The 9th International Conference on Educational Data Mining (EDM2016)

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

香港浸會大學
HONG KONG BAPTIST UNIVERSITY

Outline

- Introduction & Related Work
- Our Methodology
- Experiment & Results
- Conclusions & Future Work

What is MOOC?

M

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

O

OPEN
Anyone, anywhere can register for these courses.

O

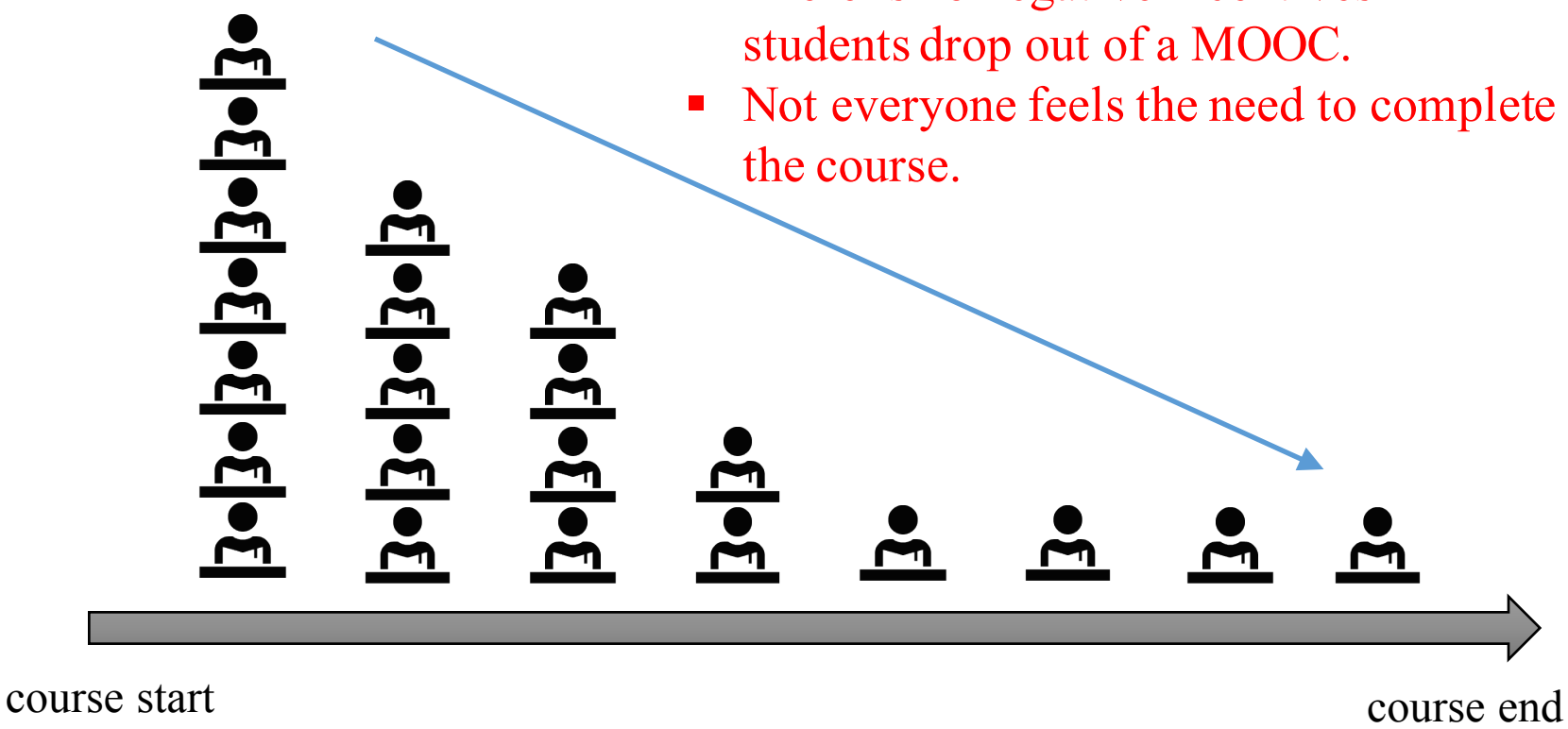
ONLINE
Coursework is delivered entirely over the Internet.

C

COURSE
MOOCs are very similar to most online college courses.

Introduction

- **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 $y_{i,t} = 1$.

