

적응형 부스팅을 이용한 파산 예측 모형: 건설업을 중심으로*

허준영

한성대학교 컴퓨터공학과
(jyheo@hansung.ac.kr)

양진용

한성대학교 컴퓨터공학과
(jyang0112@gmail.com)

2013년 건설 경기 전망 보고서에 따르면 주택건설경기 침체 상황의 지속으로 건설 기업의 유동성 위기가 지속될 것으로 전망된다. 건설업은 파산으로 인한 사회적 파급효과가 다른 산업에 비해 큰 편이지만, 업종의 특성상 다른 산업과는 상이한 자본구조와 부채비율, 현금흐름을 가지고 있어서 기업의 파산 예측이 더 어려운 측면이 있다. 건설업은 레버리지가 큰 산업으로 부채비율이 매우 높은 업종이며 현금흐름이 프로젝트 후반부에 집중되는 특성이 있다. 그리고 경기사이클에 따른 부침이 매우 심하여 경기하강국면에선 파산이 급증하는 양상을 보인다. 건설업이 레버리지 산업인 이상 건설업체의 파산을 증가는 여신을 공여한 은행에 큰 부담으로 작용한다. 그럼에도 그간의 파산예측모델이 주로 금융기관에 집중되어 왔고 건설업종에 특화된 연구는 드물었다.

기업의 재무 자료를 바탕으로 한 파산 예측 모델에 대한 연구는 오래 전부터 다양하게 진행되었다. 하지만, 일반적인 기업 전체를 대상으로 하는 모델이기 때문에, 건설 기업과 같이 유동성이 큰 기업의 예측에는 적절하지 못할 수 있다. 건설 산업은 오랜 사업 기간과 대규모 투자, 그리고 투자금 회수가 오래 걸리는 특징을 갖는 자본 집약 산업이다. 이로 인해 다른 산업과는 상이한 자본 구조를 갖기 마련이고, 다른 산업의 기업 재무 위험도를 판단하는 기준과 동일한 적용이 곤란할 수 있다.

최근에는 기계 학습을 바탕으로 한 기업 파산 예측 연구가 활발하다. 기계 학습의 대표적 응용 분야인 패턴 인식을 기업의 파산 예측에 응용한 것이다. 기업의 재무 정보를 바탕으로 패턴을 작성하고 이 패턴이 파산 위험 군에 속하는 지 안전한 군에 속하는지 판단하는 것이다.

전통적인 Z-Score와 기계 학습을 이용한 파산 예측과 같은 기존 연구들은 특정 산업 분야가 아닌 일반적인 기업을 대상으로 하기 때문에 기업들의 특성을 전혀 고려하고 있지 못하다. 본 논문에서는 건설 기업을 규모에 따라 각 기업들의 예측 능력을 비교하여 적응형 부스팅이 가장 우수함을 확인하였다. 본 논문은 건설 기업을 자본금 규모에 따라 세 등급으로 분류하고 각각에 대해 적응형 부스팅의 예측력을 분석하였다. 실험 결과 적응형 부스팅이 다른 기법에 비해 예측 결과가 좋았고, 특히 자본금 규모가 500억 이상인 기업의 경우 아주 우수한 결과를 보였다.

논문접수일 : 2013년 9월 2일 논문수정일 : 2013년 12월 11일 게재확정일 : 2014년 1월 8일

투고유형 : 국문일반 교신저자 : 허준영

1. 서론

근래 들어 부동산 경기침체 여파로 건설기업

이 심각한 위기를 겪으며 파산이 급증하였다. 2008년 이후 건설공사의 이윤율이 지속적으로 하락함에 따라 건설기업의 유동성에 부정적인

* 본 연구는 한성대학교 교내학술연구비 지원과제 임.

영향을 주었으며, 2013년 건설 경기 전망 보고서에 따르면 주택건설경기 침체 상황의 지속으로 건설 기업의 유동성 위기가 지속될 것으로 전망된다(CERIK, 2013). 건설업은 파산으로 인한 사회적 파급효과가 다른 산업에 비해 큰 편이지만, 업종의 특성상 다른 산업과는 상이한 자본구조와 부채비율, 현금흐름을 가지고 있어서 기업의 파산 예측이 더 어려운 측면이 있다.

우리나라 전체 GDP에서 건설업이 차지하는 비중은 OECD 국가평균 5.1%에 비해 5.9%로 높은 편이다. 건설 투자는 대규모 자금이 투입되어 경제성장률을 끌어올리는 효과가 있는데 미국의 경우 최근 건설투자의 증가에 힘입어 뚜렷한 경기회복세를 보이고 있는 반면, 우리나라의 경우 부동산침체 여파로 건설투자가 큰 폭으로 감소하고 있어 경기회복의 발목을 잡고 있다. 건설업은 레버리지가 큰 산업으로 부채비율이 매우 높은 업종이며 현금흐름이 프로젝트 후반부에 집중되는 특성이 있다. 그리고 경기사이클에 따른 부침이 매우 심하여 경기하강국면에선 파산이 급증하는 양상을 보인다. 건설업이 레버리지 산업인 이상 건설업체의 파산을 증가는 여신을 공여한 은행에 큰 부담으로 작용한다. 실제 부동산 프로젝트 파이낸싱에 집중했던 저축은행들이 부동산경기침체로 인한 건설업체의 파산으로 동반 몰락한 것이 최근의 일이다. 저축은행뿐만 아니라 시중은행도 상당한 충당금을 쌓아야 하는 등 그 여파를 피해갈 수 없었다. 아울러 복잡한 하도급 관계로 얽혀 있어 건설업체의 파산은 다수회사의 연쇄 파산으로 이어지기 쉽다. 또한 건설업은 후방산업이라 할 수 있는 시멘트, 철강 등의 업종과 산업연관효과가 커서 건설업의 침체는 산업전반에 영향을 미친다. 한편 건설업은 고용유발효과가 매우 큰 산업으로 건설업체의 파

산은 일반서민경제에 미치는 영향이 매우 크다. 그럼에도 그간의 파산예측모델이 주로 금융기관에 집중되어 왔고 건설업종에 특화된 연구는 드물었다.

