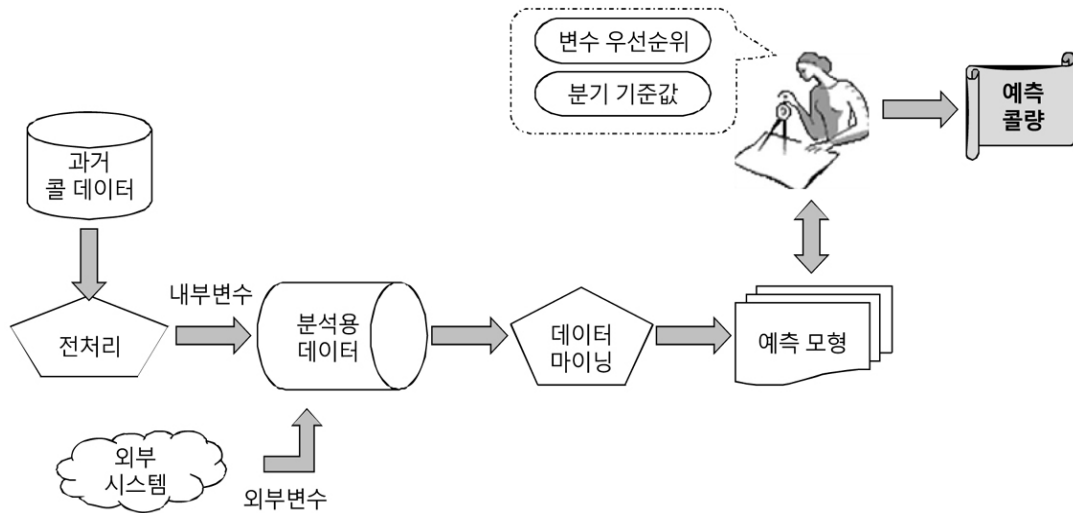
<그림 3> 일별 콜량의 주별 주기성



<그림 4> 동요일 콜량의 시간대별 변화(월요일)

변수를 설정하고 이 값을 예측하기 위한 다양한 응용에 널리 활용되고 있는데, 콜 예측 분야에 있어서는 아직 그 활용도가 낮다고 할 수 있다. 타 분야에 데이터 마이닝의 예측 기법을 적용한 최근 연구는 다음과 같다. 인공지능망 모델을 이용한 예측의 사례

로는 항만 물동량 예측을 위한 연구(전찬영, 송주미, 2006)와 제조 공정의 불량율 예측 연구(정영수, 강창욱, 2004) 등이 있으며, 의사결정나무 모형을 이용한 사례로는 산재보험 부정 수급 식별을 위한 연구(함승오, 홍정식, 2008)와 토양의 온도 및 습도를 예측하기



<그림 5> 콜량 예측 모형 개발의 전체 구조

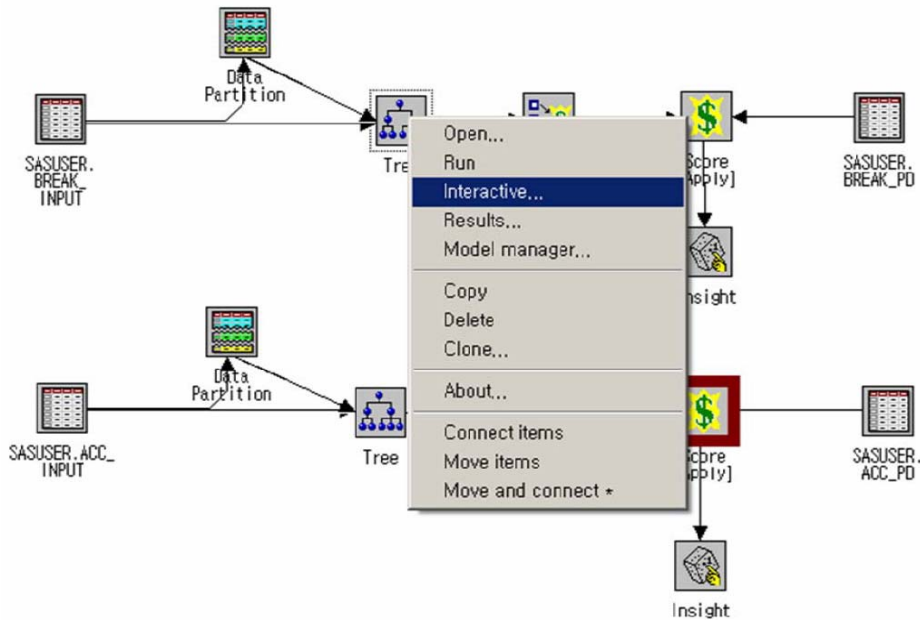
위한 연구(Myers et al., 2008) 등을 들 수 있다. 또한 신경망, 시계열 분석 및 판단 보정 기법을 이용한 교통량 예측 모델(장석철 외, 2005), 통신 사용자의 부적절한 사기 행위 예측을 위한 모델(Wu et al., 2007) 등도 데이터마이닝의 예측 기법을 이용한 사례로 들 수 있다.

3. 대화식 의사결정나무를 활용한 예측 모형 개발

본 연구에서 개발하고자 하는 예측 모형의 전체 구조는 <그림 5>와 같다. <그림 5>에서 과거 콜 데이터는 요일, 날짜, 콜량 등을 포함한 과거 데이터로, 전처리 과정을 통해 분석의 내부변수로 사용된다. 또한 각종 이벤트 및 기상 데이터 등이 콜량에 영향을 미칠 수 있으므로, 이들을 외부변수로 정의하여 분석에 사용한다. 내부 및 외부 변수에 대한 데이터 마이닝 분석을 통해 다양한 예측 모형이 도출할 수 있으며, 특히 대화식 의사결정나무 모형을

사용함으로써 최종 콜량 예측 과정에서 변수 우선 순위 및 분기 기준값 설정 과정에 전문가의 지식을 반영하고자 한다.

데이터 마이닝의 예측 모형은 예측에 필요한 변수와 규칙을 통계적 또는 수학적 알고리즘에 의해 찾아내는 분석방법이다. 하지만 통계적 또는 수학적으로 최적인 알고리즘이 일상 생활이나 실제 환경에서 항상 최적의 해를 제공하는 것은 아니다. 또한 데이터 마이닝을 통해 만들어진 모형을 데이터와 근무요건이 지속적으로 변화하는 업무 환경에 적용할 경우, 모형과 현실의 불일치가 자주 발생하게 될 우려도 있다. 데이터 마이닝의 대표적 분석 기법인 의사결정나무 모형 또한 마찬가지로 통계적, 수학적으로 최적의 분기 규칙을 생성하지만, 이 역시 실무 환경에서의 경험적 지식과는 무관하게 나타날 가능성이 있다. 따라서 현장 업무에 대한 경험과 실제 적용상의 문제점을 함께 고려하여 동적으로 분기 규칙을 생성하는 접근법이 더욱 바람직하다고 할 수 있다. SAS Enterprise Miner의 의사



<그림 6> 대화식 의사결정나무의 수행 예

결정나무(Decision Tree) 노드는 동적 분기를 위한 “Interactive Training Tree” <그림 6> 기능을 제공한다. 이 기능은 Tree 옵션에 따라 기본 모형을 생성한 후, 실무 지식을 가진 사용자가 보다 변수의 영향력 우선순위 및 분기 기준점을 대화식으로 수정함으로써 의사결정나무의 형성 과정에 개입할 수 있는 기회를 제공한다.

