

RNN(Recurrent Neural Network)을 이용한 기업부도예측모형에서 회계정보의 동적 변화 연구

권혁건

한양대학교 일반대학원
비즈니스 인포메틱스학과
(old.kwon@gmail.com)

이동규

한양대학교 일반대학원
비즈니스 인포메틱스학과
(harry8256@naver.com)

신민수

한양대학교 경영대학 경영학부
(minsooshin@hanyang.ac.kr)

기업의 부도는 이해관계자들뿐 아니라 사회에도 경제적으로 큰 손실을 야기한다. 따라서 기업부도예측은 경영학 연구에 있어 중요한 연구주제 중 하나로 다뤄져 왔다. 기존의 연구에서는 부도 예측을 위해 다변량판별분석, 로짓분석, 신경망분석 등 다양한 방법론을 이용하여 모형의 부도 예측력을 높이고 과적합의 문제를 해결하고자 시도하였다. 하지만 기존의 연구들이 시간적 요소를 고려하지 않아 발생할 수 있는 문제점들을 갖고 있음에도 불구하고 부도 예측에 있어서 동적 모형을 이용한 연구는 활발히 진행되고 있지 않으며 따라서 동적 모형을 이용하여 부도예측모형이 더욱 개선될 여지가 있다는 점을 확인할 수 있었다.

이에 본 연구에서는 RNN(Recurrent Neural Network)을 이용하여 시계열 재무 데이터의 동적 변화를 반영한 모형을 만들었으며 기존의 부도예측모형들과의 비교분석을 통해 부도 예측력의 향상에 도움이 된다는 것을 확인할 수 있었다. 모형의 유용성을 검증하기 위해 KIS Value의 재무 데이터를 이용하여 실험을 수행하였고 비교 모형으로는 다변량판별분석, 로짓분석, SVM, 인공신경망을 선정하였다. 실험 결과 제안된 모형이 비교 모형에 비해 우수한 예측력을 보이는 것으로 나타났다. 따라서 본 연구는 변수들의 변화를 포착하는 동적 모형을 부도 예측에 새롭게 제안하여 부도예측 연구의 발전에 기여할 수 있을 것으로 기대된다.

주제어 : 순환 신경망, 부도 예측, 시계열 모형

논문접수일 : 2017년 7월 31일 논문수정일 : 2017년 9월 12일 게재확정일 : 2017년 9월 20일
원고유형 : 일반논문 교신저자 : 신민수

1. 서론

기업의 부도는 사회에 경제적으로 큰 영향을 미치는 사건이다. 일련의 경제위기들을 겪으면서 기업의 부도위험을 정확하게 판단하는 것에 대한 중요성은 더욱 더 강조되고 있으며 부도 예측을 통해 사회·경제적으로 발생하게 되는 손실을 줄일 수 있다는 점에서 부도예측 모형을 개발하는 것은 중요한 의미와 가치를 갖는다. 따라서 학계와 산업계에서는 기업의 부도를 예측하기

위한 연구를 중요하게 인식하고 이에 관련된 다양한 연구들을 지속적으로 수행해왔다.

부도예측연구는 1960년대부터 시작되었다. 초기의 연구에서는 판별분석(discriminant analysis), 로짓분석(logistic regression), 프로빗분석(probit analysis)등과 같은 통계적기법을 활용한 모형들이 주로 연구되었고 1990년대에 이르러서는 부도예측에 인공신경망(artificial neural network)과 같은 머신러닝 기법을 적용하여 예측의 성능을 향상시키고자 하는 시도들이 존재했다. 이후로

도 회계정보와 시장정보를 통합하여 부도 예측력을 높인 헤저드모형, 예측력의 일반화에 있어서 높은 성능을 보이는 SVM(Support Vector Machine)을 활용한 모형 등의 연구가 꾸준히 이어져오고 있다.

통계적 기법을 활용한 연구에서 (Altman, 1960)은 기업의 재무변수를 사용하여 판별분석으로 부도 기업을 구분하는 Z-Score 모형을 연구하여 부도를 예측하였다. (Ohlson, 1980)은 Altman의 연구와 달리 로짓분석을 이용하여 부도 예측 확률을 계산하는 모형을 만들었다. 머신러닝 기법을 적용한 연구에서는 (Odom and Sharda, 1990)이 기존 부도 예측에 사용된 변수들을 인공신경망에 적용하여 분석을 진행하였고 (Shin et al, 2005)은 SVM을 활용하여 부도 예측 모형을 만들었다. 국내에서도 부도 예측을 주제로 한 연구가 활발히 진행되었다. (Kim et al, 2017)은 이진 분류문제에서의 딥러닝 활용성과 비즈니스 애널리틱스에 적용 가능성을 평가하였으며, (Ok and Kim, 2009)는 유전자 알고리즘을 활용하여 다양한 기법의 결과를 유기적으로 통합하는 부도 예측 모형을 제안했다. (Kim and Ahn, 2015)에서 이진 분류 방법 성능 향상에서 분류 기법 보단 전처리 단계에서 성능을 향상 시키려 노력하였다. K-RNN(K-Reverse Nearest Neighbor)을 이용하여 Outlier를 제거 하였고 OCSVM(one-class support vector machine)을 이용하여 전체 데이터 중 training data를 sampling하였다. (Jo et al, 2015; Lee et al, 2015)은 군집 분석의 한 종류인 SOM(self-organizing map)을 이용하여 부도의 범주를 세분화 하였다. 특히 (Jo et al, 2015)에서는 인공신경망과 성능을 비교 하였으며 SOM을 통한 부도의 세분화가 일반적인 신경망 보다 성능이 뛰어났다. (Min, 2014)는 사례

선택(Instance Selection)을 활용한 배깅(bagging) 모형을 연구하였으며, 분류기들을 다양화시키는 앙상블 기법을 제안함으로써 부도 예측 모델의 성능 향상에 기여하였다.

