

신용카드 매출정보를 이용한 SVM 기반 소상공인 부실예측모형

윤종식^{1*} · 권영식¹ · 노태협²

¹동국대학교 산업시스템공학과 / ²서울여자대학교 경영학과

SVM based Bankruptcy Prediction Model for Small & Micro Businesses Using Credit Card Sales Information

Jongsik Yoon¹ · Young S. Kwon² · Tae Hyup Roh³

¹Department of Industrial & Systems Engineering, Dongguk Univ.

²Department of Business Management, Seoul Women's Univ.

The small & micro business has the characteristics of both consumer credit risk and business credit risk. In predicting the bankruptcy for small-micro businesses, the problem is that in most cases, the financial data for evaluating business credit risks of small & micro businesses are not available. To alleviate such problem, we propose a bankruptcy prediction mechanism using the credit card sales information available, because most small businesses are member store of some credit card issuers, which is the main purpose of this study.

In order to perform this study, we derive some variables and analyze the relationship between good and bad signs. We employ the new statistical learning technique, support vector machines (SVM) as a classifier. We use grid search technique to find out better parameter for SVM. The experimental result shows that credit card sales information could be a good substitute for the financial data for evaluating business credit risk in predicting the bankruptcy for small-micro businesses. In addition, we also find out that SVM performs best, when compared with other classifiers such as neural networks, CART, C5.0 multivariate discriminant analysis (MDA), and logistic regression.

Keyword: small-micro business, credit evaluation, credit card sales, bankruptcy prediction, support vector machines, grid search

1. 서론

2000년 이후 소자본 창업의 확대에 의해 정부에서는 소상공인들을 위해 창업지원과 경영컨설팅뿐만 아니라 신용보증을 통한 소자본 창업자금 및 운용자금 대출 지원을 강화하고 있다. 또한 금융계는 정부의 부동산 정책 변화와 금리인상으로 인한

개인 및 기업 여신의 저조로 새로운 니치 마켓인 소상공인을 타겟으로 한 다양한 파생상품을 개발하고 있다. 이로 인해 정부와 금융기관은 소상공인의 신용평가에 더욱 관심을 높이고 있으며 소상공인을 위한 정확하고 과학적인 신용평가모형 개발을 위해 노력하고 있다.

그러나 소상공인의 신용위험은 개인신용위험과 기업신용

본 논문은 2006년 정부(교육인적자원부)의 재원으로 한국학술진흥재단의 지원을 받아 수행된 연구임(KRF-2006-511-D00436).

*연락처 : 윤종식 박사, 100-715 서울시 중구 필동3가 26 동국대학교 산업시스템공학과, Fax : 02-2269-2212, E-mail : king337@dgu.edu
2007년 01월 접수, 1회 수정 후 2007년 03월 게재확정.

위험의 특성을 갖고 있어 신용평가모형 개발이 쉽지 않다 (Park, 2001; Yang, 2003). 지금까지 소상공인을 위한 신용평가 모형은 주로 개인 신용위험만을 고려한 개인 신용평가모형에 심사관이 기업 위험적 특성을 주관적으로 고려하는 정도에 그치고 있다(Yang, 2003; Gim *et al.*, 2004). 이는 가맹점의 예상 매출액, 입지조건 등 사업성에 관해 심사관이 객관적으로 심사할 수 있는 기준의 부재가 그 원인이 되고 있다.

개인 신용평가모형이나 기업 신용평가모형의 경우, 설명력이 좋은 주요 변수들을 사용하여 선형 확률 및 조건부 확률, 통계적 모형, 인공신경망, 의사결정나무 등 다양한 분류 알고리즘들을 통해 높은 정확도를 보이고 있다. 소상공인 신용평가 모형에도 객관적인 설명변수를 활용하여 분류예측력을 높이는 것이 필요하다.

1990년대 이후 신용카드 산업이 급속히 성장하고, IMF 이후 정부가 신용카드 사용 활성화 정책을 주도하면서 소상공인 중제조업을 제외한 대부분의 업종에서 신용카드를 통한 매출이 발생하고 있는 실정이다. 특히, 최근에는 현금영수증 발행 또한 확대되어 가맹점의 매출을 추정할 수 있는 자료의 객관성도 더욱 높아지고 있다.

본 논문의 목적은 소상공인 가맹점에서 발생한 신용카드 매출정보가 소상공인 부실모형에서 가맹점의 재무적 요인과 심사관의 판단정보를 보완할 수 있는 주요한 정보임을 확인하고 이를 활용하여 소상공인의 부실을 예측하는 모형을 개발하는 것이다. 또한 다양한 분류문제에서 우수한 성능을 보이는 SVM(Support Vector Machines)을 소상공인 부실예측에 적용하였다.

본 논문은 다음과 같이 구성되어 있다. 제 2장에서는 소상공인 부실예측 모형과 SVM에 대한 이론적 고찰을 수행하고, 제 3장에서는 본 논문에서 제시한 SVM을 적용한 소상공인 부실예측 모형을 설명한다. 제 4장에서는 설계된 절차에 따른 실험과 분석 결과를 설명하고, 그 의의를 알아보고, 제 5장에서는 전체 결과 및 시사점을 기술하고, 연구의 한계와 향후 활용방안을 논의한다.

2. 이론적 고찰

2.1 소상공인 부실예측모형

소상공인 부실예측에 관한 연구는 기업과 개인에 대한 부실예측에 비해 아직 일천한 상황이다. 이는 소상공인 부실예측을 위한 소상공인의 신용위험이 기업위험과 개인위험의 특성을 동시에 내포하고 있고, 이에 따라 객관적 재무자료의 입수가 어렵기 때문이다. 또한 국가마다 그 특성 차이가 크기 때문에 국제적인 연구의 대상에서는 소외되어 왔다.

그러나 통계청의 2004년 「사업체 기초통계」에 따르면 국내 소상공인 업체수는 2,784,116개로 우리나라 전체 사업체의 87.2%를 차지하고, 소상공인 종업원 수(종사자 수)는 5,421,021

명으로 우리나라 전체 종업원의 36.6%를 차지하고 있는 것으로 나타났다. 이는 소상공인이 국가 경제의 상당부분을 담당하고 있고, 국민 경제에 많은 기여를 하고 있음을 의미하는 것이다. 이로 인해 정부와 은행에서는 최근 소상공인들을 위해 자금지원과 파생상품 개발에 노력하고 있다.

Altman *et al.*(1977) and Kim(2006) 등은 기업규모가 기업의 가치를 결정하는데 중요한 요소이며, 소규모 기업일수록 부도확률이 높다고 주장하면서 소기업과 소상공인 신용평가에 관한 연구의 필요성을 강조하였다. Berger *et al.*(2005)는 \$100,000이하의 소기업 신용을 위한 소기업 신용평가의 경제적 효과를 검증하면서, 소기업 신용평가 모형의 중요성을 강조하였다. 또한 Frame *et al.*(2004)은 대출기관들이 소상공인의 대출을 꺼리는 이유는 불투명한 재무적 안정성 때문이라고 주장했고, 수익이 적은 소기업일수록 신용평가 모형을 통한 정보 비대칭 문제의 완화하여 대출을 효과적으로 할 수 있고, 대출금액도 상승시킨다고 주장하여 소기업과 소상공인의 신용평가 모형의 효용성을 제시했다.

