

이진 분류문제에서의 딥러닝 알고리즘의 활용 가능성 평가

김기태

한양대학교 일반대학원 경영학과
(smash1005@hanyang.ac.kr)

이보미

한양대학교 일반대학원
비즈니스인포매틱스학과
(entic@hanyang.ac.kr)

김종우

한양대학교 경영대학 경영학부
(kjiw@hanyang.ac.kr)

최근 알파고의 등장으로 딥러닝 기술에 대한 관심이 고조되고 있다. 딥러닝은 향후 미래의 핵심 기술이 되어 일상생활의 많은 부분을 개선할 것이라는 기대를 받고 있지만, 주요한 성과들이 이미지 인식과 자연어처리 등에 국한되어 있고 전통적인 비즈니스 애널리틱스 문제에서의 활용은 미비한 실정이다. 실제로 딥러닝 기술은 Convolutional Neural Network(CNN), Recurrent Neural Network(RNN), Deep Boltzmann Machine (DBM) 등 알고리즘들의 선택, Dropout 기법의 활용여부, 활성화 함수의 선정 등 다양한 네트워크 설계 이슈들을 가지고 있다. 따라서 비즈니스 문제에서의 딥러닝 알고리즘 활용은 아직 탐구가 필요한 영역으로 남아있으며, 특히 딥러닝을 현실에 적용했을 때 발생할 수 있는 여러 가지 문제들은 미지수이다.

이에 따라 본 연구에서는 다이렉트 마케팅 응답모델, 고객이탈분석, 대출 위험 분석 등의 주요한 분류 문제인 이진분류에 딥러닝을 적용할 수 있을 것인지 그 가능성을 실험을 통해 확인하였다. 실험에는 어느 포르투갈 은행의 텔레마케팅 응답여부에 대한 데이터 집합을 사용하였으며, 전통적인 인공신경망인 Multi-Layer Perceptron, 딥러닝 알고리즘인 CNN과 RNN을 변형한 Long Short-Term Memory, 딥러닝 모형에 많이 활용되는 Dropout 기법 등을 이진 분류 문제에 활용했을 때의 성능을 비교하였다.

실험을 수행한 결과 CNN 알고리즘은 비즈니스 데이터의 이진분류 문제에서도 MLP 모형에 비해 향상된 성능을 보였다. 또한 MLP와 CNN 모두 Dropout을 적용한 모형이 적용하지 않은 모형보다 더 좋은 분류 성능을 보여줌에 따라, Dropout을 적용한 CNN 알고리즘이 이진분류 문제에도 활용될 수 있는 가능성을 확인하였다.

주제어 : 이진분류, 딥러닝, 다층 퍼셉트론, 합성곱 신경망, 장단기 기억

논문접수일 : 2016년 11월 18일 논문수정일 : 2017년 1월 2일 게재확정일 : 2017년 1월 9일
원고유형 : 일반논문 교신저자 : 김종우

1. 서론

2016년 3월, 서울에서 열린 바둑 대국에서 Google의 DeepMind가 만든 바둑 인공지능 프로그램인 알파고(AlphaGo)가 이세돌 9단을 상대로 4대 1의 대승리를 거두었다. 바둑은 기존의 장기나 체스에 비해 계산해야 하는 경우의 수가 우주의 원자수보다 많아 인공지능 분야에서의 큰 도

전과제였는데, 이번 대국을 통해 아직까지 기계가 바둑에서는 사람을 이기지는 못할 것이라 생각한 많은 사람들의 예상을 크게 뒤엎으며 전 세계의 사람들에게 인공지능 기술 연구의 성과를 각인시켰다(Chu et al., 2016). 이 대국 이후 인공지능 기술은 4차 산업혁명의 핵심으로 불리면서 많은 주목을 받게 되었는데, 알파고의 알고리즘에 사용된 여러 인공지능 기술들 중 특히 딥러닝

(Deep Learning) 기법과 그 활용 방안에 대해 기업들뿐만 아니라 일반인들까지도 큰 관심을 보이고 있다.

이 딥러닝 기법의 가장 큰 특징으로는 그 동작 방식이 마치 사람의 뇌와 유사하다는 점을 들 수 있다. 사람의 뇌는 단순한 뉴런들로 구성되어 있으나 수많은 연결들을 통해 복잡한 사고를 하는 것과 같이, 딥러닝 역시 이를 모방해 사람의 뇌를 흉내 낸 가상의 뉴런들을 층층이 쌓아올려 수많은 연결들을 만들고 이를 훈련시키는 것이다. 이 결과, 딥러닝 모형은 마치 인간처럼 데이터로부터 보다 복잡한 개념의 특징들을 찾아낼 수 있게 되었으며, 기존 머신러닝 기법들이 풀지 못하던 어려운 문제들을 풀 수 있게 되었다(Kim, 2016b). 이러한 특징 때문에 뉴런의 층들을 더욱 많이 쌓을수록 딥러닝 모형의 성능은 점점 좋아지고 있으며, 이로 인해 알파고의 대국 이후로 조만간 기계가 인간의 일자리를 넘보게 될 것이라는 전망이 힘을 얻고 있다.

