

사례기반추론을 이용한 다이렉트 마케팅의 고객반응예측모형의 통합

홍태호* · 박지영**

〈 목 차 〉

I. 서론	4.1 데이터
II. 문헌연구	4.2 실험설계 및 결과
2.1 고객반응예측에 관한 연구	V. 결론
2.2 고객반응예측을 위한 방법론	참고문헌
III. 연구프레임워크	<Abstract>
IV. 실험 및 결과분석	

I. 서론

통신, 유통, 금융 등과 같은 산업은 새로운 고객을 확보하기 어려운 시장이 포화된 환경에서 고객확보를 위한 기업들의 경쟁은 더욱 극심해지고 있다. 이러한 시장의 변화를 이끈 주요한 요인들 중 하나는 고객의 니즈가 더욱 다양해지고 있으며 또한 서비스에 대한 기대수준이 높아진 점이라고 할 수 있다. 이러한 고객들의 변화에 따라서 기업들은 고객들의 다양한 요구에 대해 맞춤형 서비스와 같은 전략적 대응을 하기 위해 다양한 정보기술을 기반으로 고객 정보의 체계적인 분석을 실행하고, 고객대

응전략을 수립하고 있다. 또한 기업들은 기존의 마케팅 활동도 보다 더 고객 중심적이고 고객 지향적인 방향으로 접근하고 있다. 1980년대 초부터 마케팅 영역에서 매우 중요한 이슈가 되어 온 CRM(Customer Relationship Management)은 경쟁력이 치열해진 경영환경에서 하나의 선도적인 전략이 되고 있다. 고객의 니즈가 다양해지고 그 변화 속도가 매우 빨라지고 있는 등 급변하고 있는 시장 환경을 고려할 때 CRM은 기업 활동의 필수적인 요소로 부각되고 있으며, 이에 기업들은 고객 관계가 고객 가치를 구축하는 근원적인 도구이며, 고객 가치의 성장이 곧 기업 가치를 극대화하는 핵심요소라는 것을 인지하게 된 것이다(Rogers, 2005). 고객들 개개

* 홍태호, 부산대학교 경영학부 부교수, hongth@pusan.ac.kr

** 박지영, 부산대학교 경영학과 박사과정, 교신저자, jiyoung@pusan.ac.kr

인이 기업에 기여하는 가치는 상이하기 때문에 기업들은 고객의 구매정보에 따라 마케팅 전략을 차별화 하는 등 보다 효과적으로 고객과의 관계를 관리하면서 수익창출을 극대화하려고 한다. 결과적으로 기업들의 마케팅전략은 제품이나 브랜드 중심에서 자연스럽게 고객중심의 접근법으로 이동하고 있다(Reinarts et al., 2004).

기업은 제품이나 서비스 등과 관련된 여러 가지 마케팅 프로모션을 통해 고객들로 하여금 긍정적인 반응을 이끌어내고자 하고, 이를 토대로 수익창출을 기대한다. 기업의 마케팅 프로모션은 고객이 긍정적으로 반응했을 때 수익이 창출되지만, 프로모션 진행과정에서 카탈로그 제작비, 우편물 발송비, 프로모션 운영비 등과 같은 여러 가지 비용도 함께 발생된다. 특히 우편물 발송을 주요 마케팅 수단으로 활용하고 있는 다이렉트 마케팅(Direct Marketing)에서는 프로모션 대상을 잘 정의하여 프로모션 수행과정에서 발생하는 비용을 최소화하여 수익을 극대화할 수 있어야 한다. 기업의 마케팅 프로모션의 대상을 잘 정의한다는 의미는 마케팅 프로모션에 대해 긍정적으로 반응할 고객들이 누구인가를 찾는 문제이다. 즉, 기업은 제품이나 서비스에 관련된 정보에 관심이 있는 고객에게는 집중적으로 마케팅 프로모션을 실행하고, 그렇지 않은 고객에게는 정보수신에 대해 귀찮음을 느끼지 않도록 프로모션을 자제함으로써 결과적으로 더 나은 효과를 창출할 수 있다. 만약, 기업의 마케팅 프로모션에 대해 긍정적인 반응이 기대되는 고객들을 예측할 수 있다면, 프로모션 수행과정에서 발생하는 비용을 최소화하면서 최대의 수익을 기대할 수 있다. 본 연구에서 제안하는 고객반응예측모형은 기업의 마케팅

프로모션에 대해 긍정적인 반응이 예상되는 고객을 선별할 수 있도록 하고, 기업들로 하여금 선별된 고객들에 대해 비용 대비 효과적인 마케팅 프로모션을 수행하도록 할 수 있다.

마케팅 프로모션에 대한 고객반응을 예측하는 것은 기업부도예측, 침입탐지시스템, 이탈고객예측과 같은 이진분류의 문제이다. 특히 로짓, 인공신경망, SVM(Support Vector Machine) 모형은 이진분류문제를 해결하는데 많이 적용되고 있으며, 인공신경망과 SVM은 예측성가에 있어서도 탁월한 성과를 보여주고 있다. 그러나 의사결정문제에서 각각의 모형들은 어느 것이 항상 최선이라고 단정하기 어렵고, 결과적으로는 예측성고가 높은 모형을 선택하게 된다. 이런 관점에서 통합모형은 단일모형이 보여주는 성과를 더욱 높일 수 있는 방안으로 인식되고 있다. 통합모형에 대한 연구는 크게 두 가지로 나누어질 수 있는데, 첫 번째는 각 단일모형의 대등한 통합에 대한 것이고, 두 번째는 다른 단일모형에 대한 보조적 통합에 대한 것이다. 단일모형간의 대등한 통합은 서로 다른 지식의 통합과 유사하다. 이러한 모형은 추천시스템에서 주로 사용되는 협업필터링(Collaborative Filtering)에서 서로 다른 추천결과를 최종적으로 하나의 추천결과로 이끌어내는 것과 유사하다(Adomavicious and Tuzhilin, 2005; Goldberg et al., 1992; Konstan et al, 1997; Resnick et al., 1997).

본 연구에서는 로짓, 인공신경망, 그리고 SVM을 이용하여 고객반응예측을 위한 단일모형을 개발하고, 단일모형이 보여주는 성과를 더욱 높일 수 있는 방안으로 사례기반추론을 이용하여 단일모형을 통합(Integration)하는 방법

론을 제시한다. 사례기반추론은 로짓, 인공신경망, SVM과는 다르게 학습이 필요 없는 기존의 사례를 참조하여 새로운 사례에 대한 해를 제공한다. 추천시스템과 같은 다른 의견을 가진 여러 유사 사용자들의 추천결과를 통합하는 것과 같이 본 연구의 통합모형에서는 고객의 반응여부에 대해 로짓, 인공신경망, SVM이 서로 다른 결과를 낼 때, 이를 조정하거나 통합하는 것이 필요하다. 단일모형의 서로 다른 결과의 통합은 협업필터링과 같은 방식 즉, 사례기반추론을 이용하여 각 모형의 서로 다른 결과값을 통합함으로써 훨씬 나아진 결과를 기대할 수 있다. 따라서 새로운 사례에 대한 결과를 예측할 경우 단일모형의 결과만을 참조한 경우보다 이미 구축된 여러 개의 사례집합을 참조함으로써 더욱 우수한 성과를 제시할 수 있게 되는 것이다.

일반적으로 고객반응예측모형의 성과는 반응고객을 얼마나 잘 예측했는가에 대한 적중률(Hit Ratio)로 평가할 수 있으며, 마케팅 프로모션의 결과로 얻을 수 있는 수익과 프로모션 수행과정에서 발생하는 비용문제를 고려한 경제성 분석으로 평가할 수도 있다. 결국 적중률이 높은 고객반응예측모형을 개발하려는 의도는 첫째 다이렉트 마케팅 비용을 감소시키고, 둘째 가망고객에게 마케팅 역량을 집중하고, 셋째 마케팅 프로모션의 효익을 극대화하기 위함이다. 따라서 본 연구에서는 실제 다이렉트 마케팅 과정에서 수집된 Direct Marketing Educational Foundation의 데이터셋을 이용하여 고객반응예측모형을 개발하고, 개발한 모형을 비용 및 기대가치 측면에서 경제성 분석을 실시하였다.

본 연구의 구성은 다음과 같다. 다음 2장에서

는 고객반응예측과 관련된 연구와 함께 고객반응예측과 관련된 연구에 적용된 기법들에 대해 알아본다. 3장에서는 연구프레임워크를 제시하고, 4장에서는 본 연구에서 제안하는 모형에 대한 실험설계 및 결과를 제시한다. 마지막 5장에서는 본 연구의 의의 및 한계점을 제시한다.

II. 선행연구

본 장에서는 고객반응예측과 관련된 문헌들을 정리해보고, 주로 이진분류 문제해결에 적용되어 온 모형 가운데 본 연구에서 고객반응예측을 위해 사용하게 될 인공신경망 및 SVM을 소개하고, 통합모형에 대해 알아보기로 한다.

2.1 고객반응예측에 관한 연구

기업들은 제품 및 서비스 등에 대해 마케팅 프로모션을 실시함으로써 고객들로 하여금 프로모션에 대한 긍정적인 반응을 얻도록 하고, 제품이나 서비스에 대한 구매를 유도함으로써 결과적으로 수익을 창출하려고 한다. 기업의 마케팅 프로모션에 대한 고객반응이란, 고객들이 프로모션 관련 제품이나 서비스를 구매하는 것과, 관련 항목에 대한 고객들의 긍정적인 관심도 포함하는 포괄적인 의미이며, 고객반응이 반드시 구매활동만을 의미하는 것은 아니다. 그러므로 반응고객은 기업이 제공하는 제품 및 서비스 정보에 대해 호기심을 가지고 오프라인 매장을 찾는든지, 관련 항목에 대한 검색이나 문의 등이 가능한 상태의 고객을 포함한다.

