

일반화가속모형을 이용한 기술신용평가 주요 지표 분석

Analysis of Important Indicators of TCB Using GBM

전우정(Woo-Jeong(Michael) Jeon)*, 서영욱(Young-Wook Seo)**

초 록

기술력 기반의 중소벤처기업에 대한 기술금융 지원을 위해 정부는 2014년 7월부터 기술보증기금 및 일정 자격을 갖춘 민간 기술신용평가사에게 일종의 기술력 등급평가인 기술신용평가를 실시하여 은행의 여신에 활용토록 하였다. 본 논문에서는 최근까지의 기술신용평가 현황 및 한국신용정보원에서 축적하고 있는 기술평가 관련 가용 지표들에 대한 선행 연구를 개략적으로 살펴본 후 기술평가등급점수에 유의적인 영향을 미치는 지표(indicator)를 통상적인 다중회귀기법으로 탐색할 것이다. 본 논문의 관심 대상인 지표 별 등급 영향도와 모형의 적합도는 대표적인 기계학습 분류기(classifier)인 일반화가속모형(Generalized Boosting Model; GBM)을 적용하여 분석하였는 바, 주요 지표를 독립변수(feature)로 투입하여 지표의 상대적 중요성 및 분류 정확도를 산출하였다. 분석결과 회귀모형과 기계학습 모형 간 지표별 상대적인 중요도는 크게 차이나지 않는 것으로 분석되었으나, GBM 모형의 경우 회귀모형에 비해서 이노비즈 인증, 연구소 및 연구개발전담부서 보유, 특허등록건수, 벤처확인 지표 등 기술개발역량이 상대적으로 기술등급에 더 큰 영향을 미치는 것으로 분석되었다.

ABSTRACT

In order to provide technical financial support to small and medium-sized venture companies based on technology, the government implemented the TCB evaluation, which is a kind of technology rating evaluation, from the Kibo and a qualified private TCB. In this paper, we briefly review the current state of TCB evaluation and available indicators related to technology evaluation accumulated in the Korea Credit Information Services (TDB), and then use indicators that have a significant effect on the technology rating score. Multiple regression techniques will be explored. And the relative importance and classification accuracy of the indicators were calculated by applying the key indicators as independent features applied to the generalized boosting model, which is a representative machine learning classifier, as the class influence and the fitness of each model. As a result of the analysis, it was analyzed that the relative importance between the two models was not significantly different. However, GBM model had more weight on the InnoBiz certification, R&D department, patent registration and venture confirmation indicators than regression model.

키워드 : 기술신용평가, 기계학습, 일반화가속모형, 손실함수, 경사하강법, 카트나무모형
TCB, Machine Learning, Generalized Boosting Model, Loss Function, Gradient Descent, CART

* First Author, MOT Major Department of Business Consulting, Daejeon University(jeonmichael201809@gmail.com)

** Corresponding Author, Department of Business Consulting, Daejeon University(ywseo@dju.kr)

Received: 2017-10-22, Review completed: 2017-11-22, Accepted: 2017-11-26

1. 서 론

기술력 기반의 중소벤처기업에 대한 기술금융 지원을 위해 정부는 2014년 7월부터 기술보증기금 및 일정 자격을 갖춘 민간 기술신용평가사(Technology Credit Bureau; 이하 TCB사)에게 일종의 기술력 등급평가인 기술신용평가를 실시하여 은행의 여신에 활용토록 하였다[1]. 기술신용평가를 통해 기술신용등급이 결정되면 해당 기업은 대출금리 및 한도 책정에 있어서 우대를 받을 수 있다[2].

이러한 정책당국의 취지에 맞춰 기술신용평가모형(이하 TCB 모형)이 개발되었으며, 모형은 크게 기존 신용평가와 동일한 기업평가부분과 기업의 기술사업역량 및 기술경쟁력을 평가하는 기술평가부분으로 구성되어 있다.

기술신용대출의 결정요인은 기술금융의 정책목적 상 기술평가등급이 절대적이다. 기술금융이라는 정책목적외에 실무적으로도 기술평가부분이 상대적으로 더 큰 영향력을 가지는 중요한 요인인데, 이는 실제 은행 여신이 발생하는 과정을 살펴보면 알 수 있다. 기술신용평가는 기업이 아니라 해당 기업과 거래하는 은행이 신청하는 구조이다. 신용평가에 대한 자체 심사팀과 노하우를 보유한 시중 은행이 TCB사에 평가 의뢰하는 기업은 이미 일정수준 이상의 신용담보력을 보유하거나 오랜 거래관계 등으로 우량 고객일 가능성이 크다. 따라서 기술신용평가에서 기업(신용)평가 부분은 상대적으로 변동성이 작고 이에 따라 변동성이 상대적으로 큰 기술평가등급에 따라 금리 및 한도 전략이 결정되는 구조이기 때문이다.

기술평가의 항목은 크게 기업의 사업역량 측정과 기술경쟁력 측정의 2개 대항목으로 구성되어 있다. 사업역량은 보유기술의 주체인 기업의 경영주 능력, 조직관리능력, 기술개발능력, 제품화역량, 수익전망 등으로 구성되어 있고, 기술경쟁력은 기업이 보유한 기술의 혁신성, 시장현황, 기술(제품)의 우위성 등에 대한 평가항목으로 구성되어 있다[3]. 즉, 기술신용평가는 기술을 보유한 주체인 기업에 대한 평가와 함께 보유 기술 자체의 경쟁력에 대해서 평가하는 것이다.

