# R기반의 딥 러닝을 이용한 데이터 예측 프로세스에 관한 연구

정세훈* · 김종찬** · 박홍준*** · 소원호*** · 심춘보*

*순천대학교 멀티미디어공학과, **순천대학교 컴퓨터공학과, ***순천대학교 컴퓨터교육과

# A novel on Data Prediction Process using Deep Learning based on R

Se-hoon Jung* · Jong-chan Kim** · Hong-joon Park*** · Won-ho So*** · Chun-bo Sim*

*Dept. of Multimedia Engineering, Sunchon National University

**Dept. of Computer Engineering, Sunchon National University

***Dept. of Computer Education, Sunchon National University

E-mail : iam1710@hanmail.net, cbsim@sunchon.ac.kr

## 요 약

최근 신경망 분석의 향상된 성능을 보여주는 심화 신경망 기술인 딥 러닝(Deep learning)이 각광을 받고 있는 실정이다. 이에 본 논문에서는 딥 러닝을 기반으로 분석 시각화 툴인 R을 이용한 특정 변수의 오류율 검증과 빅 데이터 예측 프로세스 설계를 제안한다. 딥 러닝에 적용된 알고리즘은 RBM(Restricted Boltzmann Machine)을 적용하였다. 특정 입력 변수에 대한 종속 변수 구분 후 각 종속 변수의 가중치를 적용한다. RBM 알고리즘을 통해 최종 데이터의 검증 및 오류율 검출과정을 R 프로그래밍에 적용하여 설계한다.

## ABSTRACT

Deep learning, a deepen neural network technology that demonstrates the enhanced performance of neural network analysis, has been getting the spotlight in recent years. The present study proposed a process to test the error rates of certain variables and predict big data by using R, a analysis visualization tool based on deep learning, applying the RBM(Restricted Boltzmann Machine) algorithm to deep learning. The weighted value of each dependent variable was also applied after the classification of dependent variables. The investigator tested input data with the RBM algorithm and designed a process to detect error rates with the application of R.

## 키워드

Deep Learning, Neural Network, R, Prediction Process, RBM

## Ⅰ. 서 론

지난 4월 개봉한 영화 “어벤져스: 에이지 오브 울트론”에서 스스로 학습하며 진화하는 주인공의 인공지능(Artificial Intelligence, AI) 학습 소재가 관심을 받고 있다. 영화에서 나타나는 인공지능의 학습법이 딥 러닝(Deep Learning)이다. 딥 러닝은 현실의 사물이나 각종 빅 데이터를 분류하거나 분석하는데 사용하는 일종의 기술적인 방법론이며, 이를 신경망을 통해 구현하는 기술[1-2]이다. 딥 러닝은 인공신경망(Artificial Neural Network)의 단점을 극복하기 위하여 제안된 방법이며, 인공신경망은 분류(Classification) 정확도에 비해 속도가 느리며, 과적합(Overfitting) 상황에도 단점을 보유하고 있었다. 그러나 Geoffery Hinton[3]는 인공신경망의 단점들을 해결하기 위하여 신경망 층의 깊은 곳까지 학습할 수 있는 RBM(Restrict Boltzmann Machine)을 이용하여 분석 데이터를 전처리학습(Pretraining)을 처리하고 순차적으로 학습하는 계층구조의 학습방식을 개발하였다. 현재 이러한 딥 러닝에 적용된 알고리즘을 통해 애플의 시리(Siri), 구글의 나우등과 같은 음성인식에 적용되고 있으며, 인공지능과 빅 데이터 분석 기술에 적용되는 범위가 확대되고 있는 실정이다. 이에 본 논문에서는 딥 러닝을 활용하여 분석 툴인 R[5]기반의 변수 오류율 검증 및 빅 데이터 예측 프로세스 설계를 제시한다.

## II. 관련 연구

딥 러닝[2-4]은 정형 및 비정형 데이터를 군집화하거나 분류하는데 사용되는 일종의 신경망 기술이며, 인공신경망의 단점을 보완하기 위하여 제안된 기계학습법의 일종이다. R[5]은 S 언어에서 확장된 통계 및 분석학적 프로그램 언어로 데이터마이닝, 통계 등을 포함한 분석 프로그램을 수행한다. R은 다양한 패키지를 제공하며 사용자는 패키지를 활용하여 고급 분석 및 분산, 병렬 처리 등 다양한 분석 환경을 구성할 수 있다.