이메일 및 다이렉트 메일링 등은 기업에서

활용하고 있는 대표적인 마케팅전략으로 기업들은 이러한 방법으로 마케팅 프로모션을 실행함으로써 고객들로 하여금 긍정적인 반응과 함께 최종적으로 구매를 유도하게 된다. 이메일 마케팅과 다이렉트 메일링은 마케팅 프로모션을 한다는 주요 목적은 동일하지만, 비용, 신속성, 수신자의 피드백 등에 있어서 차이가 있다. 이메일과 다이렉트 메일의 장·단점을 비교해보면 다음과 같다. 우선, 이메일은 인터넷을 이용하여 상대적으로 저렴한 비용으로 한 번에 많은 양의 홍보물을 발송할 수 있다는 장점이 있다. 그러나 너무 자주, 너무 많이 발송되고 있는 마케팅 프로모션 관련 이메일은 고객들로 하여금 원치 않는 메일로 인식되어(Cranor and LaMacchia, 1998), 기업의 마케팅 프로모션에 대한 집중력 있는 정보전달력이 현저히 떨어지므로 오히려 정보 전달의 유용성을 방해하는 역할을 하고 있는 것이다. 반면, 다이렉트 메일은 고객이 개별적으로 마케팅 프로모션 정보를 우편으로 직접 받아볼 수 있도록 하기 때문에 프로모션에 대한 정보를 집중력 있게 전달할 수 있다는 장점을 가진다. 그러나 이메일 마케팅에 비해 상대적으로 비용이 크고, 고객반응을 빠른 시간 내에 측정할 수 없다는 단점을 가진다(신성훈 등, 2009). 특히, 카탈로그, 통신, 방문 등을 통해 중간상 없이 소비자에게 직접 제품을 판매하는 다이렉트 마케팅의 경우 고객 개인에게 우편물을 발송하는 것을 주요 마케팅 도구로 활용하기 때문에, 마케팅 프로모션과정에서 소요되는 비용문제를 고려해야 한다. 기업들은 고객으로 하여금 긍정적인 반응과 함께 제품이나 서비스에 대한 구매를 유도하기 위해 다양한 마케팅 프로모션 전략을 수행하게 되는

데 그 과정에서 카탈로그 제작비, 우편물 발송비, 프로모션 운영비 등을 포함한 여러 가지 마케팅 비용을 필요로 하기 때문이다.

다이렉트 마케팅 영역에서 고객반응모형은 프로모션 목표대상을 선별하여 어떤 프로모션을 얼마나 자주 진행할 것인가와 같은 여러 가지 중요한 의사결정 문제를 수반하고 있다(Prinzie and Van der Poel, 2005). 특히 이 가운데에서도 프로모션 대상을 예측하는 문제는 프로모션 과정에서 수반되는 여러 가지 의사결정 문제를 해결하는데 기반이 될 수 있으며, 다이렉트 마케팅 영역에서 고객반응예측모형의 개발이 필요한 가장 중요한 이유라 할 수 있다(Desarbo and Ramaswamy, 1994; Suh et al., 1999).

고객반응예측모형은 기업이 제안하는 프로모션 활동에 대한 고객의 반응여부에 대한 예측정보를 제공할 수 있으며, 자칫 고객들로 하여금 귀찮은 정보로 인식될 수 있는 마케팅 프로모션에 대한 정보가 필요한 고객에게만 제안할 수 있다. 또한 고객의 반응을 예측함으로써 고객에 대한 마케팅전략을 보다 구체화할 수 있다. 일반적으로 우편물을 통해 제품구매제안을 하는 기업들은 구매 가능성이 높은 고객들을 선별하여 카탈로그를 발송하게 된다(Baesens et al., 2002). 만약 고객의 반응여부를 제대로 예측할 수 있는 모형이 있다면, 프로모션 대상 고객들을 잘 선별하여 해당 고객들에게만 집중적인 마케팅 역량을 발휘함으로써 필요이상으로 소요되는 비용을 줄일 수 있게 된다. 어떤 마케팅 프로모션에 대해 고객이 반응할지 그렇지 않을지에 대한 정보를 제공하는 고객반응예측모형은 고객정보, 구매정보, 상품

정보 등을 이용하여 어떤 마케팅 프로모션에 대해 얼마나 많은 고객들이 반응할 것인지에 대한 정보를 포함하여, 어떤 고객들이 구매할 것인지 그렇지 않을 것인지에 대한 예측이 가능하다(Baesens et al., 2002; Kim et al., 2008; Suh et al., 1999).

2.2 고객반응예측을 위한 방법론

기업의 마케팅 프로모션 활동에 대한 고객반응은 이진분류 예측의 한 분야로, 가장 대표적인 분야는 기업부도예측, 침입탐지시스템, 그리고 고객이탈관리 등이 있다. 이러한 이진분류 문제를 해결하기 위해 로짓, 인공신경망, 그리고 SVM 등의 기법들이 가장 많이 적용되고 있

다. <표 1>은 이진분류의 문제에 적용된 연구를 분야 및 모형별로 정리한 것이다. 로짓, 인공신경망, SVM 등은 기업부도예측(Ohlson, 1980; Lee et al., 2005; Tam, 1991; 민재형, 이영찬, 2005; 박정민 등, 2005; Shin et al, 2005), 침입탐지시스템(Debar et al., 1992; Endler, 1998; 홍태호, 김진완, 2006), 이탈고객예측(이건창 등, 2002; 이훈영 등, 2006; 김상용 등, 2005) 분야에 많이 적용되고 있다. 또한 마케팅 분야에서도 고객반응예측(송수섭, 이의훈, 2001; Ha and Cho, 2005; Zahavi and Levin, 1997, Prinzie and Van der Poel, 2005; Van del Poel, 2003, 안현철 등, 2005b, Shin and Cho, 2006; Ciaene et al. 2001), 고객 및 시장 세분화(Agawal and Scholing, 1996; Fish et al., 1995), 신제품개발

<표 1> 이진분류의 문제에 적용된 모형 및 관련연구

분야	모형	연구
기업부도예측	로짓	Ohlson(1980)
	인공신경망	Lee et al.(2005), Tam(1991)
	SVM	민재형, 이영찬(2005), 박정민 등(2005), Shin et al.(2005)
	인공신경망, 러프셋	Ahn et al.(2000)
침입탐지시스템	인공신경망	Debar et al.(1992), Endler(1998)
	인공신경망, 귀납적학습방법, 러프셋	홍태호, 김진완(2006)
이탈고객예측	C5.0, 로짓, 인공신경망	이건창 등(2002),
	로짓	이훈영 등(2006), 김상용 등(2005)
고객선택/반응/(재)구매	인공신경망	송수섭, 이의훈(2001), Ha and Cho(2005), Zahavi and Levin(1997)
	로짓	Prinzie and Van der Poel(2005), Van der Poel(2003)
	SVM	안현철 등(2005b), Shin and Cho(2006), Viaene et al.(2001)
고객/시장세분화	인공신경망, 로짓	Agawal and Scholing(1996), Fish et al.(1995)
신제품 개발	인공신경망	Thieme et al.(2000)
시장반응/점유율예측	인공신경망	Dasgupta et al.(1994)

(Thieme et al., 2000), 시장반응 및 점유율 예측 (Dasgupta et al., 1994) 등의 연구에 로짓, 인공신경망, SVM 모형이 적용되어 좋은 성과를 보여주고 있다.

부도예측, 침입탐지시스템, 이탈관리, 그리고 고객반응예측은 모두 이진분류 기반의 예측분야이지만, 모형 구축 이후 분류결과를 통한 모형의 평가과정에서는 해당 영역의 특성을 충분히 고려할 필요가 있다. 특히 분류 대상이 되는 집단의 중요도가 같지 않은 경우에 적용분야의 특성을 고려하여 비용측면에서 모형들의 성과를 확인해야 하는 것이다. 여기에서 비용이란, 실제값과 다른 값을 예측했을 때에 발생할 수 있는 경제적 손실에 대한 규모가 어느 정도 되는가 하는 것을 의미한다.

침입탐지 또는 부도예측 분야는 예측결과와 오류를 통해 발생할 수 있는 경우 손실이 매우 크기 때문에 분류모형을 선별하는데 매우 신중할 필요가 있다. 기업을 포함한 여러 기관에서 네트워크 침입을 발견하지 못해서 실제로 침입이 발생한 경우에는 네트워크 침입 경보에 대처하면서 발생할 수 있는 기회비용과는 비교할 수 없을 정도로 엄청난 비용이 발생된다. 기업의 부도예측을 하는 경우에는 건전기업을 예측하는 것 보다 파산 가능성이 있는 기업을 예측하는 것이 훨씬 중요한 일이다. 그리고 고객반응모형의 경우에는 기업의 마케팅 프로모션에 반응할 고객임에도 불구하고 그렇지 않은 고객으로 분류했을 때에는 많은 기회비용이 발생할 수 있으며, 반대로 반응하지 않을 고객을 반응할 고객으로 분류했을 경우 기업의 입장에서 수익과 연관될 수 있는 긍정적인 반응이 기대되지 않는 고객에게 여러 가지 마케팅 비용을

소비하게 되는 결과를 낼 수 있다. 위의 내용은 분류모형의 성과에서 상대적으로 중요한 집단의 오분류에 대한 위양성률(false positive rate)과 위음성률(false negative rate)을 통해 알아볼 수 있다. 만약, 분류하고자 하는 두 개의 집단 C_0 , C_1 가운데 중요한 집단을 C_0 라고 하면 위양성률은 C_1 집단을 C_0 집단으로 잘못 분류한 비율을 의미하고, 위음성률은 C_0 집단을 C_1 집단으로 잘못 분류한 비율을 의미한다.

본 연구에서는 마케팅 영역의 고객반응예측 모형을 중심으로 다음에서 제시될 이진분류 모형 및 통합모형의 결과를 토대로 비용손실을 최소화 하고 최대의 기대가치를 만들어내는 모형이 어떠한 것인지 탐색해 보고자 한다. 다음은 본 연구에서 적용할 방법인 인공신경망, SVM 그리고 통합모형에 대한 간략한 소개이다.

