

데이터 분석적 사고력 향상을 위한 딥러닝 기반 학습 시스템 개발 연구

이영호 · 구덕회

서울교육대학교 컴퓨터교육과

요 약

본 연구의 목적은 학습자의 데이터 분석적 사고력 향상을 위한 딥러닝 기반 학습 시스템 개발 연구이다. 연구의 내용은 다음과 같다. 첫째, 데이터 분석적 사고력 향상을 위해 발견학습 모형에 딥러닝 기법을 적용하였다. 이는 데이터의 관계를 나타내주는 모델을 딥러닝 기법을 사용하여 생성하고, 새로운 데이터를 이 모델에 적용하여 데이터를 분석하는 과정을 경험할 수 있는 학습 방법이다. 둘째, 이 학습 방법에 따른 수업을 위한 딥러닝 기반 학습 시스템을 개발하였다. 딥러닝 기법을 사용하여 학습자가 입력한 데이터의 모델을 생성하고 적용할 수 있는 시스템을 개발하였다. 딥러닝을 적용한 발견학습 및 시스템 설계 연구는 데이터의 중요성이 더욱 커지는 미래 사회에서 학습자의 데이터 분석적 사고력을 향상시킬 수 있는 새로운 접근이 될 것으로 기대한다.

키워드 : 데이터 분석적 사고력, 딥러닝, 발견학습, 딥러닝 기반 학습 시스템, 인공지능

A Study on Development Deep Learning Based Learning System for Enhancing the Data Analytical Thinking

Young-ho Lee · Duk-hoi Koo

Department of Computer Education, Seoul National University of Education

ABSTRACT

The purpose of this study is to develop a deep learning based learning system for improving learner's data analytical thinking ability. The contents of the study are as follows. First, deep learning was applied to the discovery learning model to improve data analytical thinking ability. This is a learning method that can generate a model showing the relationship of given data by using the deep learning method, then apply the model to new data to obtain the result. Second, we developed a deep learning based system for DBD learning model. Specifically, we developed a system to generate a model of data using the deep learning method and to apply this model. The research of deep learning based learning system will be a new approach to improve learner's data analytical thinking ability in future society where data becomes more important.

Keywords : Data analytic thinking, Deep learning, Discovery learning, Deep Learning based System, Artificial intelligence

교신저자 : 구덕회(서울교육대학교 컴퓨터교육과)

논문투고 : 2017-06-06

논문심사 : 2017-06-19

심사완료 : 2017-06-26

1. 서론

빅데이터, 인공지능이 세계적으로 큰 영향을 미치면서 데이터화(Datafication)에 대한 논의 또한 지속적으로 이루어지고 있다[17]. Kenneth Cukier와 Viktor Mayer-Schoenberger는 데이터화를 ‘삶의 모든 측면을 파악하여 그것을 데이터로 바꾸는 과정’이라 정의하였다[5]. 데이터화를 통해 그것의 사용목적을 바꾸고 그 정보를 새로운 형식의 가치로 전환할 수 있다[5]. 또한 데이터를 사용하여 사회 현상 및 문제를 새로운 시각으로 바라볼 수 있고 문제 해결의 새로운 통찰의 도구로서 사용할 수 있기 때문에 데이터화의 중요성이 더욱 커지고 있다. 이러한 데이터화를 통해 금융, 의료/약제 산업, 생명공학, 사회복지 등 거의 모든 사회 전반적인 분야에서 다양한 문제를 해결하고 있다[12]. 또한 영국의 HSC, 싱가포르의 RAHS와 같이 세계 주요국 정부에서는 데이터 분석을 기반으로 사회 현안 및 미래 이슈들에 대한 선제적 대응 방안 마련을 위한 정책을 추진하고 있다[8]. 데이터를 실제 문제 상황에 적용하기 위해서는 데이터에 대한 전반적인 이해의 토대 위에 이를 적절히 사용할 수 있는 능력이 필요하다. 그리고 CSTA(Computer Science Teachers Association)와 ISTE(International Society for Technology in Education)에서는 정보과학적 사고의 핵심 개념으로 자료 수집, 자료 분석, 자료 표현을 제시하고 있다[4]. 이러한 상황에서 데이터 능력의 사고력으로서의 고찰과 이를 향상시킬 수 있는 방법적인 연구가 필요하다.