실제로 이러한 딥러닝 기법은 현재 실생활에서 이미 활용이 되고 있는 중이다. 특히 딥러닝은 이미지 인식 분야에서 우수한 성능을 보이고 있는데, 그 활용 예로는 사람의 얼굴을 인식하거나, 교통 신호를 인식하거나, 의료 사진의 세부 분류, 사진 내 사람 인식 등이 있다(LeCun et al., 2015). 그러나 이처럼 딥러닝이 이미지나 음성, 영상, 자연어와 같이 기존에 머신 러닝을 사용해 학습하기 어려웠던 고차원 데이터 영역에서 좋은 성능을 보이며 많은 연구가 이루어지고 있는 것과는 다르게, 기존에 이미 다루고 있던 비즈니스 데이터 및 기타 구조화 되어 있는 데이터들에 대한 딥러닝 활용 연구는 상대적으로 찾아보기가 어렵다.

이에 본 연구는 지금까지 연구된 딥러닝 기법

들이 고차원 데이터의 인식 문제뿐만이 아니라, 허위 거래 식별, 부도 예측, 고객들의 이탈 분석, 마케팅 응답 예측 등 기존의 비즈니스 데이터의 이진 분류 문제에서도 활용이 가능한지를 확인해 보고자 한다. 그리고 그 성능을 측정하여 전통적인 인공신경망 모형과 비교했을 때, 보다 나은 성능을 보이는지를 알아보하고자 한다.

2. 관련 연구

2.1 딥러닝 연구의 방향

딥러닝은 갑자기 나타난 완전히 새로운 이론이 아니라 기존에 인공신경망이 지니던 한계점 중 하나인 심층 구조 학습을 해결하며 심층 신경망을 활용한 것들을 딥러닝이라 칭하면서 등장한 용어이다(Kim et al., 2015; LeCun et al., 2015). 이러한 딥러닝은 1969년에 밝혀진 입력층과 출력층만을 가지는 퍼셉트론(Perceptron)은 간단한 XOR 문제도 해결하지 못한다는 한계를 극복하기 위해, 은닉층을 낀 다층 구조의 퍼셉트론들을 만들면서 시작되었다고 볼 수 있다. 이러한 다층의 퍼셉트론(Multi Layer Perceptron; MLP)은 역전파 알고리즘(Back Propagation Algorithm)의 등장과 맞물려 층수를 늘릴수록 보다 복잡한 경계를 가지는 데이터도 구별해 낼 수 있었지만, 층이 깊어질수록 역전파 알고리즘으로 의미 있는 오류를 잘 전달하지 못하는 ‘사라지는 경사 문제(Vanishing Gradient Problem)’나 ‘과적합(Overfitting)’이 일어나는 문제 등을 가지고 있었다(Jo et al., 2015; Zhang, 2015; Choi and Min, 2015).

이러한 심층 구조 학습의 문제는 Hinton 등

(2006)이 층별로 사전학습(Pretraining)을 통해 문제를 해결한 DBN(Deep Belief Networks)을 제시하며 해결의 가능성을 보였고, 이후 딥러닝의 시대가 열리기 시작하였다(Choi and Min, 2015). 여기에 더해 층이 늘어나면서 많아진 연산량을 코어가 많은 GPU(Graphic Processing Unit)를 활용해 학습시킬 수 있게 되면서 기존 대비 몇 십 배 빠른 학습이 가능해져 본격적으로 연구에 박차를 가하게 되었다(LeCun et al., 2015).

이러한 딥러닝이 본격적으로 주목을 받기 시작한 것은 2012년 국제 대규모 이미지 인식 기술 대회인 ILSVRC(Imagenet Large Scale Visual Recognition Challenge)에서 처음으로 출전한 Hinton 교수의 토론토대학 팀이 CNN(Convolutional Neural Networks)을 사용한 딥러닝 모형을 만들어 출전한 결과, 2위 팀보다 10% 가까운 인식률 차이를 보이면서 우승하면서이다(Krizhevsky et al., 2012; Matsuo, 2015). 또한 Hinton 교수는 2013년에 미국의 DARPA(Defense Advanced Research Projects Agency)의 음성 데이터인 TIMIT 데이터를 다층의 LSTM(Long Short-Term Memory)을 사용하여 인식한 결과, 기존 알고리즘들을 제치고 인식률 80%의 벽을 깨며 음성 인식 영역에서도 딥러닝의 활용 가능성을 다시 알렸다(Choi and Min, 2015; Graves et al., 2013). 이러한 사건들로 인해 각종 이미지 인식 대회에서는 Google, Microsoft 등 세계적인 IT 대기업들이 보다 깊은 딥러닝 모형을 만들어 내면서 점차 그 성능을 높여 나가고 있으며, 딥러닝 구조의 개선과 새로운 알고리즘 개발을 통해 동영상 분류나 이미지-텍스트 멀티모달 등 여러 부분으로 활용 범위를 넓혀나가고 있다(Kim et al., 2015).

2.2 딥러닝의 주요 알고리즘

딥러닝의 모형은 크게 샘플 데이터를 생성할 수 있는 생성 모형(Generative Model)과 주어진 데이터를 구별할 수 있는 판별 모형(Discriminative Model)으로 구분할 수 있다. 이 중 본 연구는 비즈니스 데이터의 이진 분류 문제를 푸는 것을 목적으로 하기에 판별 모형에 사용되는 주요 알고리즘에 대하여 설명한다.