기업의 재무 자료를 바탕으로 한 파산 예측 모델에 대한 연구는 오래 전부터 다양하게 진행되었다. 하지만, 일반적인 기업 전체를 대상으로 하는 모델이기 때문에, 건설 기업과 같이 유동성이 큰 기업의 예측에는 적절하지 못할 수 있다. 건설 산업은 오랜 사업 기간과 대규모 투자, 그리고 투자금 회수가 오래 걸리는 특징을 갖는 자본 집약 산업이다. 이로 인해 다른 산업과는 상이한 자본 구조를 갖기 마련이고, 다른 산업의 기업 재무 위험도를 판단하는 기준과 동일한 적용이 곤란할 수 있다(Sun et al., 2013).

전통적인 기업 파산 예측 연구로는 알트만의 Z-score가 대표적이다(Altman, 2000). Z-score는 1968년 처음 발표되었으며, 간단한 식을 통해 기업의 파산 가능성을 예측한다. Z-score는 계산된 값을 세 구간으로 나누어 위험, 중간, 안전으로 판단한다. 위험 구간에 속하면 향후 2년 내에 파산 가능성이 높으며 안전 구간에 속하면 파산 가능성이 작다고 보는 것이다. 하지만 중간 구간에 속하는 경우 판단을 하지 못하며, 본 논문에서 사용한 건설 기업의 경우 많은 수가 이 구간에 속하는 단점이 있다.

컴퓨터를 이용한 기계 학습이 발전함에 따라 최근에는 기계 학습을 바탕으로 한 기업 파산 예측 연구가 활발하다. 기계 학습의 대표적 응용 분야인 패턴 인식을 기업의 파산 예측에 응용한 것이다. 기업의 재무 정보를 바탕으로 패턴을 작성하고 이 패턴이 파산 위험 군에 속하는지 안전한 군에 속하는지 판단하는 것이다.

파산 예측에 사용되는 대표적인 기계 학습으

로는 인공 신경망(Artificial Neural Networks) (Tae and Shin 2010; Wilson and Sharda, 1994) 과 적응형 부스팅(AdaBoost)(Sun et al., 2013; Altman, 2000)이 있다. 이 외에도 SVM(Support Vector Machine)(Min and Lee, 2005; Shin et al., 2005)을 사용한 연구도 있다. 이 기법들을 결합한 하이브리드 연구도 다양하다(Kim, 2009; Shin and Hong, 2011; Verikas et al., 2010).

전통적인 Z-Score와 기계 학습을 이용한 파산 예측과 같은 기존 연구들은 특정 산업 분야가 아닌 일반적인 기업을 대상으로 하기 때문에 기업들의 특성을 전혀 고려하고 있지 못하다. 본 논문에서는 건설 기업을 규모에 따라 각 기법들의 예측 능력을 비교하여 적응형 부스팅이 가장 우수함을 확인하였다.

본 논문은 건설 기업을 자본금 규모에 따라 세 등급으로 분류하고 각각에 대해 적응형 부스팅의 예측력을 분석하였다. 적응형 부스팅의 상대적 성능 측정을 위해 다른 기계 학습 방법인 인공 신경망, SVM(Support Vector Machine), 결정 트리와 비교를 수행하였다. 실제로 2008년부터 2012년까지 파산한 기업과 2012년 정상 운영중인 기업의 재무 자료를 사용하여 각 기법들의 예측력을 측정하였다. 실험 결과 적응형 부스팅이 다른 기법에 비해 예측 결과가 좋았고, 특히 자본금 규모가 500억 이상인 기업의 경우 아주 우수한 결과를 보였다.

본 논문은 다음과 같이 구성되었다. 2장에서는 파산 예측 문제에 대한 설명을 하고, 3장에서는 연구에 사용한 건설 기업 데이터를 설명한다. 4장에서는 본 논문의 주요 기계 학습 방법인 적응형 부스팅을 설명하고 5장에서 실험 결과와 6장에서 결론을 맺는다.

2. 문제 기술

기업 파산 예측은 경영에서 중요한 문제이며 이 문제의 목표는 건설한 기업과 향후 파산 가능성이 높은 기업을 가려내는 것이다. 즉, 기업의 위기 시점을 예측하는 모델을 세우고, 이 모델을 통해 적절한 의사 결정을 내리는 것이다. 파산 예측을 위해서 기업의 현재 재무 상황을 아는 것이 필수적이다. 이 정보는 기업의 규모에 따라 값의 차이가 매우 크기 때문에 보통 파산 예측 모델에서는 비율을 사용한다(Alfaro et al., 2008). 예를 들어 유동 자산(current assets)의 경우 총자산(total assets)으로 나눈 비율을 사용한다.