미국의 경우, 신용평점을 이용해 소상공인(small-micro business)에게 \$250,000까지 대출을 허용해 주고 있다. 그러나 여전히 소기업과 소상공인 신용평점은 기존의 기업평가와 개인 신용평가에 비해 상대적으로 새로운 현상으로 받아들여지고 있다. 1995년 미국의 Fair, Isaac의 모텔(FICO: Fair Isaac & Company)이 소개된 이후 소기업 혹은 소상공인 신용평가 모형에 관심을 갖기 시작하면서 1998년에서야 미국의 몇몇 대규모 은행에서 소기업 신용모형을 도입하기 시작했다. Cole *et al.*(2004)은 미국 내 대형은행은 2/3이상이 소기업을 위한 신용평가모형을 사용하는 반면 소형은행은 12%만이 소기업 신용평가모형을 사용하기 때문에 소기업의 재무정보보다는 기업 특성으로 대출업무를 진행하여 정확한 대출금액 산출이 이루어지고 있지 않다고 주장했다. 이를 통해 소기업 신용평가 모형의 도입의 중요성을 주장했다.

국내에서 Gim and Jong(1999)은 중소기업과 소기업을 분리한 부실예측모형 설계의 유용성을 주장하였다. 특히 소상공인이 포함되어 있는 도·소매·서비스업을 AHP(Analytic Hierarchy process), 다변량판별분석(MDA), 의사결정나무분석(CHAD), 신경망(NN) 알고리즘으로 개발하여 각각 63.8%, 65.5%, 55.2%, 70.7%의 예측 성능을 보이는 모형을 개발했다. Park(2001)은 판별분석(Discriminant Analysis)을 통해 소상공인 중 사고기업 100개와 정상기업 100개를 판별하는 예측모형을 도출하였다. 대출금연체, 부동산, 차입금 규모, 생활안정성과 같은 변수가 통계적으로 유의한 변수로 선택되었으며, 사고기업은 62%, 정상기업은 60%의 예측률을 보였다. Yang(2003)는 소상공인 중 사고기업 1,004개와 정상기업 1,995개를 로짓모형과 신경망으로 예측모형을 개발하였다. 업력, 보증금액, 종업원수, 기차입금 등이 주요 변수로 선택되었으며, 로짓모형은 57%, 신경망은 58.09%의 낮은 예측력을 보였다. 이런 낮은 예측력을 평가항목들의 효율성 부족으로 보고 분석적 계층기법(AHP)을 통해

재무정보의 가중치 24.7%과 비재무정보의 가중치를 75.3%으로 나누고 재무정보는 사업성, 재정상태, 안정성으로 구분하고 비재무정보는 사업주에 대한 경영능력, 재정상태, 신뢰성, 안정성을 사실정보로 정의하고 사업 전망에 관한 사업성과 경영능력은 판단정보로 구분하는 모형을 제시하였다. Gim et al. (2005)는 AHP를 통해 소상공인 신용평가모형을 위해 재무정보의 가중치를 57.5%, 비재무정보의 가중치를 42.5%로 나누고, 재무정보를 시설투자비용, 재정상태, 재무안정성, 자금조달능력으로 구성하고 비재무정보를 경영능력, 신뢰성, 사업안정성, 사업성, 영업경쟁력의 사실정보와 고용창출능력, 우대가점, 자격증소지여부, 경영자 가족관계의 판단정보로 구성하였다.

소상공인 신용평가모형에서 주로 개인평가모형에 기업평가모형을 결합하는 형태의 연구가 진행된 이유는 사업자의 신용상태를 통해 위험을 차별화하는데 효과적이라고 판단하기 때문이다. 이는 소상공인의 자금조달이 사업자의 신용카드와 같은 개인 금융 매개체를 통해 많이 이루어진다는 점에서 개인 신용평가 모형이 기업신용평가 모형보다 더 잘 적용된다고 볼 수 있다(Lee, 2005). 다른 관점으로는, 기업신용평가 모형에서 사용되는 재무적 변수가 소기업이나 소상공인에서는 획득이 어렵고, 재무제표의 신뢰성도 보장하기 어렵기 때문에 개인 신용(consumer risk)을 나타낼 수 있는 신용카드 대금 연체, 대출금 연체, 부동산 보유, 신용불량규제 여부, 차입금 등과 같은 변수들을 많이 이용한다고 볼 수 있다(Mester, 1997). 그리고 재무적 변수의 부재는 결국 비재무변수를 위주로 소상공인의 대출 심사를 담당하고 있는 담당자들의 노하우를 활용한 AHP 기법을 위주로 연구되고 있다.

2.2 SVM(Support Vector Machines)

최근 SVM을 적용한 기업부실예측에 관한 연구가 많이 진행되고 있다. Shin et al.(2005)은 기존에 기업부실예측모형에서 가장 좋은 성능을 보이지만 모형 개발시 한계를 가지는 신경망의 대안으로 SVM을 채택하여 다른 알고리즘과 비교할 때, 더 좋은 성능을 보임을 입증했다. Min and Lee(2005) and Lee(2007)는 학습집단을 k-겹 교차타당성을 활용한 격자탐색 방법으로 SVM에서 사용되는 주요한 변수의 최적값을 찾는 방법을 제안하여 기업의 부실예측모형과 신용평가 모형을 개발했다. Min et al.(2006) and Wu et al.(2007)은 RBF 커널을 사용하는 SVM에서 최적의 변수 및 변수 선택법을 유전자 알고리즘과 SVM을 결합하여 기업의 부실을 예측하였다.

SVM은 주어진 학습집단에만 적합한 함수를 추정하는 것이 아니라, 함수의 클래스를 고차원(high dimension)의 특징공간(feature space)으로 사상하여 간단한 선형판별함수의 사용이 가능하도록 변환하여 학습함으로써 일반화 오류(generalization error)를 최소화한다(Burges, 1998).

학습집단이 $\{x_i, y_i\}$, $i = 1, \dots, l$ 이고 $x_i \in R^d$ 의 입력벡터를 가지고 $y_i \in \{-1, 1\}$ 인 두 개의 집단을 분류하는 문제에서 두

집단을 분류하는 초평면(hyperplane)이 있다고 가정할 때, 초평면은 $W^T \cdot \Phi(x) + b = 0$ 로 표현할 수 있다. 여기서 W 는 단위 길이를 갖는 초평면과 직교하는 가중치 벡터이며, b 는 편차(bias)이다. $\Phi(\cdot) : R^d \rightarrow R^{d_k}$ 은 입력벡터를 고차원의 특징공간으로 이동시키는 역할을 수행한다.