현재 판별 모형에 사용되는 가장 유명한 알고리즘으로는 이미지나 영상 처리에서 강점을 보이는 CNN(Convolutional Neural Network)을 들 수 있다. CNN은 Fukushima(1980)의 연구인 'Neocognitron(신인식기)'를 기반으로 하는데, 이 연구는 손으로 작성한 문자들의 인식을 위한 인식기를 만드는 것으로 각 문자들의 특징을 학습하기 위해 문자의 조그만 부분들을 인식하는 셀(Cell) 부분과 인식한 셀 부분들을 다시 인식하는 셀들의 복층 구조로 이루어져 있다(Kim, 2016a). 추후 LeCun 등(1989)은 이 특징 추출 구조에 역전파 알고리즘을 더하여 우편 번호를 자동으로 인식하는 3층의 인공신경망 인식기를 만들며 현재 사용되는 CNN의 근원을 완성하였다. 이 CNN은 미세한 특징들을 잡아내는 부분적인 신경망들의 가중치를 공유하는 방식으로 기존의 완전 연결 신경망(Fully Connected Neural Network) 방식보다 구조를 단순화 할 수 있었기에 사전학습이 없어도 역전파 알고리즘을 통한 다층 구조 학습이 비교적 용이하였다(Kim et al., 2015; LeCun et al., 2015). 이렇게 다층 구조에 강점을 지니고 있던 CNN은 딥러닝 연구가 본격적으로 활성화되면서 수십 층으로 이루어진 깊은 신경망 모형으로 만들어졌고, CNN의 장점인 특징들의 위치 변화에 강하다는 특성이 더해져 이미

지 인식 부분에서 좋은 성능을 보이고 있다(Kim, 2010; Krizhevsky et al., 2012). 현재 보다 깊고도 효과적인 CNN 신경망을 만들기 위한 연구가 계속되고 있다.

다음으로 음성과 같은 시계열 데이터를 판별하는데 있어서는 RNN(Recurrent Neural Networks)를 기반으로 한 알고리즘들이 주로 사용되고 있다. 기본적인 RNN은 다른 신경망 알고리즘들과는 다르게 자신의 출력을 다음 층에 전달하는 것과 함께 자기 자신에게도 전달하는 재귀 구조를 가지고 있다. 이런 재귀 구조는 메모리와 같은 역할을 수행해 시간의 흐름에 따른 변화를 모델링 할 수 있게 한다(Choi and Min, 2015). 그러나 RNN은 입력하는 데이터의 길이가 길어질수록 먼저 입력한 데이터의 정보가 점차 잊히는 문제를 가지고 있다. 이러한 문제를 개선하기 위하여 Hochreiter와 Jurgen(1997)은 RNN의 입력과 출력 부분에 Gate라는 셀들을 붙여 값을 조절하는 LSTM(Long Short-Term Memory)의 원형 알고리즘을 만들어 이를 해결하였다. 이후 LSTM은 추후 Forget Gate 등이 추가되는 등 여러 연구자들에 의해 개선이 되었고, 현재에는 개선된 LSTM이나 GRU(Gate Recurrent Unit)로 불리는 LSTM의 변형 모형 등이 주로 사용되고 있다(Ahn, 2016; Cho et al., 2014).

이러한 학습 알고리즘 외에도 학습 속도 향상이나 다층 구조 학습의 문제점들을 해결하는데 기여한 주요 연구들이 있는데, 이의 대표적인 예로는 ReLU(Rectified Linear Units)라는 활성화 함수를 들 수 있다. 이 ReLU 활성화 함수를 식으로 나타내자면 다음 <Formula 1>과 같다.

$$f(x) = \max(0, x) \quad (1)$$

ReLU 활성화 함수를 간단하게 살펴보면, 입력 값이 음수면 0으로, 양수면 입력 값을 그대로 출력해 주는 함수이다. 이는 기존 인공신경망 연구에서 주로 사용되던 Sigmoid 활성화 함수에서 일어나던 오류 역전파시 경사도가 사라지는 문제를 줄일 수 있으며, 결과에 보다 빠르게 수렴하는 등의 장점이 있어 널리 사용되고 있다(Kim, 2016b; Zhang, 2015).

또한 강건한 학습을 위한 방안으로 Dropout이라는 기법을 들어볼 수 있는데, 이는 학습 도중에 뉴런의 일정 비율을 임의로 제거하면서 학습을 시키는 기법이다. 이 기법을 사용하면 비슷한 가중치를 가지고 있는 뉴런들이 계속하여 하나의 뉴런처럼 같이 움직이는 상호적응(co-adaptation) 현상을 막는 것과 함께 과적합을 막는데 효과가 있다(Choi and Min, 2015; Srivastava et al., 2014).