데이터 분석적 사고력은 데이터의 수집, 분석, 변환의 과정을 아우르는 사고력이다[14]. 이를 향상시키기 위해서는 데이터의 전반적인 과정을 수업 프로세스에서 경험해 볼 수 있는 학습 과정과 방대한 양의 데이터를 분석할 수 있는 시스템의 개발이 필요하다. 이에 본 연구에서는 데이터 분석적 사고력을 향상시키는 수업을 위한 방법으로서 발견학습 수업 모형에 기초한 수업 방법을 제시하였다. 발견학습 모델은 주어진 자료에 대한 통찰의 과정을 통해 자료에 대한 모델을 도출하고 이를 일반화하는 학습 모델이다[10]. 본 연구에서 제시한 딥러닝 기반 발견학습 수업 방법은 주어진 데이터를 기반으로 딥러닝 분석을 통해 데이터의 모델을 형성하고, 이를 다양한 데이터에 사용하는 학습 방법이다. 데이터를

분석하는 기법에는 통계적 방법, 머신 러닝(SVM, KNN 등)을 사용한 방법 등 다양한 방법이 존재한다[16]. 본 논문에서는 중단 간 기계학습(end-to-end machine learning)이라 불리는 딥러닝 기법을 사용하여 학습 시스템을 개발한다.

본 논문은 다음과 같은 순서로 구성되어 있다. 2절에서는 본 논문의 이론적 배경인 데이터 분석적 사고력과 딥러닝에 대해서 기술하였다. 3절에서는 딥러닝 기법을 발견학습 수업 모형에 적용하는 방법을 제안하였으며, 4절에서는 이 수업 방법에 적합한 수업 시스템을 설계 및 구현하였다. 5절에서는 본 연구에 대한 결론과 시사점을 기술하였다.

2. 이론적 배경

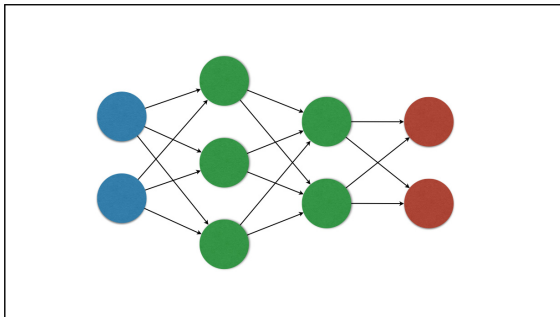
2.1. 데이터 분석적 사고력

데이터 기술은 정형 또는 비정형 데이터를 수집, 분석, 변환의 과정과 관련된 제반 기술을 의미한다[9]. 이러한 전반적인 데이터 기술을 보유한 데이터 전문가 또는 데이터 과학자는 데이터를 문제 해결 목적에 부합하도록 데이터를 분석하고 이를 활용할 수 있는 데이터 분석적 사고력을 가진 인력이다[9]. 본 논문에서는 데이터 분석적 사고력을 데이터에 대한 전반적인 이해를 바탕으로, 분석한 데이터를 사용하여 합리적인 의사결정을 할 수 있는 사고로 정의한다. 데이터의 수명 주기는 수집, 분석, 변환의 과정을 거친다[3]. 각각의 단계는 다음과 같다. 최근 인터넷과 세상을 연결하는 장치의 종류가 다양해지고 있으며, ‘사물 인터넷’의 형성으로 데이터의 생성 속도뿐만 아니라 생성 방식 또한 크게 증가하고 있다. 수집 단계에서는 이와 같이 다양한 경로를 통해 생성되는 데이터 중 문제 해결에 적합한 데이터를 수집하는 과정이다. 다음은 데이터 분석 과정이다. 데이터는 이해할 수 있을 경우에만 가치가 있으며, 그렇지 않을 경우에는 임의적인 관찰 결과에 불과하다[3]. 이와 같이 데이터 분석 과정은 데이터를 통한 문제해결에서 가장 중요한 위치를 차지하고 있다. 이 과정은 수집한 데이터 중에서 의미 있는 데이터를 바탕으로 결과를 해석하고 현상을 이해하기 위한 모델을 찾아가는 과정이다. 직관

적인 방법으로 방대한 데이터에서 하나의 모델을 찾는 것은 불가능하다. 그러므로 다양한 데이터 과학 기법을 사용하여 데이터를 분석을 실시한다. 마지막으로 데이터의 변환 과정은 데이터 분석 과정을 통해 얻은 모델을 사용하여, 새로운 데이터를 재정의 하거나 미래를 예측하는 과정이다. 이러한 수집, 분석, 변환의 과정을 통해 데이터에 기반한 의사결정을 할 수 있는 데이터 분석적 사고가 가능하다.

2.2. 딥러닝

딥러닝이란 계층이 깊은 인공 신경망을 사용한 알고리즘을 말한다[11]. 학계에서는 인공 신경망이 2~3개의 층으로 되어 있는 신경망을 Shallow Learning이라 하고 그 이상인 것을 Deep Learning이라 구분한다. 이러한 이유로 딥러닝을 심층 신경망(Deep Neural Network)이라고도 한다.