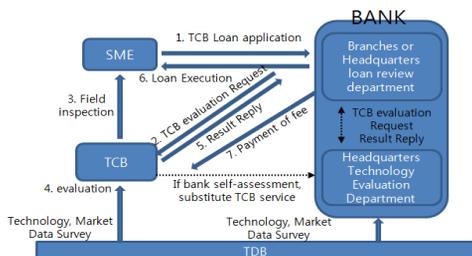
본 논문은 먼저 최근까지의 기술신용평가 현황 및 한국신용정보원의 TDB(Tech Data Base)의 기업(기술) 데이터 분석 및 가공 서비스인 TPS에서 제공하는 기술평가 관련 가용 지표들에 대해서 개략적으로 살펴본다. 이후 주요 기술평가등급점수에 유의적인 영향을 미치는 지표(indicator)를 통상적인 다중회귀기법으로 탐색할 것이다. 이러한 결과를 바탕으로 본 논문의 주요 관심 대상인 지표 별 등급 영향도 및 모형의 적합도는 대표적인 기계학습 분류기(classifier)인 일반화가속모형(Generalized Boosting Model; 이하 GBM)을 적용하여 산출하였다. GBM은 주요 지표를 독립변수(feature)로 투입하여 최적 횟수만큼 반복 학습시킨 후 각 지표의 상대적 중요성 및 분류의 정확도를 산출하는 기계학습의 한 종류로서 회귀와 분류의 특성을 동시에 지니는 모형이다.

최종적으로 기술신용평가의 주요 지표에 대한 전통적인 다중회귀 모형의 분석 결과와 GBM의 결과를 비교분석하여 두 모형의 차이점 및 추후 시사점을 도출하였다.

2. 기술신용평가 현황

기술신용평가는 크게 대출집행 기관인 시중 은행과 평가기관인 TCB사 및 평가에 수반되는 기술 및 시장정보를 제공하는 평가 인프라 성격의 TDB로 구성되어있다. TDB는 평가프로세스의 효율성과 평가활용 정보의 일관성을 확보하기 위해서 한국신용정보원이 2014년 상반기에 설립한 일종의 기술 및 시장 정보 데이터베이스로서 TDB가 생산하는 자체 기술 및 시장 보고서, TCB 평가서, 특허, 논문, 기업인증정보, 공공 및 민간 연구소에서 생성하는 기술가치평가서 및 시장정보 등 약 1,000만 건 이상의 데이터를 축적중이며 계속 업데이트 하고 있다[4].

기술신용평가 프로세스는 아래 <Figure 1>과 같다. A기업이 B은행(영업점 및 본점)에 TCB 대출을 신청하면 B은행은 거래이력 등을 고려하여 TCB사(2017. 10 기준 NiceD&B, Kibo, KED, eCredible, NICE 평가정보, SCI 평가정보 등 총 6개社) 혹은 B은행 자체 기술평가부서에 TCB 평가를 요청한다. 현재 대부분의 기술신용평가는 민간 TCB사에 의해서 진행되고 있는데 정부는 은행자체의 기술평가 인력 및 시스템 등 기술신용평가 인프라가 성숙되면 은행자체 평가 비중을 최대 100%까지 늘리는 방안을 수립한 상태이다[5].



<Figure 1> TCB Loan Process

평가요청을 받은 TCB사는 B은행이 제출한 A기업의 재무제표 및 특허리스트 등 관련 제반 서류를 확인한 후 A기업에 대한 현장실사 및 TDB 등을 활용하여 관련 시장 및 기술 조사를 통해 기술신용평가를 진행한다. 자체 검수과정 등을 거쳐 최종 기술신용평가 결과가 확정되면 TCB사는 이를 신청기관인 B은행 영업점 및 본점에 회신하고, B은행은 A기업에게 TCB 대출을 실시하게 된다. 최종적으로 B은행은 A기업에 대한 평가를 진행한 TCB사와 기술신용평가 수수료를 정산함으로써 A기업에 대한 TCB 대출 프로세스를 종료하게 되는 것이다.

2.1 평가 현황

2017년 6월 말 현재 금융위에 따르면 기술신용 대출 잔액은 112조 8,172억 원으로서, 2014년 7월 말 최초 1,922억 원의 대출잔액에 비해서 3년 동안 약 580배 폭증한 규모이다. 이는 중소기업 총 대출의 약 40% 정도가 기술신용 대출로 이루어지고 있음을 의미한다. 연장, 대환, 재계약 등을 제외한 순수 대출평가액은 71조 7,797억 원 수준이다.

은행별 대출잔액을 살펴보면, 기업은행이 34조 5,706억 원으로 가장 많은 대출잔액을 보유하고 있는데, 이는 기업은행이 정책목적상 특수목적 은행 성격이 강하기 때문인 것으로 보인다. 신한은행이 16조 4,442억 원 수준으로 두 번째로 많은 대출잔액을 보유중이며, 이어서 국민은행이 15조 4,655억 원, 우리은행 13조 3,062억 원, KEB하나은행 12조 7,255억 원 순인 것으로 나타났다. 소규모 지방은행의 경우 부산은행이 3조 8,604억 원, 대구은행이 3조 2,351억 원, 경남은행이 2조 9,198억 원으로 3개 은

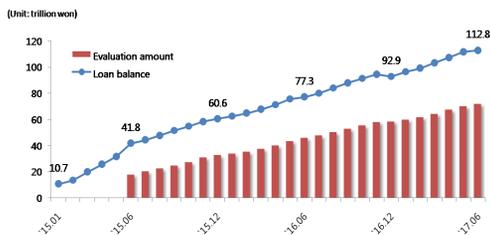
행이 약 10조 원 규모의 기술신용대출을 실행한 것으로 조사되었다.

한편, 기술신용대출 건수도 폭발적으로 증가하였는데 2014년 7월 최초 486건에서 현재 252,295건으로 약 500배 이상 증가하였다. 기업은행이 73,210건, 신한은행이 43,542건, 국민은행 40,229건 순으로 대출건수가 많은 것으로 나타났다. 지방은행의 경우 부산은행이 7,512건으로 가장 많았으며, 대구은행 6,505건, 경남은행 6,349건으로 비슷한 수준을 보였다.

이상을 참고하면, 금년 상반기 TCB 정책금융을 둘러싼 외부 여건의 큰 변화에도 불구하고 기술신용대출규모는 지속적으로 증가하고 있음을 알 수 있다. 대출 시장은 기업은행을 선두로 신한과 국민은행이 2그룹, 우리 및 KEB 하나은행이 3그룹을 형성하여 경쟁하고 있으며, 지방은행은 부산은행이 다소 높은 점유율을 보이는 가운데 대구은행 및 경남은행이 2그룹으로서 지방 중소기업의 기술신용대출시장을 주도하고 있는 것으로 나타났다.