기업의 파산 예측 중요성에 대해서 누구나 동의하지만 기업의 파산 시점을 어떻게 정의하는지는 이견이 있을 수 있다. 본 논문에서는 워크아웃, 법정관리, 파산의 경우를 모두 파산 시점으로 정의하였다. 기업이 재무적 곤경에 처하면 주주, 채권자, 피고용인 등 여러 이해당사자들이 큰 고통에 직면한다. 위 세 가지 형태의 재무적 곤경 모두 경제적 이해당사자들에게 큰 손실을 주게 되므로 이를 예측을 위한 시점으로 정의하는데 무리가 없을 것으로 판단된다.

기업의 파산 예측 모델에 사용할 재무 정보는 다양한 변수들이 사용될 수 있다. 예측이란 것은 현재의 자료를 토대로 미래의 상태를 판단하는 것으로, 본 논문의 예측 문제에서는 현재 재무 정보를 토대로 1년 후 기업의 파산 여부를 예측하도록 한다. 모델을 설정하고 모델의 예측력을 확인하기 위해서 파산 기업이 파산 시점에서 1년 전 재무 정보와 정상 기업의 1년전 재무 정보를 사용하였다.

3. 데이터 설명

기업 파산 예측 모델에 적용할 기업의 재무 정보는 최근 5년 간 기업들의 재무제표를 참고로 하였다. 대한민국 모든 건설 기업에 대해 다음과 같은 재무 정보를 수집하였다. 이 데이터 (주)나이스디앤비(NICE, 2013)에서 보유하고 있는 자료에서 원하는 재무 정보만 추출한 것이다. (주)나이스디앤비는 국내 최대 기업신용정보 제공업체로 기업 신용 평가와 정보 서비스를 제공하고 있다.

기업을 파산 기업과 정상 기업으로 나누고 파산 기업은 2008년부터 2012년 사이에 워크아웃, 법정관리, 파산한 기업으로 정의하고, 정상 기업은 2012년 12월 기준으로 파산하지 않은 기업으로 정의하였다. 결과적으로 파산 기업은 총 1381개, 정상 기업은 총 28481개를 선정하였다. 그리고 파산 기업과 정상 기업의 재무 정보를 수집하였다. 이 때 파산 기업 재무 정보는 파산 기업의 파산 시점에서 직전 년도의 재무 정보를 사용하였고, 정상 기업의 재무 정보는 정상 기업의 2011년도 재무 정보를 사용하였다. 즉, 5년 간 파산한 기업의 파산 직전 1년 전 재무 정보와 정상 기업의 1년 전 재무 정보를 사용하여 1년 후 파산 유무를 예측하는 모델에 사용하였다.

수집한 기업의 재무 정보를 이용하여 <Table 1>과 같은 변수를 생성하여 모델에 적용하였다. 기업의 재무 정보에서 매출과 이익, 유동성, 채무, 자산을 설명할 수 있는 변수를 기존 유사연구에서 사용한 변수를 바탕으로 선정하였다 (Wilson and Sharda, 1994; Altman, 2000; Min and Lee, 2005; Shin et al., 2005; Verikas et al., 2010; Sun et al., 2013). 또한 변수들 간에 상관계수가 낮은 변수를 선정함으로써, 서로 독립 가능성이 높은 변수를 채택하였다. 이는 상관계

수가 높은 두 변수 중 하나를 제거해도 파산 예측 모델에 영향을 주지 않고 모델을 더 단순화할 수 있기 때문이다. 각 변수는 기업의 규모나 매출과 무관하게 적용할 수 있도록 비율로 계산하였다. 예를 들어 유동 자산(Current assets)과 관련한 변수인 CA/TA와 CA/CL의 경우 각각 총자산(Total assets)과 유동부채(Current liabilities)로 나눈 값을 사용한다.

<Table 1> Model Variables

Variables	Description
<i>EBIT/TA</i>	Earnings before interest and taxes /Total assets
<i>EBT/CAP</i>	Earnings before taxes/Capital
<i>WC/TA</i>	Working capital/Total assets
<i>WCS</i>	Working capital/Sales
<i>CATA</i>	Current assets/Total assets
<i>CA/CL</i>	Current assets/Current liabilities
<i>C/TA</i>	Cash/Total assets
<i>C/CL</i>	Cash/Current liabilities
<i>lnTA</i>	Natural logarithm value of total assets
<i>S/CAP</i>	Sales/Capital
<i>S/CA</i>	Sales/Current assets
<i>S/TA</i>	Sales/Total assets

<Table 2>와 <Table 3>에 전체 건설 기업의 모델 변수들에 대한 기술적 분석과 상관관계 행렬을 표시하였다. 앞서 말했듯이 <Table 3>을 통해 변수들 간에 상관계수가 낮음을 확인할 수 있다. <Table 2>에서 WC/S나 C/CL, WC/TA, EBT/CAP, S/CAP과 같은 변수들의 편차가 평균의 10배가 넘는 매우 큰 편차를 보이고 있다. 이는 건설 기업들의 재무 구조가 서로 매우 상이함을 나타낸다고 볼 수 있고, 결국 Z-score와 같은 통계적 방법의 예측력을 낮게 만든다.