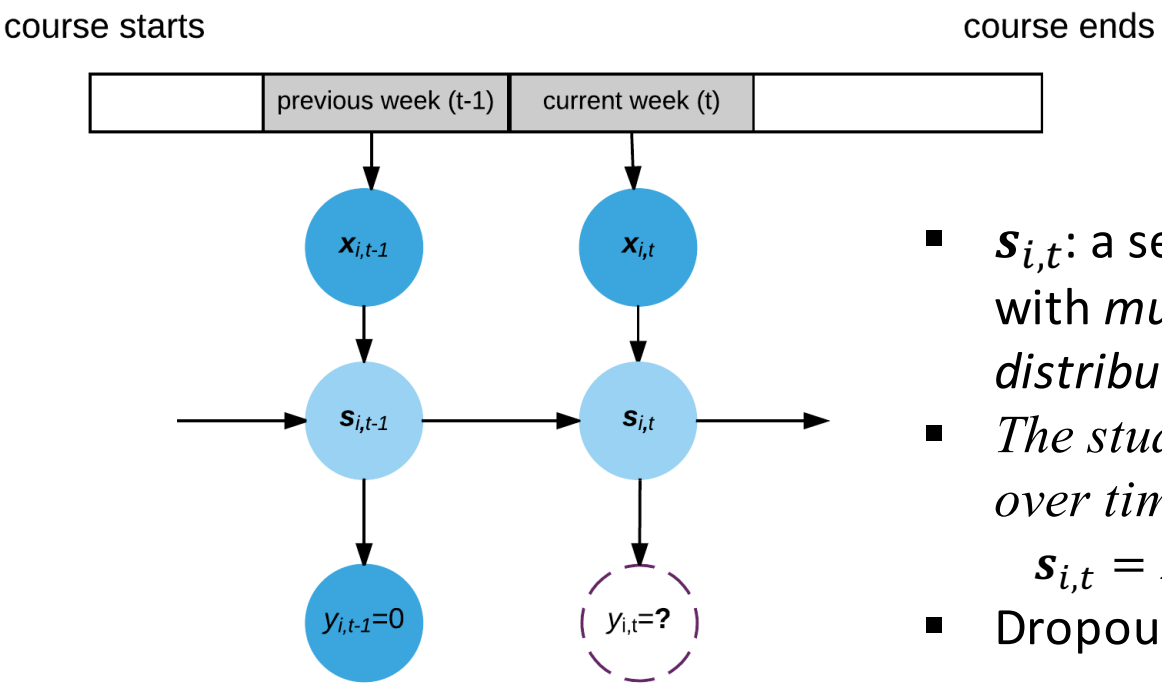
Nonlinear State Space Model (NSSM)

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)



- $\boldsymbol{s}_{i,t}$: a set of random variables with *multivariate Gaussian distribution*
- *The student's latent states evolving over time*

$$\boldsymbol{s}_{i,t} = \boldsymbol{F}\boldsymbol{s}_{i,t-1} + \boldsymbol{G}x_{i,t} + \boldsymbol{w}_{i,t} \quad (1)$$

- Dropout probability $\pi_{i,t}$:

$$\boldsymbol{\pi}_{i,t} = \boldsymbol{\sigma}(\boldsymbol{h}_t^T \boldsymbol{s}_{i,t} + \boldsymbol{\beta}_t^T \boldsymbol{x}_{i,t}) \quad (2)$$

- Input feature sequence: $(\boldsymbol{x}_{i,1}, \boldsymbol{x}_{i,2}, \dots, \boldsymbol{x}_{i,n_i})$
- Dropout label sequence: $(y_{i,1}, y_{i,2}, \dots, y_{i,n_i})$
- Latent state sequence: $(\boldsymbol{s}_{i,1}, \boldsymbol{s}_{i,2}, \dots, \boldsymbol{s}_{i,n_i})$

Table 1: List of features derived from each student's learning activities by the week t

Features	Description
x_1	The average number of activities per week by the week t.
x_2	The total number of activities in week t.
x_3	The average number of sessions per week by the week t. [3]
x_4	The total number of sessions in week t.
x_5	The average number of active days per week by the week t. [4]
x_6	The total number of active days in week t.
x_7	The average time consumption per week by the week t.
x_8	The total time consumption in week t.
x_9 - x_{15}	The average number of 7 different types of activity per week by the week t.
x_{16} - x_{22}	The total number of 7 different types of activity in week t.
$x_{23} - x_{25}$	The average number of videos watched, wiki viewed and problem attempted per session by the week t respectively.
$x_{26} - x_{28}$	The average number of videos watched, wiki viewed and problem attempted per session in week t respectively.

States & Parameters Estimation - EM algorithm

- **Initialize** each student's starting latent state $s_{i,0}$ and model parameters $\Phi = \{\boldsymbol{F}, \boldsymbol{G}, \boldsymbol{h}_t, \boldsymbol{\beta}_t\}$
- **Expectation step (E-Step)**
  - *Extended Kalman filter*
    - For $t = 1,2,\dots,n_i$
      - correct student state $\boldsymbol{s}_{i,t}$ based on the previous $t-1$ observations
  - *Extended Kalman smoother*
    - For $t = n_i, n_i - 1, \dots, 2,1$
      - smooth student state $\boldsymbol{s}_{i,t}^{(t)}$ by considering the entire sequence of the student's observations
- **Maximization step (M-Step)**: update parameters of model Φ 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 2: Statistics of xuetangX dataset for the experiment

Item	Statistical description
# courses	39
# students	79,186
# enrollments	120,542
# activity logs	8,157,277
# longest lifetime of enrollment	5 weeks

[1] 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.

	LR	LR-SIM	LSTM	NSSM
Week 1	0.812	0.886	0.891	**0.900**
Week 2	0.819	0.876	0.887	**0.891**
Week 3	0.807	0.854	0.861	**0.870**
Week 4	0.768	0.778	0.786	**0.796**
Week 5	0.673	0.679	0.689	**0.702**

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.

	LR	LR-SIM	LSTM	NSSM
Week 1	0.835	0.933	0.936	**0.936**
Week 2	0.911	0.915	0.915	**0.919**
Week 3	0.868	0.872	0.867	**0.871**
Week 4	0.782	0.784	0.785	**0.789**
Week 5	0.655	0.662	0.673	**0.686**

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