(Fig. 1) Deep Learning Network Overview

인공 신경망은 생물의 뇌 신경망을 모델로 제작한 컴퓨터 처리 시스템이다[15]. 뇌 신경망에는 뉴런이라는 신경 세포가 신호 전달을 담당하며, 뉴런은 상호 연결되어 네트워크 구조를 형성한다. 뉴런과 뉴런을 연결하는 시냅스는 신호의 수신 또는 송신의 역할을 수행하며, 신호를 송신할 때 뉴런에서 보내는 전기 신호의 양이 임계치 이상일 때 활성화되어 다음 뉴런으로 신호를 전달한다.

인공 신경망 또한 이와 유사한 구조로 설계되어 있으며, (Fig. 1)과 같이 가장 왼쪽의 입력층, 중간은 은닉층, 가장 오른쪽의 출력층이라는 세 종류의 층으로 구성된다. 이 때 은닉층의 수에 따라 인공 신경망이 딥러닝

로 구분된다[11]. 아래 그림은 은닉층이 2개인 3층 인공 신경망을 나타낸 그림이다.

각각의 노드는 웨이트라는 가상의 선으로 연결되어 있으며, 각각의 웨이트는 연결 가중치라는 값을 가진다. 시냅스에서와 마찬가지로 노드에 입력되는 값이 임계치를 넘어서면 다음 노드로 값을 전달하는데, 이 때 입력 신호의 총 합을 하나의 출력 신호로 변환하는 활성화 함수를 사용한다. 이 때 사용되는 활성화 함수는 주로 Sigmoid함수, ReLU함수, Tanh함수 등을 사용한다[15]. 사용자가 어떠한 특정한 값을 입력하였을 때 각 노드와 노드를 연결하는 웨이트의 연결가중치의 값에 의해 결과가 도출되기 때문에 딥러닝에서의 핵심을 각 웨이트가 가지는 연결 가중치의 값이라 할 수 있다. 은닉층의 수가 깊어지면 깊어질수록 노드와 노드를 연결하는 웨이트의 수가 그에 비례하여 증가하게 되며 이는 더 정교한 예측 모델의 생성이 가능하다. 인공 신경망에서는 웨이트의 값을 확률적 경사 하강법(Stochastic Gradient Descent Method) 등과 같은 방법을 사용하여, 최적의 웨이트의 값으로 수렴해 나간다. 이 과정을 ‘딥러닝을 통해 데이터를 학습한다’고 표현하며, 데이터의 학습이 완료된 모델을 도출할 수 있다[15].

2.3. 발견학습 모형

발견학습은 어떠한 현상이나 자료에서 직관적인 사고인 통찰의 과정을 통해 지식의 모델을 형성해나가는 학습으로 정의할 수 있다.

Bruner(1961)는 발견학습에서의 발견을 주위의 여러 대상을 관찰하여 그 과정에서 공통성과 이질성을 찾아 하나의 유목(category) 또는 여러 개의 유목으로 나누는 것이라고 보았다[7]. 이러한 발견학습의 배경은 형태 심리학(gestalt psychology)에서 살펴볼 수 있다. 형태 심리학에 따르면 학습은 학습자가 문제 사태에 대한 통찰(insight)을 획득하는 과정이다. 통찰은 사상들을 연결하고 있는 조직의 원리 또는 관계를 발견하는 것을 의미한다[2]. 이 통찰은 직관적 사고와 밀접한 관련을 가지고 있다. 직관적 사고(intuitive thinking)는 분석적인 지적 과정에 의존함이 없이 어떤 문제 사태의 의미, 의의, 또는 구조를 파악하는 행위를 의미한다[7]. 학습자가 학습 문제를 해결하기 위해 주어진 자료를 바탕으로 일반

적인 모델을 생성하는 것이 발견학습의 목표이다. 그리고 획득한 지식을 적용 또한 발전적으로 적용한다[7].

3. 딥러닝 기반 발견학습 설계

3.1. 딥러닝 기반 발견학습

딥러닝 기반 발견학습은 발견학습 모형(Discovery model)에 기초하여 고안된 학습 방법이다. 제시된 학습 자료에서 법칙 또는 원리를 발견하고 이를 다른 상황에 적용하는 것이 발견학습의 주된 목표이다. 이와 유사하게 딥러닝 기반 발견학습은 많은 양의 데이터에서 데이터 간의 관계 및 원리를 나타낼 수 있는 모델을 생성하고, 이를 새로운 데이터에 적용하여 의미를 도출하는 학습 방법이다.

