Student Academic Performance Prediction Using Deep Multi-Source Behavior Sequential Network

Xiang Li, Xinning Zhu, Xiaoying Zhu, Yang Ji and Xiaosheng Tang

Beijing University of Posts and Telecommunications, Beijing, China

2020/4/29

Introduction

Observation

- ◆ Online education is popular
- ◆ Identifying at-risk students is one of the most important issues
- ◆ Large amounts of data are generated

Problem & Motivation

- ◆ Limited data source
- ◆ Ignore the difference of Internet usage pattern

北京郵電大學
Beijing University of Posts and Telecommunications

Introduction

Task

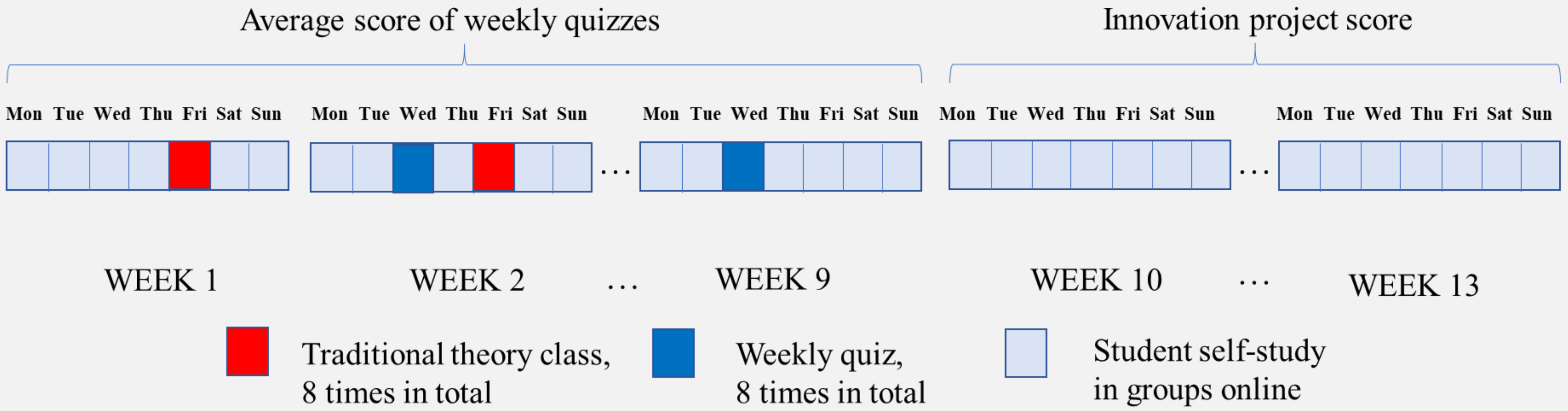
- Extract 505 students' online behavior sequences from two datasets
  - Online learning activities
  - Internet access activity
- Mine distinct behavior patterns & Predict students' performance

北京郵電大學
Beijing University of Posts and Telecommunications

Dataset Description and Insight

Online learning activities *(O)*

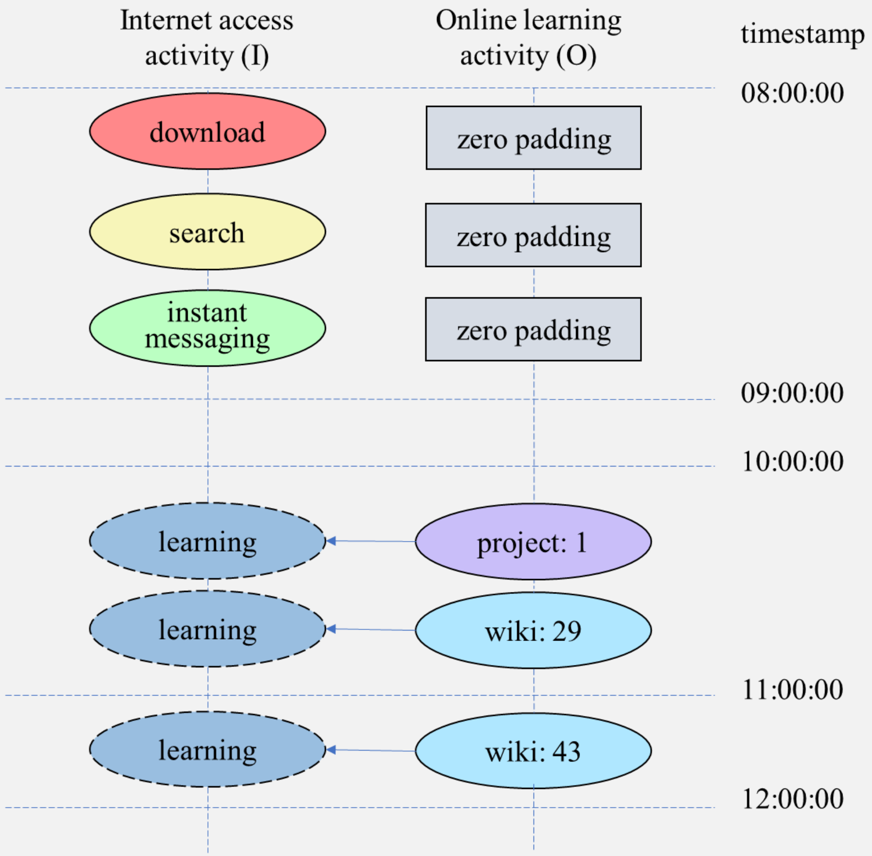
- From a university **project-based course** website
  - Involve View & Create activities
  - Group work
  - At-risk: last 25% of the whole grade