초평면과 평행하면서 반대쪽으로 두 개의 경계초평면(bounding hyperplane)은

$$\begin{cases} W^T \cdot \Phi(x_i) + b \geq +1 \text{ for } y_i = +1 \\ W^T \cdot \Phi(x_i) + b \leq -1 \text{ for } y_i = -1 \end{cases} \quad (1)$$

또는

$$y_i (W^T \cdot \Phi(x_i) + b) - 1 \geq 0 \quad \forall i \quad (2)$$

조건을 만족한다.

이때 원점에서 이 초평면과의 수직거리는 $|b| / \|W\|$ 이며, 따라서 두 경계초평면의 수직거리는 $2 / \|W\|$ 가 되고 이 수직거리를 마진(margin)이라고 표현한다. 이 마진을 최대로 하는 초평면을 찾는 것이 SVM의 목적이다.

두 집단을 완전히 분리시킬 수 없는(non-separable) 대부분의 분류 문제를 해결하기 위해 식 (3)과 같이 오분류(misclassification)을 허용하여 여유변수(slack variable)인 오차항(ξ_i)을 도입한 후 가중치 벡터를 찾아내는 방법을 적용하게 된다.

$$\begin{aligned} & \text{Min } \frac{1}{2} W^T \cdot W + C \sum_{i=1}^l \xi_i \quad (3) \\ & \text{subject to } \begin{cases} y_i (W^T \cdot \Phi(x_i) + b) \geq 1 - \xi_i, & i = 1, \dots, l \\ \xi_i \geq 0, & i = 1, \dots, l \end{cases} \end{aligned}$$

C 는 마진폭에 대응하는 분류오차의 가중치이며, 양의 값을 가지는 실수로 값을 크게 할수록 오차항을 최소화하는 역할을 하고, 값을 적게 할수록 오차항을 최대화하는 역할을 하면서 SVM의 분류성능을 조정하게 된다. 식 (3)은 선형제약식을 가진 이차계획(quadratic programming: QP)모형으로 최적해는 라그랑지승수인 α_i 를 이용하여 구하게 된다. 이때 승수인 α_i 는 학습집단에 각각 곱해지는데, 비음인 α_i 가 존재하는 데이터를 SV(support vector)라고 한다. 결국 SV만을 이용해서 모수값을 추정하게 된다.

$$\text{Max } L_D \quad (4)$$

$$= \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j \Phi(x_i)^T \Phi(x_j)$$

subject to

$$\sum_{i=1}^l \alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq C, \quad i = 1, \dots, l$$

최소화 문제인 식 (3)은 식 (4)와 같은 쌍대문제(Dual problem)로 전환할 수 있으며 $K(x_i, x_j) = \Phi(x_i)^T \Phi(x_j)$ 는 커널함수(kernel function)를 나타낸다. 이는 학습집단의 데이터들이 커널함

수를 통해 고차원의 특징공간으로 이동하게 되는 것을 의미한다. 이는 데이터를 고차원 공간으로 이동시켜 특징 공간내에서 선형으로 분리 가능한 입력자료 집합으로 만들어 주는 역할을 하게 된다. SVM에서 $\phi(x)$ 를 실제로 계산할 수 없기 때문에 Mercer의 조건(Mercer's condition)을 이용하여 매핑함수인 커널함수 $K(\cdot, \cdot)$ 로 문제를 해결하게 된다.

대표적인 커널함수는 RBF (Radial Basis Function) 커널로 가우시안 분포를 활용하며 $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$, $\gamma > 0$ 인 γ 를 조정하면서 최적함수를 구한다. 그리고 polynomial Function는 $K(x_i, x_j) = (\gamma x_i^T x_j + r)^d$, $\gamma > 0$ 인 γ 와 d 를 조정하면서 최적의 함수를 찾아낸다. 여기서 $d, r \in N$ 와 $\gamma \in R^+$ 는 상수이다.

결국 식 (5)와 같은 최종 SVM을 구하게 된다.

$$\hat{y} = \text{sgn} \left(\sum_i \alpha_i y_i K(x, x_i) + b \right) \quad (5)$$

최근 다양한 SVM을 활용한 다양한 분야의 연구가 진행되고 있다. 특히 SVM은 실제 문제에 적용시 다른 알고리즘에 비해 예측력이 우수하고, 오차항을 도입하여 과대적합을 피하면서 합수근사의 문제에서 이상치에 둔감하다는 장점을 가지고 있다(Min and Lee, 2005).

2.3 부실예측에 관한 기존 연구

부실예측모형에 대한 연구는 Durand(1941)가 은행, 금융회사 등 여러 금융기관의 개인신용정보를 이용하여 우·불량 고객을 추출하고, 판별분석에 의해 주요변수의 가중치를 구하면서 시작되었다. 이후 Beaver(1966)에 의해 도산기업과 건전기업을 간의 차이를 단변량 통계분석을 통해 연구하기 시작하면서 기업부도예측모형에 대한 연구로 발전하기 시작했다. 이후 부실예측 모형에 관한 연구는 1980년대 초까지는 통계적 모형인 다변량판별분석, 회귀분석, 로짓(logit) 등에 의해 활발히 진행되었다(Ohlson, 1980; Gentry et al., 1985). 그리고 1990년대 초반부터 머신러닝(machine learning)기법인 인공신경망과 다양한 기법들의 비교를 통해 인공신경망의 예측력의 상대적 우수성에 대해 많은 연구가 진행되었다(Odom and Sharda, 1990; Tam and Kinag,

1992). 또한 최근에는 사례기반추론, 유전자 알고리즘, 의사결정나무, SVM 등과 같은 방법들이 활발하게 연구가 진행되고 있다(Miller et al., 1995; Huang et al., 2004). 본 연구에서는 다변량 판별분석, 로지스틱 회귀분석, 의사결정나무(CART, C5.0), 인공신경망을 통해 모형을 개발하여 본 연구에서 제안하는 SVM 모형과 그 예측력을 비교하고자 한다.

3. 연구 방법

3.1 실험 데이터

본 연구의 목적은 소상공인 신용 평가시 소상공인의 부정확한 재무적 요인과 심사관의 판단정보를 보완하기 위해 가맹점의 신용카드 매출정보의 활용 가능성을 제시하고, 이를 활용한 부실예측모형을 개발하는 것이다.

본 연구를 위해 K-VAN사의 서비스를 이용하는 412,773개의 가맹점에서 발생한 28개월 동안의 데이터 중 월별 신용카드 매출액과 건수 그리고 가맹점 정보를 통해 파생된 다양한 변수들을 부실예측모형의 설명변수로 사용하고, 전국은행연합회에서 얻은 가맹점과 가맹점주의 신용불량 등재정보와 신용카드 매출 중단 정보를 활용하여 목적변수인 소상공인 부실을 정의하고자 한다.