기존의 발견학습 모형과 딥러닝 기반 발견학습의 차별성은 다음과 같다. 방대한 데이터 환경에서 데이터의 관계를 추출하여 모델을 생성하기 위해서는 데이터 과학적 접근이 필요하다. 이에 딥러닝 기반 발견학습에서는 데이터를 분석함에 있어 데이터 과학적 기법을 적용하였다. 구체적으로, 데이터 간의 관계에 대한 모델을 생성해주는 딥러닝 기법을 사용하였으며, 이를 적용한 시스템을 개발하여 딥러닝 기반 발견학습에 사용할 수 있도록 하였다.

딥러닝 기반 발견학습은 학습자의 학습 과정 내에서 수집하고 접근할 수 있는 데이터를 사용하여, 데이터를 분석하고 의미를 도출하는 학습활동에 적합하다.

3.2. 딥러닝 기반 발견학습 흐름

딥러닝 기반 발견학습은 데이터를 수집하고, 수집한 방대한 양의 데이터를 딥러닝을 통해 각 데이터의 관계를 설명하는 모델을 형성하고 이를 일반화 및 적용하는 학습 방법이다. 딥러닝 기반 발견학습은 <Table 1>과 같은 단계로 구성된다. 기본적인 학습 방법은 기존의 발견학습 모형의 단계와 유사하나 데이터의 분석과정과 모델의 생성 및 적용 단계에서 본 연구에서 개발한 시스템을 사용한다.

<Table 1> Steps and details of Deep Learning Based Discovery Learning

Step	Detail
Explore and Identify Problems	Explore and identify problems, present learning problems
Data presentation and model learning	Data presentation and data scientific method (deep learning) Utilization model creation
Model evaluation and development	Evaluate generation model accuracy and improve accuracy
Generalization and application	Apply new data to derived models

첫 번째 단계는 탐색 및 문제 파악이다. 먼저 학습 문제를 제시하고, 문제 해결을 위한 데이터의 종류 및 자료의 특성에 대해서 살펴보는 단계이다.

두 번째 단계는 자료 제시 및 모델 학습 단계이다. 문제를 해결하기 위한 데이터를 제공하는 단계로, 이 단계에서는 교사에 의해 주어진 데이터 이 외에 필요에 따라 학습자가 추가적인 데이터를 수집할 수 있다. 이렇게 수집한 데이터를 바탕으로 본 연구에서 개발한 딥러닝 기반 학습 시스템을 사용하여, 데이터에 적합한 모델을 생성한다. 적은 양의 데이터에서는 직관적 혹은, 분석적인 방법으로 데이터와 데이터의 관계에 대해 파악할 수 있지만, 빅데이터라 불리는 방대한 양의 데이터에서 정확한 관계를 살펴보기 위해서는 다양한 데이터 과학적 기법이 필요하다. 본 연구에서 사용한 딥러닝 기반 학습 시스템은 4절에서 기술한다.

세 번째 단계는 모델 평가 및 발전 단계이다. 이 단계에서는 딥러닝 기반 학습 시스템을 통해 생성된 모델을 평가하는 단계이다. 딥러닝을 통해 생성된 모델의 정확도(accuracy) 향상을 위해 딥러닝에 필요한 하이퍼 파라미터 값을 수정한다.

마지막 단계는 일반화 및 적용 단계이다. 세 번째 단계에서 생성한 데이터와 데이터의 관계를 나타낸 모델이 새로운 데이터에 대해 어떠한 결과 값을 예측하는지 살펴본다.

4. 시스템 설계 및 개발

4.1. 시스템 설계

4.1.1. 시스템 설계의 방향

본 연구에서 개발하고자 하는 딥러닝 기반 학습 시스템은 초등학교 고학년 및 중등학생을 대상으로 데이터에 기초한 발견학습을 위한 시스템이다. 학습자들은 웹 페이지를 데이터를 입력하여 모델을 생성하고, 생성한 모델을 평가 및 활용할 수 있도록 다음과 같은 방향으로 설계하였다.

첫째, 데이터의 수집, 분석, 활용의 단계를 모두 경험해 볼 수 있도록 시스템을 설계한다. 데이터 분석적 사고력을 향상시키기 위해서는 데이터를 활용하는 모든 단계를 경험해볼 필요가 있다. 이 단계가 포함된 시스템을 통해, 시스템의 사용 과정과 학습의 단계가 일치되도록 설계한다.

