



Journal of Knowledge Information Technology and Systems

ISSN 1975-7700

<http://www.kkits.or.kr>

High-School Baseball Pitcher's Pitching Speed Prediction Using Linear Regression Analysis Method

Young-Hwan Oh^{*}

Division of IT Convergency, Korea Nazarene University

ABSTRACT

Recently, studies on artificial intelligence such as AlphaGo and machine learning have been actively conducted. In statistics, linear regression is a regression method that models the linear correlation between dependent variable y and one or more independent variables x . Generally, a linear regression model is established using least square method. In other words, linear regression analysis is a method of modeling the relationship between independent variables, dependent variables, and constant terms. There is a simple linear regression method that models the relationship between two variables and a multiple linear regression method based on two or more independent variables. In a baseball game, the pitcher must be good at ball speed, pitching control, pitching balance, etc., in order to get good results when dealing with batter. High school baseball player's pitching speed is important factor to grow as excellent pitcher. Also, pitcher's ball speed is one of the important factors that determine the winning or defeat of the baseball game. In this paper, we use the Deep Learning Framework(Tensorflow) to measure ball speed of pitcher among high school baseball players and use it for athlete 's exercise and training rehabilitation. In this study, we generate training data about stride and speed of the pitcher and perform the linear regression prediction method using the gradient-descent method which is the optimization algorithm.

© 2019 KKITS All rights reserved

KEYWORDS : Artificial intelligence, Machine-learning, Linear regression, Tensorflow, Gradient descent algorithm

ARTICLE INFO: Received 1 July 2019, Revised 23 July 2019, Accepted 9 August 2019.

^{*}Corresponding author is with the Division of IT Convergency, Korea Nazarene University, 456

Wolbong-ro Seobuk-gu Cheonan, 31172, KOREA.
E-mail address: yhoh@kornu.ac.kr

1. 서론

최근에 알파고(AlphaGo)와 이세돌의 경기를 알파고가 이기면서 전 세계의 인공 지능(Artificial Intelligence)에 대한 높은 관심이 있다[1]. 그 후 인공 지능에 관한 연구가 활발히 진행 중이고, 연구자와 학자들이 인간을 넘어서는 인공지능을 제작하기 위해 끊임없이 노력 중이다. 인공 지능이라는 학문적 개념은 1956년 미국 다트머스 대학(Dartmouth college)에 있던 존 매카시(John McCarthy) 교수가 개최한 다트머스 회의(Dartmouth Conference)에서 처음 등장하였으며[2], 근래에 들어서 폭발적으로 성장하고 있는 중이다. 특히 2015년 이후 빠른 속도의 병렬 처리 성능을 지원하는 GPU(Graphics Processing Unit)의 도입으로 더욱 가속화되고 있다.

한편, 머신 러닝은 기본적인 알고리즘을 이용해 데이터를 예측하고, 분석하고, 학습한다. 이 학습한 내용을 기반으로 판단이나 의사 결정을 한다. 그러나 최종적으로는 코딩 소프트웨어에 최적 알고리즘을 직접 기술해 넣는 것이 아니고 대용량의 데이터와 학습 알고리즘을 이용하여 전자기기 그 자체를 학습하도록 하여 작업 수행능력을 만드는 것을 목표로 한다[3]. 최근에는 인간의 뇌가 가진 생물학적 특성, 특히 뉴런 연결 구조를 기반으로 하는 인공 신경망(Artificial neural network) 연구에 박차를 가하고 있다. 딥 러닝(Deep learning)은 뇌의 뉴런과 유사한 정보 입출력 방법을 활용하여 방대한 양의 데이터를 학습하고 재생산하여 기존의 인공 신경망에서 발전한 다른 형태의 인공 지능 기법이다.

야구 경기는 투수가 공을 던지는 것으로부터 시작된다. 투수(Pitcher)의 송구와 포수(Catcher)의 포구는 야구 경기에서 차지하는 비중은 매우 크다[4]. 투수는 좋은 성적을 내기 위해서 투구 스피드(Ball

speed)와 제구(Ball control), 다양한 변화구(Change-up)를 가져야 한다. 훌륭한 투수의 기본 요소는 제구력과 스피드, 투구 밸런스(Pitching balance)이다. 투수의 볼 스피드가 약 140km인 공을 던질 경우 투수의 손으로부터 홈플레이트를 지나는 시간은 약 0.4~0.5초이다. 좋은 투수로 성장하기 위한 중요한 요소는 투구의 스피드이다.