일반적으로 회사의 재정상황이나 회계정보는 시간에 따라 변화하게 된다. 마찬가지로 시장상황 또한 변화하여 어느 특정 시점의 정보만을 가지고 부도를 예측하기에는 많은 어려움이 있다. (Deakin, 1972)는 예측 변수의 중요도가 시간에 따라서 변화한다는 점을 발견하였고 이는 (Grice, 2001)에 의해서 다시 한번 확인되었다. 하지만 정적인 모형의 경우 시간의 흐름에 따라 변화하는 데이터의 속성을 무시한 채 연구가 진행된다. (Shumway, 2001)은 이로 인해 정적인 모형의 부도 예측 결과는 편향되고 일관성 없는 결과를 도출 해내게 된다고 하였다. 따라서 부도예측모형의 연구에서는 시간 흐름에 따라 변화하는 효과를 고려하는 것 또한 필요하다. 이전의 부도예측 모형 연구에서는 시간에 따른 변화를 포착하는 동적 모형에 대한 연구가 부족하였고, 동적 예측 변수를 사용한 연구도 많지 않았다. (Niemann et al., 2008)은 예측 변수가 여러 기간의 정보를 포함할 경우 모형의 성능이 더 발전할 수 있다는 점을 지적하였다. (Campbell et al., 2008)의 헤저드 모형에서는 수익성 관련 설명 변수를 이전 10개월의 가중평균값을 이용하여 이러한 단점을 극복하려는 시도를 하였다. (Kim and Partington, 2015)은 Cox 헤저드 모형이 여러 기간의 데이터를 계산하도록 변형한 결과 기존의 Cox모형이나 로짓모형보다 더 좋은 예측력을 보인다는 것을 확인하였다.

본 연구에서는 시계열 데이터를 학습하는데 좋은 성능을 보이는 RNN(Recurrent Neural Network)을 사용하여 부도예측의 성능을 향상시

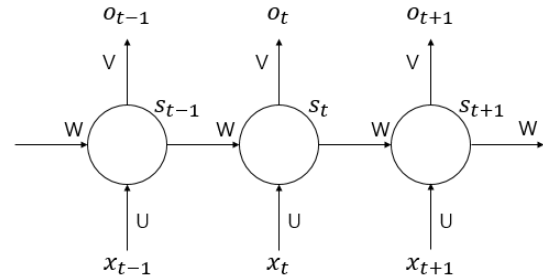
키려는 시도를 하였다. 동적인 모형과 정적인 모형의 차이를 확인하기 위해 기존 연구에서 부도 예측에 사용된 다변량판별분석, 로짓분석, SVM, 인공신경망 등 모형과 성능을 비교하였다. 모형 추정에 사용된 변수는 기업의 재무변수를 사용하였다. 또한 선행연구에서 동적 모형의 추정에 사용된 변수들로 RNN 모형을 추정할 때와 정적 모형의 추정에 사용된 변수들을 사용했을 때의 결과 차이를 확인함으로써 RNN 모형이 좋은 성능을 보이기 위해서는 적절한 변수의 선택 또한 중요하다는 것을 확인할 수 있었다. 부도 예측력 평가에는 예측정확도(accuracy), ROC곡선, 부도 적중률(hit-ratio)을 적용하였다. 예측정확도와 ROC곡선을 이용한 평가 척도에서는 본 연구의 모형이 기존의 모형들 보다 향상된 성능을 보였으나 부도적중률에서는 본 연구의 모형이 높은 1종 오류, 낮은 2종 오류를 보여 향후 개선의 여지를 남겼다.

2. 분석모형

2.1 RNN의 구조

RNN은 인공 신경망 알고리즘의 한 종류이다. 일반적인 인공신경망은 입력된 데이터가 신경망의 노드와 활성화함수(activation function)를 통해 계산되고 결과값과 주어진 정답 간의 차이를 최소화 하는 방식으로 가중치를 조절하여 최적의 가중치 값을 찾아낸다. 이 과정에서 입력 데이터는 모든 노드와 한 번씩 계산된다. 하지만 RNN은 노드가 자기 자신과 순환적으로 연결되어 있으며 이러한 순환적인 구조 때문에 순차적인 데이터를 처리하는데 유용하게 사용될 수 있다는

장점이 있다.



〈Figure 1〉 Structure of RNN

이러한 순환적 구조 때문에 RNN에 입력된 데이터의 t-1시점 출력 값은(특정 시점을 t라고 하였을 때) t시점의 출력 값에도 영향을 주게 된다.

x_t = input of time step t

$$s_t = f(Ux_t + Ws_{t-1})$$

$$o_t = \text{softmax}(Vs_t)$$

o_t = output of time step t

활성함수 f는 보통 tanh나 ReLU가 사용되며 첫 hidden state를 계산 하기 위해 s_{-1} 는 0으로 초기화 시킨다. 과거의 출력이 다시 가중치 계산에 사용되는 구조 때문에 RNN은 과거의 정보를 기억할 수 있다는 특징이 있다. 또한 RNN의 경우 일반 인공신경망의 역전파(back propagation) 알고리즘과는 약간 다르게 매 시점마다 공유된 가중치를 역전파 알고리즘을 통해 업데이트 시켜줘야 한다는 특징이 있다. 하지만 시점이 이어질 수록 기울기(gradient)의 값이 0으로 수렴하거나 무한으로 발산하게 되는 현상이 발생할 수 있다. 이러한 문제를 vanishing gradient 문제라고 하며

기울기 값이 없어지기 때문에 더 이상 학습이 진행되지 못하게 된다.