신용카드사는 자사의 카드만을 관리하므로 다양한 카드의 거래가 일어나는 가맹점의 전체 신용카드 매출에 관한 정보를 가지고 있지 않다. 하지만 신용카드 산업에서 신용카드 VAN사는 가맹점의 단말기에서 발생하는 모든 거래 정보를 관리하면서 각 신용카드회사로 거래내역을 실시간으로 분기시켜주는 역할을 한다. 이로 인해 가맹점내에서 발생하는 모든 거래 내역은 신용카드 VAN사가 모두 관리하고 있다. 또한 VAN사는 가맹점주와 정보제공동의서를 득하여 가맹점주가 자신의 가맹점에서 발생한 매출 정보를 요구할 때, 언제라도 신용카드 매출정보를 공개할 수 있게 시스템화 되어 있어 활용이 편리하다.

3.2 실험 설계

본 실험은 <그림 1>과 같은 절차에 따라 진행했다. 전체

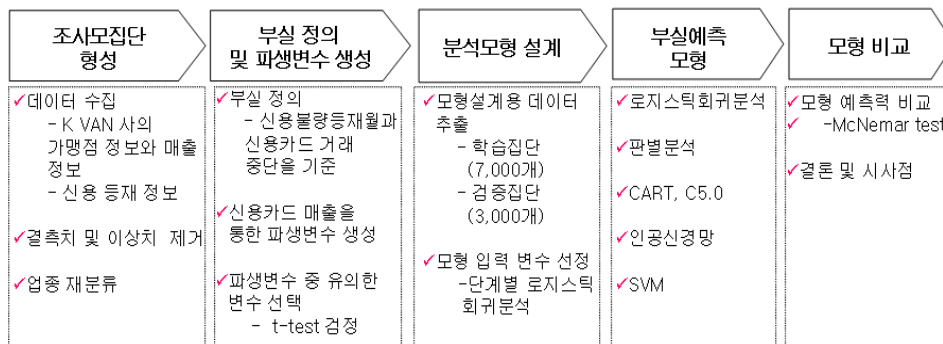


그림 1. 실험 과정

412,773개의 연구 대상으로 결측치와 이상치를 제거하고 이를 통해 부실을 정의하고, 업종, 지속거래기간, 거래평균, 표준편차, 거래 패턴 등 다양한 매출정보와 우량과 부실 가맹점과의 분포를 분석하여 다양한 파생변수를 생성한다. 이렇게 생성된 다양한 파생변수를 이용하여 모형의 입력변수로 사용하며, 다양한 기법들을 통한 부실예측모형을 개발하여 비교 분석한다.

부실 가맹점은 전체 가맹점의 4% 미만에 불과하므로 우량 가맹점과 부실 가맹점을 실제 비율로 표본 추출할 경우 우량 가맹점에 대한 학습이 과적합하여 부실 가맹점에 대한 학습이 이루어지기 어렵다. 이에 부실가맹점 예측 모형을 설계하기 위해 우량 가맹점과 부실 가맹점에서 각각 5,000개 가맹점들을 무작위로 표본 추출하고 이를 학습집합 7,000개, 검증집합 3,000개로 재구성하여 모형의 예측력을 검증하도록 한다.

신용카드 매출정보를 통해 얻은 여러 파생변수들을 t-test를 통해 우량과 불량을 선별하는 유의한 변수들을 선정하여 부실 가맹점 예측 모형 생성에 적용한다.

4. 실증분석 결과

4.1 부실정의

소상공인의 경우, 일반적인 기업들과 달리 개인 사업자가 대부분으로 사업의 주기가 짧다. 그리고 이전이나 폐업에 관한 정보가 부실하여 사업이 매출부진이나 부도에 의한 폐업인지 아니면 사업체 이전이나 다른 목적 등의 의도적 폐업인지에 대한 정보가 전무하다.

본 연구에서는 가맹점의 거래 중단 정보와 전국은행연합회의 신용불량 등재정보를 활용하여 소상공인의 부실을 정의하기로 한다. 이는 소상공인의 경우 신용불량이 발생하더라도 계속해서 매출이 발생하고 증가함으로 인해 자금회전이 회복될 수 있고, 가맹점의 신용카드 매출이 중단되었다고 하더라도 해당 신용카드 VAN사에서 이탈하여 다른 VAN사로 전환한 경우일 수도 있기 때문에 위의 2가지 정보를 조합하여 부실을 정의하고자 한다.

전체 412,773개의 연구 대상 중 27,502개인 6.7%가 신용불량 등재인 것으로 분석되었다. 신용불량 등재 가맹점 27,502개를 대상으로 신용불량 등재월과 마지막 신용거래 발생월간의 차이를 <그림 2>로 도식화한 결과, 신용불량 등재월 이전에 거래가 종료된 가맹점이 16,182개로 58.84%, 거래 종료월과 신용불량 등재월이 일치하는 가맹점이 1,602개로 5.83%를 차지하고 있으며, 신용불량 등재월 이후 거래가 종료된 가맹점은 9,718개로 35.34%로 나타났다. 또한 신용불량 등재 전후 3개월 사이에 거래가 종료된 가맹점은 전체의 32.24%, 전후 6개월 사이에 거래가 종료된 가맹점은 전체의 51.91%를 차지하는 것으로 나타났다.

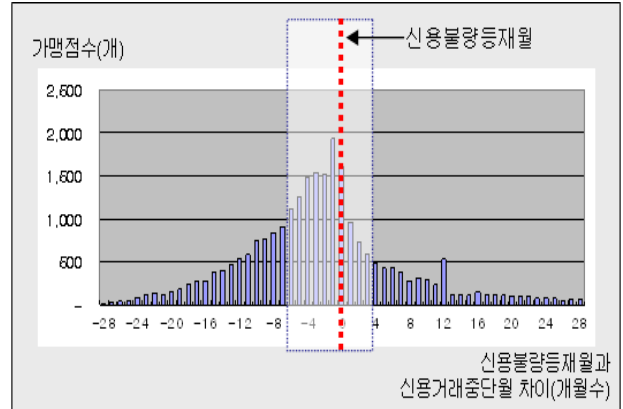


그림 2. 신용불량 등재월과 거래종단시점 차이별 가맹점 분포

소상공인의 부실을 신용불량 등재월 전후 동일한 기간으로 정의할 수도 있지만, <그림 2>의 분포를 통해 신용불량 등재월 전에 가맹점의 신용카드 거래가 중지 되는 경우가 신용불량 등재월 이후 신용카드 거래가 중지되는 경우보다 많은 것을 알 수 있다. 이는 신용불량 등재 이후 이를 극복하지 못하고 영업을 중단하는 것보다 영업 부진으로 영업을 중단하고 몇 개월 이후 신용불량 등재가 발생하는 경우가 많은 것을 의미한다. 이와 같은 이유로 소상공인의 부실 정의에서 신용불량 등재월 이전에 신용카드 거래가 중단된 경우와 이후에 신용카드 거래가 중단된 경우에 해당하는 기간을 다르게 정의하기로 했다. 또한 <그림 2>을 통해 신용불량등재가 발생하기 이전 6개월을 기준으로 가맹점의 분포가 1,000개 이상으로 증가하는 것을 볼 수 있으며, 이후 3개월을 기준으로 가맹점 분포가 500개 이하로 감소하는 것을 볼 수 있다. 이와 같은 이유로 본 연구에서는 신용불량 등재월 이전 6개월에서 신용불량 등재월 이후 3개월까지 신용카드 거래가 중단된 가맹점 12,738개, 신용불량 등재 가맹점의 46.32%를 부실로 정의한다.