2.2.1 인공신경망

인공신경망은 일반적으로 통계적인 방법이 적용되어오던 영역과, 분류 및 예측이 필요한 다양한 분야에서 응용되고 있는 방법이며, 유연한 모형개발 툴이다(Paliwal and Kumar, 2009; Zang et al., 1999). 노드(nodes) 또는 뉴런(neuron)이라는 많은 수의 처리요소(processing elements)와 여러 개의 층(layer)으로 구성되어 있다. 인공신경망 형태 중 가장 잘 알려진 형태는 다층퍼셉트론(Multi-layer perceptron)이며, 이는 입력층(input layer)과 출력층(output layer) 사이에 은닉층(hidden layer)이 존재하며, 모든 노드와 층이 전방향(feedforward)으로 연결되어 있다(Rumelhart and McClelland, 1986). 입력층, 은닉층, 출력층에 이르기까지 가중치는 신경망

결과가 목표치에 도달할 때까지 계속 변화하게 되는데, 이는 인공신경망을 목표 출력값과 인공신경망의 계산결과 산출된 출력값의 차이를 최소화하는 방향으로 변화하게 된다. 인공신경망은 인간이 새로운 것을 학습하는 것과 같은 방식으로 여러 개의 구성개념(construct)과 알고리즘들을 사용하여 학습하며, 신경망 구조를 통해 예측, 분류, 군집화, 그리고 최적화 등에 탁월한 성능을 보이지만 학습 알고리즘과 연관된 다양한 모수들에 대한 결정이 간단하지가 않고, 많은 수의 학습용 데이터와 그에 대한 반복학습을 필요로 하며 과대적합으로 인한 일반화의 어려움, 그리고 결과에 대한 명확한 해석이 어렵다는 단점이 있다(Kumar and Ravi, 2007). 인공신경망에 대한 자세한 내용은 Rumelhart and McClelland(1986)을 참고한다.

2.2.2 Support Vector Machine(SVM)

SVM은 Vapnik(1995)에 의해 개발된 것으로 통계적 학습이론을 기반으로 하고 있으며, 입력공간과 관련된 비선형문제를 고차원의 특징공간의 선형문제로 대응시켜 나타내기 때문에 수학적 분석이 용이하다(Hearst et al., 1998). 특히 SVM이 최근 각광을 받는 이유는 대부분의 학습 알고리즘은 경험적 위험최소화에 기반하는 모형인데 반해, SVM은 구조적위험최소화에 충실하므로, 분류문제에 있어 일반화 기능이 높아 많은 분야에서 응용되고 있다(Tay and Cao, 2001).

SVM은 분류 문제에 있어 훈련 데이터들을 서로 다른 두 개의 클래스로 분류할 때 분류의 기준이 되는 분리 경계면(hyper plane)을 학습 알고리즘을 통해 찾고 클래스를 구분하는 최적

분리경계면(maximum margin hyperplane)을 구하기 위해 분리경계면과 가장 가까운 점(support vector)과의 거리를 최대화한다. SVM이 두 개의 클래스 구분이 불가능 한 경우에는 커널함수를 이용하여 입력 자료를 고차원의 특징공간으로 사상시킴으로써, 고차원 공간에서는 선형 분리가 가능한 분리경계면을 생성할 수 있다. 그러므로 SVM에서는 커널함수의 역할이 매우 크다는 것을 알 수 있다. 커널함수는 원래 데이터를 고차원공간으로 사상시킴으로써 특징공간 내에 선형으로 분리 가능한 입력 데이터 셋을 만들며, 일반적으로 사용되는 커널함수는 다음의 (1), (2)와 같은 다항식 커널(Polynomial kernel)과 가우시안 RBF(Gaussian Radial Basis Function) 등이 있다.

다항식 커널:

$$K(x, x_i) = (x \cdot x_i + 1)^d \dots\dots\dots (1)$$

가우시안 RBF:

$$K(x, x_i) = \exp\left(-\frac{1}{\sigma^2}(x - x_i)^2\right) \dots\dots\dots (2)$$

SVM은 통계적 이론을 기반으로 분류 및 회귀문제를 해결할 수 있으며, 상대적으로 적은 수의 표본에서도 사용이 가능하다. 그러나 커널이나 그에 따른 모수들을 선택하는 문제가 약간은 막연할 수 있으며, 검증 단계에서 성과가 부진할 수 있다(Kumar and Ravi, 2007). SVM에 대한 자세한 내용은 Vapnik(1995)을 참고한다.

2.2.3 통합모형

로짓, 인공신경망, SVM 모형은 이진분류문제

를 해결하는데 많이 적용되고 있으며, 예측성과에 있어서도 탁월한 성과를 보여주고 있다. 로짓은 다양한 영역에서 이진분류문제 해결에 일반적으로 많이 활용되고 있는 기법이지만, 선형관계를 기본으로 하고 있기 때문에 비선형관계를 잘 설명하기가 어렵고, 이상치에 민감하지 못하고 과적합 문제를 충분히 해결할 수 없으며, 불균형데이터를 효과적으로 처리할 수 없다는 단점이 있다(송수섭, 이의훈, 2001; Geng, 2006). 인공신경망의 경우, 많은 영역에서 성과의 우수함을 보여주고 있지만, 학습 알고리즘과 연관된 다양한 모수들에 대한 결정이 어렵고, 많은 수의 학습용 데이터와 그에 대한 반복학습을 필요로 하며, 과대적합으로 인한 일반화의 어려움과 결과에 대한 명확한 해석이 어렵다는 단점이 있다. SVM 역시 통계적 기반의 설명력이나 적용 영역에 대해 성과가 우수하지만, 커널이나 그에 따른 모수들을 선택하는 문제에서 어려움이 있고, 검증 단계에서 성과 부진이 단점이 될 수 있다(Kumar and Ravi, 2007).

위의 모형들은 예측성과에 있어서 우수한 성과를 보여주고 있지만, 의사결정문제에서 어느 모형이 항상 최선이라고 단정하기는 어렵다. 그러나 결과적으로는 예측성과가 높은 모형을 선택하게 된다는 관점에서 통합모형은 단일모형이 보여주는 성과를 더욱 높일 수 있는 방안이 될 수 있다. 통합모형은 여러 개의 단일모형을 대등하게 통합하거나, 주요 단일모형을 보조하는 방편으로 통합할 수도 있다. 단일모형간의 대등한 통합은 서로 다른 지식의 통합과 유사하며, 추천시스템에서 주로 사용되는 협업필터링과 유사하다. 즉, 서로 다른 추천결과를 최종적으로 하나의 추천결과로 이끌어내는 것과 같

다.

본 연구에서는 로짓, 인공신경망, 그리고 SVM을 이용하여 고객반응예측을 위한 단일모형을 개발하고, 단일모형이 보여주는 성과를 더욱 높일 수 있는 방안으로 사례기반추론을 이용하여 단일모형을 통합(Integration)하는 방법을 제시한다. 사례기반추론은 로짓, 인공신경망, SVM과는 다르게 학습이 필요 없는 기존의 사례를 참조하여 새로운 사례에 대한 해를 제공하는 특성 때문에 서로 다른 의견을 가지고 있는 여러 유사 사용자들의 추천 결과를 통합하는데 많이 활용되고 있다. 추천시스템과 같은 다른 의견을 가진 여러 유사 사용자들의 추천 결과를 통합하는 것과 같이 본 연구의 통합모형에서는 고객의 반응여부에 대해 로짓, 인공신경망, SVM이 서로 다른 결과를 낼 때, 이를 조정하거나 통합하는 것이 필요하다. 단일모형의 서로 다른 결과의 통합은 사례기반추론을 이용하여 각 모형의 서로 다른 결과값을 통합함으로써 훨씬 나아진 결과를 기대할 수 있다. 따라서 새로운 사례에 대한 결과를 예측할 경우 단일모형의 결과만을 참조한 경우보다 이미 구축된 여러 개의 사례집합을 참조함으로써 더욱 우수한 성과를 제시할 수 있게 되는 것이다.

통합모형은 <표 2>와 같이 기업부도예측(Lin and McClean, 2001; Wu et al., 2007), 신용등급(Shin and Han, 1999; 신택수, 홍태호, 2004; 노태협 등, 2005), 금융예측모델링(Versace et al., 2004), 정보시스템모형개발(박기남, 2006), 이동통신고객분류(이극노, 이홍철, 2003) 등의 영역에서 모형의 성과향상을 위해 적용되고 있다. 또한 고객반응 및 구매예측(Suh et al., 1999; Yu and Cho, 2006; Coenen et al., 2000), 고객세

<표 2> 통합모형을 적용한 연구

분야	모형	연구자
기업부도예측	통계 및 인공지능기법	Lin and McClean(2001)
	GA, SVM	Wu et al.(2007)
기업신용등급	GA, CBR	Shin and Han(1999)
	인공신경망, 로짓	신태수, 홍태호(2004)
	러프셋, CBR	노태협 등(2005)
금융예측모델링	인공신경망, GA	Versace et al.(2004)
정보시스템모형개발	CBR	박기남(2006)
이동통신고객분류	C4.5, 인공신경망	이극노, 이홍철(2003)
Data 실험	C4.5, GA	Carvalho et al.(2004)
	회귀, 판별, CART 등	Conversano et al.(2002)
	러프셋, 인공신경망	Li and Wang(2004)
고객반응/구매예측	RFM, NN, 로짓	Suh et al.(1999)
	GA, CBR	안현철 등(2005a)
	GA, SVM	Yu and Cho(2006)
	C5.0, CBR	Coenen et al.(2000)
고객세분화	GA, CBR	Chiu(2002)

분화(Chiu, 2002) 영역에서도 사용되고 있다. 특히, 고객반응예측모형 영역의 통합모형은 주로 최적변수 선정(feature selection)과 관련하여 유전자알고리즘과 의사결정나무의 규칙생성 부분을 주로 활용하고 있으며, 이는 하나의 단일 모형에 대해 보조적인 관점의 통합모형으로 볼 수 있다.