본 연구에서는 현장 업무에 대한 경험과 실제 적용상의 문제점을 고려하여 예측 실행 단계에서 동적으로 분기 규칙을 생성하는 대화식 의사결정나무 모형을 개발하고자 한다. 이러한 대화형 의사결정나무는 과거 데이터로부터 도출된 모형 위에 담당자의 업무 배경지식을 추가함으로써 모형의 예측력을 향상시킬 수 있다는 장점을 갖는다. 하지만 대화를 통한 의사결정나무 형성 과정에서 담당자가 잘못된 지식이나 오판에 의한 결정을 내릴 경우는 오히려 일반적인 의사결정나무에 비해 예측력이 저하

될 수 있다는 단점을 갖고 있다. 한편, 순수하게 과거 데이터에만 근거한 일반 의사결정나무 모형의 경우 비일상적인 이벤트, 즉 통계적 관점에서 이상치로 파악되는 관측치에 대한 세밀한 접근이 불가능하다는 한계를 갖는다. 예를 들어 자동차 보험사 콜센터의 경우, 고장 콜량과 사고 콜량은 외부 이벤트와 기상 상황에 의해 크게 좌우되는 것으로 알려져 있다. 예를 들어 7~8월 즉 무더위가 있는 여름의 콜 예측에 가장 중요한 영향을 끼치는 요인은 강수량과 낮 최고 기온 등이며, 겨울철 콜 예측에는 강수량과 아침 최저 기온 등이 중요하게 작용한다는 것을 담당자는 업무적 경험을 통해 알고 있다. 하지만 이러한 전문 지식은 일반 의사결정나무 기반의 모형 수립에는 적용될 방법이 없다. 이와 같이 대화식 의사결정나무와 일반 의사결정나무가 서로 장단점을 갖고 있으므로, 본 논문의 이후 부분에서는 대화식 의사결정나무, 일반 의사결정나무, 그리고

WFMS의 콜량 예측력을 함께 비교하고자 한다.

4. 사례 연구 : S 자동차 보험사 콜센터의 콜량 예측 사례

본 장에서는 자동차 보험사에서의 콜 예측에 주로 사용되어 온 기존 변수들 및 본 연구의 수행 과정에서 새로 도출된 파생 변수들을 설명하고, 그 변수들에 대해 의사결정나무 기법을 적용하여 생성된 두 가지 예측 모형을 소개한다. 두 가지 예측 모형은 담당자의 업무 배경지식을 반영한 모형 및 단순한 의사결정나무 기법을 통해 생성된 모형의 두 가지이며, 이 두 모형과 기존의 WFMS의 예측력을 실험을 통해 평가하고자 한다. 본 연구에서 수행하고자 하는 실험의 개요를 요약하면 <표 1>과 같다.

<표 1> 실험 개요

제안 모형	<ul style="list-style-type: none"> 일반 의사결정나무 및 대화식 의사결정나무를 통해 도출된 예측 모형 SAS Enterprise Miner를 이용하여 도출
비교 모형	<ul style="list-style-type: none"> WFMS(Workforce Management System) 패키지의 예측 모듈
예측 대상	<ul style="list-style-type: none"> 일별 사고 콜량, 일별 고장 콜량
실험 데이터	<ul style="list-style-type: none"> 국내 최대 규모의 자동차 보험사인 S사의 4년 8개월간 실 데이터 Training and Validation : 2005. 1. 1 ~ 2008. 12. 31, Test : 2009. 1. 1 ~ 2009. 8. 31

4.1 변수 도출 및 설명

콜 예측을 위한 변수는 크게 내부 변수와 외부 변수로 나뉜다. 내부 변수는 요일, 날짜, 콜량 등을 포함한 과거 데이터를 의미하며, 외부 변수는 콜센터 외부로부터 획득되는 데이터로 주로 기상 데이터를 의미한다. 내부 변수 중 제안하는 모형을 통해 예측

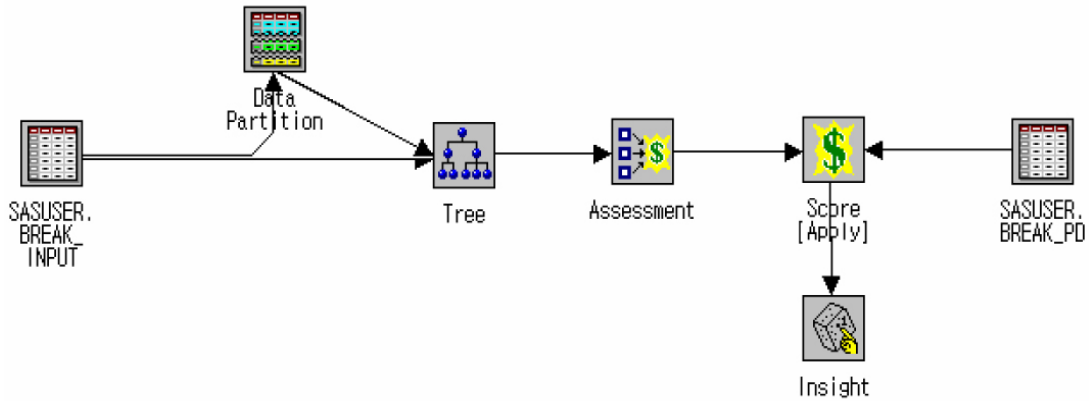
하고자 하는 목표 변수는 콜량이며, 세부적으로는 일별 고장 콜량 및 일별 사고 콜량으로 구분된다. 이외의 내부 변수는 기본적으로 연도, 월, 일, 요일 등 날짜와 관련된 변수로 구성되어 있으며, 이로부터 공휴일 관련 변수, 1주 전 콜량, 1년 전 콜량 등의 추가 내부 변수를 도출하여 사용할 수 있다. <표 2>는 콜량 예측과 관련된 변수 목록을 보여주고 있으며, 기본 변수로부터 새롭게 유도한 파생 변수는 회색 바탕으로 구분하여 나타내고 있다. 이러한 추가 변수들은 콜량의 일별, 주별, 월별, 연도별 등의 시간의 흐름에 따른 추세와 공휴일의 영향을 반영하기 위해 도출한 것이다.

<표 2> 콜량 예측을 위한 관련 변수 목록

변수 유형	변수명	Type
목표 변수	사고 콜량, 고장 콜량	Interval
입력 변수	1주전/2주전/4주전/1년전 사고 콜량	Interval
	1주전/2주전/4주전/1년전 고장 콜량	Interval
	날짜 - 연도, 월, 일, 요일	Ordinal
	전일 공휴일 여부, 공휴일 여부, 익일 공휴일 여부	Binary
외부 변수	최저기온, 최고기온, 강수량, 적설량	Interval

4.2 예측 모형 설계

예측 모형은 상용 통계 패키지인, SAS Enterprise Miner의 의사결정나무(Tree) 모형을 통해 구축하였다<그림 7>. 예측해야 할 목표 변수가 사고 및 고장의 두 영역으로 나뉘어져 있으므로, 모델링 역시 유사한 조건으로 두 가지로 구분하여 수행하였다. 입력 변수 및 목표 변수는 <표 2>에 소개된 바와 같으며, 분기 기준은 F Test에서 Significance



<그림 7> SAS Enterprise Miner를 이용한 콜 예측 모델링

level을 0.200으로 지정하여 적용하였다. 이 외의 상세 옵션은 <표 3>에 정리하였다.