4.2 파생변수 생성

신용카드 매출 정보를 이용하여 기간, 매출액, 거래건수, 중분류별 매출과 건수의 분위수, 매출패턴, 그리고 업종에 관련된 다양한 파생변수 40여개를 생성하였다. 그리고 각 파생변수들과 우량과 부실과의 교차분석을 통해 연관성을 살펴보았다. 특히 거래의 지속성, 신용카드 매출 실적, 매출패턴, 업종의 부실율 등은 소상공인 부실과 연관성이 있는 것으로 나타났다.

가맹점들의 매출액 분포는 평균이 779만원인데 비해 가맹점의 74%가 500만원 미만에 밀집해 좌측으로 치우친 형태의 분포를 보였다. 또한 건수의 분포도 매출액 분포와 동일하게 좌측으로 치우친 형태의 분포를 보였다. 이는 매출이 높고 거래가 많은 소수의 가맹점이 전체 가맹점의 평균을 끌어 올리는 형태로 구간화를 위해 금액이나 건수로 등분 구간화 할 경우 실제 데이터의 왜곡이 발생할 수 있다. 이를 보완하기 위해 기간, 매출액, 건수에 관한 변수들의 분포를 반영하여 가맹점의

수를 활용하여 등분 구간화 하였다.

매출패턴의 경우 <그림 3>과 같이 방향성과 정도를 의미하는 값으로 이루어져 방향성(+, 0, -)과 정도(+, 0, -)의 조합으로 9가지의 특성을 나타내는 명목변수임으로 더미(dummy)화하였다. 더미화하여 실험한 결과 매출패턴에 관련된 정보들이 특정 유형에 대해서만 유의한 변수로 추출되었다. 이를 모형에 그대로 사용할 경우, 특정 매출 패턴을 가진 경우의 모형과 그 외 매출패턴을 가진 모형을 나누어 설명해야 하는 번거로움이 발생했다. 본 연구에서는 매출패턴을 방향값 1~3, 정도값 1~3으로 순서형 변수 형태로 나누어 모형에 적용했다.

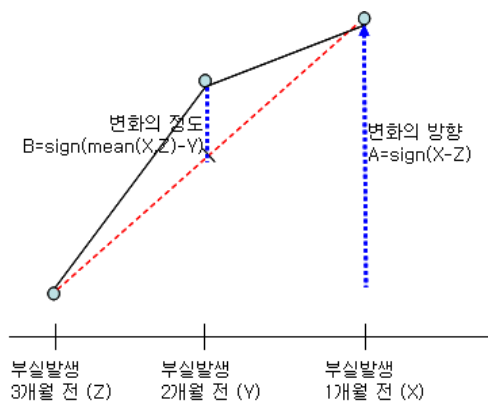


그림 3. 변화의 방향과 정도

업종은 대분류 15개, 중분류 39개, 소분류 160개로 신용카드 VAN사에서 사용되고 있는 분류체계에서 법인사업자를 제외하고 소상공인에 속하지 않는 대형 업종은 제외하여 재분류하였다. 또한 단란주점, 유흥주점, 안마시술소와 같은 유흥성이 강하여 은행권과 여신업에서 제외하는 업종을 제거하였다. 부실율이 가장 낮은 업종은 의료기관으로 0.83%였으며 레저업소, 학원, 수리서비스, 문구·완구, 농업용품, 숙박, 시계·귀금속 등으로 나타나 부실율이 2% 미만으로 낮은 업종으로 분석되었다. 부실율이 가장 높은 업종은 스넥카테일이 5.01%인 것으로 나타났으며, 맥주홀, 사무통신, 양식, 건축자재, 주방용품, 일식, 건강식품은 부실율이 4% 이상으로 높게 나타났다. 업종의 경우도 매출패턴과 같이 39개의 더미변수를 사용하기보다는 중분류 업종의 부실율을 5개의 구간으로 구간화하여 모형에 반영하였다.

이렇게 파생된 변수들을 모형의 입력변수로 사용하기 위해 구간화하여 t-test한 결과 신뢰수준 95%내에서 유의하게 판단되는 변수 22개를 <표 1>과 같이 도출하였다.

유의성 검정 결과 매출발생 개월수가 가장 유의한 변수로 나타나 가맹점의 업력이 소상공인의 부실예측모형에서 중요한 역할을 하고 있음을 알 수 있다. 또한 최대매출월과 최소매출월 사이 기간과 기준시점까지의 지속거래기간, 6개월 동안 거래중단 개월수와 같이 지속적인 영업이 이루어지는 가맹점도 부실에 중요한 역할을 하고 있음을 알 수 있다.

표 1. 파생변수의 유의성 검정

구분	변수	구간	t값
기간	6개월동안 거래중단 개월수	6	15.415**
	최대매출월과 최소매출월 사이기간	14	23.624**
	기준시점까지의 지속거래기간	14	-19.318**
매출액	매출발생개월	14	-31.384**
	월매출평균	11	7.913**
	최고매출액과 최소매출액 차이	11	9.189**
	6개월 매출액 평균	11	6.803**
	6개월 최소매출액	11	-5.192**
	6개월 매출액 표준편차	11	11.329**
건수	평균건단가 지수화	11	6.042**
	월 최고건수	19	9.273**
	월 최소건수	15	4.882**
	6개월 평균 건수	15	-2.953**
분위수	3개월 매출평균의 분위수	6	-1.772*
	3개월 매출표준편차의 분위수	6	1.699*
	3개월 건수평균의 분위수	6	-13.823**
	3개월 건수 표준편차의 분위수	6	-7.673**
매출패턴	3개월 매출 변화유형(방향)	3	1.685*
	3개월 매출 변화유형(정도)	3	-10.993**
	3개월 건수 변화유형(방향)	3	2.536*
	3개월 건수 변화유형(정도)	3	-12.632**
업종	중분류 업종부도율	5	12.693**

주) *: 유의수준 0.05이하, **: 유의수준 0.01이하.

4.3 변수 선정

앞 절에서 생성한 22개의 파생변수들을 기준으로 SVM, 인공신경망, 의사결정나무, 다변량 판별분석, 로지스틱 회귀분석 모형의 예측력을 비교하기 위해 단계적 로지스틱 회귀분석(stepwise logistic regression)을 이용하여 99% 신뢰수준하에서 유의한 12개의 변수 선택하여 분석에 사용할 최종 변수를 <표 2>와 같이 구성하였다.

