A Nonlinear State Space Model for Identifying At-Risk Students in Open Online Courses

Feng Wang and Li Chen
Department of Computer Science
Hong Kong Baptist University
{fwang, lichen}@comp.hkbu.edu.hk
Outline

• Introduction & Related Work
• Our Methodology
• Experiment & Results
• Conclusions & Future Work
What is MOOC?

MASSIVE
There may be 100k+ students in a MOOC.

OPEN
Anyone, anywhere can register for these courses.

ONLINE
Coursework is delivered entirely over the Internet.

COURSE
MOOCs are very similar to most online college courses.
• **Issue: high dropout rate:** 75% [K. Jordan, 2016]

- There is no negative incentives if students drop out of a MOOC.
- Not everyone feels the need to complete the course.
Research Question

• How to identify at-risk students of dropping out of a course?

• **Motivation**
  • So as to allow intervention before the course completes.

• **Challenges**
  • Diverse engagement patterns
  • Low-intensity participation
Related Work

Various types of feature:

- Clickstream data (e.g., *watching videos, accessing course’s modules, etc.*) [S. Halawa et al., 2014; J. He et al., 2015]
- Quiz performance [C. Taylor et al., 2014; J. He et al., 2015]
- Centrality of students in discussion forums [D. Yang et al., 2013]
- Sentiments of discussion forum posts [D.S. Chaplot et al., 2015]
Related Work, cont.

**Binary classifier:**
- Support Vector Machine (SVM) [M. Kloft, et al., 2014]
- Logistic Regression (LG) [C. Taylor, et al., 2014]
- Survival Model [D. yang, et al., 2013]
- Probabilistic Soft Logic (PSL) [A. Ramesh, et al., 2014]

**Limitation:**
- They assume a student’s dropout probabilities at different time steps are independent. However, usually a student’s state at one time can be influenced by her/his previous state.
Related Work, cont.

**Sequential classifier**
- Simultaneously Smoothed Logistic Regression (LR-SIM) [J. He et al., 2015]
- Hidden Markov Model (HMM) [G. Balakrishnan. 2013]
- Recurrent Neural Network (RNN) [F. Mi and D.-Y. Yeung 2015]

**Limitations:**
- The estimation of next state depends only on the current state;
- The estimated states are deterministic that would lead to error propagation in the estimation procedure;
- The parameters of their models are time-invariant.
Outline

• Introduction & Related Work
• Our Methodology
• Experiment & Results
• Conclusions & Future Work
Contributions

• We implement a Nonlinear State Space Model (NSSM) to address the dropout problem.
  • Students’ states vary over time
• We conduct experiment to compare our method with related ones.
Dropout Prediction Problem
Formulation

• **Sequence classification task**
  
  • *Goal:* to predict whether a student will have activities in the coming week.
  
  • *Dropout:* for current week $t$, if there are activities associated to student $i$ in the coming week, her/his dropout label in the week $t$ is assigned $y_{i,t} = 0$, otherwise $y_{i,t} = 1$. 
Nonlinear State Space Model (NSSM)

NSSM defines *continuous value states* to summarize all the information about a student’s past behavior.

**Properties:**

- Takes into account all of the current and previous states to estimate next state;
- The parameters in NSSM are time varying (*i.e.*, being different at different time steps);
Nonlinear State Space Model (NSSM)

- $s_{i,t}$: a set of random variables with **multivariate Gaussian distribution**
- The student’s latent states evolving over time
  \[ s_{i,t} = Fs_{i,t-1} + Gx_{i,t} + w_{i,t} \]  \hspace{1cm} (1)
- Dropout probability $\pi_{i,t}$:
  \[ \pi_{i,t} = \sigma(h_t^T s_{i,t} + \beta_t^T x_{i,t}) \]  \hspace{1cm} (2)

- Input feature sequence: $(x_{i,1}, x_{i,2}, ..., x_{i,n_i})$
- Dropout label sequence: $(y_{i,1}, y_{i,2}, ..., y_{i,n_i})$
- Latent state sequence: $(s_{i,1}, s_{i,2}, ..., s_{i,n_i})$
<table>
<thead>
<tr>
<th>Features</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>The average number of activities per week by the week $t$.</td>
</tr>
<tr>
<td>$x_2$</td>
<td>The total number of activities in week $t$.</td>
</tr>
<tr>
<td>$x_3$</td>
<td>The average number of sessions per week by the week $t$. $^3$</td>
</tr>
<tr>
<td>$x_4$</td>
<td>The total number of sessions in week $t$.</td>
</tr>
<tr>
<td>$x_5$</td>
<td>The average number of active days per week by the week $t$. $^4$</td>
</tr>
<tr>
<td>$x_6$</td>
<td>The total number of active days in week $t$.</td>
</tr>
<tr>
<td>$x_7$</td>
<td>The average time consumption per week by the week $t$.</td>
</tr>
<tr>
<td>$x_8$</td>
<td>The total time consumption in week $t$.</td>
</tr>
<tr>
<td>$x_9 - x_{15}$</td>
<td>The average number of 7 different types of activity per week by the week $t$.</td>
</tr>
<tr>
<td>$x_{16} - x_{22}$</td>
<td>The total number of 7 different types of activity in week $t$.</td>
</tr>
<tr>
<td>$x_{23} - x_{25}$</td>
<td>The average number of videos watched, wiki viewed and problem attempted per session by the week $t$ respectively.</td>
</tr>
<tr>
<td>$x_{26} - x_{28}$</td>
<td>The average number of videos watched, wiki viewed and problem attempted per session in week $t$ respectively.</td>
</tr>
</tbody>
</table>
**States & Parameters Estimation - EM algorithm**