2.2 평가 모형 및 지표

TCB 평가 모형은 경영주역량, 관리능력, 기술개발능력, 제품화역량, 수익전망을 평가하는 기술사업역량과 기술혁신성, 시장현황 및 제품



〈Figure 2〉 TCB Loan Balance & Evaluation amount

우위성을 평가하는 기술경쟁력의 2개 대항목으로 구성되어 있으며 각 대항목에 하위 항목으로서 앞에서 언급한 중항목 및 개별 소항목으로 구성되어 있다. 동 체계는 모든 TCB사가 동일하나 실제 평가 소항목 구성 및 배점(weight)은 각 TCB별로 상이하다. 이는 기술

〈Table 1〉 TCB Standardized Model Structure

Category	Item	Sub Item
1. Technology Business Capability	1.1 CEO	Entrepreneurial spirit
		Level of experience in the industry
		Skills in Management of Technology Management
	1.2 Management ability	Level of expertise of executives
		Degree of capital participation of executives
	1.3 Ability to develop technology	Intellectual Property Rights
		R&D investment ratio
		R&D activation level
		Technical staff expertise
	1.4 Productization capability	technology development and commercialization
		Capital procurement ability
	1.5 Earnings forecast	Production capacity
		Marketing competence
		Ability to create added value
	2. Technical Competitiveness	2.1 Technology Innovation
Differentiation of technology		
Imitation difficulty		
Completeness of technology		
Technology self-reliance		
2.2 Market Status		TCT
		Market Size/Growth
2.3 Market competitiveness		Market structure and characteristics
		Market position
		Competitive Advantage

보증기금과 같은 공공기관과 민간 신용평가사로부터 출발한 TCB사간의 평가 정책상의 차이 및 각 민간 신용평가사 간에도 기존의 신용평가모형 상 비재무항목의 차이 및 평가 관점상의 차이 때문인 것으로 판단된다.

이러한 상황을 반영하여 한국신용정보원에서는 TCB사별 중복 항목과 불필요한 항목을 제거하여 표준모형을 개발 중으로 본 논문에서는 표준모형의 평가항목을 기준으로 TDB TPS의 가용 지표를 기준으로 분석하였다.

상기 항목체계를 기반으로 기업의 기술력에 영향을 미치는 요인으로 고려되어 한국신용정보원이 TDB에 축적하고 있는 주요 지표는 아래 <Table 2>와 같으며, 이에 대한 기존 선행연구는 다음과 같다.

기업이 신용상태를 기반으로 하는 일반대출

<Table 2> Selection of the Appropriate Indicators

No	TDB holding Indicator	check
1	Technology credit loan ratio	O
2	Facility loan ratio	O
3	Industry default rate	X
4	Total default rate	X
5	Industry delinquency rate	O
6	Average sales	Operating margin O
7	Average operating profit	
8	Assets	O
9	Retention rate of Research institute/R&D department	O
10	MAINBIZ(certification ratio)	O
11	VENTURE(verification ratio)	O
12	INNOBIZ(certification ratio)	O
13	Number of Patent applications	O
14	Number of Patent registrations	O
15	Number of employees	O
16	Number of technical staff	X
17	Technical staff ratio	O

에 비해서 기술대출 비율(Technology credit loan ratio)이 높을수록 해당 기업은 높은 기술혁신역량을 보유하여 상대적으로 고(高)기술 등급을 받을 가능성이 높을 것으로 생각할 수 있다. 특히, Lee[6]는 시설자금 대출(Facility loan ratio)의 경우, 운전자금 대출에 비해서 노동과 자본의 증가이외에 기술역량 발전에 따른 생산성 증가를 의미하는 총요소생산성(TFP)이 더 높다고 주장하였다.

연체율(Industry default rate)의 경우 기업의 신용도에 영향을 미치는 지표로서 연체율이 높을 경우 자본 조달의 어려움 등으로 기술사업역량을 하락시킴으로써 대체로 기업의 기술혁신역량을 낮출 가능성이 존재한다.

R&D 스코어보드를 이용하여 영업이익률(OP margin)과 기술혁신역량간의 관계를 분석한 Cho et al.[7]은 기업규모를 상장대기업, 상장중소기업, KOSDAQ 벤처기업, KOSDAQ 일반기업으로 구분하여 각각의 집단에 대해서 연구개발집약도와 영업이익률의 상관관계를 분석하였는 바, 대체로 정의 상관관계를 나타내었으나 기업규모가 상대적으로 작은 KOSDAQ 일반기업의 경우에는 연구개발투자로 인한 수익성 악화로 영업이익률과 기술혁신역량의 투입측면인 R&D 집약도가 부의 상관관계를 나타낸다고 주장하였다.

기업의 물적규모인 자산총액(Assets)과 인적규모인 종업원 수(Number of employees)로 대표되는 기업규모와 기술혁신역량지표의 투입지표로서의 R&D 집약도와 산출지표로서의 지식재산권과의 관계를 분석한 Sung[8]에 따르면, R&D 집약도와 인적규모인 종업원수간에는 정의 관계를 나타낸 반면, 물적규모인 자산총액과는 부의 관계인 것으로 나타났다. 산

출지표로서 지식재산권과 기업규모와의 상관관계는 종업원수 및 자산총액 모두 지식재산권과 통계적으로 유의한 정의 상관관계를 보이는 것으로 분석되어 일반적으로 기업규모가 클수록 기술혁신역량이 뛰어난 것으로 주장하였다.