## III. 딥 러닝 기반의 데이터 예측 프로세스

그림 1은 데이터 예측 프로세스의 구성도이며, 제안하는 프로세스는 분류 및 분석을 위한 특정 데이터 집단의 표본값을 분석 클래스로 범위를 규정하고 각 분석 클래스 집단을 분석하기 위한 전처리 과정으로 각 집단의 분류 가중치를 적용하는 샘플링 과정을 적용한다. 분류된 가중치 데이터를 기반으로 RBM 알고리즘[3]을 적용한다.

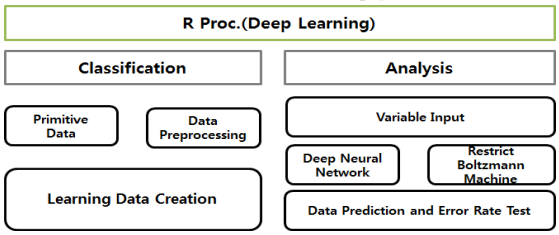


그림 1. 데이터 예측 프로세스 구성도

3.1 데이터 분류를 위한 전처리 설계

데이터 예측을 위한 분류 프로세스는 데이터 가중치 분류를 위하여 원시 데이터로부터 데이터셋(클래스)을 추출한다. 데이터 예측을 위한 데이터의 전처리 과정은 데이터의 손실을 막고 통합적인 메타데이터 분류를 위해 임시 데이터베이스에 임시로 저장하게 된다. 각 클래스의 독립 변수를 선택하여 샘플링을 위한 변수의 스레스홀드를 적용하여 학습 데이터를 생성한다.

3.2 RBM 알고리즘을 적용한 딥 러닝 설계

생성된 학습 데이터는 입력변수로 규정하고 RBM(Restrict Boltzmann Machine) 알고리즘을 적용한다. 또한 딥 러닝의 성능 확인을 위하여 심층 신경망을 적용하여 예측 데이터 오류율을 확인한다. RBM 알고리즘은 DNN(Deep Neural Network) 알고리즘의 단점인 학습 은닉 계층의 과적응(Overfitting)의 문제를 해결하기 위하여 층간의 연결을 삭제한 형태의 모델이다. 은닉층과의 연결은 무방향 이분 그래프 형태로 구성되어 있다. RBM 알고리즘의 훈련 과정 가중치 갱신은 경사 하강법을 기반으로 하고 있으며, 식 1과 같다. p(v)는 입력 변수 벡터의 확률을 의미한다.

$$\Delta w_{ij}(t+1) = w_{ij}(t) + \eta \frac{\partial \log(p(v))}{\partial w_{ij}} \quad (1)$$

식 2의 Z는 입력 변수의 정규화를 위한 분배함수이다. E(v, h)는 딥 러닝의 상태 에너지함수이고, 에너지 함수가 낮을수록 신경망 분석값이 더 적합한 상태임을 나타낸다.

$$p(v) = \frac{1}{Z}\sum_{h} e^{-E(v,h)} Z \quad (2)$$

3.3 R을 적용한 데이터 예측 프로세스 설계

데이터 예측 프로세스 성능 확인 및 구현을 위한 툴은 시각화의 장점을 포함한 R을 활용하였다. R의 *deepnet*[6]의 라이브러리를 활용하여 딥 러닝의 RBM 알고리즘을 프로세스에 적용하였다. 그림 2는 RBM 라이브러리를 활용한 데이터 예측 프로세스의 학습 데이터 훈련관련 코드이다.

```
rbm.train(primitive data, hidden=2, numepochs = 3, batchsize = 100, learningrate = 0.8, learningrate_scale = 1, momentum = 0.5, visible_type = "bin", hidden_type = "bin", cd = 2)
```

그림 2. 딥 러닝의 학습 데이터 훈련 코드(R)

## IV. 결 론

본 논문에서는 딥 러닝 및 R기반의 데이터 예측 프로세스의 설계를 제안하였다. RBM 알고리즘을 통해 은닉층을 활용하지 않고 무방향 이분 그래프 형태를 활용하였다. 향후 각종 데이터의 예측 프로세스 적용 및 데이터 오류 검증을 통해 예측 데이터 신뢰성을 높이고자 한다.

## 참고문헌

[1] T. Sainath et al., "Convolutional neural networks for LVCSR," ICASSP, 2013.

[2] T. Mikolov et al., "Recurrent neural network based language model," Interspeech, 2010.

[3] G. E. Hinton., "Learning multiple layers of representation," Trends in Cognitive Sciences, 11, pp.428–434, 2007.

[4] R. Raina, A. Madhavan, A. Y. Ng,"Large-scale deep unsupervised learning using graphics processors," ICML, pp.873-880, 2009.

[5] R, "http://www.r-project.org/"

[6] deepnet, "http://cran.r-project.org/web/packages/deepnet"