빠르고 정확한 투구과정을 위해서는 다음과 같은 과정이 필수적이다. 즉, 빠르고 정확한 투구 속도를 내기 위해서는 공을 던지는 동안 투수의 신체 분절의 원활한 협동동작이 필요하다[5]. 투수의 투구모션은 와인드업(Wind-up), 스트라이드(Stride), 팔코킹(Arm cocking), 팔가속(Arm acceleration), 팔감속(Arm deceleration), 팔로우 드로우(Follow-through)까지 총 6단계로 구분된다[6].

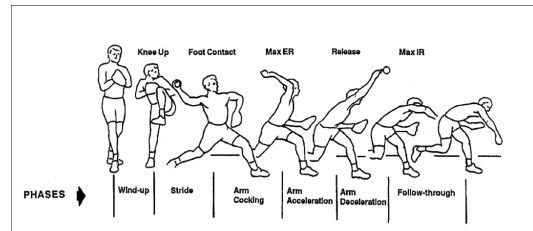


그림 1. 투구 단계
Figure 1. Pitching Phases

이와 같이 투구 동작 중 발생나는 근육 운동은 발바닥에서부터 다리, 허리, 몸통, 팔, 손목까지의 신체의 전반적인 근육 협동작용이 만들어지기 때문에 모든 신체 관절을 이용한 단계적이면서 동시적인 동작이라 할 수 있다[7]. 투수의 안정된 투구 자세와 빠른 투구 스피드를 위해서는 팔과 허리의 강한 회전 동작과 손목, 팔목의 신속한 코킹을 통하여 야구공에 큰 에너지를 전달할 수 있어야 한다. 또한 강한 하체의 힘과 빠른 상체의 회전, 안정적인 넓은 보폭은 공의 힘과 속도를 높이는 데 꼭 필요하다[8].

2019년 현재 대한야구소프트볼협회(Korea Baseball Softball Association)에 등록된 고교야구는 80개팀이 있으며 고교야구 주말리그, 전국대회 등 많은 경기를 소화하고 있다. 고교야구 주말리그제는 문화체육관광부, 대한야구소프트볼협회가 2011년부터 시행한 제도로써 ‘공부하는 엘리트 학생 스포츠’ 라는 취지를 가지고, 전국을 10-12개 권역으로 나눠 6-8개 팀을 배정하여 주중을 제외한 주말리그로 열리고 있다[9]. 고교야구 선수와 KBO 프로선수들과는 체력, 기술, 체격적인 면에서 큰 차이가 있기 때문에 이를 줄이기 위한 다양한 시도와 노력을 하고 있다. 야구 선수의 기초데이터가 되는 운동능력, 순발력, 지구력 등의 경기력을 위한 평가와 기준에 대한 연구들이 프로야구와 대학야구 선수를 대상으로 국한되어 왔으며, 고교야구 선수들에 대한 기본적인 연구들은 거의 없는 실정이다. 야구 운동평가에 있어 투수의 제구력과 투구 스피드는 감독이나 코치, 트레이너에게 투수 훈련 계획을 수립하고 재활훈련을 하는 데 있어 매우 중요한 데이터로서 활용될 수 있다.

좋은 투수로 성장하기 위해서는 자신이 던질 수 있는 구종에 대한 제구와 구속, 투구 밸런스가 중요하다. 제구를 잡기 위해서는 강한 집중력과 든든한 하체가 필수적이다. 또한 투구 밸런스는 홈플레이트를 향하여 중심 이동을 하는 경우 하체, 상체, 팔꿈치, 머리 모양이 홈플레이트와 일직선상에 놓여야 한다. 그리고 투수의 중요한 요소는 볼 스피드에 있다. 투구 모션과 손가락 형태, 유연성 그리고 몸통 회전과 보폭의 상관성이 투구 스피드를 올리는 중요한 인자이다. 오른손 투수 인 경우 공을 던지기 위해서 왼발을 앞으로 쪽 내밀게 되는 데 이때 멀리 디딜 수 있다면 그만큼 타자와 투수의 간격을 좁힐 수 있고 야구공에 더욱 큰 힘을 줄 수 있으므로 볼 스피드가 늘어나는 효과를 줄

수 있다. 특히 보폭의 크기는 투수의 몸의 아크를 크게 만들기 때문에 투구 스피드 큰 영향을 미치게 된다. 본 논문에서는 딥 러닝을 이용하여 고등학교 투수의 속구 스피드를 예측하기 위한 선형회귀분석기법을 연구한다. 이를 위하여 텐서플로우(Tensorflow)를 이용하며 예측모델을 구현하고 성능을 평가한다.