기업의 부설연구소 및 연구개발부서(R&D department)와 같은 기술개발인프라와 기술혁신역량간의 관계를 분석한 Lee et al.[9]에 따르면, 기업부설연구소 보유 유무는 기술사업화 역량과 통계적으로 유의미한 상관관계가 존재하지 않는 것으로 나타났는데, 이는 분석 대상 기업이 이미 연구소 및 연구개발전담부서와 같은 기술개발인프라를 보유한 경우가 대다수인 이노비즈인증 및 벤처확인 기업을 표본모집단 대상으로 하였기 때문인 것으로 판단된다. 즉, 표본 편의(sample bias)로 인해 상관관계 감소 효과가 발생했기 때문이다. 실제 기업부설연구소 및 연구개발전담부서 관련 인정제도를 관리하는 한국산업기술진흥협회(KOITA)에 따르면, 연구소 인정 제도의 취지는 “과학기술분야 또는 지식기반서비스분야의 연구개발 활동을 촉진하기 위하여 일정요건을 갖춘 기업부설연구소 및 연구개발전담부서를 신고하도록 하고 인정된 기업에 대해서는 조세, 자금, 인력 등의 지원혜택을 부여하여 기업의 기술혁신역량을 높이는 것이다”라고 규정하여 기업부설연구소 및 연구개발전담부서 보유가 기업의 기술역량을 높이는 것으로 간주하고 있다.

Park et al.[10]은 서울지역 194개 벤처기업에 대한 11년간의 패널데이터를 사용하여 기업의 경영성과와 학습역량간의 상관관계를 분석하였는데, 이를 위해서 기업부설연구소 보유 여부, 벤처확인(VENTURE), 이노비즈인증

(INNOBIZ) 여부를 기술기업의 학습역량(기술혁신역량)지표로 간주하였다.

Kim and Choi[11]는 국내 화학기업의 특허 자료를 이용하여 기업의 특허 등록 및 출원 등 특허 전략이 기업경영성과에 영향을 미치는 정도를 분석하였는데, 이는 기업의 기술혁신역량 즉 기술경쟁력을 특허등록건수(Number of patent registrations) 및 특허출원건수(Number of patent applications)로 간주한 것이다.

마지막으로 전문기술인력 비율의 경우, Zahra [12]의 연구에서 기업의 내부역량인 기술인력 비율(Technical staff ratio)이 높을수록 기술혁신역량이 높게 나타났으며, Kim and Lee[13] 역시 기술인력비율 및 석사 이상 전문기술인력의 비중 등 기업의 기술인력비중이 높을수록 기술혁신역량이 높은 것으로 분석하였다.

3. 실증분석

3.1 기초 통계

앞서 선정된 13개의 주요 지표에 대한 기초 통계 분석 결과는 <Table 3>과 같으며, 대부분의 지표가 정규분포를 나타내고 있음을 알 수 있다.

정규분포와 비교해서 왜도가 상대적으로 커서 왼쪽으로 집중된(right-skewed)분포를 갖는 지표는 자산(4.82), 메인비즈인증(4.46), 특허출원건수(4.65)이다. 자산의 경우, 전형적인 오른쪽 꼬리 분포를 갖는 바, 이는 개인 소득분포와 비슷하게 동 분석의 대상이 중소기업의 자산총액으로 극단적인 자산규모를 가질 수는 없기 때문인 것으로 판단된다.

〈Table 3〉 Basic statistics

Indicator	Mean	Min	Max	Std	S	K
Technology credit loan ratio(%)	39.9	14.3	47.7	0.03	-1.94	12.59
Facility loan ratio(%)	38.4	2.3	70.0	0.11	-0.14	0.16
Industry Delinquency rate(%)	1.6	0.0	10.8	1.88	2.01	4.83
OP margin(%)	4.7	-17.5	16.6	0.03	-1.89	19.67
Assets(million won)	91.6	13.2	937.6	104.0	4.82	29.36
RNDdept(%)	33.5	3.0	82.0	0.15	0.59	-0.01
MAINBIZ(%)	9.9	0.0	79.0	6.54	4.46	40.77
VENTURE(%)	26.6	1.2	73.7	13.95	0.85	0.35
INNOBIZ(%)	19.0	0.0	57.9	10.72	0.69	0.39
Number of patent applications	0.6	0.0	10.0	1.07	4.65	28.12
Number of patent registrations	2.8	0.0	24.8	2.76	2.78	14.03
Number of employees	28.5	4.7	126.1	16.24	2.45	8.97
Technical staff ratio(%)	39.7	0.0	93.3	0.15	2.36	17.03

※ Std = Standard Deviation, S = Skewness, K = Kurtosis.

메인비즈니스 비율의 경우 벤처확인이나 이노비즈니스와 달리 기업간 편차가 큰 기술력을 평가하기 보다는 시장, 사업성 위주의 평가로서 대부분 평균 좌측 근방에 집중 분포하기 때문인 것으로 보인다.

특허출원건수의 경우 평균 0.6건이며 첨도가 매우 높은 수준(28.12)으로서 대부분의 중소기업이 특허출원을 하지 않는 경향을 보이는 것으로 추정할 수 있다.

3.2 다중회귀(Linear Regression)

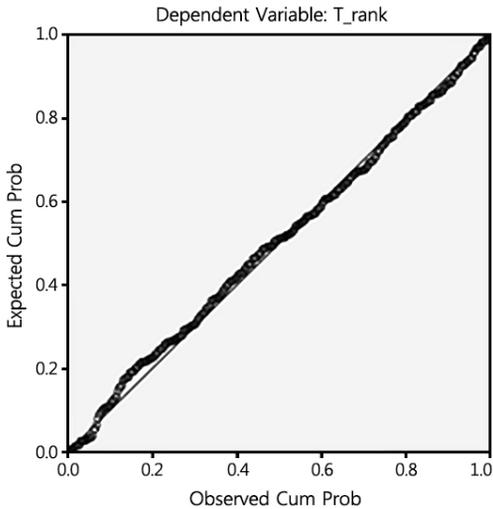
본 절 이후에서는 앞서 이론적으로 고찰된 13개 주요 지표에 대해서 실제 데이터 상으로 기술평가등급에 통계적으로 유의한 영향을 미치는 변수가 무엇인지 전통적인 다중회귀분석(Linear Regression; 이하 LR)을 이용해서 탐색하였으며, 반복 알고리즘을 이용한 기계학습의 한 종류인 부스팅 방법론을 적용하여 지표의

상대적 중요도(standardized coefficient, importance) 및 모형의 정확도(r square, pearson correlation coefficient)를 분석하였다.

