A Feasibility Study of Detecting Incomplete Learners in a Blended Learning Environment

Minoru Nakayama
Hidetomo Kanazawa
Hiroh Yamamoto
CRADLE
Tokyo Institute of Technology
Tokyo, Japan
Learning support for online learning

- Many students do not complete online learning courses.
  - A learning support system is required to avoid this occurrence and to promote effective learning.
  - To reduce impossible participants is to find out who those students are in advance during the learning process.
- Passing a number of progress tests affects a student’s possibility of completing a course (Kougo, Nojima 2004)
- Accurate prediction of incomplete participants
  - Computational models using a database of behavior (Ueno 2007)
  - (A database is required in advance)

  Most online courses are developed through try and error, and the contents are frequently modified.
Purpose

- To examine the feasibility of identifying incomplete participants using current learning behavioral data
  - Access log data without a systematic database
- Access log data for an ordinary univ. course.
  - LMS records an individual accessing history
  - Comparing the characteristics
    - complete / incomplete participants
  - To investigate the possibility of predicting instances of incomplete participants.
Method

Blended learning

- Course: Trends in Information Industry and Society
- Participants: 81 Freshman undergraduates
- Term: Autumn, 2005 (17 weeks including 2 weeks break)
- Blended learning: face to face class session + online modules
  - Learning based on F2F classes and online learning session outside of the class
  - The online module was designed to correspond with each F2F class session
  - The online module was developed for courses that are completely online
  - F2F sessions were held every week
  - Students can access the online module at any time
- Students are informed that the frequency of accessing the online module will be considered in their final assessment.
Method

Analyzing procedure

- 81 participants can be divided into two groups
  - 65 Complete participants
  - 16 incomplete participants, they did not attend the final exams

- The number of access to online modules
  - The course lasted 17 weeks in total
  - There were a number of unique events
    - Guest lecture, Quiz day, etc.
    - The events affected the number of accessing the online modules
  - To compensate for this varying distribution, the number of accesses was standardized using standard deviation
Method

Logged data

1. The number of accesses of the online course per week
2. The cumulative standardized number of accesses of the online course per week
3. The moving average of the number of accesses over three week period
4. The ratio of those taking part in online tests of the weekly modules
5. Standard deviation of the number of accesses
6. The difference in the number of accesses between two weeks
The number of access of online module
The cumulative standardized number of accesses

![Graph showing the accumulation of relative access frequency over weeks for complete and incomplete participants. The graph indicates a significant difference with p<0.05.]
Differences between complete and incomplete participants

- There were significant differences in the cumulative number of accesses between two groups after the 4th week.
- The symptom will appear around 4th week of the course.
  - Some differences in learning performance appear amongst participants who complete the course in accordance with the number of accesses to the online modules.
The relationship between final exam scores and the cumulative standardized number of accesses at 4th week

The number of accesses can be an index of the learning situation, in particular in the data after the 4th week.

The number of accesses is one of the indices indicating learning behavior.
The possibility of predicting incomplete participants

- The task can be defined as a two-class discrimination using the behavioral data
- The various information from the 4th week
  - All 7 of the variables: the number of accesses, the cumulative standardized number of accesses, the ratio of those taking part in online tests of the weekly modules, etc.
- Discrimination analysis was conducted in a stepwise fashion using regression with variable selection
  - Only the ratio of those taking part in online tests, $x_1$, was significant
  - Complete participants: $y>0$, incomplete participants $y<0$

$$y = 7.32 x_1 - 3.52$$

- The threshold of discrimination is 50% for the ratio of those taking part in online tests
Estimation result with the ratio of those taking part in online tests

<table>
<thead>
<tr>
<th>Final result</th>
<th>Estimation</th>
<th></th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Complete</td>
<td>Incomplete</td>
<td></td>
</tr>
<tr>
<td>Complete</td>
<td>55</td>
<td>10</td>
<td>65</td>
</tr>
<tr>
<td>Incomplete</td>
<td>3</td>
<td>13</td>
<td>16</td>
</tr>
<tr>
<td>Total</td>
<td>58</td>
<td>23</td>
<td>81</td>
</tr>
</tbody>
</table>

Chi-square = 27.4 ($p < 0.01$)
The precision (correct rate per estimated incomplete): 57% (13/23)
The recall rate (estimation/total): 81% (13/16)
(fatal) error rate = 0.04 (3/81)
The improvement of estimation - discriminations by three levels -