3. 실험 설계

3.1 실험 데이터

본 연구에서는 비즈니스에 관련된 이진 분류 문제에 딥러닝 알고리즘을 적용하기 위하여 2008년 5월부터 2010년 11월 사이 한 포르투갈 은행이 고객들을 대상으로 한 텔레마케팅 응답 데이터인 ‘Bank Marketing’ 데이터¹⁾를 사용하였다. ‘Bank Marketing’ 데이터는 총 45,211개의 레코드를 가지고 있으며, 입력 변수는 고객 연령, 직업, 대출 여부, 과거 텔레마케팅의 횟수 등 16개이다. 목표 변수는 고객이 계좌를 개설할 의사가 있는지를 ‘yes’ 와 ‘no’로 표기한 ‘y’ 변수이

1) <http://archive.ics.uci.edu/ml/datasets/Bank+Marketing>

다. 이 ‘y’ 변수는 긍정 응답이 5,289건, 부정 응답이 39,922건으로, 긍정 응답을 한 데이터의 비율이 전체의 약 11.70%를 차지하고 있다.

본 연구에서는 0과 1 사이의 값만 입력받을 수 있는 인공신경망 모형들의 특성상, 목록형 값들은 더미 변수들을 작성하고, 연속형 값들은 0과 1 사이로 변환하는 전처리 과정을 수행하였다. 이후 작성한 모형의 성능 평가를 위해 랜덤하게 30%의 데이터를 분리하여 테스트 데이터로 사용하고 나머지 70%의 데이터를 훈련 데이터로 사용하였다. 마지막으로 불균형 학습을 막기 위해 전처리 과정에서 훈련용 데이터에 대해 과대표본화(Oversampling)를 하여 두 응답의 데이터 건수를 균등하게 맞추었다.

3.2 실험 방법

본 연구는 현재 딥러닝 연구에서 널리 쓰이고 있는 CNN과 LSTM 알고리즘 및 Dropout 기법을

이미지나 음성 데이터가 아닌 비즈니스 데이터의 이진 분류 문제에 적용해 보고, 기존에 이진 분류 문제에 활용하던 MLP 모형과 그 성능을 비교하여 딥러닝 기법의 활용 가능성을 실험을 통해 알아보려고 하였다.

그러나 인공신경망의 특성상 수많은 조건 값들이 있어 모든 조건에 대해 실험할 수 없기에, 본 실험에서는 은닉층의 수, 은닉층 내 뉴런의 수 또는 출력되는 데이터(필터)의 개수, Dropout 기법의 적용 유무 조건을 기준으로 모형들 간의 성능 비교를 실시하였다. 다만 CNN과 LSTM의 경우 모형 생성에 다소 제약이 있어 일부 모형들이 제외되었다. 그 외의 실험 조건으로 활성화함수는 ReLU를 기본으로 하되, ReLU 활성화함수를 사용할 수 없는 LSTM 알고리즘 모형들은 Hyperbolic tangent(tanh)와 Sigmoid 함수로 대체하였다. 그 외에 이 외에 학습률, 모멘텀 등의 조건들은 모두 동일한 상태에서 진행하였다. 이를 요약하자면 아래의 <Table 1>과 같다.

<Table 1> Summary of models used in the experiment

Type of Model	Number of hidden layers	Number of neurons in the hidden layers	Number of output data(filter)	Dropout	Activation function
MLP	1	16	-	X	ReLU
		32	-	X	ReLU
	2	16	-	X	ReLU
		32	-	X	ReLU
MLP with Dropout	1	16	-	O	ReLU
		32	-	O	ReLU
	2	16	-	O	ReLU
		32	-	O	ReLU
CNN	2	-	16	X	ReLU
		-	32	X	ReLU
CNN with Dropout	2	-	16	O	ReLU
		-	32	O	ReLU
LSTM	1	-	16	X	tanh, Sigmoid
		-	32	X	tanh, Sigmoid
	2	-	16	X	tanh, Sigmoid
		-	32	X	tanh, Sigmoid

각 딥러닝 기법들을 모형에 적용시킨 자세한 방법은 다음과 같다. CNN 알고리즘은 본래 이미지 데이터에서 특정 값으로부터 인접한 값들의 특징들을 인식하고 이를 쌓아 올려 다양한 특징을 인식하는 것이 기본이나, 비즈니스 데이터는 각각의 필드들이 독립적인 값들로 얼마나 가까운 곳에 위치하는지는 의미가 없기에 기존 방식대로 특징을 인식 시키는 것에 문제가 있었다. 이에 본 실험에서는 CNN 알고리즘의 필터 크기를 한 레코드의 전체 필드 길이만큼 설정하여 데이터의 전체 특징을 한 번에 학습하도록 하였다. 이후 추출한 특징들을 바탕으로 모형이 의사결정을 할 수 있도록 은닉층을 하나 추가하여 모형을 완성하였다. 이를 도식화 하여 나타내면 아래 <Figure 1>과 같다.