둘째, 딥러닝의 전체적인 흐름을 인지할 수 있도록 한다. 딥러닝에서는 데이터 입력 이후, 연결 가중치를 최적화하기 위해 다양한 튜닝의 과정이 필요하다. 예를 들어 노드의 입력 값에 대한 출력 값을 계산하기 위한 활성화함수의 선택에서부터 연결 가중치의 값을 적절하게 수정할 수 있는 확률적 경사 하강법의 기법 선택 등 최적의 학습 환경을 만들기 위한 선택이 필요하다. 이러한 선택의 과정은 딥러닝의 세부적인 내용으로서, 시스템을 사용하는 학습자의 수준에는 부합하지 않는다. 그러므로 시스템 설계에 있어 사용자의 수준에서 직관적으로 인식할 수 있는 하이퍼 파라미터만을 선정하여, 이를 수정할 수 있도록 설계한다.

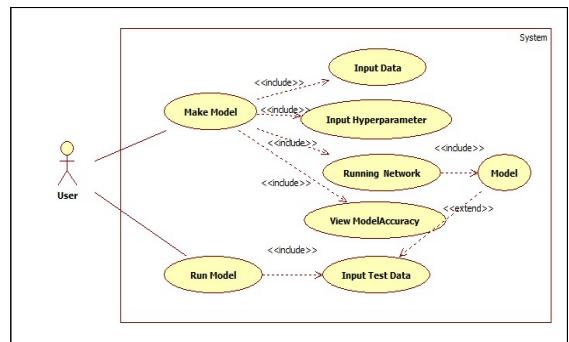
4.1.2. 시스템 구조

시스템의 Use Case 다이어그램은 (Fig. 2)과 같다. 시스템에서는 사용자는 모델 생성과 모델 실행을 선택할 수 있다.

모델 생성에서는 데이터의 입력, 하이퍼 파라미터의 입력, 딥러닝 네트워크의 실행, 모델 정확도 확인을 할 수 있다. 데이터의 입력에서는 학습자가 특징 데이터와 정답 데이터로 구성된 csv 형식의 데이터 파일을 시스템에 입력한다. 다음으로 입력 데이터 중 출력 데이터의

형태, 학습모델의 최적화를 위한 반복 횟수를 하이퍼 파라미터로 시스템에 입력한다. 다음으로 딥러닝 네트워크를 실행하여 사용자가 입력한 데이터에 적합한 학습된 모델을 생성한다. 마지막으로 시스템은 평가 데이터 파일을 사용하여 시스템의 정확도(accuracy)를 측정하여 사용자에게 제시한다.

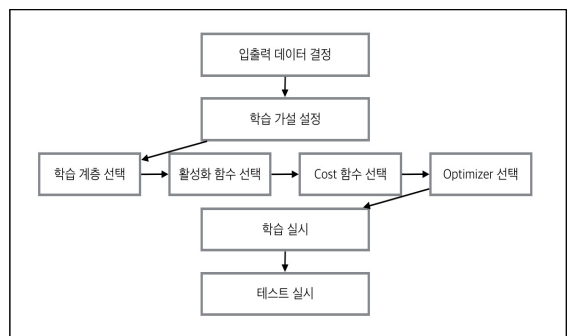
모델 실행에서는 데이터의 관계를 나타낸 모델을 사용하여, 살펴보고자 하는 데이터(실험 데이터)에 대한 분석 결과 값을 확인할 수 있다.



(Fig. 2) Deep learning based Learning System Use case

4.1.3. 딥러닝 설계

딥러닝은 머신러닝의 분야로, 머신러닝은 크게 지도 학습(Supervised Learning)과 비지도 학습(Unsupervised Learning)으로 분류된다[6]. 지도 학습과 비지도 학습의 차이는 데이터에 정답(Label)의 포함 유무이다. 본 연구에서는 지도 학습 방식의 분류 딥러닝 모델을 설계하고자 한다. 딥러닝의 전체적인 흐름은 (Fig. 3)과 같다.



(Fig. 3) Design Flow of Deep Learning

입출력 데이터 결정 단계는 다음과 같다. 입력된 데이터의 특징(feature)과 정답(label)을 정하는 단계로 본 시스템에서는 데이터의 첫 번째 열부터 마지막 전 열까지를 특징 데이터로 정의하였으며, 마지막 열을 정답 데이터로 가정하였다. 학습가설은 다음과 같이 설정하였다. W 는 노드와 노드를 연결하는 연결가중치의 값이며, b 는 편향값을 나타낸다.

$$H(x) = Wx + b$$

본 시스템에서의 학습 계층은 4개의 층으로 구성하였으며, 활성화 함수로 다음의 ReLU함수를 사용하였다. 활성화 함수로 주로 사용되는 Sigmoid함수는 층이 깊어질수록 값의 정확도를 감소시키는 한계가 있기 때문에 이를 보완하는 ReLU함수를 사용한다[15].