기술평가등급을 종속변수로 주요 지표를 독립변수로 한 다중회귀모형에서 기술등급을 나타내는 기술평가등급점수는 상기 <Table 1>의 평가항목(Sub Item)에 대한 리커트 5점 척도 스코어링을 통해 항목별 가중치(weight)를 조정한 합산점수로 산출된다. 회귀모형을 구성하기에 앞서 먼저 데이터가 고전적인 회귀모형의 제가정을 만족하는지 검토하여 한다. 회귀모형 가정은 (1)오차항의 정규분포성인데 이는 잔차의 피피플롯(P-P plot)이 우상향 직선임을 통해서 확인하였다. (2)오차항의 독립성은 더빈-왓슨(Durbin-Watson)통계량 값이 1.89로서 2 근방에 위치하여 독립성 가정을 만족함을 확인하였다. 마지막으로 (3)오차항의 등분산성(homoscedasticity) 가정은 종속변수의 추정치와 표준화 잔차를 2차원 평면에 도식한 도표인

잔차플롯(residual plot)을 검토한 결과, 무작위 분포로 추정의 무작위성을 확인하였다.

3.2.1 정규분포성



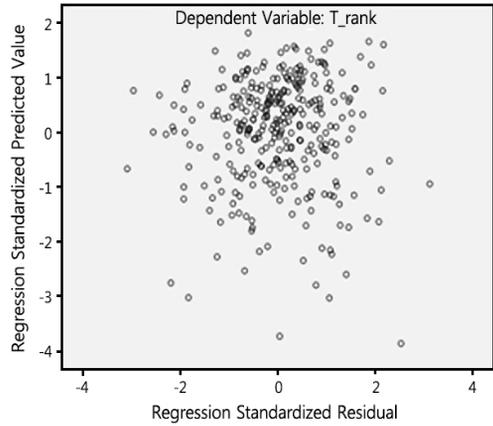
<Figure 3> Residuals Normal P-P Plot

3.2.2 오차항의 독립성

$$DW통계량 = \frac{\sum(e_i - e_{i-1})^2}{\sum(e_i^2)} = 1.891 \approx 2$$

3.2.3 오차의 등분산성

다중회귀분석 모형 추정 결과 및 계수 추정 결과는 아래와 같다. 설명변수가 모형을 예측하는 정확도인 R제곱값은 0.87로서 양호한 수준이며, 모든 회귀계수 추정치의 분산팽창요인(VIF)이 10보다 낮고, VIF의 역수인 공차한계



<Figure 4> XY Plot btw Estimates and Residuals

(Tolerance)가 0.1 이상으로 설명변수 간 상관관계(Multicollinearity)는 무시할만한 수준이며 계수 추정치는 안정적인 것으로 판단되었다.

회귀계수 중 기술평가등급에 유의적인 영향을 미치는 변수를 별 표시(**, P value = 5%)로 구분하였다. 그 결과 자산규모, 연구소 및 연구전담부서, 메인비즈인증, 벤처확인, 이노비즈인증, 특허출원건수 및 특허등록건수, 종업원수 비율이 유의적으로 분석되었다. 한편 본 논문은 사회과학 논문으로서 자연과학 혹은 공학계열 논문의 실험 정확성은 사실상 기대하기 어려워 상대적으로 관대한 유의성 판단기준인 10% 기준(*)을 적용한 경우도 표시하였는데 이 경우 메인비즈인증(0.09)과 기술인력비율(0.07)이 통계적으로 기술평가등급에 유의적인 영향을 미치는 것으로 나타났다.

<Table 4> Model Specification

Model	SSR	df	MSE	F	P	R ²	DW
Reg.	31.8	13.0	2.459	156.6	0.00	0.87	1.89
Resid.	4.7	302.0	0.016	-	-		
Total	36.6	315.0	-	-	-		

〈Table 5〉 Coefficient Estimation

Independent variable	B	Beta	t	P	VIF
(Constant)	5.869	-	58.43	0.00	-
Technology credit loan ratio	0.12	0.01	0.58	0.56	1.061
Facility loan ratio	0.01	0.02	0.22	0.82	1.154
Industry Delinquency rate	0.00	0.03	1.27	0.21	1.130
OP margin	-0.01	-0.00	-0.04	0.97	1.197
Assets	0.00	-0.07	-2.79	0.00**	1.448
RNDdept	-0.49	-0.214	-5.12	0.00**	4.096
MAINBIZ	0.00	0.037	1.67	0.09*	1.124
VENTURE	-0.01	-0.221	-5.89	0.00**	3.277
INNOBIZ	-0.01	-0.344	-7.96	0.00**	4.366
Number of patent applications	-0.02	-0.068	-2.50	0.01**	1.706
Number of patent registrations	-0.01	-0.102	-2.96	0.00**	2.799
Number of employees	-0.00	-0.187	-6.81	0.00**	1.767
Technical staff ratio	-0.11	-0.041	-1.81	0.07*	1.215

* < P value = 0.1, ** < P value = 0.05.

※ Please note that the technical rating grade, which is a dependent variable, is from T1 (highest grade) to T10 (lowest grade), and lower grade means better grade.

3.3 일반화가속모형(GBM)

다중회귀분석에서 유의적으로 분석된 지표 들은 모형의 명칭에서 알 수 있는 바와 같이 종속변수가 실수인 경우 적합되는 회귀(re-gression)기법으로 산출된 통계량을 기준으로 분석된 것이다. 본 논문의 관심은 유의적인 주 요 지표를 이용해서 기술평가등급점수를 최대 한 잘 분류하는 것이다. 따라서 이는 분류 (classification)문제로 볼 수 있으며 최적의 반 복 횟수를 통하여 학습된 알고리즘(손실함수 의 최소화)을 활용하여 우리가 추정하고자 하 는 최적화된 분류모형(분류기, classifier)의 파 라미터를 찾아내어 모형의 정확도를 최대화 하는 문제로 귀결된다.

