Exploring a Model for Predicting Achievement with LMS Activities

Hyeon Woo LEE(SANGMYUNG University)
Jong Moon LEE(UBION)
Yoon Mi CHA(SANGMYUNG University)
NMC Horizon Report Higher Education Edition(nmc.org)

The NMC Horizon Project is a global ongoing research initiative that explores the trends, challenges, and technology developments likely to have an impact on teaching, learning, and creative inquiry.

2011 / Time to Adoption(4~5 Years)

2012 / Time to Adoption(2~3 Years)

2013 / Time to Adoption(2~3 Years)

2014 / Important Developments in Educational Technology for Higher Education

2015 / Key Trend Accelerating Higher Education Technology Adoption

2016 / Fast Trends: Driving changes in higher education over the next 1 to 2 years

2017 / Fast Trends: Growing Focus on Measuring Learning
[Predictive Analytics]
Expect value of response variable through predictive modeling using learning-related variables

[Structure Discovery]
If theoretical basis or hypothesis considered in advance is not clear, get new information focusing structure or the data itself

[Relation Mining]
Figure out relationship among variables via data

[Visualization]
Deliver information from data mining and analysis with effect and help decision making
Predict learner’s achievement with LMS data.
(Focus on underachiever student)

1. **Learners**: College Student (3,674)
2. **Courses**: Offline Course using LMS (115)
3. **Achievement**: Course final grade
   
   *(Divided into A, B, C groups)*
4. **LMS Data**: Log data, Profile data, Attendance data
COURSEMOS LMS based Moodle

4주차 [9월22일 - 9월28일]
• 학습목표: 행동주의와 인지주의로 학습과정을 설명할 수 있다.
• 주요내용: 학습이론 1 (행동주의와 인지주의)
• 교재방법: 2장 3절 1, 2
• 비고: Online

4주차 학습이론 1: 행동주의 4206
4주차 학습이론 1: 인지주의 3808

토론 1: 내가 생각하는 학습은?
• 동영상을 강의 내용 후, 자신의 경험을 바탕으로 본인이 생각하는 학습을 정리해 보세요. 자신의 경험 혹은 타인의 경험을 사례로 제시하고 이를 바탕으로 본인이 생각하는 학습의 과정을 설명해보세요.
• 학생들은 자신의 과제 별 25일, 오전 9시까지(2발) 하고 다른 동료의 과제에 대해 댓글을 달아주세요(27일 정한까지).
• 자신의 과제를 제시한 후에 동료의 과제를 볼 수 있습니다.

토론 2: Chimpanzee VS Human child learning
• 아래에 제시된 Youtube 동영상을 보고, 느낀 점을 토론해 주세요.
• 자신의 과제를 제시(27일 정한까지)한 후에 동료의 과제를 볼 수 있습니다.
• 동영상 9에서 동영상의 영상이 클릭이 안되는 경우가 있습니다. (eCampus의 문제가 아니라, youtube의 explorer의 문제인듯 설네요)