〈Table 2〉 Descriptive analysis of construction companies' model variables

	<i>WC/S</i>	<i>C/CL</i>	<i>S/CA</i>	<i>EBIT/TA</i>	<i>CA/TA</i>	<i>WC/TA</i>
Mean	-13.06	0.58	3.45	-0.14	0.52	-0.29
Median	-0.02	0.03	2.34	0.03	0.52	-0.02
Std. Dev	298.03	5.45	6.01	0.96	0.27	6.65
	<i>S/TA</i>	<i>EBT/CAP</i>	<i>S/CAP</i>	<i>lnTA</i>	<i>C/TA</i>	<i>CA/CL</i>
Mean	1.40	-1.67	116.41	9.46	0.06	2.78
Median	1.02	0.08	9.85	9.44	0.02	0.96
Std. Dev	1.62	75.77	4671.31	1.67	0.10	16.81

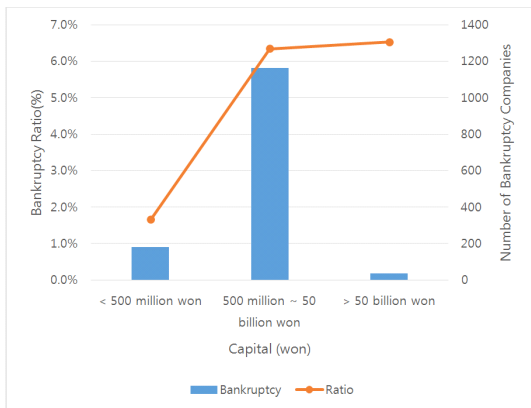
〈Table 3〉 Correlation of construction companies' model variables

	<i>WC/S</i>	<i>C/CL</i>	<i>S/CA</i>	<i>EBIT/TA</i>	<i>CA/TA</i>	<i>WC/TA</i>
<i>WC/S</i>	1.00	0.01	0.02	0.08	0.05	0.02
<i>C/CL</i>	0.01	1.00	-0.03	0.02	0.06	0.01
<i>S/CA</i>	0.02	-0.03	1.00	0.14	-0.24	-0.58
<i>EBIT/TA</i>	0.08	0.02	0.14	1.00	0.06	-0.51
<i>CATA</i>	0.05	0.06	-0.24	0.06	1.00	0.07
<i>WC/TA</i>	0.02	0.01	-0.58	-0.51	0.07	1.00
<i>S/TA</i>	0.04	-0.02	0.52	0.05	0.30	-0.32
<i>EBT/CAP</i>	0.26	0.00	0.00	0.03	0.03	0.01
<i>S/CAP</i>	0.00	0.00	-0.01	0.00	-0.01	0.00
<i>lnTA</i>	-0.03	-0.11	-0.09	-0.01	-0.18	0.01
<i>C/TA</i>	0.04	0.32	-0.03	0.08	0.32	0.01
<i>CA/CL</i>	0.01	0.68	-0.04	0.03	0.10	0.02
	<i>S/TA</i>	<i>EBT/CAP</i>	<i>S/CAP</i>	<i>lnTA</i>	<i>C/TA</i>	<i>CA/CL</i>
<i>WC/S</i>	0.04	0.26	0.00	-0.03	0.04	0.01
<i>C/CL</i>	-0.02	0.00	0.00	-0.11	0.32	0.68
<i>S/CA</i>	0.52	0.00	-0.01	-0.09	-0.03	-0.04
<i>EBIT/TA</i>	0.05	0.03	0.00	-0.01	0.08	0.03
<i>CATA</i>	0.30	0.03	-0.01	-0.18	0.32	0.10
<i>WC/TA</i>	-0.32	0.01	0.00	0.01	0.01	0.02
<i>S/TA</i>	1.00	0.01	-0.01	-0.29	0.19	-0.01
<i>EBT/CAP</i>	0.01	1.00	0.82	-0.01	0.01	0.00
<i>S/CAP</i>	-0.01	0.82	1.00	0.05	-0.01	0.00
<i>lnTA</i>	-0.29	-0.01	0.05	1.00	-0.27	-0.11
<i>C/TA</i>	0.19	0.01	-0.01	-0.27	1.00	0.14
<i>CA/CL</i>	-0.01	0.00	0.00	-0.11	0.14	1.00

규모별 분류를 위해 건설 기업의 자본금에 따라 소형, 중형 대형으로 분류하였다. 소형 건설 기업은 자본금 5억 미만, 중형 건설 기업은 자본금 5억 이상 500억 미만, 대형 건설 기업은 자본금 500억 이상인 기업으로 선정하였다. <Table 4>와 <Figure 1>은 기업 규모에 따른 파산 회사의 비율을 보여준다.

<Table 4> Bankruptcy ratio

Capital	Bankruptcy	Normal	Ratio
< 500 million won	181	10762	1.7%
500 million ~ 50 billion won	1164	17203	6.3%
> 50 billion won	36	516	6.5%
Total	1381	28481	4.6%



<Figure 1> Ratio and number of bankruptcy companies

<Table 4>와 <Figure 1>에서 알 수 있듯이 대형 규모 건설 기업의 경우 전체 숫자는 많지 않으나, 파산 비율은 상대적으로 높은 것으로 확인되었다. 또한 소규모 건설 기업의 경우 반대로 기업의 수는 많으나 파산 비율은 매우 적은 것을

알 수 있다. 대형 건설 기업의 경우 파산 비율도 높고 파산으로 인한 그 피해도 훨씬 크기 때문에 대형 건설 기업의 파산 예측이 특히 중요함을 알 수 있다.

4. 적응형 부스팅

적응형 부스팅은 기계 학습 알고리즘의 하나로 Freund와 Schapire가 수식화 한 것이다(Freund and Schapire, 1995). 다른 학습 알고리즘과 결합하여 사용하는 형태로, 보다 개선된 학습 알고리즘을 만들 수 있다. 적응형 부스팅은 약 분류기(weak classifier)를 결합하여 보다 강한 분류기 학습 알고리즘을 만든다. 결합 방식은 가중 평균 방식이고, 가중치 값을 정하는 알고리즘과 분류 방법을 본 논문에서 사용한 경우에 적용하여 설명하면 다음과 같다.