본 논문의 구성은 다음과 같다. 제 2장에서는 관련 연구, 제 3장에서는 야구선수의 모형분석, 제 4장에서는 예측모델과 학습데이터를 기술한다. 제 5장에서는 구현 및 실험결과를 설명한다. 마지막으로 제 6장에서 결론을 맺는다.

2. 관련 연구

2.1 선형회귀 분석

선형회귀 분석(Linear Regression Analysis)은 변수들 사이의 상관 관계를 분석하는 데 사용하는 통계학적 방법이다[10]. 선형회귀 분석은 독립변수 X , 종속변수 y , 상수항 b 사이의 관계를 모델링하는 방법이다. 두 변수 사이의 관계일 경우 단순 선형회귀(Simple Linear Regression)라고 하며, 여러 개의 변수를 다루는 다중 선형회귀(Multiple Linear Regression)도 있다[11].

2.2 텐서플로우(Tensorflow)

텐서플로우는 구글이 2011년에 개발을 시작하여 2015년에 오픈 소스(Open source)로 공개한 기계학습 라이브러리로 딥러닝과 기계학습 분야를 일반인들도 사용하기 쉽도록 다양한 기능들을 제공한다. 2016년 알파고와 함께 한국에서도 관심이 높아진 추세이다. 고수준의 프로그래밍 언어로 알려진 Python을 활용하여 연산처리를 작성할 수 있다[12].

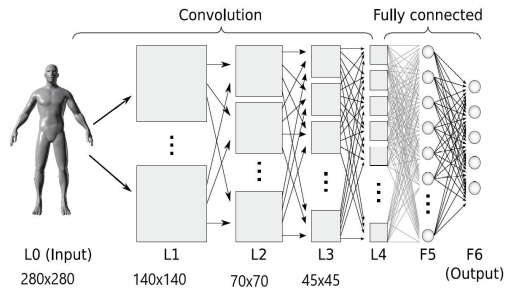


그림 2. 텐서플로우
Figure 2. Tensorflow

표 1. 투수의 신체적 특성값
Table 1. The physical characteristics of the pitcher

선수	키 (cm)	몸무게 (kg)	보폭 (cm)	볼스피드 (km)
1	183	83	162	134.5
2	181	74	167	127.3
3	173	76	163	125.7
4	187	85	174	138.3
5	184	79	171	127.4
6	174	73	161	121.7
7	178	74	166	132.8
8	185	79	172	139.1

3. 야구선수의 모형분석

야구선수의 투수 능력은 속구 스피드(Straight speed), 제구(Ball control), 집중력(Concentration)이 이를 결정한다. 특히 속구 스피드는 투수의 보폭과 몸통 회전량에 따라 다르다. 본 연구에서는 각 투수의 보폭을 측정하여 예측모델에 학습데이터(Training data)로 한다.

3.1 실험 방법

본 실험의 참여자는 P고등학교 야구부에서 운동하고 있는 투수 8명을 임의로 선택하여 속구 스피드 측정을 위한 실험을 하였다. 이들의 평균 나이는 17.4세이고, 평균 키는 177.4cm 그리고 평균 몸무게는 76.7kg이었다. 측정된 속구 스피드는 코치로 하여금 투수들의 운동량과 훈련재활에 도움을 주는 목적이 있음을 알려주었다. 아래 <표 1>은 투수의 보폭과 속구 스피드의 측정값을 나타낸다. 3회 측정하여 이중 최고 구속(중속)을 채택하였다.

3.2 측정 요소

투수에게 있어 보폭(스트라이드)은 투구 스피드와 상관하여 매우 중요한 요인이다. 개인의 신장과 하체 길이, 투구 동작, 투구 형태에 따라 차이가 있지만 대략 스트라이드는 150cm~175cm 이다. 보폭이 넓어질 경우 스피드 향상은 물론 제구(볼컨트롤)에도 도움을 준다.