선정된 12개의 변수는 업력, 지속적 영업력, 매출규모, 매출취약성으로 크게 4가지로 설명할 수 있으며 소상공인 부실예측을 위한 중요한 변수가 된다.

4.4 SVM 모형

본 연구에서는 SVM을 통한 모형을 설계하기 위해 RBF 커널과 d차 다항커널을 사용하였다. 이를 위해 학습집단을 위해 7-겹 교차타당성 분석을 통해 학습하고 실험도구로 Matlab LIBSVM을 사용하였다.

먼저 RBF 커널을 활용한 최적의 SVM을 구하기 위해 허용오차 C와 커널모수인 γ 를 격자탐색방법을 이용하였다. 격자탐

표 2. 모형비교를 위해 선택된 변수

구분	변수	설명	β 값
기간	6개월 동안 거래중단 개월수	CBP	-0.091
	최대매출월과 최소매출월 사이 기간	CBP	-0.078
	기준시점까지의 지속거래기간	CBP	-0.062
	매출발생개월	BH	0.261
매출액	매출평균	SoS	-0.205
	최고액과 최소액의 차이금액	FoS	-0.131
	6개월 최소매출액	FoS	0.112
건수	최소건수	FoS	0.047
분위수	3개월 건수평균의 분위수	SoS	0.325
매출패턴	3개월 거래의 변화 방향	SoS	-0.268
	3개월 거래의 변화 정도	SoS	0.239
업종	중분류 업종 부도율	FoS	-0.194

주) CBP=지속적 영업력, BH=업력, SoS=매출규모, FOS=매출취약성.

색방법은 C 와 γ 의 여러 조합을 통해 최적의 C 와 γ 의 조합을 찾는 방법이다. 이 방법은 모든 조합을 만들어 학습을 진행해야 하기 때문에 많은 비용이 소요되는 단점이 있지만, 휴리스틱(heuristic)하게 접근하는 방식에 비해 최적의 모수를 찾을 수 있다는 장점이 있다(Min et al., 2005). 본 연구에서는 C 와 γ 를 10의 승수로 조합하여 학습집단과 검증집단을 통해 최적의 C 와 γ 를 구했다.

표 3. 학습집단을 통한 RBF 모수별 예측력

$C \backslash \gamma$	10^0	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
10^0	58.80	76.60	78.40	78.70	77.50	70.50	69.90	70.00
10^1	59.10	68.60	78.00	79.00	78.90	77.40	70.40	70.00
10^2	59.10	67.30	75.80	78.60	79.10	78.60	77.40	70.00
10^3	59.10	67.30	68.90	79.00	78.70	79.40	78.70	77.40
10^4	59.10	67.30	67.20	78.00	79.00	79.40	79.60	78.60
10^5	59.10	67.30	66.00	76.40	79.10	78.60	79.70	79.50
10^6	59.10	67.30	65.60	74.40	78.70	79.20	78.60	79.60
10^7	59.10	67.30	66.20	74.40	76.30	78.90	78.40	79.60

<표 3>은 학습집단을 격자탐색방법을 통해 얻은 최적의 C 와 γ 는 10^5 과 10^{-6} 으로 예측률이 79.70%(우량: 79.75%, 부실: 79.64%)로 나타났다. 하지만 <표 4>와 같이 검증집단을 통해 검증한 결과 70.83%(우량: 71.4%, 부실: 86.96)로 다른 모수에 비해 낮은 예측력을 보이고 있었다.

학습집단의 격자탐색과 검증집단의 격자탐색을 조합해 볼 때, C 와 γ 는 10^3 과 10^{-3} 에서 학습집단은 79.0% (우량: 79.55%, 부실: 78.44%), 검증집단은 74.20% (우량: 74.86%, 부실: 88.66%)

으로 최적의 예측률을 갖는 것으로 분석하였다.

표 4. 검증집단을 통한 RBF 모수별 예측력

$C \backslash \gamma$	10^0	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
10^0	58.30	72.50	73.43	72.17	69.73	65.37	64.77	64.73
10^1	59.07	67.87	73.77	72.73	71.30	69.63	65.27	63.73
10^2	59.07	67.27	71.93	72.97	71.90	70.77	69.67	65.23
10^3	59.07	67.30	69.33	74.20	72.67	71.00	70.70	69.63
10^4	59.07	67.30	66.13	73.63	72.67	71.90	70.43	70.80
10^5	59.07	67.30	65.30	72.77	73.63	72.87	70.83	70.33
10^6	59.07	67.30	64.77	67.37	73.20	72.87	71.03	70.17
10^7	59.07	67.30	64.67	67.37	73.00	72.17	71.57	69.73

d 차 다항커널을 활용한 최적의 SVM을 구하기 위해 허용오차 C 와 커널모수인 γ , 그리고 차수 d 를 여러개의 조합으로 실험한 결과 <표 5>와 <표 6> 같이 나타났다.

표 5. 학습집단의 d 차 다항커널 모수별 예측력

C	d	γ				
		10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10^2	1	79.9	79.9	79.6	79.1	76.2
	2	79.0	78.5	76.8	60.7	60.7
	3	76.8	76.8	61.8	54.2	54.2
	4	74.6	73.7	51.6	51.6	51.6
	5	73.1	71.7	51.0	51.0	51.0
10^3	1	79.9	80.0	79.9	79.6	79.1
	2	79.2	79.0	78.2	69.9	60.7
	3	77.4	79.2	76.5	54.2	54.2
	4	76.7	76.3	58.2	51.6	51.6
	5	74.4	73.1	51.0	51.0	51.0
10^4	1	79.9	80.0	79.9	79.9	79.6
	2	79.2	79.0	78.5	76.8	60.7
	3	76.9	78.3	75.9	54.2	54.2
	4	76.8	76.3	64.8	51.6	51.6
	5	74.6	74.0	53.8	51.0	51.0

RBF 커널을 통해 최적의 C 와 γ 인 10^3 과 10^{-3} 를 중심으로 C 는 10^2 에서 10^4 까지 확대하고 γ 는 10^{-2} 에서 10^{-6} 까지 확대하면서 d 는 1에서 5까지 늘려 실험하였다.

학습집단을 통해 d 차 다항커널의 최적 모수를 선정할 경우 C 는 10^3 , γ 는 10^{-3} , d 는 1일 경우 최적의 SVM을 구성한다. 하지만 <표 6>과 같이 검증집단의 예측력과 비교해 볼 때, C 는 10^3 , γ 는 10^{-3} , d 는 3일 경우 학습집단의 예측력은 79.2%(우량: 79.95%, 부실: 78.44%), 검증집단의 예측력은 73.6%(우량: 74.46%, 부실: 72.8%)로 최적의 예측력을 갖는 것으로 나타났다.