약 분류기 알고리즘으로는 결정 트리를 사용하였고, 깊이(depth)는 1로 하였다. 즉 12개의 결정 트리 알고리즘은 각각 변수에 대해 분류기 학습을 수행한다. 각각의 결정 트리 알고리즘은 하나의 변수만 사용하고 깊이가 1이기 때문에 파산 예측력은 매우 낮다. 적응형 부스팅은 이 약한 예측력을 갖는 결정 트리를 조화시켜 높은 예측력을 보이게 된다. 이 12개의 약 분류기의 집합을 H라고 한다.

훈련 샘플 m 개를 $(x_1, y_1), \dots, (x_m, y_m)$ 이라고 하자. 여기에서 x 는 분류 대상의 특징을 나타내고, y 는 클래스로 -1 또는 1의 값을 가질 수 있다. 본 논문에서는 한 기업의 모델 변수 집합이 x 가 되고 파산 기업은 -1, 정상 기업은 1로 분류 된다. 즉 $x=(EBIT/TA, EBT/CAP, WC/TA, WC/S, CA/TA, CA/CL, C/TA, C/CL, \ln TA, S/CAP,$

S/CA, S/TA)인 특징 벡터가 된다. 각각의 약 분류기는 특징 x 에서 하나의 값만으로 분류를 시도한다.

$D_1(i) = \frac{1}{m}$ 로 가중치의 분포를 초기화한다.

$t=1$ 부터 T 까지 총 T 회에 대하여 다음을 반복 수행한다.

- 오차가 가장 적은 약 분류기를 골라서 h_t 라고 한다. 여기에서 오차는 가중치 분포에 따라 결정되는 가중 오차이다.

$h_t = \operatorname{argmax}_{h_t \in H} |0.5 - e_t|$, 여기에서 $e_t = \sum_{i=1}^m D_t(i) I(y_i \neq h_t(x_i))$ 함수 I 는 인자가 참일 경우 1, 거짓일 경우 0을 리턴 한다.

- $|0.5 - e_t| \leq \beta$ 이면 루프를 종료한다. 상수 β 는 미리 정해 둔 임계값으로 오차가 이 값 이하로 낮아지면 훈련을 종료한다.
- α_t 를 다음과 같이 계산한다. $\alpha_t = \frac{1}{2} \ln \frac{1-e_t}{e_t}$
- 그리고 m 개의 가중치 분포 D 를 다음과 같이 변경한다.

$$D_{t+1}(i) = \frac{D_t(i) e^{\alpha_t(2I(y_i \neq h_t(x_i)) - 1)}}{\sum_i D_t(i) e^{\alpha_t I(y_i \neq h_t(x_i))}}$$

최종 분류기는 다음과 같이 계산된다.

$$H(x) = \operatorname{sign} \left(\sum_{t=1}^T \alpha_t h_t(x) \right)$$

여기에서 sign 함수는 인자가 양수일 경우 +1 (정상 기업), 음수일 경우 -1(파산 기업)을 리턴 한다. 본 논문에서 T 는 1000으로 설정하였다. 즉 12개의 약 분류기가 1000개의 가중치 합으로 계산되어 정상(+1)과 파산(-1)으로 분류된다.

5. 실험

정상 기업의 수가 파산 기업의 수에 비해 절대적으로 다수이기 때문에 둘의 수를 비슷하게 조정할 필요가 있다. 다만, 정상 기업에서 어떤 방식으로 표본을 추출하는가에 따라 모델의 파산 예측이 달라질 가능성이 있기 때문에 공정한 방법이 필요하다. 따라서 정상 기업에서 무작위 표본 추출을 하여 파산 기업의 수와 일치하도록 하였다.

정상 기업과 파산 기업을 자본금 규모에 따라 분류한 후 파산 기업의 수만큼 정상 기업을 무작위로 표본 추출하여 파산 기업의 수와 동일하게 하였다. 전체 기업 샘플 수는 정상 기업 샘플 수와 파산 기업 샘플 수가 각각 1381로, 둘을 합하면 2762개이고, 이 중 80%인 2208개를 훈련 샘플로 사용하고 나머지 20%인 554개를 테스트 샘플로 사용하였다. 훈련 샘플과 테스트 샘플에는 각각 정상 기업과 파산 기업의 수가 동일하다. 기업 규모에 따른 실험에서도 80%를 훈련 샘플로 사용하고 20%를 테스트 샘플로 사용하였다.

적응형 부스팅(AdaBoost)의 상대적 성능 비교를 위해 다른 기계 학습을 사용한 모델로 인공 신경망과 SVM, 결정 트리를 실험하였다. 그리고 전통적 모델로 Z-score를 실험하였다.

적응형 부스팅의 약 분류기로 결정 트리(Decision tree)를 사용하였고, 단순 결정 트리와 비교를 위해 결정 트리를 사용한 모델에 대한 실험도 함께 수행했다. 인공 신경망 모델은 변수를 모두 입력으로 하고, 히든 레이어의 노드 수는 10개, 출력은 1개로 하였다. SVM은 모델 변수를 특징(feature)로 하고, 커널 함수는 RBF를 사용하였다. 모든 기계 학습 기반 모델에서 타겟 값은 파산을 -1, 정상 기업을 1로 하였다.