<표 3> Tree 노드의 주요 옵션

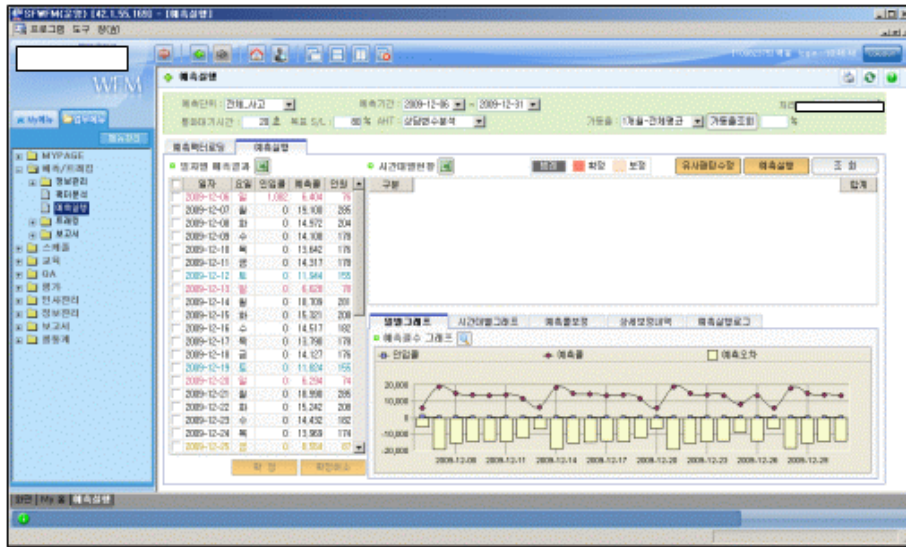
항 목	설정값
Minimum number of observations in a leaf	5
Observations required for a split search	15
Maximum number of branches from a node	12
Minimum depth of tree	100

본 실험에는 국내 자동차 보험사 S사 콜센터의 4년 8개월 간의 실 데이터가 사용되었다. 즉 2005년 1월 1일부터 2008년 12월 31일까지의 사고 콜 및 고장 콜에 대한 데이터로부터 예측 모델을 수립하고, 이 모형을 통해 2009년 1월 1일부터 2009년 12월 31일까지의 콜을 예측함으로써, 모형의 예측력을 평가하였다. 예측 모형의 정확도 평가를 위해 평균제곱오차 (Mean Squared Error, MSE), 평균절대백분비오차 (Mean Absolute Percentage Error, MAPE), 절대백분비오차의 중위수 (Median Absolute Percentage Error, MAPE) 등의 척도가 사용될 수 있지만, 본 실험에서는 다음과 같은 기준을 정의하고 사용하였다.

$$(\text{적중률}) = (\text{실제 값이 예측 값의 허용 오차범위 내에 위치한 건수}) / (\text{전체 예측 건수})$$

이와 같은 적중률 기준을 정의하여 사용하는 것은, 구간 변수인 실제 콜량을 오차 없이 정확하게 예측하는 것은 큰 의미가 없을 뿐 아니라 불가능에 가깝고, 예측 콜량에 따라 적정인력 규모의 여유 구간을 설정하여 운영하는 실제 업계 특성을 반영하기 위한 것이다. 이러한 정의에 따라 정확도를 산출하는 경우 적중률은 허용 오차범위 값에 따라 큰 영향을 받을 수 있다. 따라서 오차 허용률을 다양하게 변화시키면서 각 모형의 예측 정확도를 비교할 예정이며, 일반적으로 콜센터 업무에서 여유 인력 규모를 예측 인력의 약 10% 내외로 유지하는 관행을 고려하여 10%의 오차 허용률에 대한 실험도 포함할 것이다.

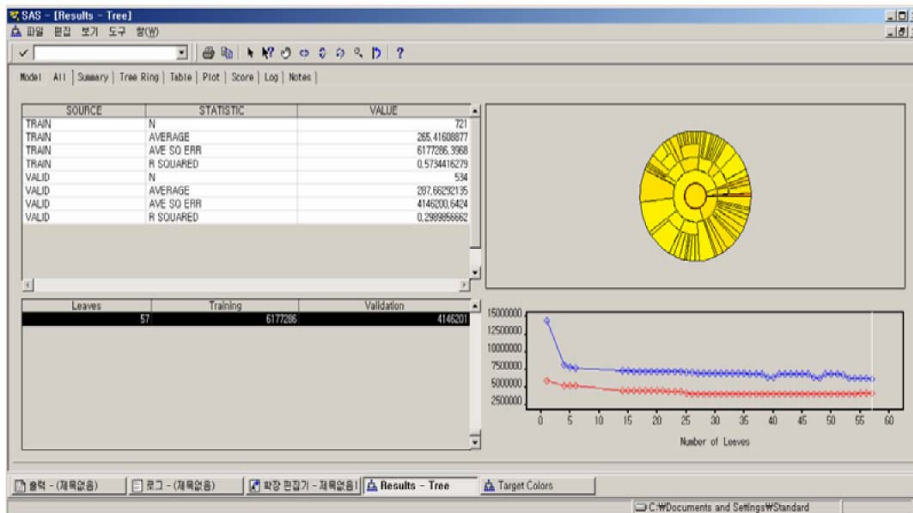
예측력의 비교 실험은 업계에서 가장 널리 사용되고 있는 예측 시스템 중의 하나인 WFMS를 대상으로 수행한다. <그림 8>은 WFMS의 사고콜 예측 화면의 한 예이며, 예측을 원하는 일자를 선택하고 예측 실행 버튼을 클릭하면, 예측 콜량 및 그에 따른 필요인력을 계산하여 결과로 보여준다.



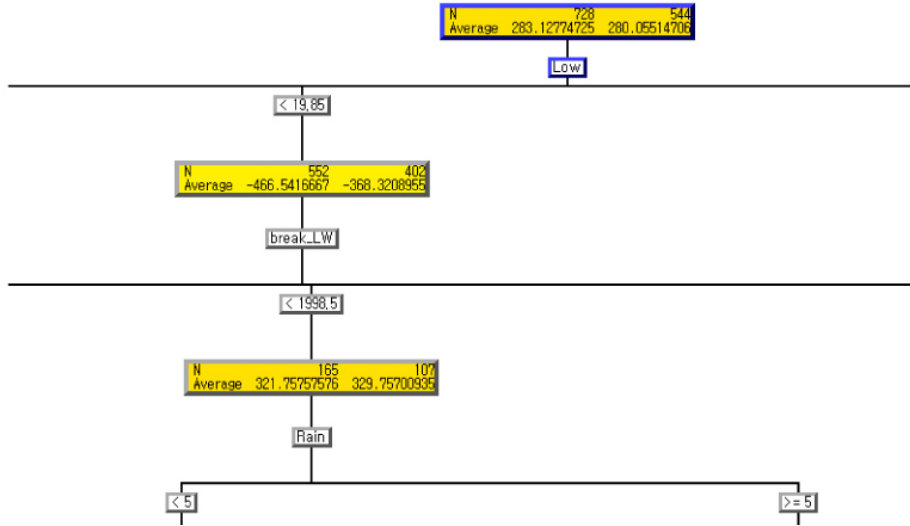
<그림 8> WFMS의 사고콜 예측 화면 예시

<그림 9>은 SAS Enterprise Miner의 의사결정 나무를 사용하여 예측모형을 수립하는 과정 및 결과의 일부이다. 그림 (a)는 생성된 의사결정나무 모형을 요약한 화면으로 Score Table, Tree Ring, Training

과 Validation결과에 따른 Leaves 개수를 보여준다. (b)는 이렇게 생성된 의사결정나무 모형의 한 부분이며, (C)는 2009년 1월부터 8월까지의 예측을 실행한 결과 값을 보여준다.