Ⅲ. 연구프레임워크

다이렉트 마케팅 영역에서 고객반응예측모형은 프로모션 활동의 대상이 될 목표고객은 누구인지, 그들에게 어떠한 판매 제안을 할 것인지, 어떻게 또는 무슨 활동을 해 나갈 것인지, 언제, 그리고 얼마나 자주 행할 것인지에 대한

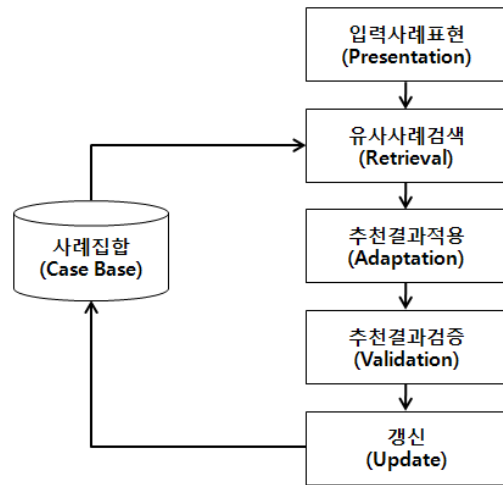
예측과 같은 여러 가지 의사결정 문제를 해결하기 위해 필요하다(Prinzie and Van der Poel, 2005). 특히 반응고객을 예측하는 문제는 마케팅 프로모션 과정에서 여러 의사결정 문제의 기반이 되기 때문에 고객반응예측모형의 개발은 다이렉트 마케팅 영역에서 매우 중요한 내용이라 할 수 있다(Desarbo and Ramaswamy, 1994; Suh et al., 1999). 마케팅 프로모션에서는 프로모션에 반응하는 고객들을 통한 수익창출 영역과 프로모션 전략 수행과정에서 발생하게 되는 여러 가지 비용문제를 고려해야 한다. 만약 기업이 전략적으로 선점해야 할 목표고객이 잘 정의되어 있다면, 비용문제를 경감시키는데 큰 도움이 될 수 있으며, 마케팅 프로모션 이전에 고객의 반응여부를 알 수 있는 예측모형을 갖는다면, 이와 같은 문제의 답을 쉽게 찾을 수

있다. 결과적으로 고객반응예측에 있어서 좋은 모형이란, 제품에 대한 마케팅 프로모션에 반응할 수 있는 고객과 정보수신을 귀찮아하는 고객들을 잘 선별하여 불필요한 비용을 줄이고 반응고객을 통해 수익을 극대화할 수 있도록 하는 것이다.

본 연구에서는 다이렉트 마케팅에서 고객반응과 관련하여 보다 예측력이 우수한 고객반응예측모형을 개발하기 위해 단일모형과 사례기반추론을 활용하여 단일모형을 통합하는 모형을 제시한다. 사례기반추론은 새로운 사례에 대한 해를 제공하기 위해 학습이 불필요한 기존의 사례를 참조한다. 기존의 이진분류문제에 적용되어 오던 인공지능망과 SVM은 기법 자체를 이용한 단일모형만으로도 우수한 성과를 보여 주었지만, 고객이 반응할 것인가에 대해 서로 다른 의견을 갖는 로짓, 인공지능망, SVM에 대한 결과를 사례기반추론으로 통합함으로써 더 나은 성과를 기대할 수 있다. 사례기반추론은 문제의 해답을 찾기 위한 추론 과정에서 불필요한 참조사례를 포함하여 모든 사례들이 사용된다는 단점을 가지고 있지만, 인공지능망, SVM과 같이 학습이 필요하지 않고, 잘 알려지지 않은 영역에서도 문제해결을 위해 빠른 해답을 제공받을 수 있으며, 특정 알고리즘을 적용할 수 없는 영역에 대해서도 최적의 값을 기대할 수 있고, 다양한 상황 가운데에서 중요한 부분들에 초점을 맞출 수 있는 장점을 가지므로 많은 영역의 의사결정 문제해결에 적용되고 있다(Chiu, 2002; Kolodner, 1993).

본 연구에서 제안하는 사례기반추론을 이용한 고객반응예측 통합모형은 두 단계를 거쳐 이루어진다. 첫 번째 단계에서는 로짓, 인공신

경망, SVM 등 각각의 단일모형의 예측값을 먼저 구하여 성과를 비교하고, 두 번째 단계에서는 사례기반추론을 이용하여 각 단일모형의 예측값에 대해 유사도를 고려하여 최종 예측값을 결정한다. 사례기반추론과정은 <그림 1>과 같이 모두 다섯 단계의 프로세스로 구성된다(Bradley, 1994). 먼저 새로운 입력사례가 주어지면, 사례집합(case base)에서 일정한 유사성 척도에 부합되는 유사사례를 추출(retrieve)한 다음, 현재의 문제에 대한 추천결과를 적용(adaptation)하게 된다. 다음으로 추천결과에 대한 검증(validation) 단계를 거쳐 새롭게 제시된 해답은 다시 새로운 사례로 저장되어 미래의 문제해결을 위해 사용될 수 있도록 사례집합이 갱신(update)된다.



<그림 1> 사례기반추론 과정(Bradley, 1994)

본 연구에서 사례기반추론은 고객반응예측 결과를 통합하는데 활용되며, 로짓, 인공지능망, 그리고 SVM을 이용한 단일예측모형의 고객반응 예측값들이 사례집합에 저장되며, 새로운 값의 특성에 대해 유사도(similarity)를 측정하여

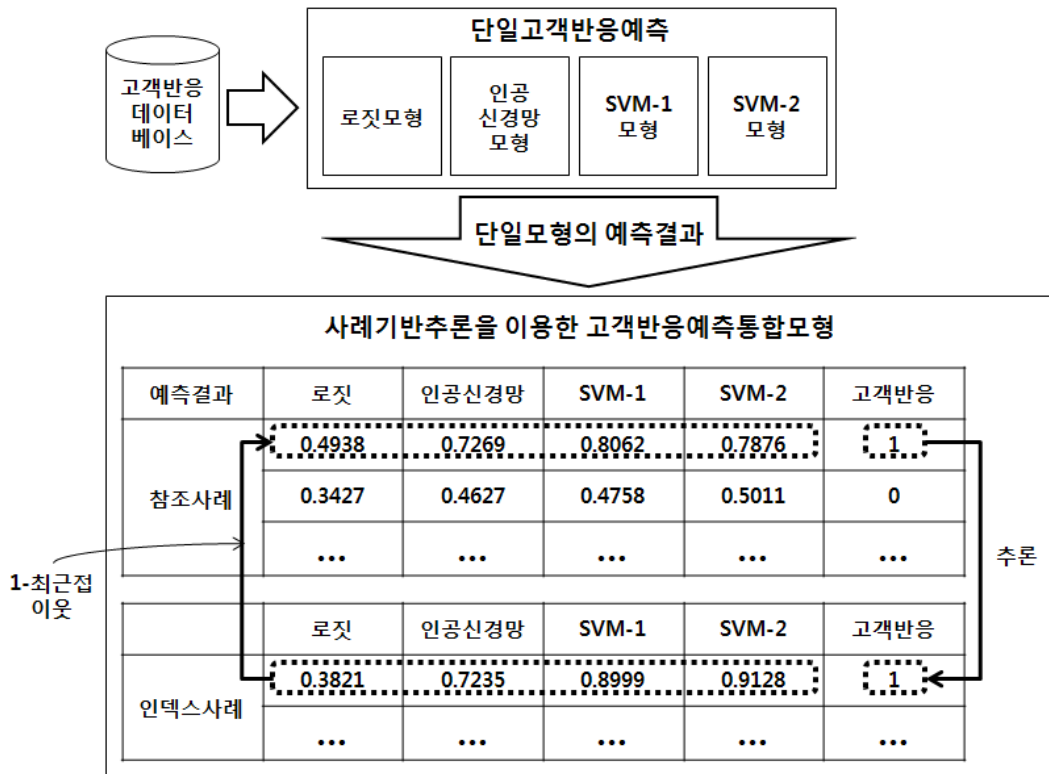
사례집합에서 가장 가까운 사례를 찾는다. 여기서 유사도는 다음의 식(3)과 같다(Kolodner, 1993).

$$\frac{\sum_{i=1}^n W_i \times sim(f_i^I, f_i^R)}{\sum_{i=1}^n W_i} \dots\dots\dots (3)$$

여기서, W_i 는 i 번째 변수의 가중치, f_i^I 는 입력사례에 대한 i 번째 값, f_i^R 는 추출한 유사사례에 대한 i 번째 값, 그리고 $sim()$ 는 f_i^I 와 f_i^R 에 대한 유사도 함수이다. 유사도를 측정하여 유사도가 가장 높은 사례 즉, 가장 가까운 거리에 있는 사례 1개를 추출(1-nearest neighbor)하

여, 그 사례에 저장되어 있는 값을 제시하는 과정이 진행되는데, 일반적으로 이런 유사도 측정은 최근접이웃방법(nearest-neighbor method)이 가장 많이 활용되고, 이와 동시에 새로운 사례에 대한 입력결과는 추천결과 적용과정과 추천결과에 대한 검정과정을 거쳐서 다시 다음 사례의 매칭에 이용될 수 있도록 갱신되는 절차를 따른다.

<그림 2>는 고객반응예측 성과를 향상시킬 수 있도록 고객거래데이터를 기반으로 기존의 이진분류 문제에 적용되고 있는 기법을 이용하여 고객반응예측모형을 개발하고, <그림 1>과 같은 사례기반추론과정을 거쳐 통합고객반응예측모형을 개발하는 본 연구의 프레임워크이다.



<그림 2> 연구프레임워크

본 연구에서 고객반응예측결과를 통합하는데 사례기반추론을 활용하고, 유사성 함수로 유클리디언 거리(Euclidean distance)를 사용하여 사례간의 거리가 가장 작은 참조 셋을 첫 번째 근접이웃으로, 그 다음 거리가 작은 셋을 두 번째 근접이웃으로 찾는다. 일반적으로 가중치를 구하는 문제에 대해서는 첫째 가중치를 동일하게 두는 가장 일반적인 방법과, 둘째 전문가의 판단에 의해 가중치를 개별적으로 두거나, 셋째 유전자알고리즘과 같은 방법을 통해 최적화하기도 하는데, 본 연구에서는 유사성을 구할 때 가중치를 동일하게 두는 가장 일반적인 방법으로 실험하였다.