4. 예측모델과 학습데이터

본 장에서는 고교야구 투수의 보폭과 관련되어 투구 스피드를 텐서플로우를 이용하여 예측모델을 산정한다. 먼저 학습데이터를 정하고 최소제곱법(Least squares method)을 이용하여 학습률(Learning rate)을 구한다[13].

4.1 학습데이터

학습은 기본적으로 지도 학습(Supervised learning)과 비지도 학습(Unsupervised learning)으로 구분되며, 지도 학습은 이미 정해진 데이터를 사용하여 학습을 한다. 이 때 기초가 되는 데이터를

학습데이터라고 한다[14]. 본 연구에서는 첫 번째 변수 x 로부터 두 번째 변수 y 를 예측하려고 한다. 다음 <표 2>는 독립 변수와 종속 변수를 보여 준다.

표 2. 독립변수와 종속변수
Table 2. Independent and dependent variables

x(보폭)	y(볼 스피드)
162	134.5
167	127.3
163	125.7
174	138.3
171	127.4
161	121.7
166	132.8
172	139.1

여기에서 x 데이터는 보폭, y 데이터는 볼 스피드이다. 이러한 데이터를 학습데이터라고 한다. 위의 표를 그래프로 나타내면 다음 <그림 3>과 같다.

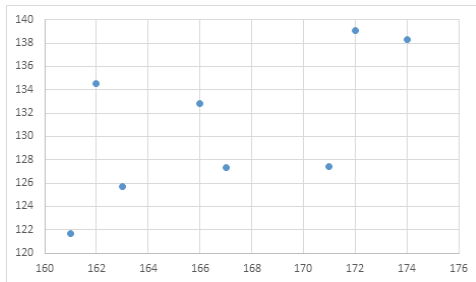


그림 3. 학습데이터(x:보폭, y:볼스피드)
Figure 3. Training data(x:stride, y:ball speed)

보폭과 볼 스피드의 상관관계를 표현할 수 있는 하나의 직선을 만들어 보면 다음 <그림 4>와 같다. $Y = X$ 라는 직선함수를 그리면 좌표 평면에 위치하는 데이터를 볼 수 있다. 선형회귀 알고리즘의 목표는 학습데이터를 관통하는 하나의 직선을 찾아내는 것이다. 일단 $Y = X$ 인 일차함수를 검색하면 이 함수를 이용하여 예측할 수 있다.

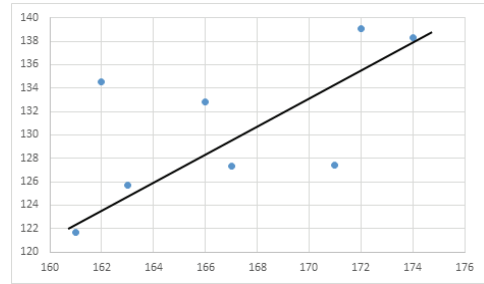


그림 4. 선형회귀 일차함수
Figure 4. Linear regression linear function

4.2 가설 방정식

4.1 절에서 언급한 보폭과 볼 스피드 학습데이터를 가지고 학습을 진행하면 하나의 직선 방정식이 표현된다.

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

이 방정식을 가설(Hypothesis)라고 한다. 선의 모양은 W (weight)와 b (bias)에 따라서 다르게 나타나며 어떤 그래프가 가장 정확한 값을 예측할 수 있는지 알아내는 방법이 중요하게 된다. W 는 직선의 기울기가 되고, b 는 y 축을 지나는 y 절편이 된다. 여기서 x 값은 독립변수이고, y 값은 종속변수이다. 즉, x 값에 따라 y 값은 반드시 다르게 된다. 다만, 이 방정식의 값을 정확하게 계산하려면 상수 W 와 b 의 값을 알아야 한다. 즉, 직선을 정확하게 표현하려면 직선의 기울기 a 값과 y 절편 b 값을 정확히 예측해 내야 한다. 이러한 학습 데이터를 바탕으로 정확한 직선을 만들어 낸다면 투수의 보폭에 대한 볼 스피드를 예측할 수 있다.

4.3 최소 제곱법

본 연구에서는 여러 개의 가설 중에서 가장 적합한 가설을 구하기 위하여 최소 제곱법을 이용한다. 즉, 가설과 학습데이터 사이에 오차가 크지 않

다면 적합한 가설이 된다. 다음 <그림 5>는 최소제곱법을 이용한 예이다.