표 6. 검증집단의 d차 다항커널 모수별 예측력

C	d	γ				
		10-2	10-3	10-4	10-5	10-6
10^2	1	70.3	70.3	70.3	70.6	68.9
	2	72.7	72.2	70.5	58.7	58.7
	3	72.2	71.6	59.0	53.6	53.6
	4	71.7	68.4	51.7	51.7	51.7
	5	70.8	66.7	51.2	51.2	51.2
10^3	1	70.4	70.3	70.3	70.3	70.6
	2	72.4	72.1	71.5	64.4	58.7
	3	71.1	73.6	69.5	53.6	53.6
	4	70.7	70.4	55.9	51.8	51.8
	5	70.4	68.0	51.2	51.2	51.2
10^4	1	70.3	70.3	70.3	70.3	70.3
	2	72.8	72.0	72.2	70.5	58.7
	3	70.8	73.0	69.2	53.6	53.6
	4	70.6	71.8	60.9	51.8	51.8
	5	70.3	69.9	53.5	51.2	51.2

SVM의 실험 결과 RBF 커널이 d차 다항커널 보다 1% 정도의 높은 예측력을 보였다. 또한 학습 속도면에서도 RBF 커널을 통한 학습이 d차 다항커널에 비해 빨랐다. 위의 결과를 통해 본 연구에서는 SVM은 RBF 커널의 C와 γ 는 10^3 과 10^{-3} 인 모형을 선정하였다.

4.5 다른 부실예측 모형과의 비교

본 연구에서는 SVM의 예측률을 기존에 사용되었던 부실예측 모형과 비교하기 위해, 인공신경망, 의사결정나무(CART, C5.0)은 클레멘타인(clementine) 8.5를 이용하고, 다변량판별분석, 로지스틱회귀분석 모형은 SPSS 12.0을 이용하여 설계하였다.

인공신경망 모형은 입력층의 뉴런수 12개, 은닉층 1개, 출력층의 뉴런수 1개, 전이함수는 sigmod 함수를 사용하였고, 학습률과 모멘텀은 0.3으로 고정하였으며, <표 7>과 같이 학습회수를 1,000번, 2,000번, 3,000번으로 증가시키고 은닉층의 뉴런수를 6에서 24까지 2씩 증가시키면서 최적의 모형을 구하기 위해 반복 실험하였다.

실험결과 학습 회수가 3,000회이고 은닉층의 뉴런수가 8개 일 때 가장 우수한 예측력을 나타내는 것으로 분석되었다. 학습 집단은 76.20%, 검증집단은 73.13%(우량: 73.76%, 부실: 72.53%)의 예측률을 보였다.

의사결정나무 모형을 설계하기 위해 CART와 C5.0 알고리즘을 활용하여 최적의 모형을 설계하였다. CART를 이용한 모형은 가지의 깊이를 4에서 하나씩 늘려가면서 실험하였으며, 깊이가 8에서 학습집단은 74.70%, 검증집단은 70.33%(우량: 71.60%, 부실: 69.06%)로 가장 높은 예측력을 보였다. C5.0을 이용한 모

형은 가지의 깊이가 14일 때 학습집단은 78.20%, 검증집단은 70.83%(우량: 70.67%, 부실: 71.0%)로 가장 높은 예측력을 보였다.

표 7. 인공신경망의 예측정확도

학습 회수	은닉노드	정확도(%)	
		테스트집단	검증집단
1,000	6	77.20	72.20
	8	79.10	73.03
	12	77.30	72.73
	14	76.70	72.07
	16	78.60	72.90
	18	77.10	72.73
	20	79.10	72.63
	24	78.50	72.83
2,000	6	77.80	72.90
	8	77.80	73.10
	12	78.90	73.07
	14	79.70	71.80
	16	79.20	72.17
	18	79.40	72.47
	20	79.90	72.27
	24	78.30	72.30
3,000	6	78.60	72.80
	8	78.50	73.13
	12	77.70	71.87
	14	78.50	72.37
	16	78.10	72.63
	18	78.20	72.70
	20	78.50	72.23
	24	80.20	71.23

표 8. 알고리즘별 예측정확도

모형	정확도(%)	
	학습집단	검증집단
SVM	79.00	74.20
인공신경망	78.50	73.13
의사결정나무(C 5.0)	78.20	70.83
의사결정나무(CART)	74.70	70.33
다변량 판별분석	68.96	70.10
로지스틱 회귀분석	68.89	70.10

다변량판별분석과 로지스틱회귀분석에서는 학습집단의 예측률과 검증집단의 예측률이 거의 비슷하게 나타났다.

4.6 예측성과 비교

앞 절에서 실험한 기법들의 예측정확도를 비교하면 SVM이 다른 기법들에 비해 약 4% 높게 나타나고 있으며, <표 8>과 같이 SVM, 인공신경망, 의사결정나무, 다변량판별분석, 로지스틱회귀분석의 순으로 예측정확도가 나타났다.

SVM과 나머지 분류기법들간의 정확도의 차이가 통계적으로 유의한지 여부를 검증하기 위해 명목변수의 두 집단간의 대응표본 차이 검정에 사용되는 맥네마(McNemar) 검정을 실시하였다. 맥네마 검정은 관련된 두 가지 이분형 변수에 대한 비모수 검정으로 카이제곱 분포를 사용하여 응답 변화를 확인하며, 전후 계획에서 실험적 조정에 따른 응답 변화 내용을 추적하는데 유용하다.

표 9. SVM과 알고리즘간의 McNemar 검정

모형	NN	DT	MDA	logit
SVM	2.684 (0.101)	18.416 (0.0000*)	35.021 (0.0000*)	35.021 (0.0000*)

주) McNemar 통계량 값(유의확률), * : P<0.01.

검증집단의 예측과 비예측 결과를 통해 맥네마 검정 실험 결과는 <표 9>와 같이 SVM과 인공신경망은 유의한 차이가 없는 것으로 나타났으나 SVM과 의사결정나무, 다변량판별분석, 로지스틱회귀분석의 모형간에는 유의수준 1%하에서 예측 결과의 통계적인 차이가 있는 것으로 나타나 다른 모형에 비해 예측력이 뛰어난 것을 확인할 수 있다.

5. 결론

신용카드 매출정보가 소상공인의 부실예측 모형에서 부실한 재무정보를 보완할 수 있는 대안으로써 중요한 정보를 보여주는 이번 연구에서 가맹점의 신용카드 매출의 특성을 분석하고, 우량과 부실에 영향을 주는 주요한 변수인 업력, 지속적 영업력, 매출규모, 매출취약성으로 대변되는 변수들을 추출하였다. 업력을 대변하는 매출발생개월, 지속적 영업력을 대변하는 기준시점까지의 지속거래기간, 6개월 동안 거래중단 개월수, 최고 매출발생월과 최소 매출발생월 사이 기간이 선택되었다. 또한 가맹점의 매출규모를 대변하는 평균 매출액, 3개월간 건수평균 분위수, 3개월간 거래 패턴(방향, 정도) 그리고 매출취약성을 대변하는 최고액과 최소액의 차이금액, 월 최소건수, 6개월 최소매출액, 중분류 업종의 부도율과 같은 중요한 변수를 찾아냈다.