또한, 개발한 고객반응예측모형을 현실 데이터에 적용해 봄으로써 고객반응모형을 마케팅 전략으로 활용했을 때 예측성과 및 비용문제를 함께 모색해보았다. 즉, 마케팅 프로모션의 주요 대상이 될 수 있는 반응고객을 잘 예측할 수 있는 모형의 성과에 대한 중요성과 함께 반응고객을 잘 예측하지 못했을 때 발생하게 되는 여러 가지 마케팅 비용 측면에서 모형의 실효성을 함께 고려해볼 수 있도록 하였다. 결국 고객반응예측에 대해 적중률을 높일 수 있는 통

합모형을 개발하려는 의도는 첫째 다이렉트 마케팅 비용을 감소시키고, 둘째 가망고객에게 마케팅 역량을 집중하고, 셋째 마케팅 프로모션의 효익을 극대화하기 위함이다.

IV. 실험 및 결과분석

4.1 데이터

본 연구에 사용된 데이터는 DMEF04 데이터 셋(Direct Marketing Educational Foundation, <http://www.directworks.org>)이며, 이 데이터셋은 101,532명의 고객들에 대한 다이렉트 마케팅 관련 데이터로, 기업에서 마케팅 프로모션을 위해 정기적으로 고객들에게 카탈로그를 보낸 후 주문으로 이어진 고객 즉, 고객반응률이 약 9%인 반응고객 및 비반응고객의 구조가 불균등한 데이터셋이다. DMEF04 데이터셋은 고객반응예측모형에 관련된 많은 연구(Kim et al., 2008; Lee and Cho, 2007; Malthouse, 2001, 2002; Shin and Cho, 2006)에서 사용되었으며, 각 고객에 대해 주문 수, 주문금액, 주문항목 등에 대

<표 3> 모형에 사용된 변수 (Malthouse, 2001; Kim et al., 2008)

변수명	설명	변수명	설명
Purseas	구매가 일어난 시즌 수	Tran54	$270 \leq recency < 366$
Falord	현재까지 가을 시즌 주문 수	Tran55	$366 \leq recency < 730$
Ordtyr	올해 주문 수	Comb2	$\sum_{i=1}^{14}$ 올해 구매한 제품그룹 _i
Puryear	구매가 일어난 해의 수	Tran25	$1/(1 + \text{최근 시즌 아이템})$
Sprord	현재까지 봄 시즌 주문 수	Tran42	$\log(1 + \text{올해 주문 수} \times \text{지금까지 가을시즌 주문 수})$
Recency	1992년 10월까지 주문 일	Tran44	$\sqrt{\text{지금까지 주문수} \times \text{지금까지 봄시즌 주문수}}$
Tran38	$1/recency$	Tran46	$\sqrt{comb2}$
Tran52	$90 \leq recency < 180$	Y	1: 반응고객
Tran53	$180 \leq recency < 270$		0: 비반응고객

한 91개의 입력변수들로 구성되어 있다. 본 실험에서 사용된 변수는 <표 3>과 같으며, 이는 DMEF04 데이터셋의 91개 입력변수 중 Malthouse(2001)와 Kim et al.(2008) 등의 연구에서 사용된 변수로 구성되어 있다. 종속변수 Y는 마케팅 프로모션 관련 카탈로그를 받고 난 이후 한번 이상 주문한 고객들은 “반응고객” 그렇지 않은 고객은 “비반응고객”으로 나타내었다.

4.2 실험설계 및 결과

4.2.1 모형구축

고객반응예측모형을 개발하고 개발한 예측모형을 실제 데이터에 적용하기 위해 다음과 같은 절차로 진행하였다. 고객반응예측모형 개발의 경우, 고객의 응답률 자체가 매우 낮기 때문에 집단 간에 심각한 데이터 불균형이 발생한다(Shin and Cho, 2006). 우선 실험 데이터에 대해 본 연구에서는 집단의 분산 등을 고려하여 전체 데이터셋에서 반응고객 데이터 5,000개와 비반응고객 데이터 5,000개 등 10,000개의 데이터를 랜덤으로 추출하여 균형데이터로 실험을 실시한다. 데이터셋은 학습용, 평가용, 검증용으로 각각 6,000개, 2,000개, 2,000개로 분할하여 사용하고, 로짓모형은 10,000개의 데이터를 각각 8,000개, 2,000개로 분할하여 학습용 및 검증용 데이터셋으로 사용하였다. 많은 연구에서 학습용 데이터셋은 모형을 구축하는데 사용되고, 검증용 데이터셋은 예측력을 평가하는데 사용된다. 데이터셋을 주로 랜덤하게 학습용 및 검증용으로 분할하여 사용하고 있지만, 이는 모형선택과 평가에 있어서 바이어스를 다수 내

포하고 있기 때문에, 타당성을 확보하는데 어려움이 따를 수 있다. 이를 해결하기 위해 cross-validation 분석을 사용할 수 있는데 이는 샘플링 변동에 따른 실험결과의 신뢰성을 보여줄 수 있다(Zang et al., 1999). 본 연구에서는 각 모형에 대해 5-fold cross validation으로 수행하였다. 5-fold cross validation은 전체 표본을 모두 5개의 셋으로 분할하여 실험하게 되므로, 실험결과에 대한 타당성을 확보할 수 있다. 균형 데이터를 이용하여 단일 이진분류 기반 고객반응예측모형을 개발한 이후, 개발 모형을 실험데이터와 유사한 불균형데이터에 적용하여 마케팅 비용을 고려한 모형들 간의 성과를 비교할 수 있도록 했다. 불균형데이터는 전체 데이터의 수가 91,532개이며 반응고객이 4,707명, 그리고 비반응고객이 86,825명인 반응률이 약 5%인 데이터셋이다.

인공신경망모형 구축을 위해 은닉층을 4, 8, 12, 16, 20, 24, 28, 32으로 변화해가면서 시행착오(trial-and-error) 방법으로 성과가 좋은 모형을 찾으며, 이때 학습률과 모멘텀은 각각 0.1로 두고, 각 층에 대한 출력함수로 시그모이드 함수를 사용하였으며 실험은 Neuroshell2로 실행하였다.

SVM은 커널함수로 다항식 커널과 가우시안 RBF 커널을 사용하였다. SVM 모형을 구축하기 위해서는 상한값을 나타내는 C와, 사용되는 커널함수의 모수 즉, 다항식 커널의 차수를 나타내는 d, 가우시안 RBF 커널의 σ 값이 필요하다. 각 모수의 설정값에 따라 SVM 모형에 대한 전체성과가 차이가 나기 때문에, 모수에 대한 결정은 시행착오방법으로 탐색해 나간다. 최적화 모형을 구축하기 위해, 다항식 커널을 이용

한 모형은 SVM-1으로, 가우시안 RBF 커널을 이용한 모형은 SVM-2로 각각 명명하고, 모수 $C=\{1, 25, 50, 75, 100, 150, 200, 300, 500, 1000\}$, $d=\{1, 2, 3, 4, 5\}$, $\sigma=\{0.25, 0.5, 1, 2, 5, 10, 15\}$ 로 설정하여 실험 후 결과를 비교한다. 실험은 공개소프트웨어인 Libsvm-2.83 (<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>)으로 실행하였다.

고객반응예측모형을 구축하기 위해 로짓, 인공신경망, 그리고 SVM을 사용하여 각각의 단일예측모형을 구축한 이후, 각각의 모형에 대한 실험결과를 이용하여 복잡하고 비구조적인 의사결정문제를 효과적으로 해결하는데 유의성을 보여주는 방법인 사례기반추론을 적용하여 통합모형을 다음과 같이 구축한다. 앞서 구축한 단일예측모형의 고객반응예측값들은 사례집합에 저장되고, 새로운 값의 특성에 대해 유사도를 측정하여 사례집합에서 가장 가까운 거리에 있는 사례 1개를 추출하여, 그 사례에 저장되어 있는 값을 제시하도록 하였다. 유사도 측정은 최근접이웃방법을 활용하였고, 예측결과에 대한 최적의 k는 1부터 10까지의 범위에서 탐색할 수 있도록 하였으며, Excel VBA와 SPSS

17.0을 사용하여 실험하였다.

4.2.2 실험결과

(1) 고객반응예측모형 개발

로짓, 인공신경망, 그리고 SVM 등의 방법을 적용한 단일모형과 이들 모형의 예측결과를 기반으로 구축한 통합모형에 대해 5-fold cross validation으로 실험한 결과는 <표 4>와 같다. 실험설계 시 사례기반추론을 이용하여 개발한 통합모형에서 예측결과에 대한 최적의 k는 1부터 10까지의 범위에서 탐색하였으며, Set1, Set2, Set3은 k가 1인 경우, Set4와 Set5는 k가 3인 경우이다. 통합방법론의 성과의 우수성을 검증하기 위해 다수결 방법을 적용한 결과도 함께 제시하였다. MV 모형은 다수결 방법을 적용한 결과이며, 로짓, 인공신경망, SVM-1, SVM-2 등의 4개의 모형으로 voting을 수행하였다. 여기서 2:2의 동률을 이루는 경우는 반응 고객인지 아닌지 판단이 불확실한 경우를 나타내며, Set1부터 Set5까지 2:2의 동률을 이루는 경우는 각각 126개, 72개, 76개, 71개, 104개이다. MV모형에 대해 동률을 이루는 경우를 제외하였음에도 불구하고 연구에서 제안한 통합모

<표 4> 고객반응예측에 대한 단일모형 및 통합모형 성과표

(단위: %)

모형	로짓		인공신경망		SVM-1		SVM-2		MV ¹⁾		통합모형
	학습용	검증용	학습용	검증용	학습용	검증용	학습용	검증용	검증용	검증용	
Set1	76.79	75.50	78.03	77.70	80.33	79.45	83.88	80.45	75.85	86.55	
Set2	77.91	76.40	78.35	77.35	79.63	78.95	83.82	80.35	76.60	85.30	
Set3	77.35	77.25	78.53	78.55	80.13	79.65	85.13	79.80	77.25	86.25	
Set4	77.06	77.80	78.67	79.20	80.72	79.85	84.23	80.35	78.00	85.15	
Set5	77.39	78.10	78.37	78.55	80.42	79.15	84.08	79.65	76.55	85.30	
평균	77.30	77.01	78.39	78.27	80.25	79.41	84.23	80.12	76.85	85.71	

주) 1) MV: 다수결에 의한 검증모형(동률인 경우는 제외)

<표 5> 맥네마 테스트 결과

모형	인공신경망	SVM-1	SVM-2	MV	통합모형
로짓	31.126**	64.763**	76.696**	241.241**	308.986**
인공신경망		21.940**	33.160**	5.309*	237.134**
SVM-1			6.473*	5.440*	180.823**
SVM-2				2.161	165.180**
MV					141.156**

주) 숫자는 검정통계량

*95% 신뢰수준에서 통계적으로 유의함. **99% 신뢰수준에서 통계적으로 유의함.