Chimpanzee VS Human child learning
## Analytics Method Categories

<table>
<thead>
<tr>
<th>Method</th>
<th>based Statistics</th>
<th>Machine Learning</th>
<th>Deep Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>- Regression</td>
<td>- Logistic Regression</td>
<td>- Perceptron</td>
</tr>
<tr>
<td></td>
<td>- Linear Discriminant Analysis</td>
<td>- Cluster Analysis</td>
<td>· Multi-Layer Perceptron</td>
</tr>
<tr>
<td></td>
<td>- Quadratic Discriminant Analysis</td>
<td>- Decision Tree</td>
<td>- Restricted Boltzmann Machine</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Support Vector Machine</td>
<td>- Autoencoder</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Naive Bayes</td>
<td>- Convolutional Neural Network</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Ensemble Learning</td>
<td>- Recurrent Neural Network</td>
</tr>
<tr>
<td></td>
<td></td>
<td>· Bagging</td>
<td>- Long Short Term Memory</td>
</tr>
<tr>
<td></td>
<td></td>
<td>· Boosting</td>
<td>- Generative Adversarial Networks</td>
</tr>
<tr>
<td></td>
<td></td>
<td>· Random Forest</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Feature</th>
<th>- Delicate methodology in mathematics</th>
<th>- Modeling for purpose</th>
<th>- Enable prediction and distribution of complex model, through various activation function and n-tier architecture</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>- Deductive approach</td>
<td>(distribution, cognition)</td>
<td>- Blackbox model(difficult to read)</td>
</tr>
<tr>
<td></td>
<td>- Verifying theory/model</td>
<td>- Deal with new data via learning/evaluation for its purpose</td>
<td>- High predictability</td>
</tr>
<tr>
<td></td>
<td>- Main idea</td>
<td>- Enable to handle large amount of data</td>
<td>- Require data on a large scale</td>
</tr>
<tr>
<td></td>
<td>probability, distribution, example, probability</td>
<td>- Construable model</td>
<td>- Danger of overfitting</td>
</tr>
</tbody>
</table>
Pre Research #1

01. Any difference on LMS interaction(log data) according to learning achievement?

Learning achievement Predictability Search of University Students by Utilizing LMS Interaction Data
(Presented in Joint Symposium of KAEIM* and KSET** in 2018)

Result
1) Interaction(action in LMS) had difference according to grade(grade group).
Group with higher grade had more activities in LMS than that with lower grade.

* Korean Association for Educational Information and Media
** Korean Society for Educational Technology
WEEK 1

C: 11.31
B: 16.18
A: 19.42
02. Is the learning achievement related to activities (learning status check, reading and writing) in LMS? If so, when does the difference appear?

An Analysis on the Difference of LMS Activities according to Academic Achievement of College Students in Off-line Courses
Journal of KAEIM (2019), 25(1)

Result
1) Activity degree such as learning status check, reading and writing was different depending learning achievement. Group with higher grade had more activities in LMS than that with lower grade.
2) From the first week, learning status check and reading activity had meaningful difference. As for writing activity, the difference was clear from the third week.
01. Writing

Obtained Final Grade
- A+: A0
- B+: B0
- C+: C0, D, F

Graph showing grade distribution for W01 to W16.
02. Read

Obtained Final Grade
- A: A+, A0
- B: B+, B0
- C: C+, C0, D, F
03. learning-status check

Obtained Final Grade
- A: A+, A0
- B: B+, B0
- C: C+, C0, D, F
There was difference in LMS action data according to learning achievement group. Then, is it possible to predict learning achievement using action data?
Weekly Model Accuracy

Logistic Regression
Weekly Model Accuracy

Frequency of Groups with grades

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>631</td>
<td>(29.3%)</td>
<td>850</td>
<td>(39.5%)</td>
</tr>
<tr>
<td>671</td>
<td>(31.2%)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Week 9, Writing Graph by Achievement Group
Distribution Chart: Writing in Week9 by Achievement Group
### Analytics Techniques
- Student profile
- Course attendance
- Pre-Semester grade

### Add independent variables
- Logistic Regression
- Decision Tree
- Random Forest
- Support Vector Machine
- Deep Learning
Week 9 / Logistic Regression

Model Accuracy: 43.3% → 57.5%
Recall vs Precision

Recall: 60.0% (3/5)
How many actual underachievers are selected?

Precision: 50.0% (3/6)
How many selected learners are actual underachiever?
Results

Week 9 / Logistic Regression

Model Accuracy: 43.3% → 57.5%

Recall for C group : 49.71% → 72.86%

Precision for C group : 47.91% → 62.65%
Recall: 60.0% (3/5)

How many actual underachievers are selected?

Precision: 50.0% (3/6)

How many selected learners are actual underachiever?