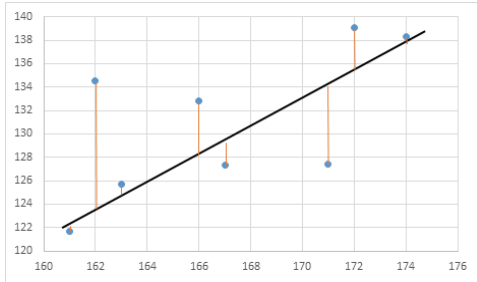


그림 5. 최소 제곱법
Figure 5. Least squares method

$H(x) - y$ 가 실제 데이터와 가설로 나온 데이터 간의 차이가 된다. 이 데이터는 +, -값이 둘 다 나오기 때문에 제곱을 해준다.

$$(H(x) - y)^2$$

비용(cost)은 아래 수식과 같이 정의할 수 있다. m개의 학습데이터가 주어졌을 때, 각각의 학습데이터에서 그려진 직선이 예측한 값($H(x) =$ 보폭)과 실제 학습데이터에서의 값($y =$ 볼스피드)의 차이를 제공한 값의 평균이다.

$$cost(W, b) = \frac{1}{m} \sum_{i=1}^m (H(x^{(i)}) - y^{(i)})^2$$

즉, 선형회귀 알고리즘은 이러한 학습데이터가 주어졌을 때, 비용값을 최소화하는 가설의 파라미터(W, b)를 찾는 알고리즘이다. 즉, cost를 가장 작게 만드는 W 와 b 를 구하는 것이 선형회귀분석의 목표가 된다. 이 때, 일반적으로 경사하강(Gradient descent) 알고리즘을 이용하여 최저값을 검색한다.

4.4 경사 하강법

경사 하강법은 미분의 기울기를 이용하여 그래프의 차이를 비교한 후 가장 작은 방향으로 이동시키는 방법이다.

$$cost(W) = \frac{1}{2m} \sum_{i=1}^m (Wx^{(i)} - y^{(i)})^2$$

$$W := W - \alpha \frac{\partial}{\partial W} cost(W)$$

즉, 미분은 순간변화율을 의미하고 한 점에서의 기울기를 뜻한다. 임의의 점에서 미분을 구하면 그래프(가설)의 기울기를 알 수 있다. 여기서 구하고자 하는 값은 오차가 가장 작을 때의 기울기 0을 의미한다. 경사 하강법에서 미분 값이 0을 찾는 경우 기울기가 양수와 음수를 번갈아 반복됨으로 기울기의 부호를 바꿀 때 이동거리를 정해주는 것이 중요하다. 이를 학습률(learning rate)라고 한다.

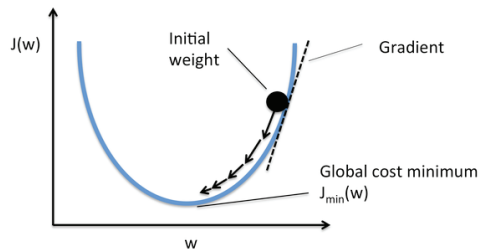


그림 6. 경사 하강법
Figure 6. Gradient descent method

5. 구현 및 실험결과

5.1 텐서플로우를 이용한 선형회귀분석

본 논문에서는 텐서플로우를 이용하여 $y = W * x + b$ 형태의 단순한 선형회귀분석 모델을 작성한다 [15]. 먼저 2차원 좌표계에 학습데이터를 생성하고 데이터에 맞는 최적의 직선을 텐서플로우를 이용

하여 실행한다.

<그림 7>에서는 투수의 측정된 학습 데이터를 이용하여 데이터셋(Data set)을 지정한다. 그 다음에 W값과 b값을 지정하여 학습을 위해 임의의 값을 할당한다. 텐서플로우에서는 가중치와 바이어스에 해당하는 변수를 Variable로 이용하여 나타낸다. 그리고 텐서플로우는 그 값을 자동으로 조절하여 주며 비용함수가 최소화 되도록 가중치와 바이어스 값을 조정하게 된다.

```
# X(보폭) 와 Y(볼 스피드)의 상관관계를 분석하는 선형 회귀 모델을 만들고 실행한다.
import tensorflow as tf

#학습데이터( x_data : 보폭, y_data : 볼 스피드 )
x_data = [1.62,1.67,1.63,1.74,1.71,1.61,1.66,1.72]
y_data = [1.345,1.273,1.257,1.383,1.274,1.217,1.328,1.391]

# name: 나중에 텐서보드등으로 값의 변화를 추적하거나 살펴보기 용이하기 위해 이름을 붙인다.
X = tf.placeholder(tf.float32, name="X")
Y = tf.placeholder(tf.float32, name="Y")

# tf.Variable(변수)에 x_data, y_data 를 각각 할당한다.
W = tf.Variable(tf.random_uniform([1], -1.0, 1.0), name='w')
b = tf.Variable(tf.random_uniform([1], -1.0, 1.0), name='b')
```