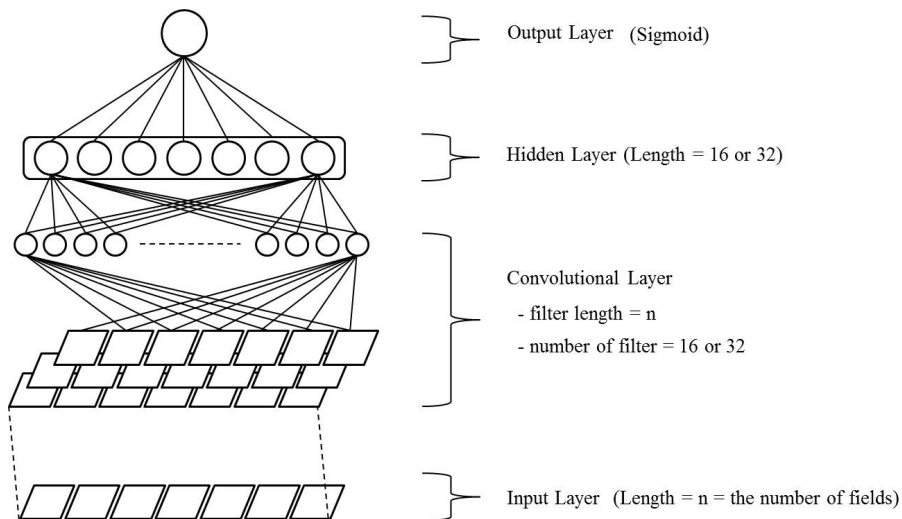
다음으로 LSTM 알고리즘의 경우 데이터를 적용하는데 큰 문제가 없었으나, 은닉층을 두 층으로 쌓은 모형의 경우 각 값들의 위치에 따른 영향을 줄여보고자 두 번째 층의 입력은 역방향으

로 받도록 설정하였다. 이를 도식화 하여 나타내면 <Figure 2>와 같다.

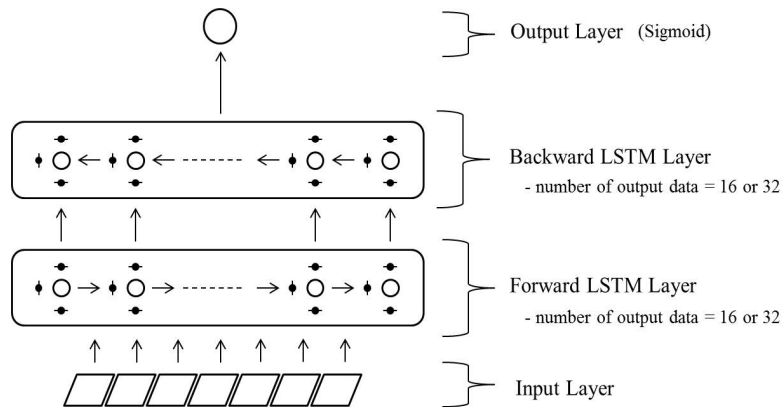
마지막으로 Dropout 기법의 경우 각 모형들의 층마다 적용하였으며, 0.5의 확률로 뉴런이 사라지도록 설정하였다.

실험을 통해 산출한 모형들의 성능 비교에는 재현율(Recall) 값과 정밀도(Precision) 값의 조화 평균인 F1 Score를 사용하였다. 이는 기존 이미지 인식 문제에서는 성능 평가 기준이 단순히 정확도였던 것에 반해, 이진 분류 문제에서는 소수 집단인 목표 변수를 얼마나 잘 맞추는지가 중요하기 때문이다. 이에 본 연구에서는 이러한 이진 분류 문제에서 평가 척도로 자주 쓰이고 있는 F1 Score를 평가 기준으로 사용하였다. F1 Score를 구하는 식은 다음 <Formula 2>와 같다.

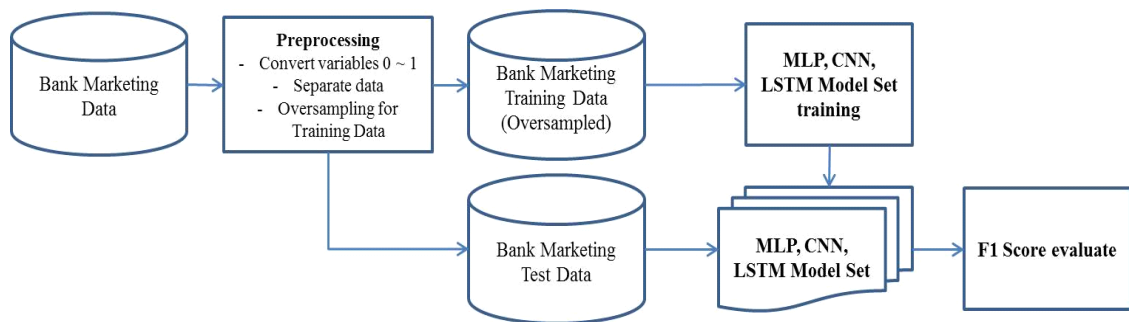
$$F1\ Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (2)$$



<Figure 1> Schematic diagram of CNN model used in experiment



<Figure 2> Schematic diagram of LSTM model used in experiment



<Figure 3> Schematic diagram of experiment progress

지금까지의 실험 과정들을 간단히 도식화하여 표현하면 위의 <Figure 3>과 같다.

4. 실험 설계

이전 장에서 설계한 실험을 수행한 결과를 정리한 결과는 다음 <Table 2>와 같다. 가장 높은 F1 Score를 보인 예측 모형 집합은 Dropout 기법을 적용한 CNN 모형으로, 32개의 필터를 가지도록 설정했을 때 제일 좋은 0.4283의 값을 보였다.