MV모형과의 맥네마 테스트는 MV가 동물인 사례를 제외하고 검증함.

형 및 단일모형보다 예측성도가 낮음을 볼 수 있다. 반응고객과 비반응고객에 대한 분류성도는 각 모형별로 각 데이터셋에 대한 성과가 차이를 보이고 있지만, 전체 모형들의 평균성도는 사례기반추론을 이용한 통합모형이 가장 우수하며, SVM-2, SVM-1, 인공신경망, 로짓모형, MV모형 순이다.

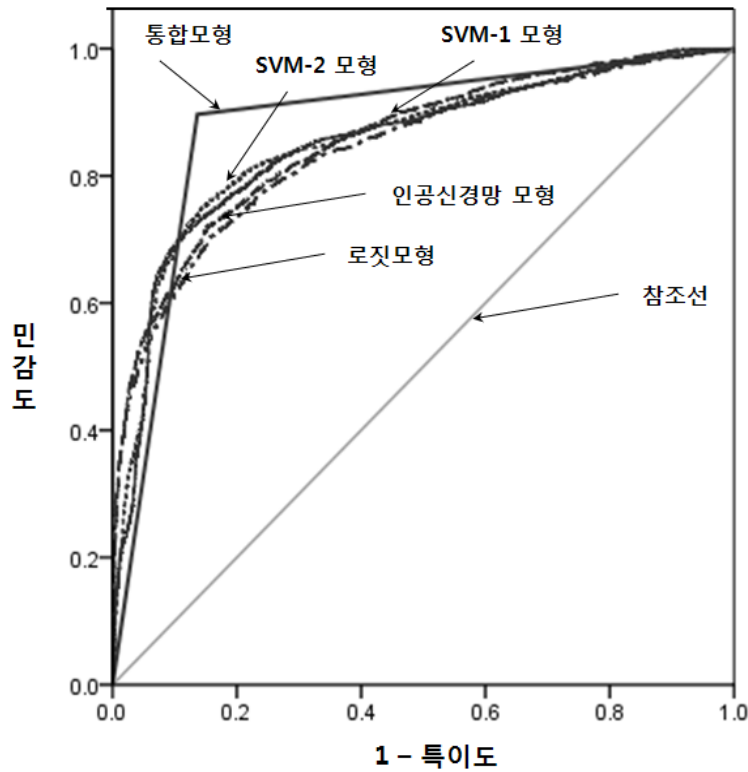
본 연구에서 제시한 모형간 성과 차이가 통계적으로 유의한지 검증하기 위해 맥네마(McNemar) 테스트를 수행하였으며, 결과는 <표 5>와 같다. 통합모형은 로짓, 인공신경망, SVM-1, SVM-2 모형과 99% 신뢰수준에서 통계적으로 모형간 성과차이를 보이고 있다. MV 모형과의 맥네마 테스트 결과는 동물인 경우를 제외하고 모형 간의 성과를 비교한 결과이다.

<그림 3>은 예측모형에 대한 ROC(Receive Operation Characteristic) 곡선을 나타낸 것이다. ROC 곡선은 예측모형의 성과를 정확도 관점에서 평가할 수 있으며, 모형의 성과를 민감도(sensitivity)와 특이도(specificity)에 의해 판단할 수 있는 정보를 제공한다. 만약, 분류하고자 하는 두 개의 집단 C_0 , C_1 가운데 중요한 집단을 C_0 라고 하면 분류모형의 민감도는 실제 C_0 집단을 C_0 집단으로 정확하게 분류할 확률이고,

특이도는 C_1 집단을 C_1 집단으로 정확하게 분류할 확률을 의미한다. ROC 곡선에서 y축은 분류모형의 민감도를 나타내며 주요 집단을 얼마나 정확하게 판별하는가에 대한 확률값을 나타내고, x축은 분류기준값에 따라 계산된 특이도에 대한 (1-특이도)의 값을 나타낸다. 그래프에서 참조선은 예측력이 전혀 없는 모형에 대한 ROC 곡선 즉, 우연에 의한 ROC 곡선을 의미한다. 좋은 성과를 나타내는 모형일수록 ROC 곡선은 참조선 위쪽으로 위치하고, ROC 곡선의 아래쪽 면적은 예측모형의 성과를 측정할 수 있는 값으로 사용될 수 있다. 즉, 그래프가 좌상향으로 많이 치우쳐 있을수록 좋은 모형이며 참조선에 가까울수록 나쁜 모형이라 할 수 있는데, 기준모형의 성과를 나타내는 참조선을 기준으로 할 때 본 연구에서 제안하고 있는 사례기반추론을 이용한 통합모형이 다른 모형들에 비해 좋은 모형이라고 볼 수 있다.

(2) 고객반응예측모형 적용

본 연구에서 고객반응예측모형구축은 균형 데이터를 이용하였으며, 여기서 균형데이터란 실험을 위한 데이터셋에서 반응고객과 비반응고객의 비율을 50:50으로 구성한 것이고, 불균



<그림 3> ROC 곡선

형데이터는 고객반응률이 약 5%인 데이터이다. 특히, 다이렉트 마케팅에서 고객에 대한 긍정적인 반응을 유도하기 위해 고객 개개인에게 우편물을 발송하는 다이렉트 메일링을 도구로 할 경우, 반응할 고객을 그렇지 않은 고객으로 잘못 예측했을 때 기회비용이 발생하고, 제품에 대한 판매제의를 위한 카탈로그 제작, 우편물 발송, 프로모션을 운영 등에 여러 비용들이 발생된다. 마케팅 프로모션의 비용관점에서 반응 고객을 그렇지 않은 고객으로, 그리고 반대로 예측할 경우 불필요한 비용을 발생시키므로, 모형에 대한 경제성측면의 성과를 평가할 수 있는 주요 요소가 된다. 즉 기업들은 고객으로 하여금 긍정적인 반응을 유도하기 위한 여러 가

지 마케팅 프로모션 진행과정에서 프로모션에 대한 반응고객을 잘 예측할 수 있다면, 불필요한 비용은 최소화하면서 수익을 극대화할 수 있다는 것이다.

그러므로 마케팅 프로모션에서 고객반응예측모형의 성과에 대한 중요성과 함께, 프로모션 과정에서의 비용 최소화 및 수익 극대화 측면에서 모형에 대한 경제성 측면을 분석할 필요가 있다. 다이렉트 마케팅 활동에서 고객반응예측모형을 적용한 경우 반응고객을 통해 얻을 수 있는 기대가치와 마케팅 프로모션에 대한 비용측면의 이점은 Shmueli et al.(2007)를 참고하기 바란다.

비용과 이익을 고려하여 고객들에게 카탈로

그를 발송하는 마케팅 프로모션에 대한 기대가치를 일반화하면 다음의 식(4)와 같다.

$$NEV_i = RRC_i \cdot P - PRC_i \cdot C \dots\dots\dots (4)$$

NEV_i : 모형 i 에 의해 카탈로그 발송을 통한 마케팅 프로모션 후 얻을 수 있는 기대가치

P : 고객 1인이 카탈로그에 반응함으로써 기업에 기여하는 가치

C : 카탈로그 1건당 발송비용

RRC_i (Real Response Customer): 모형 i 에 의해 실제 반응한 고객 수

PRC_i (Predicted Response Customer): 모형 i 에 의해 반응할 것으로 예측된 고객 수

$i=1$: Naive I. 전체고객발송, Naive II. 10% 랜덤고객발송, Naive III. 20% 랜덤고객발송

$i=2$: 로짓

$i=3$: 인공신경망

$i=4$: SVM-1

$i=5$: SVM-2

$i=6$: 통합모형

<표 6>은 좋은 예측모형을 선별하기 위해 마케팅 프로모션을 위한 카탈로그 발송 후에 예측할 수 있는 기대가치를 모형별로 비교한 것이며, 이때 적용한 데이터셋은 DMEF04 데이터셋이며, 전체 91,532개의 데이터 중 반응고객이 4,707명인 반응률이 약 5%인 불균형데이터 셋이고, 고객 1명이 카탈로그에 반응함으로써 기업에 기여하는 가치는 1,000원, 카탈로그 1건당 발송비용은 100원이라고 가정한 결과이다.

결과에서, 비용 및 기대가치가 차이가 나는 이유는 분류모형에서 반응고객에 대한 예측성과가 모형별로 다르기 때문이다. 또한 마케팅 프로모션을 위해 카탈로그를 발송하는 활동에 있어서, 모형을 사용하지 않고 랜덤하게 고객들을 선택하여 발송한 경우와, 예측모형을 사용하여 발송한 경우 비용 및 기대가치 측면에서 모형을 사용한 쪽이 더욱 큰 이익을 가져다 줄 수 있음을 알 수 있다. 특히, 본 연구에서 제안한 사례기반추론을 이용하여 로짓, 인공신경망, SVM-1, SVM-2 등의 단일모형을 통합한 모형이 단일모형보다 최소의 비용으로 기대가치를 극대화할 수 있는 모형임을 알 수 있다.