그림 7. 학습데이터 정의
Figure 7. Definition of training data

다음 <그림 8>에서 비용 함수(cost function)는 위와 같이 가설 데이터와 학습데이터의 \hat{y} 값의 자승의 합을 구해 데이터의 개수만큼 나누도록 한다.

```
# X 와 Y 의 상관 관계를 분석하기 위한 가설 수식을 작성한다.
# y = W * x + b
# W 와 X 가 행렬이 아니다(tf.matmul 이 아니라 기본 곱셈 기호를 사용한다)
hypothesis = X * W + b

# 손실 함수를 작성한다.
# mean(h - Y)^2 : 예측값과 실제값의 거리를 비용(손실) 함수로 정한다.
cost = tf.reduce_mean(tf.square(hypothesis - Y))
# 텐서플로우에 기본적으로 포함되어 있는 함수를 이용해 경사 하강법 최적화를 수행한다.
optimizer = tf.train.GradientDescentOptimizer(learning_rate=0.1)
# 비용을 최소화 하는 것이 최종 목표
train_op = optimizer.minimize(cost)
```

그림 8. 가설과 비용함수
Figure 8. Hypothesis and cost function

이를 위하여 텐서플로우의 tf.square() 함수와 tf.reduce_mean() 함수를 사용한다. 텐서플로우에서는 비용을 최소화하기 위해 최적화 알고리즘인 경사 하강법을 활용한다. 여기에서는 학습도를 0.1로 하여 경사 하강법 알고리즘을 실행한다.

다음 <그림 9>는 텐서플로우의 변수값을 초기화하기 위해 global_variables_initializer() 함수를 사용한다. 최종적으로 그래프가 구현한다. 본 연구에서는 최적화를 위하여 100번을 수행하여 학습을 진행한다.

```
# 세션을 생성하고 초기화한다.
with tf.Session() as sess:
    sess.run(tf.global_variables_initializer())

# 최적화를 100번 수행한다.
for step in range(100):
    # sess.run 을 통해 train_op 와 cost 그래프를 계산한다.
    # 이 때, 가설 수식에 넣어야 할 실제값을 feed_dict 을 통해 전달한다.
    .. cost_val = sess.run([train_op, cost], feed_dict={X: x_data, Y: y_data})

    print(step, cost_val, sess.run(W), sess.run(b))

# 최적화가 완료된 모델에 결과 값을 넣고 결과가 잘 나오는지 확인한다.
print("\n=== Result ===")
print("X: 1.60, Y: ", sess.run(hypothesis, feed_dict={X: 1.60}))
print("X: 1.77, Y: ", sess.run(hypothesis, feed_dict={X: 1.77}))
```

그림 9. 세션 생성과 최적화
Figure 9. Session creation and optimization

5.2 실행 결과

선형회귀 학습을 실행하면 cost의 값이 점점 최소화되고 이에 따라 hypothesis의 W와 b의 값을 조정하게 된다. W와 b의 값에 무작위 데이터 값을 주게 되고, 아래의 학습모델로 봤을 때 W와 b가 수렴하는지 지켜보면 된다. 그 결과 값으로 각 열은 트레이닝 횟수, cost, W, b의 값을 나타낸다. 다음 <그림 10>은 실행 결과를 나타낸다.

```

0 1.153378 [0.56851083] [-0.63250107]
1 0.9854511 [0.6016397] [-0.6126693]
2 0.8420199 [0.6322569] [-0.5943407]
3 0.71951115 [0.6605529] [-0.5774013]
4 0.6148728 [0.6867037] [-0.5617457]
5 0.52549803 [0.7108718] [-0.5472767]
6 0.44916052 [0.73320764] [-0.53390425]
7 0.38395828 [0.75385004] [-0.5215453]
....
93 0.0021755053 [1.005051] [-0.37077117]
94 0.0021754315 [1.0050704] [-0.37075445]
95 0.00217537 [1.0050882] [-0.37073871]
96 0.0021753162 [1.0051045] [-0.37072387]
97 0.0021752683 [1.0051194] [-0.3707099]
98 0.0021752268 [1.005133] [-0.3706967]
99 0.0021751928 [1.0051454] [-0.3706842]
....
=== Result ===
X: 1.60, Y: [1.2375486]
X: 1.77, Y: [1.4084232]
    
```