다음으로 좋은 성능을 보인 예측 모형 유형은 Dropout 기법을 적용한 MLP 모형으로, 32개의 뉴런으로 구성된 은닉층을 두 층으로 구성했을 때 0.4231의 값을 보였다. 이 외에 각 모형 유형 별로 제일 높은 F1 Score 값은 LSTM 모형 집합이 0.4072, CNN 모형 집합이 0.3975, 그리고 일반적인 MLP 모형 집합이 0.3923이었다.

본 연구에서는 실험을 진행하면서 몇 가지 특이 사항들을 발견할 수 있었다. 우선 첫 번째로는 Dropout 기법을 적용한 MLP, CNN 모형 집합들이 Dropout 기법을 적용하지 않은 MLP, CNN

〈Table 2〉 F1 score, precision and recall values of MLP, CNN, and LSTM models

Type of Model	Number of hidden layer	Number of neurons or output data (filters)	F1 Score	Precision	Recall
MLP	1	16	0.3788	0.2731	0.6184
		32	0.3923	0.2953	0.5841
	2	16	0.3650	0.2586	0.6203
		32	0.3551	0.2458	0.6391
MLP with Dropout	1	16	0.4070	0.3138	0.5789
		32	0.4034	0.3124	0.5692
	2	16	0.3929	0.2939	0.5925
		32	0.4231	0.3388	0.5634
CNN	2	16	0.3975	0.3046	0.5718
		32	0.3876	0.2940	0.5686
CNN with Dropout	2	16	0.4150	0.3300	0.5589
		32	0.4283	0.3566	0.5362
LSTM	1	16	0.3423	0.2298	0.6701
		32	0.3545	0.2435	0.6514
	2	16	0.3614	0.2582	0.6022
		32	0.4072	0.3184	0.5647

모형 집합들에 비해 전체적으로 더 나은 분류 성능을 보였다는 것이다. 실험 결과를 보면, Dropout 기법을 적용한 모형들은 그렇지 않은 모형들에 비해 전반적으로 재현율은 떨어졌지만 대신 정밀도가 올라가면서 F1 Score가 상승한 것을 확인할 수 있었다. 이는 Dropout 기법을 사용하였을 경우 모형들이 조금 더 보수적인 예측을 하였다는 것으로 볼 수 있는데, Dropout 기법이 이미지나 음성 인식 문제 외에 이진 분류 문제의 모형 성능 개선에도 도움을 줄 수 있을 것이란 가능성을 확인할 수 있었다.

두 번째로는 MLP 모형 집단들보다 CNN 모형 집단들이 전반적으로 더 좋은 분류 성능을 보였다는 것이다. CNN 알고리즘이 이미지 인식에서

다른 알고리즘보다 좋은 성능을 보여준다는 것은 널리 알려져 있었으나, 입력하는 데이터의 특징이 다른 이진 분류 문제에서도 나은 성능을 보였다는 것은 흥미로운 점으로 보인다. 추후 다른 데이터를 활용하여 CNN 알고리즘을 사용한 이진 분류 문제를 진행해 추가적인 확인을 해 볼 필요가 있다고 보인다.

세 번째로는 LSTM 알고리즘의 경우 다른 알고리즘들과의 성능 차이에 비해 연산 소요 시간이 너무나 오래 걸린다는 점이였다. Dropout 기법을 적용하지 않고 은닉층을 두 층으로 설정하였을 때, 한 훈련 횟수 당 MLP 모형이 약 22초, CNN 모형이 약 26초가 소요된 반면에 LSTM 모형은 약 24분가량이 소요되었다. 게다가 은닉층

의 수를 늘릴수록 소요 시간이 늘어난다는 점을 고려해 볼 때, 보여준 성능에 비해 LSTM 알고리즘을 이대로 사용하는 것은 이진 분류 문제를 푸는데 적합하지 않다고 보았다.

또한, 본 연구에서는 거론하지 않았으나 ‘Bank Marketing’ 데이터 이외에 크기가 다른 데이터 집합들을 같은 조건 하에서 실험한 결과 레코드 수가 적은 데이터에서는 CNN 모형 집단이 MLP 모형 집단에 비해 더 개선된 성능을 보여주지 못한다는 점을 발견하였다. 예를 들어, 레코드 수가 71,047개인 데이터 집합을 사용한 실험에서 CNN 모형의 F1 Score는 0.4330, MLP 모형의 F1 Score는 0.4029로 CNN 알고리즘이 향상된 성능을 보여주었으나 레코드 수가 1,477개인 데이터 집합에서 CNN 모형의 F1 Score는 0.6650, MLP 모형은 0.6733으로 보다 오히려 미세하게 저하된 성능을 보였다. 이어진 실험에서 10,000개의 레코드를 가진 데이터 집합은 CNN 모형에서 0.1531, MLP 모형에서 0.1594의 F1 Score를 보여 레코드 수가 적은 데이터 집합에서 CNN 알고리즘을 적용하는 것은 MLP에 비해 성능 향상을 기대할 수 없다는 결론을 도출하였다.