(a) 의사결정나무 모형 생성 요약

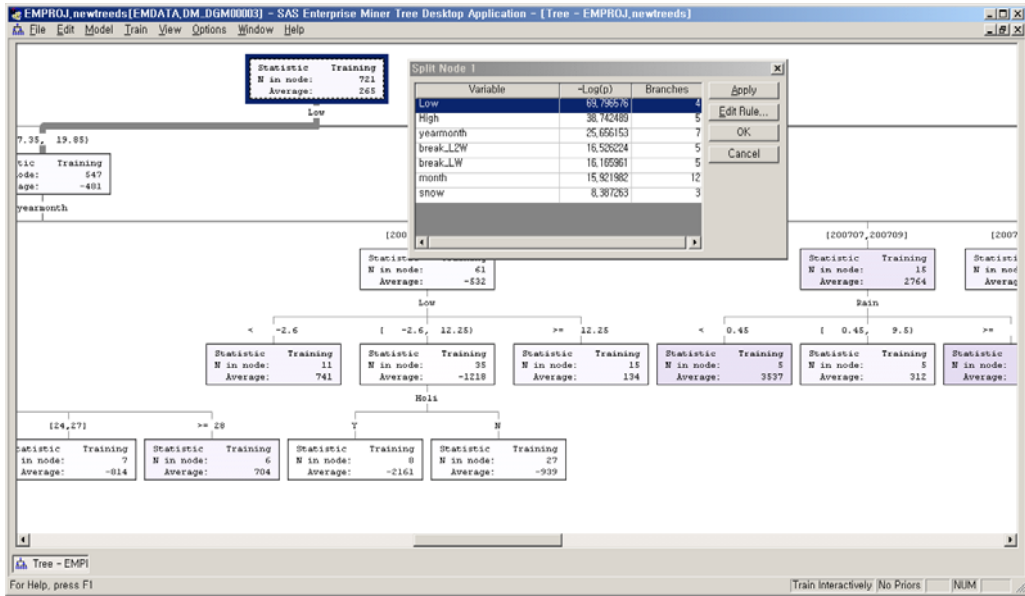


(b) 의사결정나무 모형의 일부분

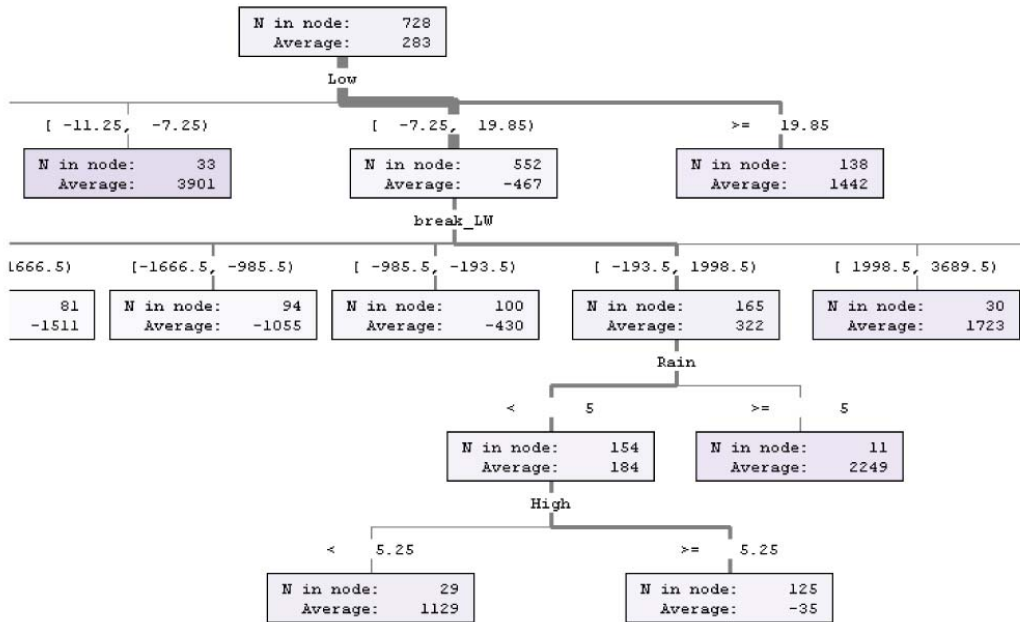
영역	구간	구간	구간	구간	구간	구간	구간	구간	구간	구간	구간	영역	구간	구간	구간	구간	구간	구간	구간
hSat	Break_Vd	Break_LW	Break_LW	Break_LM	Break_LY	Low	High	Rain	snow	WARN	NODE	LEAF	P_BREAK_D	V_BREAK_D	구간	구간	구간	구간	구간
N	3384	-2571	-1455	-1196	1545	-9.5	-2.5	0.0	0.0		32	3	4505.3750	3415.2857					
N	577	4001	-357	7182	1128	-6.9	1.2	0.0	0.0		53	36	1712.8750	138.6667					
N	4926	2464	-262	44872	233	-5.6	4.0	0.0	0.0		53	36	1712.8750	138.6667					
N	483	-1875	-1212	5831	-2395	-2.8	4.5	0.0	0.0		54	37	-557.9792	202.2963					
N	-2443	527	4218	5193	1748	-3.4	1.2	0.0	0.0		53	36	1712.8750	138.6667					
N	845	-330	1505	-110	13	-5.4	1.4	0.0	0.0		53	36	1712.8750	138.6667					
N	-758	3384	291	-1264	-940	-4.2	4.2	0.0	0.0		53	36	1712.8750	138.6667					
N	-1474	577	-2571	-1613	-785	-4.0	2.5	0.0	0.0		53	36	1712.8750	138.6667					
N	-1866	4926	4001	-349	3528	-6.8	-0.5	0.0	0.0		53	36	1712.8750	138.6667					
Y	-739	483	2464	-763	1293	-9.9	-4.3	0.0	0.0		32	3	4505.3750	3415.2857					
N	1911	-2443	-1875	-1009	1222	-9.6	-4.2	0.0	0.0		32	3	4505.3750	3415.2857					
N	1816	845	527	792	3669	-11.0	-3.2	0.0	0.0		32	3	4505.3750	3415.2857					
N	9701	-758	-330	-554	2204	-9.0	-1.5	0.0	0.0		32	3	4505.3750	3415.2857					
N	3164	-1474	3384	-1914	3264	-9.4	-3.2	0.0	0.0		32	3	4505.3750	3415.2857					
N	824	-1866	577	-1455	13946	-11.2	-2.0	0.0	0.0		32	3	4505.3750	3415.2857					
N	8452	-739	4926	-357	4203	-4.8	0.4	3.3	5.1		53	36	1712.8750	138.6667					
N	1623	1911	483	-262	2813	-5.9	3.1	0.0	0.0		53	36	1712.8750	138.6667					
N	1107	1816	-2443	-1212	-1234	0.1	6.1	0.0	0.0		54	37	-557.9792	202.2963					
N	-244	9701	845	4218	1810	-2.7	4.7	0.0	0.0		54	37	-557.9792	202.2963					
N	1153	3164	-758	1505	1032	-2.4	7.2	0.0	0.0		54	37	-557.9792	202.2963					
N	-1496	824	-1474	291	1473	1.0	5.9	0.0	0.0		54	37	-557.9792	202.2963					
N	-1741	8452	-1866	-2571	6039	-2.0	9.3	0.0	0.0		54	37	-557.9792	202.2963					
N	-1046	1623	-739	4001	5856	-11.0	-2.0	0.0	0.0		32	3	4505.3750	3415.2857					
Y	3100	1107	1911	2464	2380	-11.4	-5.1	2.0	3.5		2	1	27187.2000	11095.5000					
N	8921	-244	1816	-1875	393	-9.3	1.0	0.4	0.4		32	3	4505.3750	3415.2857					
N	4079	1153	9701	527	1727	-5.3	-0.2	0.0	0.0		53	36	1712.8750	138.6667					
N	662	-1496	3164	-330	-131	-6.8	3.9	0.0	0.0		53	36	1712.8750	138.6667					
N	1690	-1741	824	3384	69	-3.0	7.8	0.0	0.0		53	36	1712.8750	138.6667					
N	4641	-1046	8452	577	1441	-1.2	8.3	0.0	0.0		54	37	-557.9792	202.2963					
N	-1535	3100	1623	4926	2580	0.7	12.2	0.0	0.0		54	37	-557.9792	202.2963					
N	-1758	8921	1107	483	805	3.4	11.7	0.0	0.0		54	37	-557.9792	202.2963					
N	-2363	4079	-244	-2443	-1354	1.6	10.5	0.0	0.0		61	44	-2779.8899	-2234.5556					
N	-3704	662	1153	945	2499	2.1	8.0	0.0	0.0		60	43	-391.9000	-1312.7500					
N	-1465	1690	-1496	-758	2837	2.9	11.1	0.0	0.0		56	39	-2416.1657	-2520.4000					
N	-2677	4641	-1741	-1474	226	1.5	8.9	0.0	0.0		57	40	-2485.0909	-2426.4000					
N	-2591	-1535	-1046	-1866	866	3.2	6.0	0.2	0.0		59	42	-2026.8750	-1886.4000					
N	-3002	-1758	3100	-739	875	0.6	6.8	0.0	0.0		58	41	-680.6000	-1679.6000					
N	-1666	-2363	8921	1911	527	0.4	5.0	0.0	0.0		55	38	-1711.9000						
N	-2096	-3704	4079	1816	-1098	-2.2	8.2	0.0	0.0		61	44	-2779.8899	-2234.5556					
N	-2635	-1465	662	9701	2286	2.6	8.1	0.0	0.0		60	43	-391.9000	-1312.7500					
N	-2091	-2677	1690	3164	581	2.0	6.5	0.0	0.0		56	39	-2416.1657	-2520.4000					
N	-2942	-2591	4641	824	4207	3.2	9.2	0.0	0.0		57	40	-2485.0909	-2426.4000					
N	-2988	-3002	-1535	8452	384	1.8	10.6	0.2	0.0		59	42	-2026.8750	-1886.4000					
N	-2690	-1666	-1758	1623	-260	4.3	16.1	34.5	0.0		58	41	-680.6000	-1679.6000					
Y	94	-2096	-2363	8921	1107	852	2.1	5.3	0.0	0.0	55	38	-1711.9000						
N	-1160	-2635	-3704	-244	-1606	-6.4	2.4	0.4	0.5		61	44	-2779.8899	-2234.5556					
N	-3272	-2091	-1465	1153	2432	-8.6	-2.3	0.0	0.0		32	3	4505.3750	3415.2857					
N	2067	-2942	-2677	-1496	-166	-8.6	-0.5	0.0	0.0		32	3	4505.3750	3415.2857					
N	-223	2988	-2591	-1741	-838	-6.0	1.4	0.0	0.0		57	40	-2485.0909	-2426.4000					
N	-1587	-2690	-3002	-1046	-1540	-1.8	3.9	0.4	0.5		59	42	-2026.8750	-1886.4000					
N	-2910	94	-1666	3100	-1479	-6.8	-0.1	1.0	1.3		58	41	-680.6000	-1679.6000					