Z-score는 개인 회사(private firm)를 위한 모델을 사용하였다. 전체 기업 샘플에 비상장 기업도 상당수 포함되어 있기 때문이다. 이 경우 Z값은 다음 식으로 계산된다.

$$\text{Z-score-private} = 0.717T1 + 0.847T2 + 3.107T3 + 0.420T4 + 0.998T5$$

여기에서 T1=(유동자산-유동부채)/총자산, T2=유보이익/총자산, T3=세전영업이익/총자산, T4=장부가치/총부채, T5=매출액/총자산 이다. T4에

서 장부가치는 수집한 데이터에서 자본금(Capital)을 대신 사용하였다. 계산 결과 Z-score-private가 2.9 초과이면 정상 기업, 1.23 미만이면 파산 기업, 그리고 중간 범위의 경우 예측 보류로 하였다.

각 기업 규모별, 모델별 예측 결과는 <Table 5>와 같다. 예측 보류는 Z-score에서 정상, 파산 판단이 되지 않는 범위에 속하는 것으로 성공이나 오류 어디에도 속하지 않는 것이다. 정상 기업의 샘플은 파산 기업의 샘플과 수를 동일하게 하기 위해 실험 때마다 무작위로 표본 추출을 하

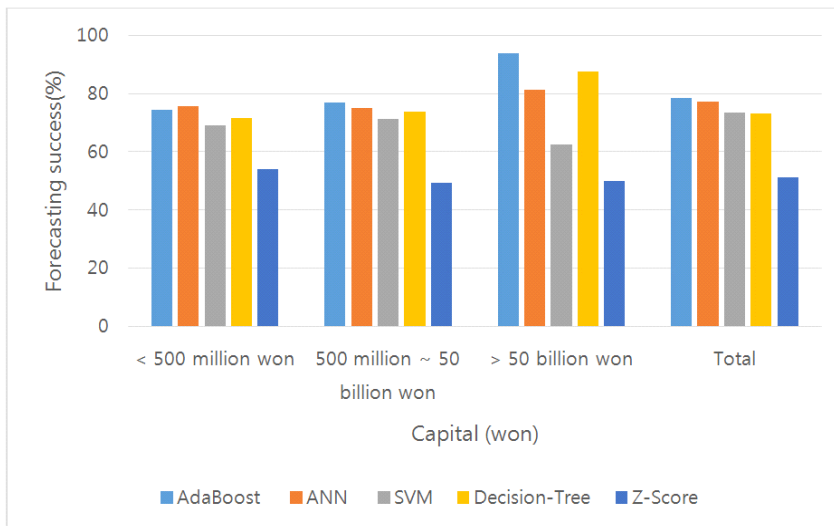
<Table 5> Bankruptcy forecasting results

	Success	Error	Unknown	Error Type	
				Type- I	Type- II
< 500 million won (Training samples:288, Test samples:74)					
AdaBoost	55	19	0	5	14
ANN	56	18	0	9	9
SVM	51	23	0	6	17
Decision-Tree	53	21	0	7	14
Z-Score	40	7	27	3	4
500 million ~ 50 billion won (Training samples: 1862, Test samples: 466)					
AdaBoost	358	108	0	43	65
ANN	350	116	0	57	59
SVM	332	134	0	61	73
Decision-Tree	343	123	0	57	66
Z-Score	230	96	140	78	18
> 50 billion won (Training samples: 56, Test samples: 16)					
AdaBoost	15	1	0	1	0
ANN	13	3	0	1	2
SVM	10	6	0	0	6
Decision-Tree	14	2	0	2	0
Z-Score	8	5	3	5	0
Total (Training samples: 2208, Test samples: 554)					
AdaBoost	435	119	0	48	71
ANN	427	127	0	57	70
SVM	406	148	0	59	89
Decision-Tree	405	149	0	60	89
Z-Score	284	102	168	80	22

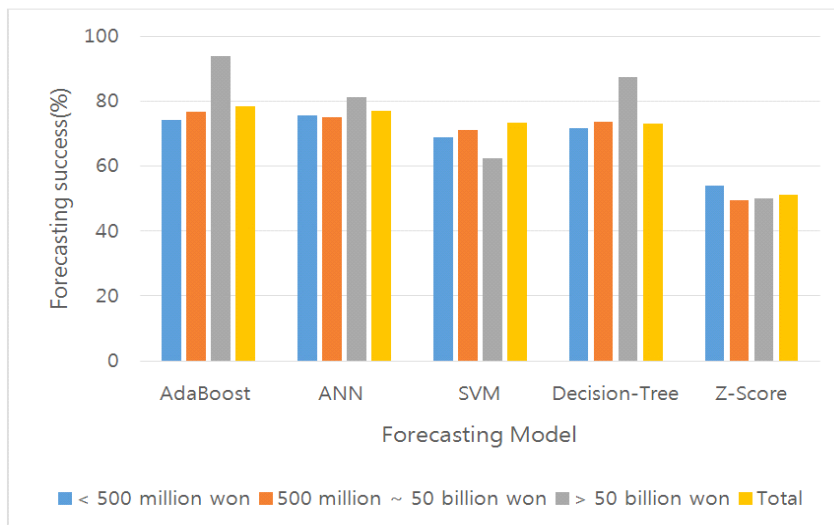
기 때문에 각 기업 규모별 실험의 결과 합계가 전체 규모의 합과 일치하지 않을 수 있다. 이를 보완하기 위해 실험을 10차례 반복하여 각 모델에 대해 중간값에 해당하는 결과를 선별하였다. <Table 5>의 결과를 바탕으로 모델별, 기업 규

모별 예측 성공률을 계산하여 <Figure 2>와 <Figure 3>에 표시하였다. 예측 성공률은 예측 성공 수를 테스트 샘플 수로 나눈 것이다.