$$y = \begin{cases} x & (x > 0) \\ 0 & (x \leq 0) \end{cases}$$

딥러닝 네트워크로 계산한 결과 값과 훈련데이터의 정답 값의 차이를 나타내주는, Cost함수로는 아래의 Cross-entropy함수를 사용한다. Cross-entropy함수는 정답일 때 Cost값을 최소화하고, 오답일 때 그 값을 최대화하는 함수이다.

$$Cost(W) = \frac{1}{m} \sum C(H(x), y)$$

$$C(H(x), y) = y \log(H(x)) - (1 - y) \log(1 - H(x))$$

마지막으로 딥러닝 네트워크의 연결가중치 값을 최적화하는 Optimizer는 확률적 경사 하강법을 사용한다.

$$W \leftarrow W - \eta \frac{\partial L}{\partial W}$$

본 시스템에서는 다양한 확률적 경사 하강법 중 모멘텀 기법과 AdaGrad의 장점을 융합한 AdamOptimizer를 사용한다[13]. 그리고 정확도 향상을 위한 Dropout기법을 사용한다.

4.2. 시스템 개발

4.2.1. 시스템 개발 및 테스트 환경

본 연구에서 딥러닝 기반 학습 시스템을 개발하기 위해 사용한 환경은 <Table 2>와 같다.

딥러닝 프레임워크는 캘리포니아 대학교 버클리 캠퍼스(UC Berkeley)에서 개발한 Caffe, 토론토대학교(Univ. Toronto)의 Cuda-convert, 구글(Google)에서 개발한 Tensorflow 등이 있다[1]. 본 연구에서는 Python 언어 기반의 Tensorflow를 사용하여 딥러닝을 구현하였다. 웹 프로그래밍 환경은 Python 웹 프레임워크인 Django를 사용하여 구축하였다.

<Table 2> System development environment

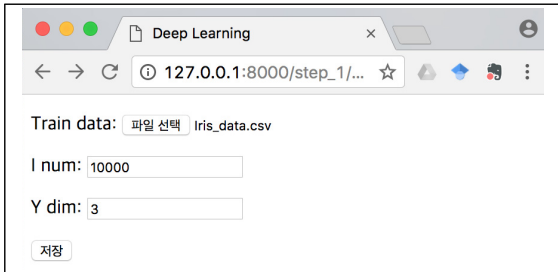
Category	Programs and Versions	Use
OS	mac OS Sierra	Development environment
Deep-Learning Framework	tensorflow1.0.1	Creating a Deep Learning Model
Web Framework	Django 1.11	Web development
Browser	Chrome 58	System validation and application

4.2.2. 개발의 실제

딥러닝 기반 학습 시스템은 데이터 입력, 딥러닝 기반 모델 생성, 결과 확인, 데이터 실험의 단계로 구성된다.

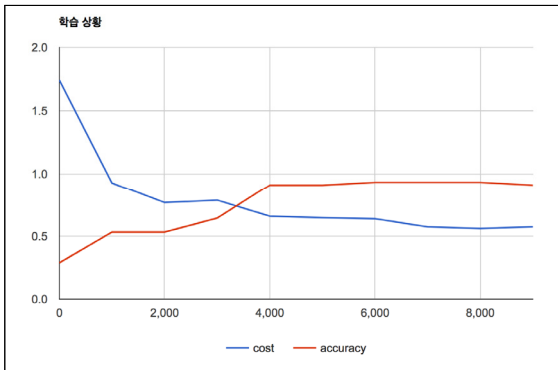
데이터 입력 및 딥러닝 기반 모델 생성 단계는 딥러닝 기반 발전학습의 2단계와 3단계에 해당되며 교사가 제시한 자료 혹은 학습자가 수집한 자료를 시스템에 입력하여 모델을 생성하는 과정이다. 이 단계에서는 자료의 입력과 더불어 사용자가 하이퍼 파라미터를 입력할 수 있다. 학습 데이터(Train data), 학습 반복 횟수(I num), 구분하고자 하는 결과의 종류(Y dim)를 입력하는 창으로 구성된다.

학습의 반복 횟수는 도출한 모델이 지나치게 훈련 데이터에 과적합(overfitting)되는 문제를 해결하기 위해 적절한 값으로 조절이 필요하다. 저장버튼을 클릭하면 사용자가 작성한 하이퍼 파라미터를 사용하여, 연구자가 설계한 딥러닝 알고리즘에 의해 학습을 시작한다. 데이터 입력 단계의 화면 구성은 아래 (Fig. 4)와 같다.