(c) 의사결정나무 모형을 사용한 예측 결과의 일부분

<그림 9> SAS Enterprise Miner를 이용한 의사결정나무 모형 생성 및 예측



(a) 대화식 의사결정나무에서의 분기 기준 설정 예



(b) 도출된 대화식 의사결정나무 모형의 일부

<그림 10> 대화식 의사결정나무를 이용한 콜 예측 모형 개발

```

IF 최저기온 < -11.25
THEN
    NODE : 2
    N : 5
    AVE : 27187.2
    SD : 23758.6
IF 월 EQUALS 12 AND -11.25 <= 최저기온 < -7.25
THEN
    NODE : 7
    N : 8
    AVE : 9522.13
    SD : 8705.85
IF 년 IS ONE OF: 2005 2006 2007 AND 월 IS ONE OF: 01 02 AND -11.25 <= 최저기온 < -7.25
THEN
    NODE : 17
    N : 17
    AVE : 971.765
    SD : 915.158
IF 년 IS ONE OF: 2008 2009 AND 월 IS ONE OF: 01 02 AND -11.25 <= 최저기온 < -7.25
THEN
    NODE : 18
    N : 8
    AVE : 4505.38
    SD : 3240.74:
    :
    :
    :
    
```

<그림 11> 규칙화된 콜량 예측 모형

<그림 10>는 <그림 9>에서 생성된 의사결정나무를 대화식으로 재실행하는 화면으로, <그림 (a)>는 분기 기준 설정을 위해 대화를 수행하는 과정을, <그림 (b)>는 그 결과로 수정된 새로운 Tree의 일부를 보여준다. 또한 결과 Tree를 규칙으로 변환하여 나타낸 일부가 <그림 11>에 제시되어 있다.

4.3 실험 결과의 분석 및 비교

본 연구에서 제안한 두 예측모형, 그리고 업계에서 널리 사용되고 있는 WFMS의 성능을 비교한 결과가 <표 4>에 요약되어 있다. <표 4>는 세 가지

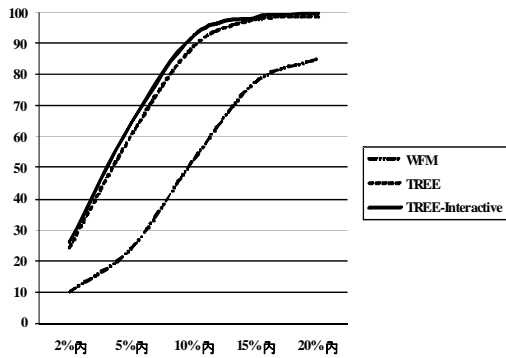
모형에 대해서 오차 허용률을 2%에서 20%까지 변화시켜가며 측정된 콜량 예측 적중률을 나타낸다. 또한 모형간의 성능 비교를 위해 <표 4>의 결과를 도식화한 그래프가 <그림 12> ~ <그림 16>에 나타나 있다.