V. 결론

본 연구에서는 다이렉트 마케팅 영역에서 고객구매와 관련하여 보다 예측력이 우수한 고객반응예측모형을 제안하기 위해 사례기반추론을 활용한 통합모형을 제시하고, 개발한 모형을 현실데이터에 적용하여 마케팅 영역에서의 비용 및 이익문제를 바탕으로 모형의 성과에 대해 고찰해 보았다. 분류 및 예측문제에 적용되어 온 여러 가지 방법론들은 어느 것이 항상 최선의 성과를 제시한다고 단정 짓기는 어렵다. 그

<표 6> 비대칭 오분류비용에 따른 모형별 광고물발송 후 얻는 기대가치

	모형을 사용하지 경우			모형을 사용한 경우				
	Naive I ¹⁾	Naive II ²⁾	Naive III ³⁾	로짓	인공신경망	SVM-1	SVM-2	통합모형
고객기여가치(원)	4,707,000	470,700	941,400	3,100,000	3,283,000	3,413,000	3,629,000	4,128,000
발송비용(원)	9,153,200	915,320	1,830,640	1,578,600	1,667,200	1,706,300	1,868,100	1,679,300
순기대가치(원)	-4,446,200	-444,620	-889,240	1,521,400	1,615,800	1,706,700	1,760,900	2,448,700

주) 1) 전체고객발송, 2) 10% 랜덤고객발송, 3) 20% 랜덤고객발송

러나 문제해결을 위해 개발한 모형에 대해서는 예측성고가 높은 모형을 항상 선호하기 때문에 최근연구에서는 단일기법보다 더 우수한 성과를 제시할 수 있는 통합모형을 많이 적용하고 있는 추세이다. 본 연구에서 적용한 사례기반추론은 참조해야 하는 지식이 여러 개가 있을 때 최적의 값을 찾아서 결과를 보여줄 수 있기 때문에 더욱 우수한 성과를 기대할 수 있게 된다. 고객이 반응할 것인가에 대한 문제에서 다른 의견을 갖는 로짓, 인공신경망, SVM 등의 결과를 사례기반추론으로 통합함으로써 훨씬 나아진 결과를 기대할 수 있으므로, 새로운 사례에 대한 결과를 예측할 경우 단일기법만 참조한 경우보다 이미 구축된 여러 개의 사례집합을 참조함으로써 더욱 우수한 성과를 제시할 수 있음을 알 수 있다.

모형이 제시하는 성과는 얼마나 반응고객들을 잘 예측할 수 있느냐 하는 문제가 되는데, 특히 마케팅 영역에서는 고객에 대한 기업의 마케팅 프로모션에서 카탈로그 발송 및 광고 등과 같은 비용에 대해 반응하는 고객이 기업에 기여하는 가치가 어느 정도 되는가 하는 문제를 충분히 고려할 수 있어야 한다. 고객반응 예측모형은 기업의 프로모션에 대해 반응할 고객들을 예측하여 그들에게 유용한 마케팅 프로모션을 하는데 큰 역할을 할 수 있다. 최근에는 기업들이 이메일, 카탈로그 등과 같은 도구로 불필요하게 과도한 마케팅 프로모션을 행하는 경향이 많아서 기업의 마케팅 프로모션이 유용한 정보가 아니라 오히려 성가신 것으로 작용하기도 한다. 이러한 관점에서 기업이 제안하는 정보에 대해 반응할 만한 고객을 찾는 일은 매우 중요하며 고객반응예측모형의 필요성은 더

욱 커진다 할 수 있다. 즉, 비용과 기대가치 관점에서 고객반응예측모형은 기업에게 있어서 불필요한 광고비용을 줄이면서 그들의 제안에 반응할 고객들에게 집중적인 프로모션을 행할 수 있다는 데 더 많은 이익을 기대할 수 있으며, 고객들은 기업의 마케팅 프로모션에 대한 정보가 필요한 그룹들은 적절한 정보를 얻을 수 있고, 그렇지 않은 그룹들은 그들의 의도에 상관없이 불필요하다고 생각되는 성가신 정보를 받지 않아서 좋다.

본 연구를 기반으로 하여 향후 연구에서는 반응 및 비반응 고객에 대한 이진분류 형태에서 확장하여, 고객세분화를 통한 다분류 문제로의 접근이 가능할 것이며, 성과측면에서 예측률 향상을 위해 다양한 통합모형개발이 가능할 것이다. 또한, 다양한 분류 기법을 적용하여 예측 성과를 도출함에 있어서 적용 분야의 특성을 고려하여 현실을 명확하게 반영하여 전략적으로 활용할 수 있는 설명력 있는 모형을 개발할 수 있도록 분석을 위한 샘플링 및 영향력 있는 최적 변수 선택(feature selection) 영역에 대한 고찰이 필요하다.

〈참고문헌〉

- 김상용, 송지연, 이기순, “CRM 고객데이터 분석을 통한 이탈고객 연구,” *한국마케팅 저널*, 제7권, 제1호, 2005, pp. 21-42.
- 노태협, 유명환, 한인구, “러프집합이론과 사례기반추론을 결합한 기업신용평가 모형,” *정보시스템연구*, 제14권, 제1호, 2005, pp. 41-65.

- 민재형, 이영찬, "Support Vector Machine을 이용한 부도예측모형의 개발: 격자탐색을 이용한 커널 함수의 최적 모수 값 선정과 기존 부도예측모형과의 성과 비교," *한국경영과학회지*, 제30권, 제1호, 2005, pp. 55-74.
- 박기남, "사례기반추론을 이용한 정보시스템 가치평가 모형개발에 관한 연구," *정보시스템연구*, 제15권, 제2호, 2006, pp. 95-123.
- 박정민, 김경재, 한인구, "Support Vector Machine을 이용한 기업부도예측," *경영정보학연구*, 제15권, 제2호, 2005, pp. 51-63.
- 송수섭, 이의훈, "인공신경망을 이용한 소비자 선택 예측에 관한 연구," *한국경영과학회지*, 제26권, 제4호, 2001, pp. 55-70.
- 신성훈, 정수연, 박철, "이메일 마케팅 성과에 관한 연구: 신용카드 고객을 대상으로 한 캠페인을 중심으로," *Information Systems Review*, 제11권, 제1호, 2009, pp. 49-67.
- 신택수, 홍태호, "인공신경망과 로짓모형을 통합한 부실확률맵기반 신용등급화에 관한 연구," *회계저널*, 제13권, 제3호, 2004, pp. 1-26.
- 안현철, 김경재, 한인구, "효과적인 고객관계관리를 위한 사례기반추론 동시 최적화 모형," *한국지능정보시스템학회논문지*, 제11권 제2호, 2005a, pp. 175-195.
- 안현철, 김경재, 한인구, "Support Vector Machine을 이용한 고객구매예측모형," *한국지능정보시스템학회논문지*, 제11권, 제3호, 2005b, pp. 69-81.
- 이건창, 정남호, 신경식, "신용카드 시장에서 데이터 마이닝을 이용한 이탈고객분석," *한국지능정보시스템학회논문지*, 제8권, 제2호, 2002, pp. 13-35.
- 이극노, 이홍철, "이동통신고객 분류를 위한 의사결정나무(C4.5)와 신경망 결합 알고리즘 연구," *한국지능정보시스템학회지*, 제9권, 제1호, 2003, pp. 139-155.
- 이훈영, 양주환, 류치환, "고객의 이탈가능성과 LTV를 이용한 고객등급화 모형개발에 관한 연구," *한국지능정보시스템학회 논문지*, 제12권, 제4호, 2006, pp. 109-126.
- 홍태호, 김진완, "데이터 마이닝 비대칭 오류비용을 이용한 지능형 칩입탐지시스템 개발," *정보시스템연구*, 제15권, 제4호, 2006, pp. 211-224.
- Adomavicious, G., and Tuzhilin, A., "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 17, No. 6, June, 2005, pp. 734-749.
- Agrawal, D., and Schorling, C., "Market share forecasting: An Empirical Comparison of Artificial Neural Networks and Multinomial Logit model," *Journal of Retailing*, Vol. 72, No. 4, 1996, pp. 383-407.
- Ahn, B.A., Cho, S.S., and Kim, C.Y., "The integrated methodology of rough set

- theory and artificial neural network for business failure prediction," *Expert Systems with Applications*, Vol. 19, 2000, pp. 65-74.
- Ahn, H., and Kim, K.-j., "Using genetic algorithms to optimize nearest neighbors for data mining," *Annals of Operations Research*, Vol. 163, 2008, pp. 5-18.
- Baesens, B., Viaene, S., Van den Poel, D., Vanthienen, J., and Dedene, G., "Bayesian neural network learning for repeat purchase modelling in direct marketing," *European Journal of Operational Research*, Vol. 138, No. 1, 2002, pp. 191-211.
- Bradley, P., "Case-based reasoning: Business applications," *Communications of the ACM*, Vol. 37, No. 3, 1994, pp. 40-43.
- Carvalho, D.R., and Feitas, A.A., "Hybrid Decision Tree/ Genetic Algorithm Method for Data Mining," *Information Science*, Vol. 163, 2004, pp. 13-35.
- Chiang, W.K., Zhang, E., and Zhou, L., "Predicting and explaining patronage behavior toward web and traditional stores using neural networks: A comparative analysis with logistic regression," *Decision Support Systems*, Vol. 41, 2006, pp. 514-531.
- Chiu, C., "A case-based customer classification approach for direct marketing," *Expert Systems with Applications*, Vol. 22, 2002, pp. 163-168.
- Coenen, F., Swinnen, G., Vanhoof, K., and Wets, G., "The improvement of response modeling: combining rule-induction and case-based reasoning," *Expert Systems with Applications*, Vol. 18, 2000, pp. 307-313.
- Conversano, C., Siciliano, R., and Mola, F., "Generalized Additive Multi-mixture Model for Data Mining," *Computational Statistics & Data Analysis*, Vol. 38, No. 4, 2002, pp. 487-500.
- Cranor, L.F., and LaMacchia, B.A., "Spam!," *Communications of the ACM*, Vol. 41, No. 8, 1998, pp. 74-83.
- Dasgupta, C.G., Dispensa, G.S., and Ghose, S., "Comparing the predictive performance of a neural network model with some traditional market response models," *International Journal of Forecasting*, Vol. 10, 1994, pp. 235-244.
- Debar, H., Becker, M., and Siboni, D., "A Neural Network Component for an Intrusion Detection Systems," *IEEE Computer Society Symposium Research in Security and Privacy*, 1992, pp. 240-250.
- Desarbo, W.S., and Ramaswamy, V., "Customer response based iterative segmentation procedures for response modeling in direct marketing," *Journal of Direct Marketing*, Vol. 8, No. 3, 1994, pp. 7-20.
- Endler, D., "Intrusion Detection Applying Machine Learning to Solaris Audit Data," *Proceedings of Computer Security Applications Conference*, 1998, pp.