1. Which is more important between recall and precision?
2. Trade-off
3. F1-Score (Harmonic Mean)
Precision by Week

Logistic Regression
Recall by Week

Logistic Regression
Results

Precision and Recall by Week

Logistic Regression

recall

precision
F1-Score by Week

Logistic Regression
Results

Naive Bayes
Decision Tree
Random Forest
Gradient Boosted Trees
Support Vector Machine
Deep Learning (Multi Layer Perceptron)
Results

Precision by Week

Deep Learning (Multi Layer Perceptron)

- Week 1: 68.8
- Week 2: 69.63
- Week 3: 68
- Week 4: 66.39
- Week 5: 68.27
- Week 6: 67.2
- Week 7: 63.93
- Week 8: 60.95
- Week 9: 63.08
- Week 10: 61.52
- Week 11: 59.31
- Week 12: 61.64
- Week 13: 61.57
- Week 14: 60
- Week 15: 49.41
- Week 16: 48.3
Recall by Week

Deep Learning (Multi Layer Perceptron)
Results

Precision and Recall by Week

Deep Learning (Multi Layer Perceptron)

recall

precision

PAGE 30
Results

F1-Score by Week

Deep Learning (Multi Layer Perceptron)

<table>
<thead>
<tr>
<th>Week</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>67.72</td>
</tr>
<tr>
<td>02</td>
<td>65.99</td>
</tr>
<tr>
<td>03</td>
<td>68.00</td>
</tr>
<tr>
<td>04</td>
<td>67.60</td>
</tr>
<tr>
<td>05</td>
<td>68.56</td>
</tr>
<tr>
<td>06</td>
<td>69.78</td>
</tr>
<tr>
<td>07</td>
<td>68.35</td>
</tr>
<tr>
<td>08</td>
<td>66.15</td>
</tr>
<tr>
<td>09</td>
<td>67.98</td>
</tr>
<tr>
<td>10</td>
<td>66.58</td>
</tr>
<tr>
<td>11</td>
<td>65.73</td>
</tr>
<tr>
<td>12</td>
<td>68.18</td>
</tr>
<tr>
<td>13</td>
<td>68.93</td>
</tr>
<tr>
<td>14</td>
<td>68.15</td>
</tr>
<tr>
<td>15</td>
<td>62.33</td>
</tr>
<tr>
<td>16</td>
<td>61.71</td>
</tr>
</tbody>
</table>
Logistic Regression vs Deep Learning (Precision by Week)

Deep Learning (Multi Layer Perceptron)

Logistic Regression

Deep Learning
Results

Logistic Regression vs Deep Learning (Recall by Week)

Deep Learning (Multi Layer Perceptron)

Deep Learning

Logistic R
Results

Logistic Regression vs Deep Learning (F1-Score by Week)

Week

F1-Score

Logistic R
Deep Learning
## Results

<table>
<thead>
<tr>
<th></th>
<th>Logistic Regression</th>
<th>Deep Learning</th>
<th>Decision Tree</th>
<th>Random Forest</th>
<th>Support Vector Machine</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1-Score for C group</td>
<td>68.34%</td>
<td>68.56%</td>
<td>64.51%</td>
<td>68.06%</td>
<td>67.04%</td>
</tr>
<tr>
<td>Precision for C group</td>
<td>66.49%</td>
<td>68.27%</td>
<td>74.64%</td>
<td>62.80%</td>
<td>66.39%</td>
</tr>
<tr>
<td>Recall for C group</td>
<td>70.29%</td>
<td>68.86%</td>
<td>56.80%</td>
<td>74.29%</td>
<td>67.71%</td>
</tr>
<tr>
<td>Model Accuracy</td>
<td>58.7%</td>
<td>58.6%</td>
<td>55.0%</td>
<td>57.1%</td>
<td>57.3%</td>
</tr>
</tbody>
</table>
1. Range of precision is about 61% to 69%.
2. Range of recall is about 58% to 85%.
3. Recall and precision are trade-off relationship.
4. F1-score can be used to evaluate model.
5. In spite of using many analytical techniques, the difference was not significant.