2.2 LSTM(Long short-term Memory)

기울기가 0이 되는 문제, 즉 vanishing gradient 문제를 해결하기 위해 LSTM 알고리즘이 사용된다. LSTM은 역전파 과정에서 기울기가 잘 전달될 수 있도록 고안된 가중치의 구조이다. LSTM 유닛은 여러 개의 게이트가 붙어있는 셀로 이루어져있으며 셀의 정보들을 저장, 불러오기, 유지하는 기능이 있다. 이러한 기능 때문에 RNN에서 보였던 기울기가 사라지는 문제를 해결할 수 있으며 오래전 시점의 정보도 더 잘 기억할 수 있게 한다. LSTM셀의 구조는 다음과 같다.

$$h_t = \tanh(W_{hh}h_{t-1} + W_{hv}v_t + W_{hm}\tilde{m}_{t-1})$$

$$i_t^g = \text{sigmoid}(W_{igh}h_t + W_{igv}v_t + W_{igm}\tilde{m}_{t-1})$$

$$i_t = \tanh(W_{ih}h_t + W_{iv}v_t + W_{im}\tilde{m}_{t-1})$$

$$o_t = \text{sigmoid}(W_{oh}h_t + W_{ov}v_t + W_{om}\tilde{m}_{t-1})$$

$$f_t = \text{sigmoid}(b_f + W_{fh}h_t + W_{fv}v_t + W_{fm}\tilde{m}_{t-1})$$

$$m_t = m_{t-1} \odot f_t + i_t \odot i_t^g$$

$$\tilde{m}_t = m_t \odot o_t$$

$$z_t = g(W_{yh}h_t + W_{ym}\tilde{m}_t)$$

본 연구모형에서는 Many to One의 RNN 구조를 이용하였다. Many to One 구조란 여러 데이터가 입력되고 하나의 값이 출력되는 구조를 말한다. Figure 1의 x_t 부분 time step은 3개로 동일하며 마지막의 출력값 o_{t+1} 만 존재하고 앞의 두 출력값 o_t , o_{t-1} 은 존재하지 않는 구조이다. 3년간의 재무

데이터를 이용하여 모형에 적합을 하였기 때문에 시퀀스는 3으로 설정하였다. 즉 과거 3개 시점 이전, 2개 시점 이전, 1개 시점 이전의 재무정보가 부도예측의 결과에 모두 영향을 주는 구조의 모형이다.

3. 자료

본 연구는 부도예측모형의 추정과 예측력 비교를 위해 2010년부터 2016년까지 유가증권시장(KOSPI), 코스닥시장(KOSDAQ) 그리고 코넥스시장(KONEX)에 상장된 비 금융기업을 분석 대상으로 선정하였다. 분석에 이용된 기업별 재무정보는 KIS Value를 이용하여 수집하였다. 기업의 재무상태 악화가 이미 반영된 재무정보를 이용하여 부도를 예측하는 오류를 방지하기 위하여 2년의 시차를 두고 재무정보를 수집하였으며 부도 관측 기간은 당해 년도 1월부터 12월까지로 정의하였다.

실증분석을 위해 모형의 추정을 위한 추정표본은 2010년부터 2011년까지로 설정하였고, RNN 모형의 최적의 파라미터 설정을 위해 2012년 데이터를 검정표본으로 사용하였다. 예측력 평가를 위한 예측표본은 2013년부터 2016년으로 설정하였다. 부도기업에 대한 정의는 (Lee, 2015)의 연구와 같이 실적부진의 사유로 상장폐지된 기업을 부도기업으로 정의하였다. KIND(한국거래소 전자공시)를 통해 상장폐지 사유를 확인하였으며 자진등록취소, 합병, 지주회사 신규상장, 주식분산기준 미달, 기타 등록 취소 등 실적부진과 관계없이 상장폐지된 기업은 정상기업으로 간주하였다.

정상 기업에 비해 부도 기업의 수는 상당히 적

〈Table 1〉 Data Set

	Altman's Variable		Kim and Partington's Variable	
	Training	Test	Training	Test
Normal	152	88	148	81
Bankruptcy	153	73	153	73

은 편이며 이러한 비 대칭 데이터는 부도예측모형이 정상 기업만 편향되게 예측하도록 하여 예측표본에서 부도기업을 찾아내지 못하는 문제를 발생시킬 수 있다. 본 연구는 정상기업을 언더샘플링 하여 부도예측모형에 학습시킨 결과가 예측 표본의 예측력에 있어 다른 샘플링 결과와 큰 차이가 없다는 (Zhou, 2012)의 연구에 근거하여 정상기업을 부도기업과 비슷한 비율로 랜덤하게 추출하여 데이터의 비율이 대칭되도록 하였다. <Table 1>은 실증분석에 사용된 2010년부터 2016년전체기업과 부도기업의 수의 비중을 보여 준다.

4. 분석 결과와 해석

본 연구에서는 2개의 서로 다른 변수 집합에 대해서 RNN과 나머지 4개 모형의 예측 성능을 비교하였으며, 성능평가는 예측정확도, ROC곡

선, 부도적중률 3가지 기준에 의해서 진행하였다. 나머지 4개의 모형은 다변량판별분석, 로짓 모형, SVM, 인공신경망이다. SVM의 kernel은 linear로 하였고 인공신경망에서 히든노드의 수는 10개 decay값은 0.1로 설정하여 연구를 진행하였다.