그림 10. 실험 결과
Figure 10. Experiment results

결과적으로 학습 데이터를 이용한 선형회귀분석을 통하여 한 투수의 보폭이 160cm인 경우 볼 스피드는 123.7km, 또 다른 투수의 보폭이 177cm인 경우 볼 스피드는 140.8km 로 예측할 수 있다.

```

# Show the linear regression result
plt.figure(1)
plt.title('Linear Regression')
plt.xlabel('x')
plt.ylabel('y')
# 주어진 데이터들을 점으로 표시
plt.plot(x_data, y_data, 'ro')
# 예측한 일차함수를 직선으로 표시
plt.plot(x_data, sess.run(W) * x_data + sess.run(b), 'b')
# X=1.60 일때의 계산 값
plt.plot([1.60], answer1, 'go')
# X=1.77 일때의 계산 값
plt.plot([1.77], answer2, 'go')
plt.show()
    
```

그림 11. 선형회귀분석을 통한 예측값 결과
Figure 11. Linear regression analysis results

다음 <그림 12>은 matplotlib 라이브러리를 이용하여 예측된 선형 방정식을 구한 후 그래프를 그려낸다. 빨간 점들은 학습데이터 점(x, y) 이고 그려진 직선은 선형회귀를 통해 구한 모델이다. 녹색 점은 각각 X값이 1.60, 1.77일 경우 예측된 값을 나타낸다.

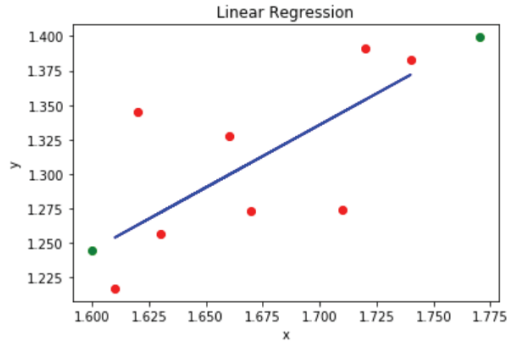


그림 12. matplotlib 라이브러리를 이용한 선형방정식
Figure 12. Linear equation using matplotlib library

6. 결론

본 논문에서는 고등학교 투수의 속구 스피드를 예측하기 위한 선형회귀 분석기법을 연구하였다. 이를 위하여 딥러닝 프레임워크인 텐서플로우를 이용하며 예측모델을 구현하고 성능을 평가하였다. 야구 경기에서 투수의 투구 스피드는 경기의 승과 패를 나눌 수 있는 중요한 요소 중의 하나이다. 그중 투수의 보폭은 투구 스피드와 계구, 벨런스와 밀접한 관계가 있다. 본 연구에서는 투수의 보폭과 투구 스피드에 대한 학습데이터를 생성하고 최적화 알고리즘인 경사 하강법을 이용하여 선형회귀 예측모델을 수행하였다. 이를 통하여 고교야구 선수에 대한 훈련과 평가, 재활에 대해서 도움을 준다.

References

- [1] AlphaGo versus Lee Sedol, https://en.wikipedia.org/wiki/AlphaGo_versus_Lee_Sedol, May 2019.
- [2] J. Moore, *The dartmouth college artificial intelligence conference: the next fifty years*, AI Magazines, Vol. 27, No. 4, pp. 87-91,