<그림 12>은 사고와 고장 문의 콜 전체 합계에 대한 예측 적중률을 보여주며, <그림 13>와 <그림 14>은 각각 사고, 고장 문의 콜에 대한 예측 적중률을 보여준다. <그림 12>과 <그림 13>에서 WFMS의 콜 예측력은 단순 의사결정나무와, 대화형 의사결정나무에 비해 오차 허용률의 전 구간에서 현저하게 낮게 나타나는 것을 확인할 수 있고, <그림 12>,

<표 4> 월별, 예측 모형별, 허용오차 범위별 적중률

월	예측모형	2% 내 적중률			5% 내 적중률			10% 내 적중률			15% 내 적중률			20% 내 적중률		
		사고	고장	합계	사고	고장	합계	사고	고장	합계	사고	고장	합계	사고	고장	합계
1	WFM	6.5	9.7	3.2	6.5	19.4	16.1	16.1	25.8	25.8	29.0	45.2	38.7	38.7	61.3	45.2
	Tree	6.5	9.7	16.1	25.8	12.9	35.5	58.1	35.5	71.0	77.4	71.0	83.9	83.9	77.4	90.3
	Interactive	9.7	9.7	22.6	48.4	25.8	58.1	74.2	54.8	80.6	87.1	83.9	93.5	90.3	87.1	100.0
2	WFM	0.0	7.1	3.6	3.6	14.3	3.6	3.6	32.1	10.7	3.6	60.7	21.4	3.6	85.7	39.3
	Tree	32.1	0.0	17.9	75.0	10.7	42.9	89.3	46.4	82.1	100.0	67.9	100.0	100.0	96.4	100.0
	Interactive	32.1	3.6	17.9	82.1	17.9	60.7	96.4	67.9	96.4	100.0	92.9	100.0	100.0	100.0	100.0
3	WFM	0.0	22.8	9.7	0.0	41.9	29.0	0.0	64.5	64.5	12.9	93.5	87.1	54.8	100.0	100.0
	Tree	35.5	9.7	29.0	77.4	25.8	67.7	100.0	71.0	93.5	100.0	96.8	100.0	100.0	100.0	100.0
	Interactive	51.6	12.9	32.3	80.6	29.0	61.3	100.0	71.0	93.5	100.0	96.8	100.0	100.0	100.0	100.0
4	WFM	0.0	33.3	6.7	0.0	66.7	6.7	6.7	93.3	46.7	16.7	100.0	93.3	90.0	100.0	100.0
	Tree	32.3	9.7	38.7	48.4	16.1	74.2	90.3	54.8	93.5	96.8	87.1	96.8	96.8	93.5	96.8
	Interactive	38.7	9.7	29.0	67.7	19.4	58.1	90.3	64.5	93.5	96.8	90.3	96.8	96.8	93.5	96.8
5	WFM	0.0	19.4	12.9	12.9	64.5	35.5	29.0	80.6	74.2	54.8	90.3	93.5	77.4	96.8	96.8
	Tree	16.7	6.7	16.7	40.0	30.0	70.0	76.7	76.7	86.7	83.3	86.7	100.0	100.0	96.7	100.0
	Interactive	16.7	10.0	23.3	36.7	36.7	70.0	76.7	83.3	86.7	83.3	90.0	96.7	100.0	96.7	100.0
6	WFM	0.0	20.0	20.0	3.3	43.3	30.0	26.7	83.3	70.0	53.3	90.0	93.3	90.0	96.7	100.0
	Tree	30.0	13.3	16.7	56.7	36.7	53.3	86.7	73.3	96.7	90.0	100.0	100.0	100.0	100.0	100.0
	Interactive	33.3	13.3	10.0	63.3	43.3	53.3	86.7	73.3	96.7	90.0	100.0	100.0	100.0	100.0	100.0
7	WFM	3.2	9.7	16.1	19.4	38.7	48.4	54.8	64.5	74.2	80.6	87.1	96.8	100.0	90.3	96.8
	Tree	19.4	16.1	16.1	61.3	45.2	61.3	90.3	77.4	87.1	96.8	93.5	96.8	100.0	93.5	96.8
	Interactive	22.6	12.9	22.6	64.5	45.2	64.5	93.5	77.4	90.3	96.8	93.5	96.8	100.0	93.5	96.8
8	WFM	0.0	12.9	6.5	3.2	35.5	19.4	22.6	83.9	51.6	54.8	93.5	90.3	83.9	100.0	96.8
	Tree	22.6	19.4	38.7	58.1	54.8	74.2	90.3	90.3	96.8	100.0	100.0	100.0	100.0	100.0	100.0
	Interactive	22.6	25.8	45.2	61.3	61.3	83.9	93.5	96.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0

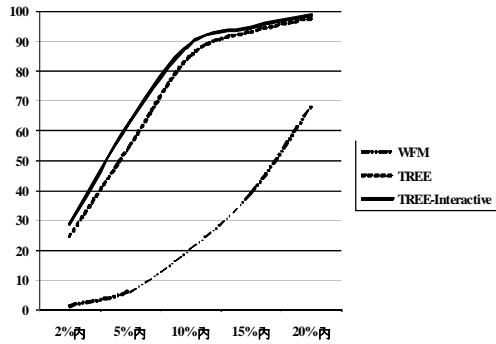
<그림 13>, 그리고 <그림 14>에서 대화형 의사결정나무 모형이 단순 의사결정나무 모형에 비해 전



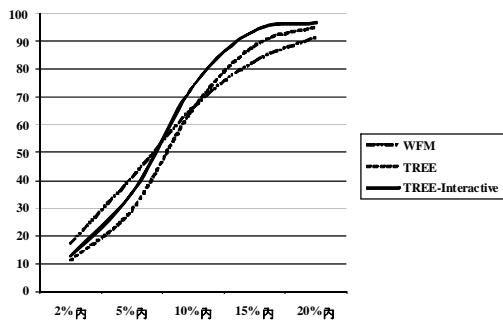
<그림 12> 오차 허용범위에 따른 예측력 비교(전체 콜)

구간에 걸쳐 보다 우수한 예측력을 나타냄을 확인할 수 있었다.

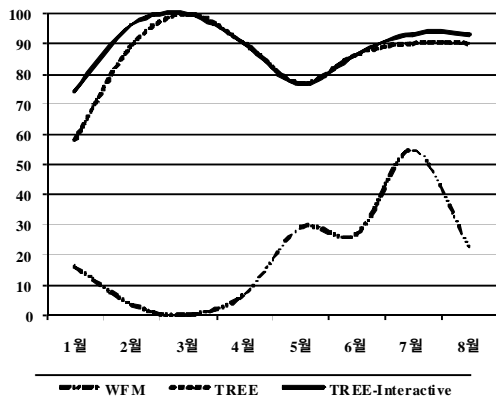
하지만 고장 관련 콜 적중률인 <그림 14>의 경우 세 모형의 적중률이 오차 허용률에 따라 상이하게 나타났는데, 오차 허용범위 7% 이내에서는 WFMS가, 7% 이후에서는 대화형 의사결정나무가 가장 높은 적중률을 보임을 알 수 있었다. 이와 같은 현상은 각 콜의 발생 특성에 기인하는 것으로 파악된다. 즉, 사고 콜의 경우 외부변수 중 강수와 적설, 최저기온의 영향으로 사고가 많이 발생할 수 있으며, WFMS의 콜 예측 모델은 이와 같은 날을 이상치로 판단, 예측 데이터에서 해당 데이터를 제외한 후 담당자의 감각과 직관으로 보정하게 된다. 따라서 <그림 13>의 사고 콜 적중률의 경우 WFMS는 전



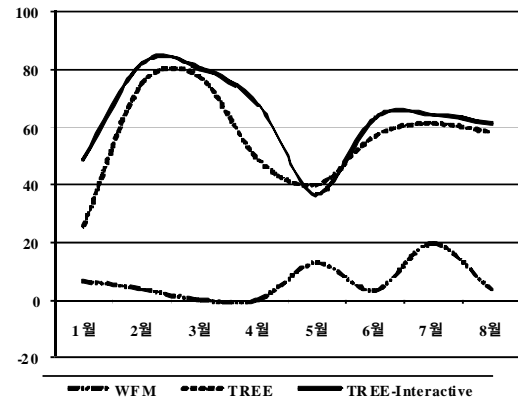
<그림 13> 오차 허용범위에 따른 예측력 비교(사고 콜)