(Fig. 4) Input data file and hyper parameter

본 시스템에서 실험적으로 적용한 데이터는 Iris Dataset으로 붓꽃에 관한 데이터이다. 이 데이터는 꽃잎과 꽃받침의 너비와 길이를 측정하였으며, 5개의 필드(field)(4개의 특징 데이터, 1개의 정답 데이터)과 150개의 레코드(record)로 구성되어 있다. 시스템에서 학습 과정을 거치게 되면 붓꽃의 특징 데이터 4개를 바탕으로 붓꽃을 분류할 수 있는 모델이 도출된다.

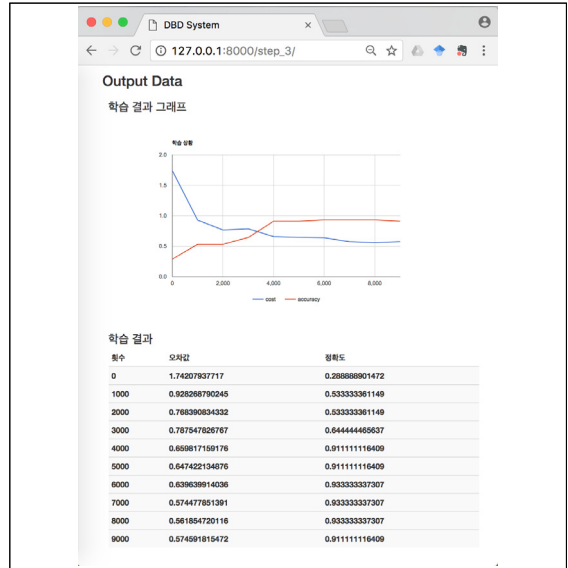


(Fig. 5) Cost and accuracy graph

학습 데이터의 30%를 시험 데이터로 설정하여 학습의 정확도를 측정하였다. 붓꽃 데이터셋(Iris Dataset)을 사용하여 10,000번 반복학습을 실시하였다. (Fig. 5)에서 볼 수 있듯이, 딥러닝 알고리즘에 의해 매 반복마다 딥러닝 데이터 모델에서 도출되는 결과 값과 정답 값의 차이를 나타내는 cost값이 아래와 같이 줄어드는 것을 확인할 수 있었으며, 실험 데이터를 사용하여 모델에 적용하였을 때 그 정확도를 나타내는 accuracy값이 점진적으로 증가하여 95%의 정확도를 보였다.

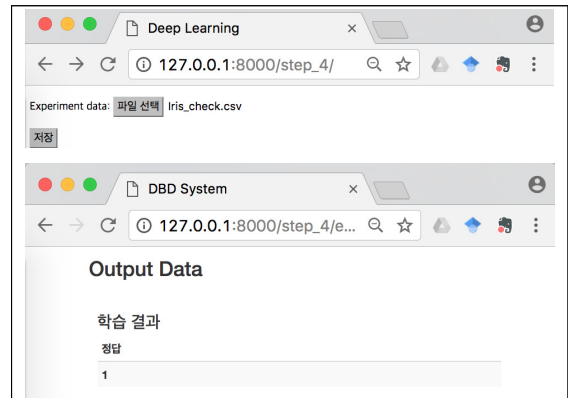
학습이 완료되면 '결과 확인하기'와 '인공지능 실험하기'창이 사용자에게 제시되며, '결과 확인하기' 버튼을 클릭하면, 반복 횟수에 따른 오차 값과 실험 데이터를

딥러닝 네트워크에 적용하였을 때의 정확도 값을 다음 (Fig. 6)와 같이 그래프와 표로 제시하여, 사용자가 시스템의 학습 과정을 볼 수 있다.



(Fig. 6) System process check by learning step

데이터 실험하기는 딥러닝 기반 발견학습의 4단계에 해당하며 학습자가 '인공지능 실험하기' 버튼을 클릭하면 (Fig. 7)과 같이 딥러닝 기법을 통해 도출한 모델을 사용할 수 있다. 사용자는 실험 데이터(Experiment data)를 선택한 후 저장 버튼을 클릭하여, 학습된 모델을 통해 결과를 얻을 수 있다. 개발한 시스템에 4개의 필드(field)와 1개의 레코드(record)로 구성된 iris_check.csv 파일을 입력하면, 딥러닝 기반으로 생성된 모델에 데이터를 적용하여 이 데



(Fig. 7) Model experiment and check results

이더가 어떤 종류의 붓꽃인지 알려준다. 본 연구에서 사용한 시스템의 실제 구현은 <https://vimeo.com/222613934>에서 동영상으로 볼 수 있다.