- 268-279.
- Fish, K.E., Barnes, J.H., and Aiken, M.W., "Artificial neural networks - A new methodology for industrial market segmentation," *Industrial Marketing Management*, Vol. 24, 1995, pp. 431-438.
- Geng, M., "A Comparison of Logistic Regression to Random Forests for Exploring Differences in Risk Factors Associated with Stage at Diagnosis between Black and White Colon Cancer Patients," the degree of M.S., Univ. of Pittsburgh, 2006.
- Goldberg, D., Nichols, D., Oki, B., and Terry, D., "Using Collaborative Filtering to Weave an Information Tapestry," *Communications of the ACM*, Vol. 35, No. 12, 1992, pp. 61-70.
- Ha, K., Cho, S., and Maclachlan, D., "Response models based on bagging neural networks," *Journal of Interactive Marketing*, Vol. 19, No. 1, 2005, pp. 17-30.
- Hearst, M.A., Dumais, S.T., Osman, E., Platt, J., and Scholkopf, B., "Support Vector Machines," *IEEE Intelligent Systems*, Vol. 13, No. 4, 1988, pp. 18-28.
- Kim, D., Lee, H.-j., and Cho, S., "Response modeling with support vector regression," *Expert Systems with Applications*, Vol. 34, 2008, pp. 1102-1108.
- Kim, H., and Shin K.-s., "A Hybrid Approach Using Case-based Reasoning and Fuzzy Logic for Corporate Bond Rating," *Korea Intelligent Information Systems Society*, Vol. 10, No. 2, 2004, pp. 91-109.
- Kolodner, J.L., *Case-based Reasoning*, Morgan Kaufmann: Los Altos, CA, 1993
- Konstan, J.A., Miller, B.N., Maltz, D., Herlocker, J.L., Gordon, L.R., and Riedl, J., "GroupLens: Applying Collaborative Filtering to Usenet News," *Communications of the ACM*, Vol. 40, No. 3, 1997, pp. 77-87.
- Kumar, P.R., and Ravi, V., "Bankruptcy prediction in banks and firms via statistical and intelligent techniques - A review," *European Journal of Operational research*, Vol. 180, 2007, pp. 1-28.
- Lee, K., Booth, D., and Alam, P., "A comparison of supervised and unsupervised neural networks in predicting bankruptcy of Korean firms," *Expert Systems with Applications*, Vol. 29, 2005, pp. 1-16.
- Lee, H.-j., and Cho, S., "Focusing on non-respondents: Response modeling with novelty detectors," *Expert Systems with Applications*, Vol. 33, 2007, pp. 522-530.
- Li, R., and Wang, Z., "Mining classification rules using rough sets and neural networks," *European Journal of Operational Research*, Vol. 157, 2004, pp. 439-448.
- Lin, F.Y., and McClean, S., "A Data Mining Approach to the Prediction of Corporate Failure," *Knowledge-Based Systems*, Vol. 14, No. 3/4, 2001, pp. 189-195.

- Malthouse, E.C., "Assessing the performance of direct marketing scoring models," *Journal of Interactive Marketing*, Vol. 15, No. 1, 2001, pp. 49-62.
- Malthouse, E.C., "Performance-based variable selection for scoring models," *Journal of Interactive Marketing*, Vol. 16, No. 4, 2002, pp. 37-50.
- Ohlson, J., "Financial ratios and probabilistic prediction of bankruptcy," *Journal of Accounting Research*, Vol. 18, 1980, pp. 109-131.
- Paliwal, M., and Kumar, U.A., "Neural networks and statistical techniques: A review of applications," *Expert Systems with Applications*, Vol. 36, 2009, pp. 2-17.
- Prinzie, A., and Van den Poel, D., "Constrained optimization of data-mining problems to improve model performance: A direct-marketing application," *Expert Systems with Applications*, Vol. 29, No. 3, 2005, pp. 630-640.
- Reinarts, W., Frafft, M., and Hoyer, W.D., "The Customer Relationship Management Process: Its Measurement and Impact on Performance," *Journal of Marketing Research*, 41, 2004, 293-305.
- Resnick, P., and Varian, H.R., "Recommender Systems," *Communications of the ACM*, Vol. 40, No. 3, 1997, pp. 56-58.
- Rogers, M., "Customer Strategy: Observations from the Trenches," *Journal of Marketing*, 69, 2005, 262-263.
- Rumelhart, D.E., and McClelland, J.I., *Parallel Distributing Processing: Exploration in the Microstructure of Cognition*, Vol. 1, Cambridge, MA: MIT Press, 1986.
- Shin, H.-S., and Cho, S., "Response modeling with support vector machine," *Expert Systems with Applications*, Vol. 30, 2006, pp. 746-760.
- Shin, K.-s., and Han, I., "Case-based reasoning supported by genetic algorithms for corporate bond rating," *Expert systems with applications*, Vol. 16, 1999, pp. 85-95.
- Shin, K.-s., Lee, T.S., and Kim, H.-j., "An application of support vector machines in bankruptcy prediction model," *Expert systems with applications*, Vol. 28, No. 1, 2005, pp. 127-135.
- Shmueli, G., Patel, N.R., and Bruce, P.C., *Data Mining for Business Intelligence*, Wiley, 2007, pp. 88-89.
- Suh, E.H., Noh, K.C., and Suh, C.K., "Customer list segmentation using the combined response model," *Expert Systems with Applications*, Vol. 17, 1999, pp. 89-97.
- Tam, K.Y., "Neural network models and the prediction of bank bankruptcy," *Omega*, Vol. 19, No. 5, 1991, pp. 429 - 445.
- Tay, F.E.H., and Cao, L.J., "Application of support vector machines in financial time series forecasting," *Omega*, Vol. 29, No. 4, 2001, pp. 309-317.
- Thieme, R.J., Song, M., and Calantone, R.J.,

- "Artificial neural network decision support systems for new product development project selection," *Journal of Marketing Research*, Vol. 37, No. 4, 2000, pp. 499-507.
- Van den Poel, D., "Predicting mail-order repeat buying: Which variables matter?" *Tijdschrift voor Economic en Management*, Vol. 48, No. 3, 2003, pp. 371-403.
- Vapnik, V. *The Nature of Statistical Learning Theory*, Springer, 1995.
- Venkatesan, R., Kumar, V., and Bohling, T., "Optimal Customer Relationship Management Using Bayesian Decision Theory: An Application for Customer Selection," *American Marketing Association*, XLIV, 2007, pp. 579-594.
- Versace, M., Bhatt, R., Hinds, O., and Shiffer, M., "Predicting the Exchange Traded Fund DIA with a Combination of Genetic Algorithm and Neural Networks," *Expert systems with applications*, Vol. 27, No. 3, 2004, pp. 417-425.
- Viaene, S., Baesens, B., Van Gestel, T., Suykens, J.A.K., Van del Poel, D., Vanthienen, J., De Moor, B., and Dedene, G., "Knowledge discovery in a direct marketing case using least squares support vector machine classifiers," *International Journal of Intelligent Systems*, Vol. 9, 2001, pp. 1023-1036.
- Wu, C.H., Tzeng, G.H., Goo, Y.J., and Fang, W.C., "A real-valued genetic algorithm to optimize the parameters of support vector machine for predicting bankruptcy," *Expert systems with Applications*, Vol. 32, No 2, 2007, pp. 397-408.
- Yu, E., and Cho, S., "Constructing response model using ensemble based on feature subset selection," *Expert Systems with Applications*, Vol. 30, 2006, pp. 352-360.
- Zahavi, J., and Levin, N., "Applying Neural Computing to Target Marketing," *Journal of Direct Marketing*, Vol. 11, No. 1, Winter, 1997, pp. 5-22.
- Zang, G., Hu, M.Y., Patuwo, B.E., and Indro, D.C., "Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis," *European Journal of Operational Research*, Vol. 116, 1999, pp. 16-32.
- Chang, C.-C., and Lin, C.-J.; LIBSVM 2.83 - A Library for Support Vector Machines, <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>
- The Direct Marketing Educational Foundation (DMEF): The Direct Marketing Association. <http://www.directworks.org>

홍태호(Taeho, Hong)



현재 부산대학교 경영학부 부교수로 재직하고 있다. KAIST에서 산업공학사를 취득하였고 경영정보시스템을 전공하여 공학석사와 박사를 취득하였다. 딜로이트 컨설팅에서 컨설턴트로 재직했으며, 주요 관심분야는 데이터마이닝, CRM, 그리고 Business Intelligence 등이다. 주요 논문을 Expert Systems, Expert Systems with Applications, Asia Pacific Journal of Information Systems, 그리고 정보시스템연구 등에 게재하였다.

박지영(Jiyoung, Park)



부산대학교 통계학과를 졸업하고, 부산대학교 대학원에서 경영학석사를 취득하였다. 현재 부산대학교 대학원 경영학과 박사과정중이며 경영정보 생산관리를 전공하고 있다. 주요 관심분야는 데이터마이닝, CRM, 그리고 벤처기업 등이다.

<Abstract>

Integrating the Customer Response Model in Direct Marketing Using Case-Based Reasoning

Taeho Hong·Jiyoung Park

In this study, we propose a integrated model of logistic regression, artificial neural networks, support vector machines(SVM), with case-based reasoning(CBR). To predict respondents in the direct marketing is the binary classification problem as like bankruptcy prediction, IDS, churn management and so on. To solve the binary problems, we employed logistic regression, artificial neural networks, SVM. and CBR. CBR is a problem-solving technique and shows significant promise for improving the effectiveness of complex and unstructured decision making, and we can obtain excellent results through CBR in this study. Experimental results show that the classification accuracy of integration model using CBR is superior to logistic regression, artificial neural networks and SVM. When we apply the customer response model to predict respondents in the direct marketing, we have to consider from the view point of profit/cost about the misclassification.

Keywords: Customer response model, Support vector machines, Artificial neural networks, Case-based reasoning, Profit/Cost analysis

* 이 논문은 2009년 8월 21일 접수하여 2차 수정을 거쳐 2009년 9월 10일 게재 확정되었습니다.