<그림 14> 오차 허용범위에 따른 예측력 비교(고장 콜)



(a) 오차 허용범위 10% 내 적중률(사고 콜)

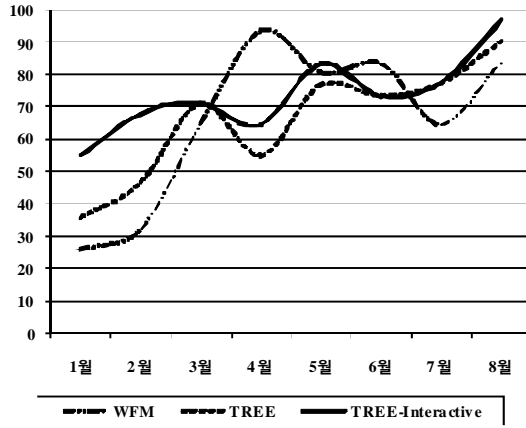


(b) 오차 허용범위 5% 내 적중률(사고 콜)

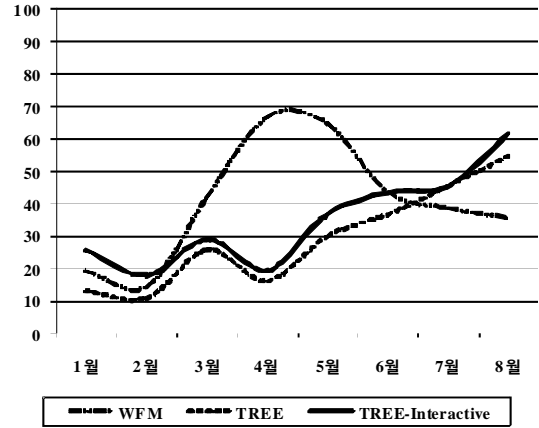
<그림 15> 사고 콜량 예측에 대한 월별 적중률

구간에 걸쳐 매우 낮게 나타난다. 반면 WFMS의 고장 콜 예측의 경우 이상치를 포함하지 않는 경우 적중률이 높게 나타나는데, 특히 기온 및 기후의 변동이 거의 없는 봄철 및 가을철의 적중률이 대단히 높게 나타나서 전체 결과에 영향을 준 것으로 판단된다. 이 현상을 보다 자세히 살펴보기 위해 각 모형의 사고 콜량 및 고장 콜량에 대한 월별 적중률을 분석하였으며, 그 결과가 각각 <그림 15>와 <그림 16>에 제시되어 있다. <그림 15>에서 사고 콜량은 전 구간에 걸쳐 데이터마이닝 기법 기반의 모형이 WFMS에 비해 우수한 적중률을 보이고 있지만, <그림 16>에서 고장 콜량의 경우 일부 구간에서 WFMS가 오히려 더 우수한 적중률을 보임을 알 수 있다. 특히 오차 허용 범위가 5%인 경우 <그림 16(b)> 3월 ~ 5월 사이에 WFMS의 적중률이 매우 높게 나타나는 것을 알 수 있으며, 이는 앞에서 설명한 바와 같이 기온 및 기후의 변동이 거의 없는 봄철의 특성이 반영된 것으로 해석된다.

종합하면, 일반적으로 고장 문의 콜센터 업무에서 예측 인력 규모의 여유 인력으로 약 10% 내외를 유지하는 상황을 고려할 때, 10%의 오차 허용범위에서 우수한 예측력을 보이는 대화형 의사결정나무



(a) 오차 허용범위 10% 내 적중률(고장 콜)



(b) 오차 허용범위 5% 내 적중률(고장 콜)

<그림 16> 고장 콜량 예측에 대한 월별 적중률

기반 모형이 가장 현실 적합성이 뛰어나다고 할 수 있다.

5. 결론

콜센터 운영비의 대부분을 차지하는 인력 운영 비용을 최소화하기 위해, 인입 콜량을 정확하게 예측하기 위한 많은 시도가 있어 왔다. 하지만 현업에서 주로 이루어지고 있는 방법은 담당자의 주관적 판단에 크게 의존한다는 한계, 또는 매우 고가의 시스템 구축을 필요로 한다는 한계를 갖는다. 이를 극복하기 위해 본 연구에서는 데이터 마이닝의 대화식 의사결정나무를 활용하여 업무 배경 지식을 충분히 활용하면서도 실무 적용이 용이한 예측 모형을 제안하였다. 또한 본 연구에서 수립한 모형의 성능 평가를 위해 국내 최고 수준의 시장 점유율을 기록하고 있는 한 자동차 보험사 콜센터의 데이터를 사용한 실험을 수행하였다. 실험 결과 본 논문에서 제안한 모형이 업계에서 널리 사용되는 WFMS에 비해 오차 허용률의 대부분의 구간에 걸쳐 상대적

으로 우수한 예측력을 갖는 것으로 나타났으며, 특히 대화형 의사결정나무 기법을 이용한 모형이 단순 의사결정나무 기법을 이용하여 구축된 모형에 비해 모든 실험 구간에서 우월한 예측력을 갖는 것으로 나타났다. 이 결과에 비추어 볼 때, 본 연구에서 제안하는 모형의 실제 업무 활용도는 매우 높을 것으로 기대된다.

본 연구에서는 날씨, 요일 등의 내부변수 이외에 외부 이벤트 효과를 반영하기 위한 외부 변수를 분석 과정에서 활용하였다. 특히 기온과 기후, 그리고 공휴일 여부에 영향을 많이 받는 자동차 보험사의 특성을 감안할 때, 이러한 외부 변수의 도입이 모형의 예측력을 향상시키는 데에 기여한 것으로 파악된다. 기존의 예측 기법은 이상치에 의한 왜곡을 방지하기 위해 이벤트가 발생한 날의 데이터는 제외하고 콜 예측을 수행하였으며, 이벤트 효과의 반영은 추후 담당자의 주관적 판단에 의해 예측치를 보정하는 형태로 이루어졌다. 본 연구에서 제안한 모형은 이러한 이벤트 효과가 데이터에 직접 반영되어 있다는 점에서 기존의 방법에 비해 보다 객관적

이고 체계적인 모형으로 인정받을 수 있다. 이러한 본 연구의 성과가 기존의 WFMS의 예측 로직과 결합하여 사용될 경우 보다 큰 시너지 효과를 발휘할 수 있을 것으로 기대되며, WFMS를 구축할 경제적 혹은 시간적 여유가 없는 업체의 경우 데이터 마이닝 기법만을 이용하여 본 연구에서 도출된 모형을 직접 적용함으로써 콜 예측을 수행할 수도 있을 것으로 사료된다.