4.1 Altman(1968)의 변수를 이용한 분석

첫 번째 변수집합에는 (Altman, 1968)의 Z-score모형에 사용된 변수들을 사용했다. Z-score 모형의 변수들은 부도예측 연구에서 오래 전부터 이용되었으며 다양한 연구들에서 그 예측 효과가 입증되었다. 또한 5개의 Z-score 모형 변수는 기업의 유동성, 안정성, 수익성, 활동성을 나타내는 부도예측 연구의 기준이 되는 변수로 자리 잡아 왔다. 이러한 이유로 (Marcus, 1990)와 같은 인공신경망을 이용한 연구에서도 Z-score 모형의 변수들을 사용하였다. 변수에 대한 설명은 <Table 2>와 같다.

〈Table 2〉 Altman's Variable

Origin	Variable	Variable Formula
Altman (1968)	X1	Working Capital / Total Asset
	X2	Retained Earnings / Total Asset
	X3	Business Profit / Total Asset
	X4	Equity / Total Debt
	X5	Sales / Total Debt

Altman의 Z-score모형에 사용된 변수 집합을 이용해 만든 데이터의 부도기업 수는 NA값을 제거 후 추정표본, 예측표본에서 각 153개, 73개였다. 정상기업의 수는 추정표본, 예측표본에서 각 152개, 88개였다. 계산에서 이상치(Outlier)가 모형에 미치는 영향을 최소화하고 오버피팅의 위험을 줄이기 위해 변수 분포에서 99%이상 1% 이하에 해당되는 값들은 99%와 1%의 값들로 대체하였다.

Grid search를 이용하여 RNN모형에 사용될 최적의 파라미터를 설정하였다. 최종적으로 RNN 네트워크의 학습률은 0.01, 2개의 은닉노드, RELU 활성화함수를 사용하였고, Fully connected layer는 한 층만 사용하고 0~1 사이의 부도 확률을 구하기 위해 sigmoid로 활성화 함수를 사용하였다. 마지막으로 800회의 반복학습을 진행했을 때 Overfitting이 가장 적은 최적의 성능을 보였다.

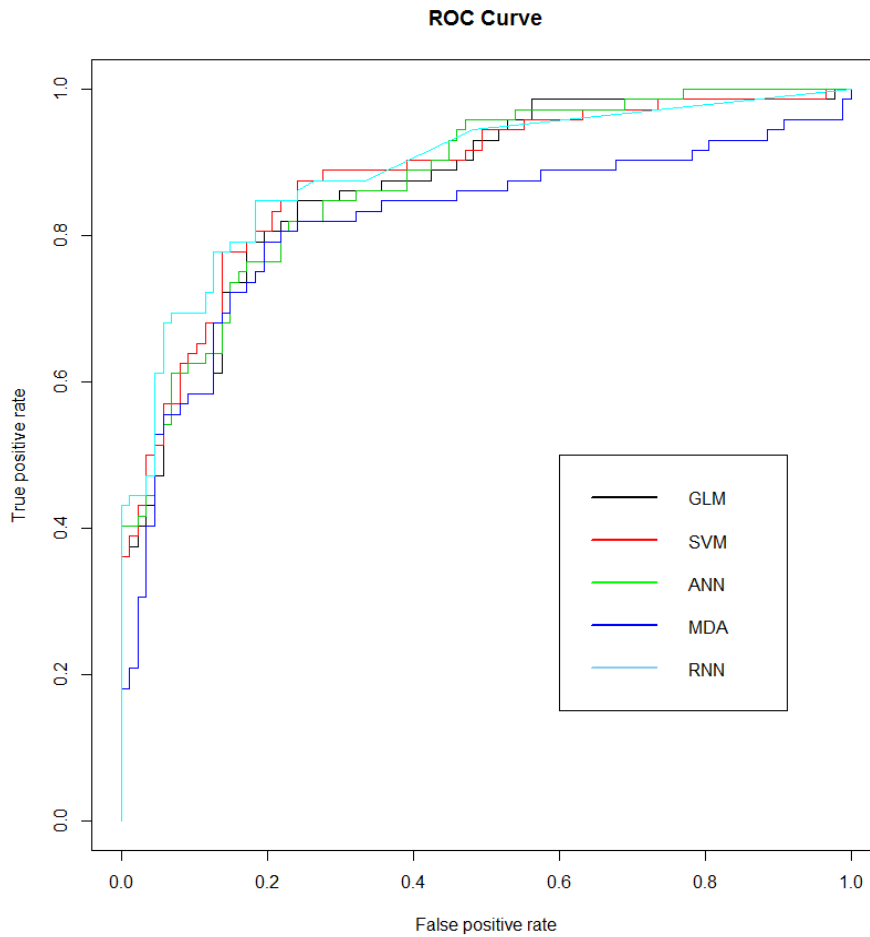
모형의 추정 결과 첫 번째 변수 집합에서 RNN의 정확도는 약 81%로 나타났으며 AUC는 약89%로 나타났다. 다른 모형들과 비교하였을 때 정확도와 AUC에서 모두 다른 모형보다 약간 높은 수준을 보였다. 이는 일반 인공신경망 모형보다 더 높은 수준이며 동적 모형의 성능이 더

우수하다는 것을 확인할 수 있다. 정확도 표는 <Table 3>, ROC커브의 그래프는 <Figure 2>와 같다.