- 2006.
- [3] J. Nilsson, *Introduction to machine learning*, Robotics Laboratory, Department of Computer Science, Stanford University, CA, 1998.
- [4] J-W. Kim, and S-H. Park, *The effect of 6-week training program on knees to increase ball speed among collegiate baseball pitchers*, Journal of The Korean Society of Sports Science, Vol. 20, No. 2. pp. 917-934, 2011.
- [5] Y-J. Lee, and J-T. Kim, *The kinematic analysis of the pitching motion for the straight and curve ball*, Journal of The Korean Society of Sports Biomechanics, Vol. 12, No. 2, pp. 109-130, 2002.
- [6] G. S. Fleisig, and R. F. Escamilla, *Kinematic and kinetic comparison between baseball pitching and football passing*, Journal of Applied Biomechanics, Vol. 12, No. 2, pp. 207-224, May 1996.
- [7] R. Whiteley, *Baseball throwing mechanics as they relate to pathology and performance – A review*, Journal of Sports Science and Medicine, Vol. 6, pp. 1-20, 2007.
- [8] D. W. Altchek, and W. R. Hobbs, *Evaluation and management of shoulder instability in the elite overhead thrower*, Journal of Orthopedic Clinics of North America, Vol. 32, pp. 423-430, 2001.
- [9] Korea Baseball Softball Association, <http://www.korea-baseball.com/>, Jun. 2019.
- [10] C. J. Peng, K. L. Lee, and G. M. Ingersoll, *An introduction to logistic regression analysis and reporting*, The Journal of Educational Research, Vol. 96, No. 1. pp. 1-15, 2002.
- [11] Simple and Multiple Linear Regression in Python, <https://towardsdatascience.com/simple-and-multiple-linear-regression-in-python-c928425168f9>, May 2017.
- [12] Tensorflow Tutorials, <https://gist.github.com/haje01/202ac276bace4b25dd3f>, Jun. 2018.
- [13] S. J. Miller, *The method of least squares*, Mathematics Department, Brown University, Rhode Island, 2006.
- [14] J. Dougherty, R. Kohavi, and M. Sahami, *Supervised and unsupervised discretization of continuous features*, Proceeding of the Twelfth International Conference on Machine Learning, pp. 194-202, 1995.
- [15] Tensorflow, <https://www.tensorflow.org/>, Jun. 2019.

선형회귀분석 기법을 이용한 고교야구투수의 투구속도 예측

오영환

나사렛대학교 IT융합학부 교수

요 약

최근에 알파고(AlphaGo)와 같은 인공 지능(Artificial Intelligence)과 기계 학습(Machine Learning)에 대한 연구가 매우 활발히 진행되고 있다. 통계학에서 선형 회귀(Linear Regression)는 종속 변수 y 와 한 개 이상의 독립 변수 x 와의 선형 상관관계를 모델링하는 회귀분석 기법으로 일반적으로 최소제곱법(Least Square Method)을 사용해 선형 회귀 모델을 정립한다. 즉, 선형회귀 분석은 독립변수와 종속변수, 상수항 사이의 관계를 모델링하는 방법으로 두 변수 사이의 관계를 모델링하는 단순 선형회귀 방법과 둘 이상의 독립 변수에 기반하는 다중 선형회귀 방법이 있다. 야구 경기에서 투수는 타자와 상대할 시에 좋은 결과를 내기 위해서 투구 스피드와 제구, 투구 밸런스 등이 좋아야 한다. 고교야구 선수가 훌륭한 투수로 성장하기 위한 중요한 요소는 투구 스피드이다. 또한 투수의 볼 스피드는 야구 경기의 승과 패를 나눌 수 있는 중요한 요

소 중의 하나이다. 본 논문에서는 딥러닝 프레임워크 (Deep Learning Framework)인 텐서플로우(Tensorflow)를 이용하여 고교야구 선수중 투수의 볼 스피드를 측정하여 선수의 운동과 훈련재활을 위해 사용하고자 한다. 본 연구에서는 투수의 보폭과 스피드에 대한 학습데이터를 생성하고 최적화 알고리즘인 경사 하강법을 이용하여 선형회귀 예측기법을 수행한다.

감사의 글

본 논문은 2019년 나사렛대학교 학술연구비 지원에 의해 연구되었음



Young Hwan Oh received the bachelor's degree in the Department of Computer Engineering from the Inha University in 1991. He received the M.S. degree and the Ph.D. degree in the Department of Computer Science and Engineering from Inha University in 1997 and 2001, respectively. From 1992 to 1994, he was a researcher at Korea Power Corporation. He has been a professor in the Department of Information and Communication at Korea Nazarene University since 2002. His current research interests include artificial intelligence, big data, database system. He is a member of the KKITS.

E-mail address: yhoh@kornu.ac.kr