향후 본 연구의 후속 연구에서 다루어야 할 과제는 다음과 같다. 우선 콜량에 영향을 줄 수 있는 외부 변수의 추가 발굴 및 변수 도입의 타당성 평가를 위해 많은 노력이 필요하다. 예를 들어 본 연구에서 채택하지 않은 일교차, 전일 강수, 강설량 등의 추가 변수의 발굴과 검증이 지속적으로 요구된다. 또한 일별 콜량 뿐 아니라 시간대별 콜량 예측을 위한 모형의 개발도 추후 시도되어야 한다. 본 연구에서는 예측의 최소 단위를 일별 콜량으로 한정하였지만 실제 콜센터 업무에서는 오전/오후, 또는 Peak/Idle 시간대에 따라 단위 시간별 평균 콜량이 크게 다르게 나타난다. 따라서 교대근무제, 파트타임제 등을 통한 탄력적 인력 배치를 위해서는 예측의 최소 단위를 시간대별 콜량으로 세밀화하는 작업이 필요하다

참고문헌

- 김윤배, 이창현, 김재범, 이계신, 이병철, “시물레이션을 통한 콜센터의 성능 개선”, *한국시물레이션학회논문지*, 12권 4호(2003).
- 김재진, 유일, 박득, “콜센터 성과에 영향을 미치는 주요 요인에 관한 연구”, *한국정보기술응용학회 추계공동학술대회*, 2004.
- 박종태, “한국형콜센터경영”, 도서출판 물푸레, 2007.
- 서창적, 이세영, “서비스 표준화 측정 모형의 개발”, *서비스 경영학회지*, 8권 3호(2007).
- 송현수, “콜센터 매니지먼트”, 새로운제안, 2002.
- 신상기, “WFMS를 이용한 생산성 관리”, *CRM and Telecommunication Mobile EXPO*, 2005.
- 오봉연, 김용석, “자치단체의 성공적인 콜 센터 구축전략 연구”, *한국정보처리학회 춘계학술대회*, 2007.
- 장석철, 석상문, 이주상, 이상욱, 안병하, “신경망, 시계열 분석 및 판단 보정 기법을 이용한 교통량 예측”, *한국경영과학회/대한산업공학회 춘계공동학술대회*, 2005.
- 전찬영, 송주미, “데이터마이닝 기법을 이용한 항만물동량 예측 활용방안 연구”, *한국해양수산개발원*, 2006.
- 정영수, 강창욱, “데이터 마이닝을 이용한 공정 불량률 예측 및 개선”, *한국산업경영시스템학회 추계학술대회*, 2004.
- (주)MPC, “콜센터 시스템의 콜량 예측 방법”, *한국특허정보원*, 2003.
- 지원철, “고객센터 Inbound Call 예측에 관한 연구”, *홍익대학교 과학기술연구소*, 2002.
- 한국능률협회컨설팅(KMAC), “2009년 한국산업의 서비스 품질지수(KSQI) 콜센터부문 조사 결과”, KMAC, 2009.
- 함승오, 홍정식, “산재보험 부정수급 식별 모형에 관한 연구”, *한국경영과학회 추계학술대회*, 2008.
- Andrews, B. H. and S. M. Cunningham, “L. L. Bean Improves Call-Center Forecasting”, *Institute for Operations Research and the Management Sciences*, 1995.
- Anton, J., D. Phelps, and M. Trotter, “Handbook for call center auditors”, *Purdue University: Center for Customer-Driven Quality*, 2001.
- Bhulai, S., W. H. Kan, and E. Marchiori, “Nearest Neighbour Algorithms for Forecasting Call Arrivals in Call Centers”, *Vrije Universiteit Amsterdam*, 2005.
- Dasari, K. M., “How to Do Call Volume Forecasting

- for Service Desk”, *iSixSigma Magazine*, 2008.
- Kungle, R. and J. Maluchik, “Call Center Forecasting at AAA Michigan”, *The Journal of Business Forecasting*. Vol. 20, No. 4(1998).
- Myers, W., S. Linden, and G. Wiener, “A Data Mining Approach to Soil Temperature and Moisture Prediction”, *National Center for Atmospheric Research*, 2008.
- Noiman, S. A., “Forecasting Demand For a Telephone Call Center: Analysis of Desired Versus Attainable Precision”, *Israel Institute of Technology*, 2006.
- Shen, H. and J. Z. Huang, “Interday Forecasting And intraday Updating of Call Center Arrivals”, *Manufacturing and Service Operations Management*, 2006.
- Wu, S., N. Kang, and L. Yang, “Fraudulent Behavior Forecast in Telecom Industry Based on Data Mining Technology”, *Communications of the IIMA*, Vol.7, No.4(2007).

Abstract

A Case Study on Forecasting Inbound Calls of Motor Insurance Company Using Interactive Data Mining Technique

Baek, Woong* · Kim, Namgyu**

Due to the wide spread of customers' frequent access of non face-to-face services, there have been many attempts to improve customer satisfaction using huge amounts of data accumulated through non face-to-face channels. Usually, a call center is regarded to be one of the most representative non-faced channels. Therefore, it is important that a call center has enough agents to offer high level customer satisfaction. However, managing too many agents would increase the operational costs of a call center by increasing labor costs. Therefore, predicting and calculating the appropriate size of human resources of a call center is one of the most critical success factors of call center management. For this reason, most call centers are currently establishing a department of WFM(Work Force Management) to estimate the appropriate number of agents and to direct much effort to predict the volume of inbound calls.

In real world applications, inbound call prediction is usually performed based on the intuition and experience of a domain expert. In other words, a domain expert usually predicts the volume of calls by calculating the average call of some periods and adjusting the average according to his/her subjective estimation. However, this kind of approach has radical limitations in that the result of prediction might be strongly affected by the expert's personal experience and competence. It is often the case that a domain expert may predict inbound calls quite differently from another if the two experts have mutually different opinions on selecting influential variables and priorities among the variables. Moreover, it is almost impossible to logically clarify the process of expert's subjective prediction. Currently, to overcome the limitations of subjective call prediction, most call centers are adopting a WFMS(Workforce Management System) package in which expert's best practices are systemized. With WFMS, a user can predict the volume of calls by calculating the average call of each day of the week, excluding some eventful days. However, WFMS costs too much capital during the early stage of system establishment. Moreover, it is hard to reflect new information onto the system when some factors affecting the amount of calls have been changed.

In this paper, we attempt to devise a new model for predicting inbound calls that is not only based on theoretical background but also easily applicable to real world applications. Our model was mainly developed by the interactive decision tree technique, one of the most popular techniques in data mining. Therefore, we expect that our model can predict inbound calls automatically based on historical data, and it can utilize expert's

* Anycar Service, Samsung Fire & Marine Insurance

** Assistant Professor, School of MIS, Kookmin University

domain knowledge during the process of tree construction. To analyze the accuracy of our model, we performed intensive experiments on a real case of one of the largest car insurance companies in Korea. In the case study, the prediction accuracy of the devised two models and traditional WFMS are analyzed with respect to the various error rates allowable. The experiments reveal that our data mining-based two models outperform WFMS in terms of predicting the amount of accident calls and fault calls in most experimental situations examined.

Key Words : Call Center Management, Classification and Prediction, Data Mining, Interactive Decision Tree, Predicting Inbound Calls.

저 자 소개



백 응

현재 애니카자동차손해사정서비스 WFM에 선임으로 재직 중이다. 한국항공대학교 컴퓨터공학과에서 학사학위를 취득하고, 국민대학교 BIT전문대학원에서 IT-MBA를 전공하여 경영정보학 석사 학위를 취득하였다. 콜센터매니저포럼의 운영위원으로 활동 중이며, 주요 관심분야는 과학적 콜센터 운영을 위한 데이터관리, 데이터마이닝 및 경영의사결정론과 경영혁신(PI)을 위한 6sigma, TRIZ 등이다.



김남규

현재 국민대학교 경영정보학부에서 조교수로 재직 중이다. 서울대학교 컴퓨터공학과에서 학사 학위를 취득하고, KAIST 테크노경영대학원에서 Database와 MIS를 전공하여 경영공학 석사 및 박사학위를 취득하였다. 한국정보기술응용학회 편집위원, 한국지능정보시스템학회 및 한국CRM학회 이사, 그리고 한국경영정보학회 및 한국정보시스템학회 종신회원으로 활동 중이다. 주요 관심분야는 시멘틱 데이터 관리, 데이터베이스 설계 및 데이터마이닝 등이다.