이러한 학습방법이 바로 기계학습이며 본 논문에서는 다양한 분류기 중 회귀모형과 분 류모형의 특성을 동시에 갖고 있는 일반화가

속모형을 사용하여 기술평가등급에 영향을 미 치는 주요 요인을 분류할 것이다.

일반화가속모형(GBM)은 Freund and Schapire [14]가 개발한 적응적 가속화모형인 Adaptive Boosting(AdaBoost) 알고리즘 기반위에서 최 적화 문제에서 대표적으로 사용되는 경사 하 강법(傾斜下降法, Gradient descent)을 손실함 수(loss function)의 최소화 기법으로 적용하여 부스팅한 모형이다[15]. 분석을 위해서 R 3.1.3 gbm package를 사용하였다. 또한 모형의 세부 적인 파라미터 조정은 Ridgeway[16]를 참조하 였다.

(1) GBM의 목적은 예측오류 즉 손실함수(L) 를 최소화하는 종속변수 y 의 예측함수 \hat{y} 를 만 드는 것이다. 수식으로 표현하면, $\hat{y} = \operatorname{argmin}$

$$\sum_i^n L(y_i, c) \text{과 같다.}$$

손실함수 L은 초기상수 C에서 출발하여 경사 하강법에 따라 개별관측점 i 에서 negative gradient가 양(+)과 음(-)을 반복하여 0으로 수렴 할 때까지 지속적으로 \hat{y} 을 T번 업데이트하여 최적 예측함수를 만든다. 손실함수는 직관적으로 예측치와 실제 관측치의 차이이다. 즉, 잔차(residual)의 합인 손실(loss)를 최소화 하는 것으로 볼 수 있다.

가우시안 손실함수의 경우 편미분을 거치면 negative gradient가 예측치와 실제치의 차이인 잔차 형태를 갖는다는 것을 확인할 수 있다. GBM은 확률모형으로서 전체 데이터 n 개를 전부 사용하지 않고 n_1 개의 부표본(sub sample)을 추출하여 사용하기 때문에 알고리즘 상 필요한 파라미터는 $\rho (= \frac{n_1}{n}, \text{부표본 추출율})$, 독립변수의 상호(교호)작용을 포함시킬지 여부인 d (나무모형의 상호작용 깊이에 해당함), 반복업데이트 횟수인 $t = 1, 2, 3 \dots T$, 예측함수를 업데이트하는 속도인 학습률 λ (learning rate)로써 수식으로 표현하면 $\hat{y} \leftarrow \hat{y} + \lambda a_k$ 상의 람다 λ 이다. 학습률은 0.1%에서 1%까지 조정 가능하며 학습률이 작을수록 반복 업데이트 횟수인 T가 커져야 일정한 정확도를 유지할 수 있기 때문에 학습률을 작게 잡으면 모형의 계산량이 비례하여 증가하는 단점이 있다.

마지막으로 종속변수의 유형에 따라 목적함수인 손실함수 유형이 달라지는데 연속형 실수인 경우에는 앞에서 언급한 가우시안 함수,

$$\frac{1}{2} \sum w_i (y_i - f(x_i))^2, f(x_i) = E(y_i|x_i)$$

을 사용하고 이산형인 경우에는 아다부스트 함수,

$$\frac{1}{\sum w_i} \sum w_i \exp(-2y_i - 1)f(x_i)$$

(여기서, $f(x_i) = \log_e \frac{P_{y_i=1|x_i}}{P_{y_i=0|x_i}}$ 로서 로지스틱 회귀모형에서 사용하는 로짓함수(sigmoid)에 해당함)을 사용한다. 이외에도 베르누이(이항 사건), 포아송(희귀한 사건 발생), 다항분포(다항 사건) 함수 등 종속변수의 유형과 함수의 목적에 맞게 다양한 손실함수를 적용할 수 있는 것이 GBM의 장점이다.

(2) GBM은 내부 알고리즘으로 CART(Classification And Regression Tree) 나무모형을 T번 반복하여 사용한다. CART는 기본적으로 분류(classification)와 회귀(regression)의 성격을 모두 갖는 나무모형이다. 이때 각 개별 독립변수의 중요도는 노드 불순도(node impurity) 감소에 대한 기여도로서 측정되는데, 독립변수 X_j 의 상대적인 중요도는 다음으로 정의된다.

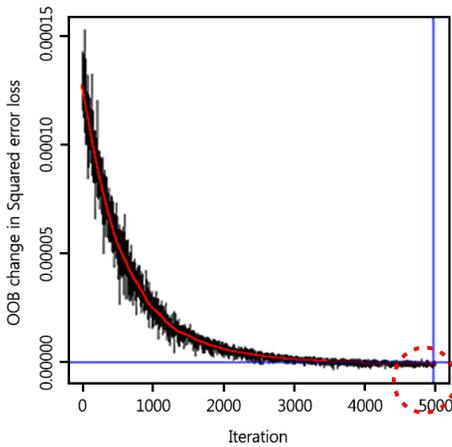
$$\text{Imp}_j^*(M) = \frac{\text{Imp}_j(M)}{\sum_j \text{Imp}_j(M)} \times 100$$

위 수식에서, $\text{Imp}_j(M)$ 은 나무모형 M_t 에서 T개의 반복나무 모형 업데이트 중 독립변수 X_j 의 분할에 의해서 감소되는 불순도의 총합이다. 본 연구에서 독립변수의 중요도는 이러한 방식으로 종속변수인 기술평가등급에 영향을 미치는 주요 지표의 상대적 중요도를 측정하였다. 이상을 종합하여 본 분석에서 사전에 결정한 모형 파라미터 집합은 다음과 같다.

$$\{\rho = 0.5, d = 2, T = 5,000, \lambda = 0.1\%, f(x_i) = \text{가우시안}\}$$

분석결과, gbm package는 해당 손실함수를 최소화하는 알고리즘을 통해 최적 반복 횟수인 나무수 T를 알려주는데 T = 4,966회로 나타나 사전에 결정한 파라미터 T = 5,000회와 거의 유사 하였다. 이는 학습률을 상대적으로 작게 잡았기 때문에(0.1%) 정확한 분류를 위해서 반복 횟수가 늘어난 상황으로 적절한 반복 횟수로 판단된다. 기술평가등급에 영향을 미치는 상대

적인 지표의 중요도는 이노비즈인증(35.3%), 연구소 및 연구개발전담부서 보유(31.6%), 특허등록건수(11.5%), 벤처확인(10.6%), 종업원수(5.3%), 자산(2.2%), 특허출원건수(1.5%) 순으로 분석되었다. 예측치와 실제치의 상관관계를 고려한 모형의 정확도는 81.5%로서 비교적 우수한 수준인 것으로 판단된다. 아래는 이러한 결과에 대해 정리한 그래프와 도표이다.