The ratio of participants taking part in online tests

- The rate is 0:
  - Incomplete participants

- The rate is less than 50%
  - The accumulated standardized number of accesses $x_2$
    
    $y = -4.14 \cdot x_2 - 2.26$

- The rate is more than 50%
  - N of days to access the online modules $x_3$ and $x_2$
    
    $y = 0.21 \cdot x_2 - 0.25 \cdot x_3 + 1.2$
The revised results using multiple discrimination functions for 3 levels

<table>
<thead>
<tr>
<th>Final result</th>
<th>Estimation</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Complete</td>
<td>Incomplete</td>
<td>Total</td>
<td></td>
</tr>
<tr>
<td>Complete</td>
<td>40</td>
<td>25</td>
<td>65</td>
<td></td>
</tr>
<tr>
<td>Incomplete</td>
<td>0</td>
<td>16</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>40</td>
<td>41</td>
<td>81</td>
<td></td>
</tr>
</tbody>
</table>

Chi-square = 19.5 (p < 0.01)
Precision rate (correct rate per estimated incomplete): 39% (16/41)
The recall rate (estimation/total): 100% (16/16)
(fatal) error rate = 0.0 (0/81), Incomplete/Complete = 38% (25/65)
Applying the procedure to other online courses

- To determine the validity, it was applied to participants of two other courses
  - Course A: a complete online course for Bachelors
  - Course B: a blended learning course for Masters

- Course A
  Participants: 15 Freshman undergraduates
  Course: complete online course without F2F sessions
  Complete participants: 9 / Incomplete participants: 6
  The number of online tests: 10
**Course A:**
a complete online course for Bachelors

<table>
<thead>
<tr>
<th>Final result</th>
<th>Estimation</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Complete</td>
<td>Incomplete</td>
</tr>
<tr>
<td>Complete</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>Incomplete</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Total</td>
<td>11</td>
<td>4</td>
</tr>
</tbody>
</table>

Chi-square = 8.18 (p < 0.01)

Precision rate (correct rate per estimated incomplete): 100% (4/4)

The recall rate (estimation/total): 67% (4/6)

(fatal) error rate = 0.13 (2/15)
Applying the procedure to Masters

- Course B
  Participants: 68 Graduate students
  Course: Blended learning
  Complete participants 59 / Incomplete 9
  The number of online test: 10

- These courses were essentially different, but same professor conducted two courses and the topics were similar.
Course B: a blended learning course for Masters

<table>
<thead>
<tr>
<th>Final result</th>
<th>Estimation</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Complete</td>
<td>Incomplete</td>
</tr>
<tr>
<td>Complete</td>
<td>33</td>
<td>26</td>
</tr>
<tr>
<td>Incomplete</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>Total</td>
<td>35</td>
<td>33</td>
</tr>
</tbody>
</table>

Chi-square = 2.89 (n.s.)
Precision rate (correct rate per estimated incomplete) : 21% (7/33)
The recall rate (estimation/total): 78% (7/9)
(fatal) error rate = 0.03 (2/68)
The performance was significant for Freshman. The performance for Masters was not significant, although the learning styles were same with the class used for the development of the procedure.

The estimation performance may depend on the type of group.

This result may be related to the differences in the characteristics of students: Freshman or Masters.

There are some differences in the types of characteristics of students for Freshman and Masters, in particular, the learning strategy is significantly different (Nakayama et al., 2007).

<table>
<thead>
<tr>
<th>Chi-square</th>
<th>Freshman UG</th>
<th>Masters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blended learning</td>
<td>19.5 (p&lt;0.01)</td>
<td>2.9 (n.s.)</td>
</tr>
<tr>
<td>Complete online</td>
<td>8.2 (p&lt;0.01)</td>
<td>N/A</td>
</tr>
</tbody>
</table>
Conclusions

The feasibility of identifying incomplete participants using current learning behavioral data.

- There is a significant difference in the cumulative standardized number of accesses between complete and incomplete participants.
- According to the results of discrimination analysis, an estimation procedure for predicting incomplete participants was developed using the ratio of those taking part in online tests.
  - This result supports the previous report.
  - The procedure was improved using multiple discrimination functions for three groups.
- The possibility of applying the procedure to other classes was also examined.
  - The performance may depend on learner group.
  - This supports that estimation procedures can predict incomplete participants.
Future works

- The development of a more robust estimation procedure.
- The development of supporting methodologies for monitoring and assisting incomplete participants.
Thank you very much for your kind attention