• **Initialize** each student’s starting latent state $s_{i,0}$ and model parameters $\Phi = \{F, G, h_t, \beta_t\}$

• **Expectation step (E-Step)**
  - *Extended Kalman filter*
    - For $t = 1,2,\ldots,n_i$
      - correct student state $s_{i,t}$ based on the previous $t-1$ observations
  - *Extended Kalman smoother*
    - For $t = n_i, n_i - 1,\ldots,2,1$
      - smooth student state $s_{i,t}^{(t)}$ by considering the entire sequence of the student’s observations

• **Maximization step (M-Step):** update parameters of model $\Phi$ by fixing the student states at different time steps
Outline

• Introduction & Related Work
• Our Methodology
• Experiment & Results
• Conclusions & Future Work
Datasets for Dropout Prediction

- From *xuetangX*\(^1\), one of popular MOOC platforms in China, released in KDD CUP 2015.

<table>
<thead>
<tr>
<th>Item</th>
<th>Statistical description</th>
</tr>
</thead>
<tbody>
<tr>
<td># courses</td>
<td>39</td>
</tr>
<tr>
<td># students</td>
<td>79,186</td>
</tr>
<tr>
<td># enrollments</td>
<td>120,542</td>
</tr>
<tr>
<td># activity logs</td>
<td>8,157,277</td>
</tr>
<tr>
<td># longest lifetime of enrollment</td>
<td>5 weeks</td>
</tr>
</tbody>
</table>

\(^1\) [http://www.xuetangx.com/](http://www.xuetangx.com/)
Compared Methods & Evaluation Metric

**Compared Methods**
- Logistic Regression (LG): a logistic regression classifier for each week [C. Taylor, et al., 2014]
- Simultaneously Smoothed Logistic Regression (LR-SIM): to minimize the difference of the predicted probabilities between two adjacent weeks [J. He et al., 2015]
- RNN with Long Short-Term Memory Cell (LSTM) [F. Mi and D.-Y. Yeung 2015]

**Evaluation Metric:**
- Area Under the Receiver Operating Characteristics Curve (AUC): widely used evaluation metric for classification problem, as it is invariant to imbalance.
- AUC measures how likely a classifier can correctly discriminate between positive and negative samples.
Results: Single Course

- We trained a separate model for each of 6 popular courses that include more than 5,000 students
- 70% early students as the training data, and remaining 30% students as the testing data.

<table>
<thead>
<tr>
<th></th>
<th>LR</th>
<th>LR-SIM</th>
<th>LSTM</th>
<th>NSSM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 1</td>
<td>0.812</td>
<td>0.886</td>
<td>0.891</td>
<td>0.900</td>
</tr>
<tr>
<td>Week 2</td>
<td>0.819</td>
<td>0.876</td>
<td>0.887</td>
<td>0.891</td>
</tr>
<tr>
<td>Week 3</td>
<td>0.807</td>
<td>0.854</td>
<td>0.861</td>
<td>0.870</td>
</tr>
<tr>
<td>Week 4</td>
<td>0.768</td>
<td>0.778</td>
<td>0.786</td>
<td>0.796</td>
</tr>
<tr>
<td>Week 5</td>
<td>0.673</td>
<td>0.679</td>
<td>0.689</td>
<td>0.702</td>
</tr>
</tbody>
</table>

Table 3: Performance comparison of LR, LR-SIM, LSTM and NSSM in terms of average AUC on 6 popular courses.
Results: Across Courses

- Would the proposed model trained on some courses can serve other courses?
- 70% courses for training and remaining 30% for testing.

<table>
<thead>
<tr>
<th></th>
<th>LR</th>
<th>LR-SIM</th>
<th>LSTM</th>
<th>NSSM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 1</td>
<td>0.835</td>
<td>0.933</td>
<td>0.936</td>
<td>0.936</td>
</tr>
<tr>
<td>Week 2</td>
<td>0.911</td>
<td>0.915</td>
<td>0.915</td>
<td>0.919</td>
</tr>
<tr>
<td>Week 3</td>
<td>0.868</td>
<td>0.872</td>
<td>0.867</td>
<td>0.871</td>
</tr>
<tr>
<td>Week 4</td>
<td>0.782</td>
<td>0.784</td>
<td>0.785</td>
<td>0.789</td>
</tr>
<tr>
<td>Week 5</td>
<td>0.655</td>
<td>0.662</td>
<td>0.673</td>
<td>0.686</td>
</tr>
</tbody>
</table>

Table 4: Performance comparison of LR, LR-SIM, LSTM and NSSM in terms of AUC on new courses across weeks.
Outline

• Introduction & Related Work
• Our Methodology
• Experiment & Results
• Conclusions & Future Work
Conclusion & Future Work

• Conclusions:
  • Take advantage of nonlinear state space model (NSSM) to discover a student’s latent state to characterize the student’s intention to perform certain activities
  • The experiment results demonstrate that our proposed model achieves higher prediction accuracy than related methods

• Future Work:
  • Try other advanced algorithms (e.g., Unscented Kalman filter) to estimate the parameters in our nonlinear state space model
  • Evaluate our proposed model on datasets collected from other MOOC platforms, such as Edx and Coursera.
Thank you