부도적중률은 부도예측모형이 구간별로 부도를 얼마나 정확히 예측하였는가를 보여주는 지표이다. 부도위험이 가장 낮은 1부터 부도위험이 가장 높은 10까지를 표로 구성하였고 전체 부도기업 중 해당 구간의 부도기업 비율을 측정하였다. 실제 부도가 발생한 기업을 부도가 발생했다고 예측할 확률이 높은 모형 일수록 부도적중률이 높다고 할 수 있으며 부도 예측에 우수한 모형이라고 할 수 있다. 6 - 10 범주를 보았을 때 RNN이 65.3%로 가장 우수한 성능을 보였지만 SVM과 ANN의 값과는 크게 다르지 않았다. 한편 부도위험이 가장 낮은 1에서도 RNN의 값이 가장 높게 나타났다. 이는 많은 부실기업을 안전하다고 예측한 것이며 이는 투자자에게 큰 손실을 초래할 수 있다. 실제 부도 기업의 부도 확률을 지나치게 낮게 예측하는 것은 RNN 모형에서 알고리즘의 복잡도가 증가하여 과적합이 발생한 것으로 보이며 이러한 과적합을 피하기 위한 개선이 필요하다는 것을 알 수 있다. 부도적중률을 나타낸 표는 <Table 4>와 같다.

<Table 3> Estimated Model

Model	GLM	SVM	ANN	MDA	RNN
Accuracy	0.755	0.780	0.761	0.755	0.811
AUC	0.868	0.879	0.874	0.818	0.889



〈Figure 2〉 Model with Altman's Variable

〈Table 4〉 Altman's Variable Hit-Ratio

Decile	GLM	SVM	ANN	MDA	RNN
1	0.042	0.028	0.097	0.042	0.263
2	0.111	0.083	0.083	0.028	0.028
3	0.069	0.083	0.069	0.056	0.014
4	0.097	0.069	0.069	0.056	0.014
5	0.069	0.097	0.041	0.208	0.028
6 - 10	0.612	0.640	0.641	0.610	0.653

〈Table 5〉 Kim and Partington's Variable

Origin	Variable	Variable Formula
Kim And Partington (2015)	X1	Net Income / Total Asset
	X2	Working Capital / Total Asset
	X3	Current Asset / Current Debt
	X4	Total Debt / Total Asset
	X5	Net Cash Flow from Operation / Total Asset

4.2 Kim et al.(2015)의 변수를 이용한 분석

두번째 변수집합에는 선행연구 중 부도예측에 동적모형을 적용하여 예측 성능을 높인 (Kim and Partington, 2015)의 연구의 재무변수들을 이용했다. 해당 변수들은 (Shumway, 2001)을 참고하고 있으며 (Shumway, 2001) 역시 재무변수와 시장변수의 동적 변화에 대해 연구하였기 때문에 마찬가지로 동적 모형을 이용한 본 연구에서 해당 변수들이 더 좋은 결과를 보일 것으로 예상할 수 있었다. Kim and Partington 모형의 변수 설명은 <Table 5>와 같다.

해당 변수 집합을 이용해 만든 데이터의 부도 기업 수는 NA값을 제거 후 추정표본, 예측표본에서 각 153개, 73개였다. 정상기업의 수는 추정표본, 예측표본에서 각 148개, 81개였다. 계산에서 이상치(Outlier)가 모형에 미치는 영향을 최소화하고 오버피팅의 위험을 줄이기 위해 변수 분포에서 99%이상 1% 이하에 해당되는 값들은 99%와 1%의 값들로 대체하였다.

첫 번째 변수집합과 마찬가지로 Grid search를 이용하여 RNN모형에 사용될 최적의 파라미터를 설정하였고, 최종적으로 RNN네트워크의 학습률은 0.01, 4개의 은닉노드, tanh 활성화함수를 사용하였고, Fully connected layer는 한 층만 사용하고 0~1 사이의 부도 확률을 구하기 위해 sigmoid로 활성화 함수를 사용하였으며, 900회의 반복학습을 진행했을 때 최적의 성능을 보였다.