<Figure 5> Optimal Number of Iteration(T = 4,966)

<Table 6> Variable Relative Importance

Indicator	%
INNOBIZ	35.35
RNDdept	31.56
Number of patent registrations	11.46
VENTURE	10.62
Number of employees	5.30
Assets	2.21
Number of patent applications	1.49
OP margin	0.41
Industry Delinquency rate	0.37
MAINBIZ	0.36
Technical staff ratio	0.31
Technology credit loan ratio	0.27
Facility loan ratio	0.22
Importance normalized to sum to 100%	

<Table 7> Comparison btw GBM and LR

	GBM	%	LR	%
1	INNOBIZ	35.35	INNOBIZ	26.80
2	RNDdept	31.56	VENTURE	17.19
3	Number of patent registrations	11.46	RNDdept	16.69
4	VENTURE	10.62	Number of employees	14.58
5	Number of employees	5.30	Number of patent registrations	7.98
6	Assets	2.21	Assets	5.42
7	Number of patent applications	1.49	Number of patent applications	5.26
8	OP margin	0.41	Technical staff ratio	3.22
9	Industry Delinquency rate	0.37	MAINBIZ	2.86
10	MAINBIZ	0.36	-	-
11	Technical staff ratio	0.31	-	-
12	Technology credit loan ratio	0.27	-	-
13	Facility loan ratio	0.22	-	-

제 3.2절의 다중회귀(LR) 분석 결과와 GBM의 결과를 비교해보면 대체로 비슷한 결과를 보여주나 세부적으로 차이를 보인다. 그에 앞서 GBM의 경우에는 모형자체에서 독립변수의 상대적 중요도를 산출해주나 다중회귀 모형은 그렇지 못하여 본 분석에서는 전체 표준화 회귀계수 대비 해당 지표의 표준화 회귀계수의 비(ratio)를 주요 지표의 상대적인 중요도로 간주하고 분석하였다.

GBM과 LR 모두 기술평가등급에 가장 큰 영향을 미치는 지표는 이노비즈인증 여부인 것으로 나타났으나, 그 영향력은 상대적으로 GBM이 더 컸다.

연구소 및 연구개발전담부서 보유 여부의 경우 LR이 16.69% 수준인 반면 GBM은 31.56%로서 약 2 배 이상 더 큰 영향을 미치는 것으로 나타났다. 이하 특허등록건수, 벤처확인, 종업원수, 자산, 특허출원건수 등은 두 모형에서 모두 주요한 지표로 분석되었는데, 대체로 GBM의 경우 상위 4개의 지표가 약 87%의 중요도를 보여 상위 순위에 집중되는 경향을 보였으나 LR의 경우 상대적으로 주요 지표가 고르게 분포되었다.

LR의 경우 통계적으로 유의하지 않은 지표는 -로 표시하였는데 GBM의 경우에도 영향력이 매우 작아 비슷한 경향을 보인 것으로 분석되었다. 한편, 앞의 <Table 1> TCB 모형의 중항목(Item)을 기준으로 기술등급에 영향을 미치는 항목의 상대적 중요도를 신경망 모형(Neural Network Model; NN)을 적용하여 분석한 Lee and Yun[17]에 따르면, TCB 모형의 8개 중항목 중에서 <Table 1>의 1.3 Ability to develop technology(기술개발역량)의 상대적 중요도가 가장 높게 나타나 지표 수준에서 분석한 본 논문의 결과와 일치함을 알 수 있었다.

4. 결론 및 제언

이상으로 TDB TPS의 주요 지표를 대상으로 기술등급에 영향을 미칠 것으로 예상되는 주요 지표의 상대적 중요도 및 모형의 적합도를 전통적인 회귀모형과 함께 기계학습의 일종인 GBM을 사용하여 비교분석하였다.

분석결과 두 모형간 주요 지표의 기술평가등급에 대한 상대적인 중요도에 있어서 대체적으로 크게 차이나지 않는 것으로 분석되었으나, GBM 모형의 경우 회귀모형에 비해서 이노비즈인증 여부, 연구소 및 연구개발전담부서 보유 유무, 특허등록건수, 벤처확인 변수가 회귀모형보다 상대적으로 큰 영향을 미치는 것으로 분석되었다.

회귀모형이 본질적으로 독립변수 단위당 종속변수의 변화량에 관심이 있는 반면에, 분류모형(나무모형)의 특성도 함께 갖는 GBM은 보다 정확한 분류에 초점을 맞추는 성질이 있다. 따라서 향후 기술등급의 분류에 있어서 상기 주요 변수들에 대한 세분화 및 정교화 작업이 필요할 것으로 생각한다. 대표적으로 특허건수와 같은 단순한 정량 변수 이외에 특허의 대표적인 질적 특성인 법적 권리성 및 권리 안전성에 대한 평가지표 도입 등 가능한 범위 내에서 고려해 볼 수 있을 것이다[18]. 이러한 시도는 기술평가의 변별력을 보다 향상시킬 수 있을 것으로 기대된다.