5. 결론

본 연구는 이미지나 음성 인식 분야에서 좋은 성능을 보이고 있는 딥러닝 기법을 전통적인 문제였던 이진 분류 문제에서도 활용이 가능한지를 알아보고자 실험을 진행하였다. 그 결과 Dropout 기법은 MLP나 CNN 모형들을 조금 더 보수적으로 훈련시켜 모형들이 좀 더 나은 이진 분류 성능을 갖는데 도움을 줄 수 있다는 가능성을 보여주었으며, Dropout 기법을 적용한 CNN

모형이 가장 좋은 F1 Score를 보였다는 점에서 추후 CNN 알고리즘을 이진 분류 문제에 활용해 볼 가치가 있다는 것을 확인할 수 있었다. 그리고 LSTM 알고리즘의 경우 다른 알고리즘 대비 모형의 훈련 시간이 지나치게 길면서 성능 차이가 크지 않아 이진 분류 문제에 활용하기에 적합하지 않음을 확인하였다. 이러한 결과들에서 미뤄볼 때, 딥러닝 알고리즘들 중 일부는 인공지능망을 사용하여 이진 분류 문제를 푸는데 도움을 줄 가능성이 있다고 판단하였다.

다만 본 연구는 인공지능망의 특성상 모형 훈련 과정에서의 수많은 변수들을 모두 통제할 수 없었기에, 변수 설정 혹은 실험 환경에 따라 다른 결과가 나올 가능성이 존재한다. 또한 이미지나 음성 데이터와는 다르게 비즈니스 데이터는 그 구성이 데이터마다 모두 다르기에, 연구 결과를 일반화하기 위해서는 보다 많은 데이터들을 사용해 검증을 해 볼 필요가 있다. 그리고 본 실험에서는 CNN 알고리즘을 적용하기 위해 필터의 크기를 한 레코드의 전체 필드 길이만큼 설정하였는데, 이는 CNN 알고리즘의 강점이었던 깊은 구조를 구현하는데 문제가 된다. 이에 CNN 알고리즘을 어떻게 하면 비즈니스 데이터에 적합하게 변형할 수 있을지도 추후 연구를 해야 할 것이다. 마지막으로 본 실험에서 사용된 알고리즘들 외에도 GRU, Batch-Normalization, Maxout 등 기존 딥러닝 기법들을 개선한 새로운 기법들이 계속하여 추가가 되고 있는데, 이들 역시 검증이 가능한지에 대해 확인해 볼 필요가 있겠다. 차후 후속 연구들에서는 이러한 한계점들을 어떻게 보완할 수 있을지 고려하여 연구를 진행할 예정이다.

참고문헌(References)

- Ahn, S. M., "Deep learning architectures and applications," *Journal of Intelligence and Information*, Vol. 22, No. 2(2016), 127-142.
- Cho, K., B. Van Merriënboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder-decoder approaches," *arXiv preprint arXiv:1409.1259*, 2014.
- Choi, H. Y., and Y. H. Min, "Introduction to deep learning and major issues[written in Korean]," *Korea Information Processing Society Review*, Vol. 22, No. 1(2015), 1-15.
- Chu, H. S., S. W. Ahn, and S. W. Kim, *AlphaGo's artificial intelligence algorithm analysis* [written in Korean], Software Policy & Research Institute, 2016.
- Fukushima, K., "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biological cybernetics*, Vol. 36, No. 4(1980), 193-202.
- Graves, A., A. R. Mohamed, and G. E. Hinton, "Speech recognition with deep recurrent neural networks," *2013 IEEE international conference on acoustics, speech and signal processing*, (2013), 6645-6649.
- Hinton, G. E., S. Osinder, and Y. W. Teh, "A fast learning algorithm for deep belief nets," *Neural computation*, Vol. 18, No. 7(2006), 1527-1554.
- Hochreiter, S., and S. Jurgens, "Long short-term memory," *Neural computation*, Vol. 9, No. 8(1997), 1735-1780.
- Jo, N. O., H. J. Kim, and K. S. Shin, "Bankruptcy type prediction using a hybrid artificial neural networks model," *Journal of Intelligence and Information*, Vol. 21, No. 3(2015), 79-99.
- Kim, H. J., "Dynamic hand gesture recognition using CNN model and FMM neural networks," *Journal of Intelligence and Information*, Vol. 16, No. 2(2010), 95-108.
- Kim, J. W., H. A. Pyo, J. W. Ha, C. K. Lee, and J. H. Lee, "Deep learning algorithms and applications," *Communications of the Korean Institute of Information Scientists and Engineers*, Vol. 33, No. 8(2015), 25-31.
- Kim, K. T., "Purchase prediction through clickstream data of internet store based on deep learning technique," *Master's Thesis*, Graduate School, Hanyang University, 2016a.
- Kim, U. J., *Introduction to artificial intelligence, machine learning, and deep learning with algorithms*[written in Korean], wikibook, Paju, Republic of Korea, 2016b.
- Krizhevsky, A., I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolution neural networks," *Advances in neural information processing systems*, Vol. 25(2013), 1097-1105.
- LeCun, Y., B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural computation*, Vol. 1, No. 4(1989), 541-551.
- LeCun, Y., Y. Bengio, and G. Hinton, "Deep learning," *Nature*, Vol. 521, No. 7553(2015), 436-444.
- Matsuo, Y., *Artificial intelligence and deep learning*(Translated by Park, K. W.)[written in Korean], Donga M&B, Seoul, Republic of Korea, 2015.

Srivastava, N., G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, Vol. 15, No. 1(2014), 1929-1958.