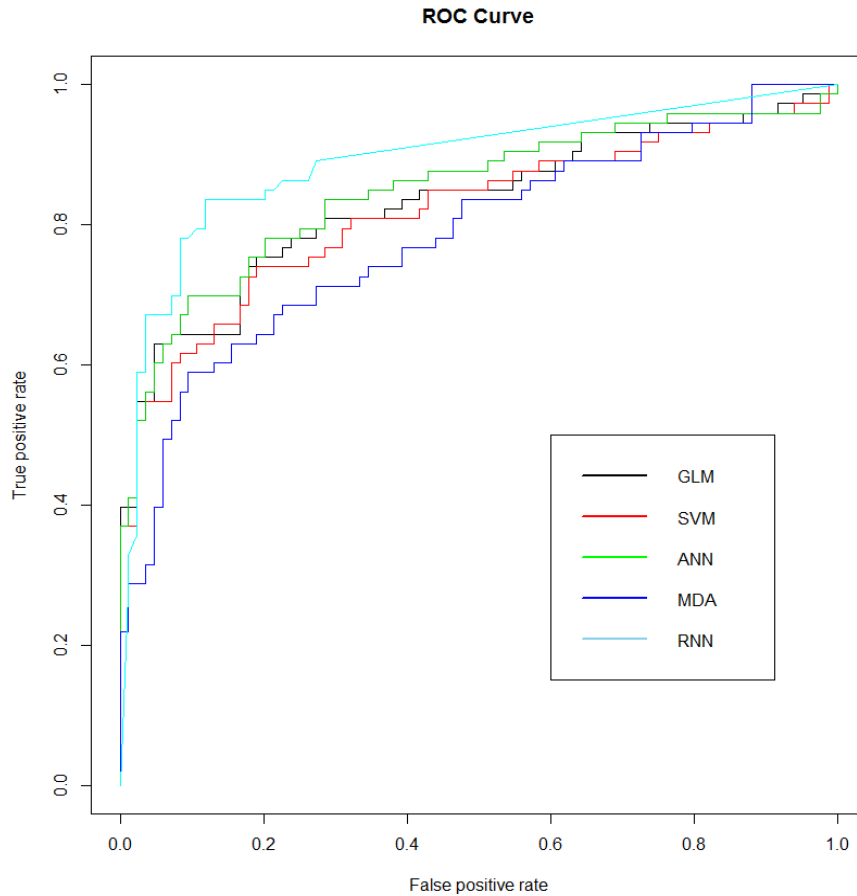
새로운 변수를 사용한 결과 모형의 예측 결과는 첫 번째 변수 집합보다 성능이 향상되었다. RNN의 정확도는 약 83%로 나타났으며 AUC는 약89%로 나타났다. 이는 첫 번째 변수 집합으로 추정했을 때 보다 기타 비교 모형들에 비해 훨씬 더 향상된 성능을 보이는 것으로 변수의 적합도가 동적 모형의 성능을 훨씬 더 향상시킨 것을 알 수 있다. 정확도 표는 <Table 6>, ROC커브의 그래프는 <Figure 3>과 같다.

부도적중률에서는 6 - 10 범주에서 RNN 모형이 67.1%로 다른 모형보다 높았다. 하지만

〈Table 6〉 Estimated Model

Model	GLM	SVM	ANN	MDA	RNN
Accuracy	0.783	0.771	0.796	0.732	0.828
AUC	0.832	0.818	0.849	0.782	0.891

Altman 변수 집합 모형과 마찬가지로 부도위험 경우가 높게 나타났다. 부도적중률을 나타낸 표이 가장 낮다고 예측한 기업에서 부도가 발생한 것은 <Table 7>과 같다.



〈Figure 3〉 Model with Kim and Partington's Variable

〈Table 7〉 Kim and Partington's Variable Hit-Ratio

Decile	GLM	SVM	ANN	MDA	RNN
1	0.068	0.055	0.055	0	0.164
2	0.110	0.096	0.096	0	0.096
3	0.082	0.096	0.096	0.110	0.014
4	0.096	0.096	0.055	0.178	0.041
5	0.055	0.110	0.068	0.205	0.014
6 - 10	0.589	0.547	0.63	0.507	0.671

4.3 실험의 결과

변수들의 동적 변화를 포착하기 위해 연속된 3년간의 과거 재무데이터를 이용한 동적 예측모형을 만들어 기존의 선행연구들의 모형과 비교한 결과 본 연구의 동적 모형이 부도 예측력에 있어 더욱 우수한 성능을 보이는 것을 확인할 수 있었다. 정확도와 ROC 커브를 이용하여 측정한 성능에서는 본 연구의 모형이 선행 연구의 모형들보다 일관되게 우수한 성능을 보였고 부도적 중률에 있어서도 부도위험이 제일 낮은 1 구간의 부도율이 높은 단점이 있었지만 나머지 구간에서는 기타 모형들보다 부도 기업을 더 잘 분류한다는 것을 확인할 수 있었다. 따라서 재무정보 분석을 통한 부도 예측에 있어서 동적 모형의 사용이 기업의 변화하는 정보를 포착하고 이를 통해 부도예측률의 향상에 도움이 된다는 것을 확인하였다.

5. 결론

본 연구에서는 RNN(Recurrent Neural Network)을 이용하여 시계열 재무 데이터의 동적 변화를 반영한 모형과 기존의 모형들을 비교함으로써 재무정보의 동적 변화를 포착한 모형이 부도예측모형의 연구에서 예측성능을 향상시킬 수 있다는 점을 확인하였다. 또한 두가지 다른 변수집합에 대해 모형을 만들어 비교함으로써 동적 모형에 적합한 변수의 사용이 동적 모형의 추정에 중요하다는 점도 확인할 수 있었다. 동적 모형은 정확도와 AUC 모두 비교모형들에 비해 성능이 우수하였고 부도 적중률에서도 RNN모형이 더욱 낮은 2종 오류를 보였다.

기존의 부도예측 연구는 정적 모형을 사용하여 기업의 변화하는 상황을 충분히 반영하지 못한 채 이루어졌다. 재무상태의 변화를 반영하지 못한 데이터와 모형을 이용한 예측은 적절하지 않을 뿐만 아니라 편향된 결과를 도출해낼 수 있다는 단점은 선행연구를 통해서 지적된 바 있다. 본 연구에서는 RNN을 이용한 동적 모형을 만들어서 실제로 연속된 3년간의 기업 재무정보의 변화를 반영하는 것이 예측력 향상에 도움이 된다는 점을 실증분석 하였다. 따라서 본 연구는 동적 모형을 이용한 부도예측 연구에 있어 학술적으로 기여할 수 있을 것으로 예상된다.