전통적 Z-score의 경우 예측 보류 구간으로 인해 성공률이 기계 학습 기반의 모델에 비해 현저



<Figure 2> Forecasting success ratio of the models by the capital size of companies



<Figure 3> Forecasting success ratio of the models by the model

히 떨어진다. 게다가 보류 구간으로 인해 예측 오류가 발생하지 않는 것은 아니다. 즉, 기계 학습 기반에 비해 Z-score의 예측력은 매우 낮다고 할 수 있다.

<Figure 2>와 <Figure 3>에서 보듯이 기업 규모에 따른 모델의 예측 성공률은 적응형 부스팅(AdaBoost)가 가장 우수함을 알 수 있다. 특히 기업 규모가 500억 이상인 대형 건설 기업의 경우 특히 상대적으로 더 우수함을 알 수 있다. 인공 신경망(ANN) 모델의 경우 소형 규모의 경우 AdaBoost보다 약간 더 예측력이 좋은 것으로 보이나, 의미를 부여할 만큼은 아닌 것으로 판단된다. SVM의 경우 실험 결과로 볼 때 파산 예측 모델로 적합하지 못한 것으로 판단된다. 결정 트리는 단순한 알고리즘에 비해 예측력이 어느 정도 우수하며 특히 대형 규모 건설 기업에서 눈에 띄는 예측력을 보인다. 적응형 부스팅은 이 결정 트리를 약 분류기로 하여 더 정확한 파산 예측이 가능했던 것으로 보인다.

6. 결론

건설기업을 대상으로 다양한 파산예측모형의 결과와 적응형 부스팅을 이용한 파산예측모형의 결과를 비교 분석해보았다. 이러한 결과는 우리에게 실무적으로 중요한 시사점을 준다. 부동산 경기침체 여파로 건설기업의 파산이 급증하고 이로 인한 경제적, 사회적 문제가 심화되고 있는 현 시점에서, 건설기업의 파산이 주주, 채권자, 피고용인 등 여러 경제주체들에게 큰 고통을 주는 만큼 건설기업에 특화된 파산예측모형의 필요성은 매우 크다고 할 수 있다.

우리가 연구결과 알 수 있었던 것은 전통적인

통계적 회귀분석 방법의 일종인 알트만의 Z-score 방법에 비해 기계학습 방법이 훨씬 더 우수한 예측력을 보였다는 점이다. 기계학습에서 종종 발생하는 과적합(overfitting) 문제가 예측력 저하로 이어지는 경우가 있다는 점을 감안할 때 우리가 사용한 기계학습 방법 모두 Z-score 방법에 비해 우수한 예측력을 보였다는 점은 특기할 만 하다. 아울러 적응형 부스팅 방법이 모든 모형 중 가장 우수한 예측력을 보여주었으며, 특히 기업 규모가 큰 경우 예측력이 더욱 우수하다는 점은 적응형 부스팅의 실무적 적용에 매우 유용한 시사점을 준다고 할 수 있다.

이러한 결과의 원인으로 건설 기업의 재무 비율 변수들의 분산이 큰 이유를 꼽을 수 있다. 이로 인해 Z-score의 예측력이 낮아지고, 마찬가지로 변동성에 영향을 받는 인공 신경망이나 SVM 또한 적응형 부스팅에 비해 상대적으로 예측력이 낮았다. 금번 연구를 통해 적응형 부스팅을 이용한 파산예측모형 - 특히 건설업 부문에서 - 이 적합함을 확인하였다. 이 결과를 통해 건설기업과 관련된 주주 및 경영진, 채권자, 피고용인 등이 이를 활용하여 파산가능성을 미리 예측하고 고통을 최소화하기 위한 합리적인 의사 결정을 하는데 도움이 되기를 기대한다.

참고문헌

Alfaro, E., N. Garcia, M. Gamez, and D. Elizondo, "Bankruptcy forecasting: An empirical comparison of AdaBoost and neural networks," *Decision Support Systems*, Vol.45, No.1(2008), 110-122.

Altman, E. I., "Predicting financial distress of

- companies: revisiting the Z-score and ZETA models,” *Stern School of Business*, New York University (2000), 9-12.
- CERIK(Construction Economy Research Institute of Korea), *2013 Construction Market Outlook Report*, 2012.
- Freund, Y. and R. E. Schapire, “A decision-theoretic generalization of on-line learning and an application to boosting,” *Computational learning theory*, Vol. 904(1995), 23-37.
- Kim, M. J., “Ensemble Learning for Solving Data Imbalance in Bankruptcy Prediction,” *Journal of Intelligence and Information Systems*, Vol.15, No.3(2009), 1-15.
- Min, J. H. and Y. C. Lee, “Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters,” *Expert Systems with Applications* Vol.28, No.4 (2005), 603-614.
- NICE, *Credit Information Service*, Available at <http://www.nicednb.com> (Accessed 10 March, 2014).
- Shin, K. S., T. S. Lee, and H. J. Kim, “An application of support vector machines in bankruptcy prediction model,” *Expert Systems with Applications*, Vol.28, No.1(2005), 127-135.
- Shin, T. S. and T. H. Hong, “Corporate Credit Rating based on Bankruptcy Probability Using AdaBoost Algorithm-based Support Vector Machine,” *Journal of Intelligence and Information Systems*, Vol.17, No. 3(2011), 25-41.
- Sun, J., B. Liao, and H. Li, “AdaBoost and Bagging Ensemble Approaches with Neural Network as Base Learner for Financial Distress Prediction of Chinese Construction and Real Estate Companies,” *Recent Patents on Computer Science*, Vol.2013, No.6(2013), 47-59.
- Tae, C. W. and K. S. Shin, “GA-based Normalization Approach in Back-propagation Neural Network for Bankruptcy Prediction Modeling,” *Journal of Intelligence and Information Systems*, Vol.16, No.3(2010), 1-14.
- Verikas, A, Z. Kalsyte, M. Beauskiene, and A. Gelzinis, “Hybrid and ensemble-based soft computing techniques in bankruptcy prediction: a survey,” *Soft Computing* Vol.14, No.9 (2010), 995-1010.
- Wilson, R. L. and R. Sharda, “Bankruptcy prediction using neural networks,” *Decision Support Systems* Vol.11, No.5(1994), 545-557.