Zhang, B. T., "Deep hypernetwork models," *Communications of the Korean Institute of Information Scientists and Engineers*, Vol. 33, No. 8(2015), 11-24.

Abstract

Feasibility of Deep Learning Algorithms for Binary Classification Problems

Kitae Kim* · Bomi Lee* · Jong Woo Kim**

Recently, AlphaGo which is Bakuk (Go) artificial intelligence program by Google DeepMind, had a huge victory against Lee Sedol. Many people thought that machines would not be able to win a man in Go games because the number of paths to make a one move is more than the number of atoms in the universe unlike chess, but the result was the opposite to what people predicted. After the match, artificial intelligence technology was focused as a core technology of the fourth industrial revolution and attracted attentions from various application domains. Especially, deep learning technique have been attracted as a core artificial intelligence technology used in the AlphaGo algorithm.

The deep learning technique is already being applied to many problems. Especially, it shows good performance in image recognition field. In addition, it shows good performance in high dimensional data area such as voice, image and natural language, which was difficult to get good performance using existing machine learning techniques. However, in contrast, it is difficult to find deep leaning researches on traditional business data and structured data analysis. In this study, we tried to find out whether the deep learning techniques have been studied so far can be used not only for the recognition of high dimensional data but also for the binary classification problem of traditional business data analysis such as customer churn analysis, marketing response prediction, and default prediction. And we compare the performance of the deep learning techniques with that of traditional artificial neural network models.

The experimental data in the paper is the telemarketing response data of a bank in Portugal. It has input variables such as age, occupation, loan status, and the number of previous telemarketing and has a binary target variable that records whether the customer intends to open an account or not. In this study, to evaluate the possibility of utilization of deep learning algorithms and techniques in binary classification problem, we compared the performance of various models using CNN, LSTM algorithm and dropout, which

* Graduate School, Hanyang University

** Corresponding Author: Jong Woo Kim

School of Business, Hanyang University

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

Tel: +82-2-2220-1067, Fax: +82-2-2220-1169, E-mail: kjw@hanyang.ac.kr

are widely used algorithms and techniques in deep learning, with that of MLP models which is a traditional artificial neural network model. However, since all the network design alternatives can not be tested due to the nature of the artificial neural network, the experiment was conducted based on restricted settings on the number of hidden layers, the number of neurons in the hidden layer, the number of output data (filters), and the application conditions of the dropout technique. The F1 Score was used to evaluate the performance of models to show how well the models work to classify the interesting class instead of the overall accuracy.

The detail methods for applying each deep learning technique in the experiment is as follows. The CNN algorithm is a method that reads adjacent values from a specific value and recognizes the features, but it does not matter how close the distance of each business data field is because each field is usually independent. In this experiment, we set the filter size of the CNN algorithm as the number of fields to learn the whole characteristics of the data at once, and added a hidden layer to make decision based on the additional features. For the model having two LSTM layers, the input direction of the second layer is put in reversed position with first layer in order to reduce the influence from the position of each field. In the case of the dropout technique, we set the neurons to disappear with a probability of 0.5 for each hidden layer.

The experimental results show that the predicted model with the highest F1 score was the CNN model using the dropout technique, and the next best model was the MLP model with two hidden layers using the dropout technique. In this study, we were able to get some findings as the experiment had proceeded. First, models using dropout techniques have a slightly more conservative prediction than those without dropout techniques, and it generally shows better performance in classification. Second, CNN models show better classification performance than MLP models. This is interesting because it has shown good performance in binary classification problems which it rarely have been applied to, as well as in the fields where it's effectiveness has been proven. Third, the LSTM algorithm seems to be unsuitable for binary classification problems because the training time is too long compared to the performance improvement. From these results, we can confirm that some of the deep learning algorithms can be applied to solve business binary classification problems.

Key Words : Binary Classification, Deep Learning, Multi-Layer Perceptron, Convolutional Neural Network, Long Short-Term Memory

Received : November 18, 2016 Revised : January 2, 2017 Accepted : January 9, 2017

Publication Type : Regular Paper Corresponding Author : Jong Woo Kim

저 자 소개



김기태

한양대학교에서 경영학사와 정보시스템학사를 취득하였으며, 한양대학교 일반대학원에서 경영정보시스템을 전공하여 경영학 석사학위를 취득하였다. 주요 연구 관심분야는 데이터마이닝, 기계학습 및 딥러닝 기법의 응용, 빅 데이터, 클라우드 컴퓨팅 서비스 등이다.



이보미

현재 한양대학교 일반대학원 비즈니스인포매틱스학과에 석사과정으로 재학 중이다. 한양대학교에서 경영학사를 취득하였다. 주요 연구 관심분야는 딥러닝, 감성 분석, 데이터마이닝 등이다.



김종우

현재 한양대학교 경영대학 경영학부 교수로 재직 중이다. 서울대학교 수학과에서 학사를 마쳤으며, 한국과학기술원에서 경영과학으로 석사학위를, 산업경영학으로 박사학위를 취득하였다. 주요 연구 관심분야는 데이터마이닝 기법과 응용, 기계학습과 딥러닝, 오피니언 마이닝, 상품추천기술, 지능형 정보시스템, 집단지성, 사회 네트워크 분석, 클라우드 컴퓨팅 서비스 등이다.