재무 정보는 산업, 기업의 규모 등에 따라 다른 특성을 가질 수 있다. 모형 추정에 있어서도 이러한 특성에 따라 구분된 모형을 이용해야 결과의 편향이나 오류를 없앨 수 있다. 하지만 본 연구에서는 부도기업의 데이터 확보의 어려움으로 인하여 데이터 수가 한정 되었고 따라서 학습에 사용할 수 있는 데이터가 부족했기 때문에 기업을 산업이나 규모별로 구분하지 않고 하나의 모델로 추정할 수 밖에 없었던 한계를 가지고 있다. 예측 성능평가에 있어서 비교분석이 선행연구에 이용된 정적모형들에 대해서만 이루어진 점 또한 아쉬운 점이다. 동적 모형을 이용한 선행연구들과도 비교가 이루어졌다면 RNN의 부도 예측에 있어서의 예측력을 더욱 잘 대비하여 확인할 수 있었을 것이다. 마지막으로 연구에 사용된 변수의 선택에 대한 추가적인 연구가 가능하다. 본 연구결과에서 보았듯 변수 집합에 따라서 모형의 예측 성능에 큰 차이가 발생할 수 있다. 변화하는 기업 정보를 더욱 잘 반영할 수 있는 변수에 대해서 연구하는 것 또한 향후 동적 모형을 활용한 부도기업 예측에 중요한 부분이 될 것이다.

참고문헌(References)

- Altman, E. I., "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy", *The Journal of Finance*, Vol.23, No.4(1968), 589~609.
- Campbell, J., J. Hilscher, and J. Szilagyi, "In Search of Distress Risk", *Journal of Finance*, Vol.63, No.6(2008), 2899-2939.
- Chava, S., and R. A. Jarrow, "Bankruptcy Prediction with Industry Effects", *Review of Finance*, Vol.8, No.4(2004), 537~569.
- Grice, J. S. and M. T. Dugan, "The Limitations of Bankruptcy Prediction Models: Some Cautions for the Researcher", *Review of Quantitative Finance and Accounting*, Vol.17, No.2(2001), 151~166.
- Kang, B. S. and J. H. Jo, "Prediction Performance Improvement of KOSDAQ Business Bankruptcy using SVM Preprocessor", *Journal of Korea Knowledge Information Technology Society*, Vol.7, No.4(2012), 21~27.
- Kiefer, N. M., "Economic Duration Data and Hazard Functions", *Journal of Economic Literature*, Vol.26, No.2(1998), 646~679.
- Kim, K. T., B. M. Lee, and J. W. Kim, "Feasibility of Deep Learning Algorithms for Binary Classification Problems", *Journal of Intelligence and Information System*, Vol.23, No.1(2017), 95~108.
- Kim, M. H. and G. Partington, "Dynamic forecasts of financial distress of Australian firms", *Australian Journal of Management*, Vol.40, No.1(2015), 135~160.
- Kim, T. H. and H. C. Ahn, "A Hybrid Under-sampling Approach for Better Bankruptcy Prediction", *Journal of Intelligence and Information System*, Vol.21, No.2(2015), 173~190.
- Lee, I. R. and D. C. Kim, "Evaluation of Bankruptcy Prediction Model Using Accounting Information and Market Information", *Journal of Korean Finance Association*, Vol.28, No.4(2015), 626~666.
- Lee, S. H., J. M. Park, and H. Y. Lee, "The Pattern Analysis of Financial Distress for Non-audited Firms using Data Mining", *Journal of Intelligence and Information System*, Vol.21, No.4(2015), 111~131.
- Min, J. H. and Y. C. Lee, "Support Vector Bankruptcy Prediction Model with Optimal Choice of RBF Kernel Parameter Values using Grid Search", *Journal of the Korean Operations Research and Management Science Society*, Vol.30, No.1(2005), 55~74.
- Min, S. H., "Integrated Corporate Bankruptcy Prediction Model Using Genetic Algorithms", *Journal of Intelligence and Information System*, Vol.20, No.4(2014), 121~139.
- Nam, O. J., H. J. Kim, and K. S. Shin, "Bankruptcy Type Prediction Using A Hybrid Artificial Neural Networks Model", *Journal of Intelligence and Information System*, Vol.21, No.3(2015), 79~99.
- Niemann, M., J. H. Schmidt, and M. Neukirchen, "Improving performance of corporate rating prediction models by reducing financial ratio heterogeneity", *Journal of Banking & Finance*, Vol.32, No.3(2008), 434~446.
- Odom, M. D. and R. Sharda, "A neural network model for bankruptcy prediction", *Proceedings of the IEEE International*

- Conference on Neural Networks*, (1990), 163~168.
- Ohlson, J. A., “Financial Ratios and the Probabilistic Prediction of Bankruptcy”, *Journal of Accounting Research*, Vol.18, No.1(1980), 109~131.
- Ok, J. K. and K. J. Kim, “Bankruptcy prediction using an improved bagging ensemble”, *Journal of Intelligence and Information System*, Vol.15, No.4(2009), 99~121.
- Park, D. J., Y. B. Yoon, and M. Yoon, “Prediction of corporate default by using machine learning method”, *Journal of the Korean Data & Information Science Society*, Vol.23, No.3(2012), 569~577.
- Park, J. W. and S. M. Ahn, “Corporate Bankruptcy Prediction Using Financial Ratios : Focused on the Korean Manufacturing Companies Audited by External Auditors”, *Korean Business Review*, Vol.43, No.3(2013), 639~669.
- Shin, K. S., T. S. Lee and Hyun-jung Kim, “An application of support vector machines in bankruptcy prediction model”, *Expert Systems with Applications*, Vol.28, No.1(2005), 127~135.
- Shumway, T., “Forecasting Bankruptcy More Accurately : A Simple Hazard Model”, *The Journal of Business*, Vol.74, No.1(2001), 101~124.
- Zhou, L., “Performance of corporate bankruptcy prediction models on imbalanced dataset : The effect of sampling methods”, *Knowledge-Based Systems*, Vol.41, No.1(2013), 16~25.