5. 결론 및 시사점

본 연구에서는 학습자의 데이터 분석적 사고력 향상을 위한 학습 모형을 설계하고 실습 시스템을 개발하였다. 구체적으로 연구자는 발견학습 모형에 딥러닝을 적용하였다. 딥러닝 기반의 발견학습은 학습 문제에 적합한 데이터를 수집하고, 수집한 데이터를 딥러닝 기법을 통해 데이터의 관계를 효과적으로 표현해주는 모델을 생성한다. 그리고 이 모델을 사용하여 새로운 데이터의 분석을 통해 학습자의 데이터 분석적 사고력을 높일 수 있는 학습 방식이다. 이 학습 방식을 구현하기 위해 딥러닝 기반 학습 시스템을 개발하였다. 방대한 데이터를 분석하기 위해서는 데이터 과학적 기법을 사용하여야 한다. 이에 딥러닝 기법을 사용하여 딥러닝 기반 발견학습에 따라 학습을 실시할 수 있는 환경을 조성하였다.

본 연구의 시사점은 다음과 같다. 첫째, 데이터화 및 데이터 분석에 대한 능력이 강조되는 현 시점에서 학습자의 데이터 분석적 사고력을 향상시킬 수 있는 딥러닝 기반 발견학습 방식을 제시하였으며, 이를 지원하는 시스템을 개발하였다.

둘째, 딥러닝 기법을 사용하여 데이터를 분석하는 과정을 통해 학습자들이 주변에서 수집할 수 있는 각종 데이터에서 하나의 모델을 만들고, 이 모델을 통해 새로운 데이터(현상)를 예측할 수 있는 경험을 제공할 수 있다.

본 연구를 통해 학습자의 데이터 분석적 사고력을 높일 수 있으며, 데이터에 대한 통찰을 통해 다양한 문제를 해결할 수 있는 새로운 관점을 제공할 수 있을 것으로 기대한다.

참고문헌

- [1] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Kudlur, M. (2016). TensorFlow: A system for large-scale machine learning. In Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI). Savannah, Georgia, USA.
- [2] Anderson, J. R. (1995). Cognitive psychology and its application (Lee Young-ae, Trans.). Seoul: Ewha Womans University Press.
- [3] BSA. (2015). What's big deal with data. [Http://data.bsa.org/#section-download](http://data.bsa.org/#section-download)
- [4] CSTA, ISTE. (2011). Computational Thinking in K - 12 Education leadership toolkit.
- [5] Cukier, K., & Mayer-Schoenberger, V. (2013). The rise of Big Data: How it's changing the way we think about the world. *Foreign Affairs*, 92(3), 28-40.
- [6] Goldberg, D. E., & Holland, J. H. (1988). Genetic algorithms and machine learning. *Machine learning*, 3(2), 95-99.
- [7] Kang Hyun Suk. (2009). Reinterpretation of knowledge in Bruner 's curriculum theory. *Educational Philosophy*, 38, 1-34.
- [8] Kim, G. H., Trimi, S., & Chung, J. H. (2014). Big-data applications in the government sector. *Communications of the ACM*, 57(3), 78-85.
- [9] Kim Han Jun. (2015). A Study on Big Data Education Focused on Data Insight. Society for Engineering Education.
- [10] Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychological science*, 15(10), 661-667.
- [11] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.
- [12] Lycett, M. (2013). 'Datafication': Making sense of (big) data in a complex world. *European Journal of Information Systems*, 22(4), 381-386.
- [13] Ngiam, J., Coates, A., Lahiri, A., Prochnow, B., Le, Q. V., & Ng, A. Y. (2011). On optimization methods for deep learning. In *Proceedings of the 28th international conference on machine learning (ICML-11)*, 265-272.
- [14] Provost, F., & Fawcett, T. (2013). Data Science

for Business: What you need to know about data mining and data-analytic thinking. " O'Reilly Media, Inc."

- [15] Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks*, 61, 85-117.
- [16] Schutt, R., & O'Neil, C. (2013). Doing data science: Straight talk from the frontline. " O'Reilly Media, Inc."
- [17] Van Dijck, J. (2014). Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & Society*, 12(2), 197-208.



구 덕 회

2000 한국교원대학교 대학원 박사
2000~2003 한국교육학술정보원
선임연구원
2003~2009 대구교육대학교
컴퓨터교육과 교수
2009~현재 서울교육대학교
컴퓨터교육과 교수
관심분야 : 컴퓨터교육이론, 프로그래밍 교육, 디지털 콘텐츠
E-mail: dhk@snue.ac.kr

저자소개

이 영 호



2010~현재 서울시 초등교사
2015~현재 서울교육대학교 생활
과학·컴퓨터교육과 박사과정
관심분야 : 인공지능 교육, 소프트
웨어 알고리즘, 컴퓨터 교육
이론
E-mail: yhbest12@naver.com