또한 이렇게 추출된 변수들을 통해 우량과 부실에 대한 최적의 예측력을 갖는 SVM 모형을 개발하였다. 시간적 제약에도 불구하고 최고의 예측력을 찾을 수 있는 격자탐색 방법을 통해 RBF 커널에서 74.2%의 예측력을 보이는 SVM 모형을 개발

하였다. 선행연구에서 제시된 모형들 중 Gim and Jung(1999)이 제시한 신경망 모형이 70.7%의 예측력을 보이고 있지만, 이는 소상공인이 포함되어 있는 소기업의 도·소매·서비스업에 대한 부실예측이었으며 소상공인을 대상으로 한 연구는 대부분 60%대 예측율을 보이고 있어 본 연구를 통해 개발된 모형의 예측력이 더 우수함을 보였다.

본 연구의 한계는 자료 수집의 한계로 인해 각 가맹점주 정보와 재정상태, 경영능력 등에 관한 정보를 모형에 적용하지 못했다.

본 연구에서 나타나는 한계점을 고려한 향후 연구로는 GIS (Geographic Information System)를 통한 상권분석 및 입지분석 정보와 신용카드 매출정보를 소상공인 신용평가 모형에 적용하고, 다양한 커널함수를 사용하여 성능과 시간적으로 최적의 모수를 갖는 SVM에 대한 연구가 진행될 것이다.

참고문헌

Altman, E. I., Haldeman, R. G. and Naraynan, P. (1977), ZETA Analysis : A new Model to Identify Bankruptcy Risk of Corporations, *Journal of banking and Finance*, 1, 29-54.

Beaver, W. (1966), Financial ratios as predictors of failure, *Empirical Research in Accounting: Selected Studies, Supplement to Journal of Accounting Research*, 4, 77-111.

Berger, A. N., Frame, W. S., and Miller, N. (2005), Credit Scoring and the Availability, Price, and Risk of Small Business Credit, *Journal of Money, Credit, and Banking*, 37(2), 191-222.

Burges, C. J. C. (1998), A tutorial on support vector machines for pattern recognition, *Data Mining and knowledge Discovery*, 2(2), 995-974.

Cole, R. A., Goldberg, L. G. and White, L. J. (2004), Cookie Cutter versus Character : The Micro Structure of Small Business Lending by Large and Small Banks, *Journal of Financial and Quantitative Analysis*, 39(2), 227-251.

Durand, D. (1941), Risk Elements in Consumer Installment Financing, *National Bureau of Economic Research*, New York.

Frame, W. S., Padhi, M., and Woosley, L. (2004), The Effect of Credit Scoring on Small Business Lending in Low-and Moderate Income Areas, *Financial Review*, 39(1), 34-54.

Gentry, J. A., Newbold, P. and Whitford, D. T. (1985), Classifying bankrupt firms with funds flow components, *Journal of Accounting Research*, 23(1), 146-160.

Gim, K. Y. and Jong, S. Y. (1999), A Study of Credit Evaluation on Small Company, *The Korean Small Business Review*, 21(1), 229-257

Gim, K. Y., Kim, S. H., and Lee, Y. H. (2005), A development of Credit Evaluation Model for Small Business by Applying the AHP, *The Korean Venture Management Review*, 8(1), 3-29.

Huang, Z., Chen, H., Hsu, C. J., Chen, W. H., and Wu, S. (2004), Credit rating analysis with support vector machines and neural networks: A market comparative study, *Decision Support Systems*, 37(4), 543-558.

Jung, W. C. (2002), An Practical Analysis of the Importance of Non-Financial Information in Evaluation of Organizations, *Yonsei Uni. Master's Thesis*.

Kim, J. H. (2006), An Empirical Study on the Credit Evaluation and Firm Size in Korea Medium-Small Firms, *Tax Accounting Research*, 18(1), 145-167.

Lee, J. H. (2005), The Financial Institution Access and Government Financial Assistance of Small Business Firms, *The Korean Small Business Review*, 27(4), 257-272.

- Lee, Y. (2007), Application of support vector machines to corporate credit rating prediction, *Expert Systems with Applications*, 33(1), 67-74.
- Miller, W., Cadden, D. T., and Driscoll, V. (1995), Bank failure and categorization- A neural network approach, *Proc. Third International Conference on Artificial Intelligence Applications on Wall Street*, 232-235.
- Min, J. H. and Lee, Y. (2005), Bankruptcy prediction using support vector machines with optimal choice of kernel function parameters, *Expert Systems with Applications*, 28(4), 603-614.
- Min, S., Lee, J., and Han, I. (2006), Hybrid genetic algorithms and support vector machines for bankruptcy prediction, *Expert Systems with Applications*, 31(3), 652-660.
- Ohlson, J. (1980), financial Ratios and probabilistic prediction of bankruptcy, *Journal of Accounting Research*, 18(1), 109-131.
- Odom, M. and Sharda, R. (1990), A neural networks model for bankruptcy prediction, *Proceedings of the IEEE International Conference on Neural Network*, 2, 163-168.
- Park, J. G.(2001), Study on the Credit Evaluation Model of the Support System of the Small Merchants and Industrialists, *Catholic University of Daegu*, Master's Thesis.
- Shin, K., Lee, T. S., and Kim, H. (2005), An application of support vector machines in bankruptcy prediction model, *Expert Systems with Applications*, 28(1), 127-135.
- Tam, K. and Kinag, M. (1992), Managerial applications of neural networks: the case of bank failure predictions, *Management Science*, 38(7), 926-947.
- Wu, C., Tzeng, G., Goo, Y. and Fang, W. (2007), A real-valued genetic algorithm to optimize the parameters of support vector machines for predicting bankruptcy, *Expert Systems with Applications*, 32(2), 397-408.
- Yang, J. H. (2003), The Development of the Credit Evaluation Model for Small-Micro Business Using AHP, *Korea Advanced Institute of Science and Technology*, Master's Thesis.



윤종식

동국대학교 정보통계학과 학사
 동국대학교 산업시스템공학과 석사
 동국대학교 산업시스템공학과 박사 수료
 현재 : Dun & Bradstreet Korea
 관심분야 : 데이터마이닝, 다변량분석,
 금융공학



권영식

서울대학교 산업공학과 졸업
 한국과학기술원 공학석사
 한국과학기술원 공학박사
 현재: 동국대학교 산업시스템공학과 교수
 관심분야: 데이터마이닝, IT서비스전략



노대협

연세대학교 응용통계학과 학사
 한국과학기술원 경영공학과 석사
 한국과학기술원 경영공학과 박사
 현재: 서울여자대학교 경영학과 교수
 관심분야: 데이터마이닝, 고객관계관리, 전자
 상거래, 회계/재무 정보시스템