Abstract

Dynamic forecasts of bankruptcy with Recurrent Neural Network model

Hyukkun Kwon* · Dongkyu Lee* · Minsoo Shin**

Corporate bankruptcy can cause great losses not only to stakeholders but also to many related sectors in society. Through the economic crises, bankruptcy have increased and bankruptcy prediction models have become more and more important. Therefore, corporate bankruptcy has been regarded as one of the major topics of research in business management. Also, many studies in the industry are in progress and important.

Previous studies attempted to utilize various methodologies to improve the bankruptcy prediction accuracy and to resolve the overfitting problem, such as Multivariate Discriminant Analysis (MDA), Generalized Linear Model (GLM). These methods are based on statistics. Recently, researchers have used machine learning methodologies such as Support Vector Machine (SVM), Artificial Neural Network (ANN). Furthermore, fuzzy theory and genetic algorithms were used. Because of this change, many of bankruptcy models are developed. Also, performance has been improved.

In general, the company's financial and accounting information will change over time. Likewise, the market situation also changes, so there are many difficulties in predicting bankruptcy only with information at a certain point in time. However, even though traditional research has problems that don't take into account the time effect, dynamic model has not been studied much. When we ignore the time effect, we get the biased results. So the static model may not be suitable for predicting bankruptcy. Thus, using the dynamic model, there is a possibility that bankruptcy prediction model is improved.

In this paper, we propose RNN (Recurrent Neural Network) which is one of the deep learning methodologies. The RNN learns time series data and the performance is known to be good. Prior to experiment, we selected non-financial firms listed on the KOSPI, KOSDAQ and KONEX markets from 2010 to 2016 for the estimation of the bankruptcy prediction model and the comparison of forecasting

* Graduate School Business Informatics, Hanyang University

** Corresponding Author: Minsoo Shin

School of Business, Hanyang University

222 Wangsimni-ro, Seongdong-gu, Seoul 04763, Korea

Tel: [REDACTED] Fax: +82-2-2220-4777, E-mail: minsooshin@hanyang.ac.kr

performance. In order to prevent a mistake of predicting bankruptcy by using the financial information already reflected in the deterioration of the financial condition of the company, the financial information was collected with a lag of two years, and the default period was defined from January to December of the year. Then we defined the bankruptcy. The bankruptcy we defined is the abolition of the listing due to sluggish earnings. We confirmed abolition of the list at KIND that is corporate stock information website. Then we selected variables at previous papers. The first set of variables are Z-score variables. These variables have become traditional variables in predicting bankruptcy. The second set of variables are dynamic variable set. Finally we selected 240 normal companies and 226 bankrupt companies at the first variable set. Likewise, we selected 229 normal companies and 226 bankrupt companies at the second variable set.

We created a model that reflects dynamic changes in time-series financial data and by comparing the suggested model with the analysis of existing bankruptcy predictive models, we found that the suggested model could help to improve the accuracy of bankruptcy predictions. We used financial data in KIS Value (Financial database) and selected Multivariate Discriminant Analysis (MDA), Generalized Linear Model called logistic regression (GLM), Support Vector Machine (SVM), Artificial Neural Network (ANN) model as benchmark.

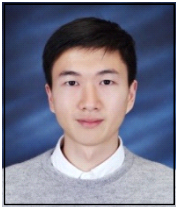
The result of the experiment proved that RNN's performance was better than comparative model. The accuracy of RNN was high in both sets of variables and the Area Under the Curve (AUC) value was also high. Also when we saw the hit-ratio table, the ratio of RNNs that predicted a poor company to be bankrupt was higher than that of other comparative models. However the limitation of this paper is that an overfitting problem occurs during RNN learning. But we expect to be able to solve the overfitting problem by selecting more learning data and appropriate variables. From these result, it is expected that this research will contribute to the development of a bankruptcy prediction by proposing a new dynamic model.

Key Words : Recurrent Neural Network, Bankruptcy Prediction, Time-Series model

Received : July 31, 2017 Revised : September 12, 2017 Accepted : September 20, 2017

Publication Type : Regular Paper Corresponding Author : Minsoo Shin

저 자 소 개



권혁건

현재 한양대학교 일반대학원 비즈니스인포매틱스학과에 석사과정으로 재학 중이다. 중국 북경대학교에서 경제학 학사 학위를 취득하였다. 주요 연구 관심분야는 데이터마이닝, 정량적마케팅분석 등이다.



이동규

현재 한양대학교 일반대학원 비즈니스인포매틱스학과에 석사과정으로 재학 중이다. 단국대학교에서 수학/통계학 학사 학위를 취득하였다. 주요 연구 관심분야는 딥러닝, 머신러닝, 비즈니스 모델링 등이다.



신민수

현재 한양대학교 경영대학 교수로서 그리고 경영대학 비즈니스인포매틱스 학과장으로 재직 중이다. 고려대학교에서 통계학/경제학사를 취득하였으며 KAIST 경영학과에서 공학 석사 학위를, 영국 University of Cambridge에서 경영정보 시스템 박사학위를 취득하였다. 주요 연구 분야는 IT전략, 정보통신/방송/콘텐츠 정책 및 비즈니스 모델 등이다.