본 모형에 있어서 분류분석 기법으로 일반화속모형을 적용하였는데, 실제 이항 및 다항 분류 문제에 있어서 가장 많이 쓰이고 있는 분류기법은 서포트벡터머신, 랜덤포레스트, 인공신경망 등의 완전한 분류 기법이다. 서포트벡터머신은 손실함수의 쌍대개념으로써 이득

함수인 마진(margine)을 제약조건하에서 최대화하는 최적화 기법이다. 각 개별 관측개체의 거리를 최대한 이격시키는 벡터를 찾는다는 점에서 집단을 나눈다는 분류(分類)문제의 가장 충실한 방법론으로 평가된다. 따라서 향후 보다 정교한 분류 모형으로 평가되는 서포트 벡터머신 등의 완전한 분류 기법을 적용하여 보다 정확한 분류 로직을 적용해볼 필요가 있을 것이다.

실제 기술평가의 주요 관심사항은 부도율(Probability of Default; PD)과 기술평가등급의 관계이다. 현재까지 TCB 평가가 본격적으로 시행된 지 3년 미만으로 부도관측기간이 충분치 못하여 기술평가등급과 부도율 간의 분석이 어려우나 추후 부도 데이터가 일정 수준 확보되면 부도율과 기술평가등급의 상관(회귀)관계에 대한 분석 및 등급 캘리브레이션의 부도율 변별력 검증이 반드시 필요할 것으로 생각된다. 부도 혹은 정상의 문제는 전형적인 이진분류의 문제로서 다양한 기계학습이 적용될 수 있는 분야이기도 하다.

마지막으로 현재 TCB 평가의 기술평가부분은 기업신용평가항목 중 비재무항목과 상당부분 겹치는 부분이 존재한다. 향후 이러한 부분에 대한 간결화 작업이 필요할 것으로 판단되며 비재무항목으로서의 기술평가 지표에 대한 탐색으로 본 연구가 의미 있을 것으로 사료된다.

References

- [1] Hong, J. B., "Development of technology evaluation model for investment," Pukyong National University, p. 2, 2015.
- [2] Cha, W. J., "Government's Role for Effective SME Financing, Legislation and Policy," National Assembly Legislation Bureau, pp. 325-326, 2016.
- [3] Lee, J. M., "Current Status of Technology Finance and Improvements for Effective Settlement," National Future Research Institute, pp. 4-5, 2015.
- [4] Jeon, W. J. and Seo, Y. W., "Analysis of TCB Grade and PD employing TCB Big Data," KIEC, pp. 403-420, 2016.
- [5] Financial Services Commission, "From "quantitative expansion" to "qualitative improvement," "systematization of technical financial system and improvement of system" is promoted," 2015.
- [6] Lee, Y. B., "An Analysis of the Relative Efficiency and the Total Factor Productivity Changes of SMEs in SME Funding Program," Korean Journal of Public Administration, Vol. 46, No. 1, pp. 199-229, 2006.
- [7] Cho, S. P., Lee, Y. H., Park, S. Y., and Bae, J. H., "Analysis of R&D Scoreboard for R&D Investment and Performance," Journal of Technology Innovation, Vol. 10, No. 1, pp. 98-123, 2002.
- [8] Sung, T. K., "A Firm Size-Innovative Activity Relationship: An Empirical Study of the Korean Manufacturing Industry," The Korean Small Business Review, Vol. 25, No. 2, pp. 305-325, 2003.

[1] Hong, J. B., "Development of technology

- [9] Lee, J. M., Noh, M. S., and Chung, S. Y., "A Study on the Effects of SME's Technology Planning Competency on the Success of Commercialization," *Journal of Technology Innovation*, Vol. 21, No. 1, pp. 253-278, 2013.
- [10] Park, S. G., Moon, H. C., and Cha, K. C., "The Effects of Venture Firm's Phased Internationalization and Learning Capability on Its Business Performance," *The Korean Small Business Review*, Vol. 35, No. 2, pp. 129-157, 2013.
- [11] Kim, S. W. and Choi, Y. H., "The Relation Between Patenting Behavior and Company Performance at the level of the Chemical Firm," *KTIS*, pp. 389-402, 2003.
- [12] Zahra, S. A., "Technology strategy and new venture performance: a study of corporate-sponsored and independent biotechnology ventures," *Journal of Business Venturing*, Vol. 11, No. 4, pp. 289-321, 1996.
- [13] Kim, Y. B. and Lee, B. H., "Patterns of Technological Learning among the Strategic groups in the Korean Electronic Parts Industry," *Research Policy*, Vol. 31, No. 4, pp. 543-567, 2002.
- [14] Freund, Y. and Schapire, R. E., "A decision-theoretic generalization of online learning and an application to boosting," *Journal of Computer and System Sciences*, Vol. 55, No. 1, pp. 119-139, 1997.
- [15] Friedman, J. H., "Greedy function approximation: a gradient boosting machine," *Annals of Statistics*, Vol. 29, No. 5, pp. 1189-1232, 2001.
- [16] Ridgeway, G., "Generalized Boosted Models: A guide to the gbm package," Update 1.1, 2007.
- [17] Lee, J. W. and Yun, J. Y., "A Study on Suitability of Technology Appraisal Model in Technology Financing," *Journal of Korea Technology Innovation Society* Vol. 20, No. 2, pp. 305-308, 2017.
- [18] Park, C. G., Roh, H. S., Choi, Y. G., Kim, H. W., and Lee, J., K., "A Study on the Application Methods of Big Data in the Technology Commercialization Process," *The Journal of Society for e-Business Studies*, Vol. 19, No. 4, pp. 73-99, 2014.

저 자 소 개



전우정 (E-mail: jeonmichael201809@gmail.com)
2000년 서강대학교 경제학과 (학사)
2002년 서강대학교 경제학과 (석사)
2015년 고려대학교 수학과 (학사, 학점은행제)
2016년~현재 대전대학교 일반대학원 융합컨설팅학과 기술경영(MOT) ABD
2016년~현재 NICE D&B TCB 평가실 책임 전문위원
관심분야 (기술가치(등급)/신용)평가, 기술금융, 행동경제학, Entrepreneurship, Mathematical Modeling



서영욱 (E-Mail: ywseo@dju.kr)
2009년 성균관대학교 경영학과 (박사)
2015년~현재 대전대학교 일반대학원 융합컨설팅학과 교수
관심분야 컨설팅 창의성과학, 의사결정, 복잡계시스템, 정보기술 성과 평가