Abstract

Bankruptcy Forecasting Model using AdaBoost: A Focus on Construction Companies

Junyoung Heo* · Jin Yong Yang**

According to the 2013 construction market outlook report, the liquidation of construction companies is expected to continue due to the ongoing residential construction recession. Bankruptcies of construction companies have a greater social impact compared to other industries. However, due to the different nature of the capital structure and debt-to-equity ratio, it is more difficult to forecast construction companies' bankruptcies than that of companies in other industries. The construction industry operates on greater leverage, with high debt-to-equity ratios, and project cash flow focused on the second half. The economic cycle greatly influences construction companies. Therefore, downturns tend to rapidly increase the bankruptcy rates of construction companies. High leverage, coupled with increased bankruptcy rates, could lead to greater burdens on banks providing loans to construction companies. Nevertheless, the bankruptcy prediction model concentrated mainly on financial institutions, with rare construction-specific studies.

The bankruptcy prediction model based on corporate finance data has been studied for some time in various ways. However, the model is intended for all companies in general, and it may not be appropriate for forecasting bankruptcies of construction companies, who typically have high liquidity risks. The construction industry is capital-intensive, operates on long timelines with large-scale investment projects, and has comparatively longer payback periods than in other industries. With its unique capital structure, it can be difficult to apply a model used to judge the financial risk of companies in general to those in the construction industry.

Diverse studies of bankruptcy forecasting models based on a company's financial statements have been conducted for many years. The subjects of the model, however, were general firms, and the models may not be proper for accurately forecasting companies with disproportionately large liquidity risks, such as construction companies. The construction industry is capital-intensive, requiring significant investments

* Corresponding Author: Junyoung Heo
Department of Computer Engineering, Hansung University
389 Samseon-dong 2-ga, Seongbuk-gu, Seoul 136-792, Korea
Tel: +82-2-760-8039, Fax: +82-2-760-4346, E-mail: jyheo@hansung.ac.kr
** Department of Computer Engineering, Hansung University

in long-term projects, therefore to realize returns from the investment. The unique capital structure means that the same criteria used for other industries cannot be applied to effectively evaluate financial risk for construction firms.

Altman Z-score was first published in 1968, and is commonly used as a bankruptcy forecasting model. It forecasts the likelihood of a company going bankrupt by using a simple formula, classifying the results into three categories, and evaluating the corporate status as dangerous, moderate, or safe. When a company falls into the “dangerous” category, it has a high likelihood of bankruptcy within two years, while those in the “safe” category have a low likelihood of bankruptcy. For companies in the “moderate” category, it is difficult to forecast the risk. Many of the construction firm cases in this study fell in the “moderate” category, which made it difficult to forecast their risk.

Along with the development of machine learning using computers, recent studies of corporate bankruptcy forecasting have used this technology. Pattern recognition, a representative application area in machine learning, is applied to forecasting corporate bankruptcy, with patterns analyzed based on a company’s financial information, and then judged as to whether the pattern belongs to the bankruptcy risk group or the safe group. The representative machine learning models previously used in bankruptcy forecasting are Artificial Neural Networks, Adaptive Boosting (AdaBoost) and, the Support Vector Machine (SVM). There are also many hybrid studies combining these models.

Existing studies using the traditional Z-Score technique or bankruptcy prediction using machine learning focus on companies in non-specific industries. Therefore, the industry-specific characteristics of companies are not considered. In this paper, we confirm that adaptive boosting (AdaBoost) is the most appropriate forecasting model for construction companies by based on company size. We classified construction companies into three groups - large, medium, and small based on the company’s capital. We analyzed the predictive ability of AdaBoost for each group of companies. The experimental results showed that AdaBoost has more predictive ability than the other models, especially for the group of large companies with capital of more than 50 billion won.

Key Words : Bankruptcy Forecasting, Construction, AdaBoost, Z-Score

저 자 소개



허준영

현재 한성대학교 컴퓨터공학과 조교수로 재직 중이다. 서울대학교에서 컴퓨터공학과 공학사를 취득하고, 동 대학원에서 컴퓨터공학 박사를 취득하였다. 주된 연구분야는 시스템 소프트웨어와 기계학습, 금융 공학 등이 있다.



양진용

현재 자유 연구가로 활동하고 있다. 한양대학교에서 경제학박사를 취득하였으며, 주된 연구분야는 계량금융, 계산금융, 경제물리이다.