Dataset Description and Insight

Internet access activity *(I)*

- From the log file of campus network
  - Involve categories of URLs and timestamp
  - 11 categories
- Converge online learning activities with current Internet access activities
  - Rename online learning activities as "learning" category
  - Zero padding

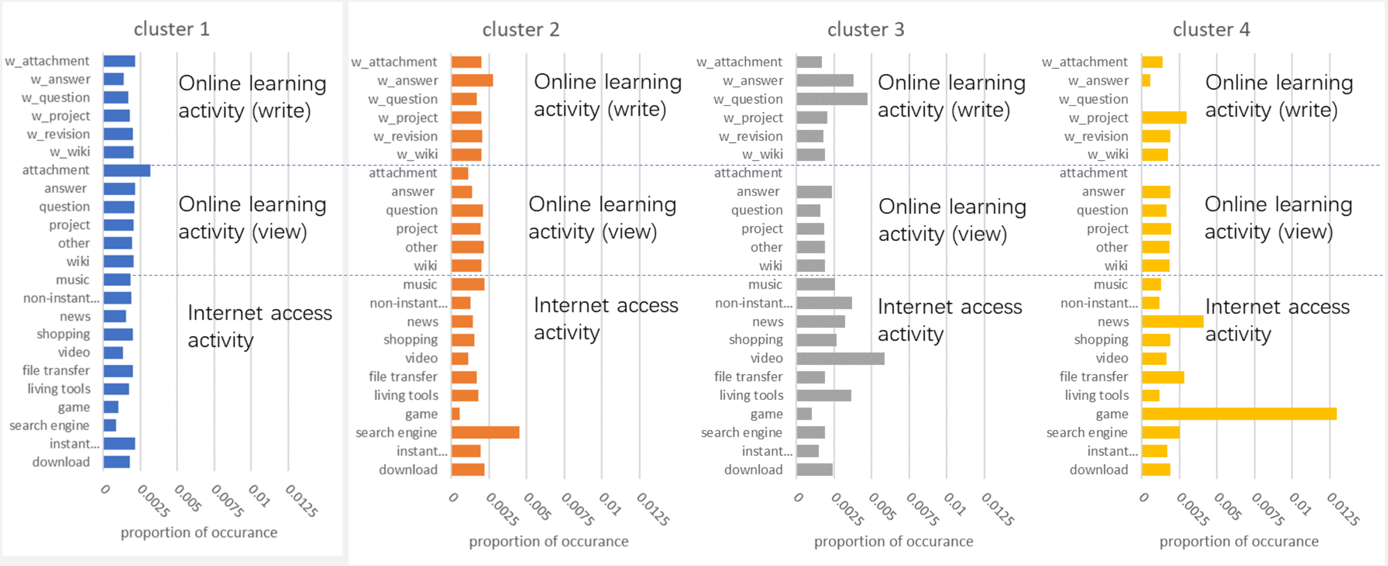


Beijing University of Posts and Telecommunications

Dataset Description and Insight

Distinct Behavior Patterns

- Hierarchical cluster algorithm
  - Feed normalized frequency counts of each activities
  - Find 4 different behavior patterns as parts of static information

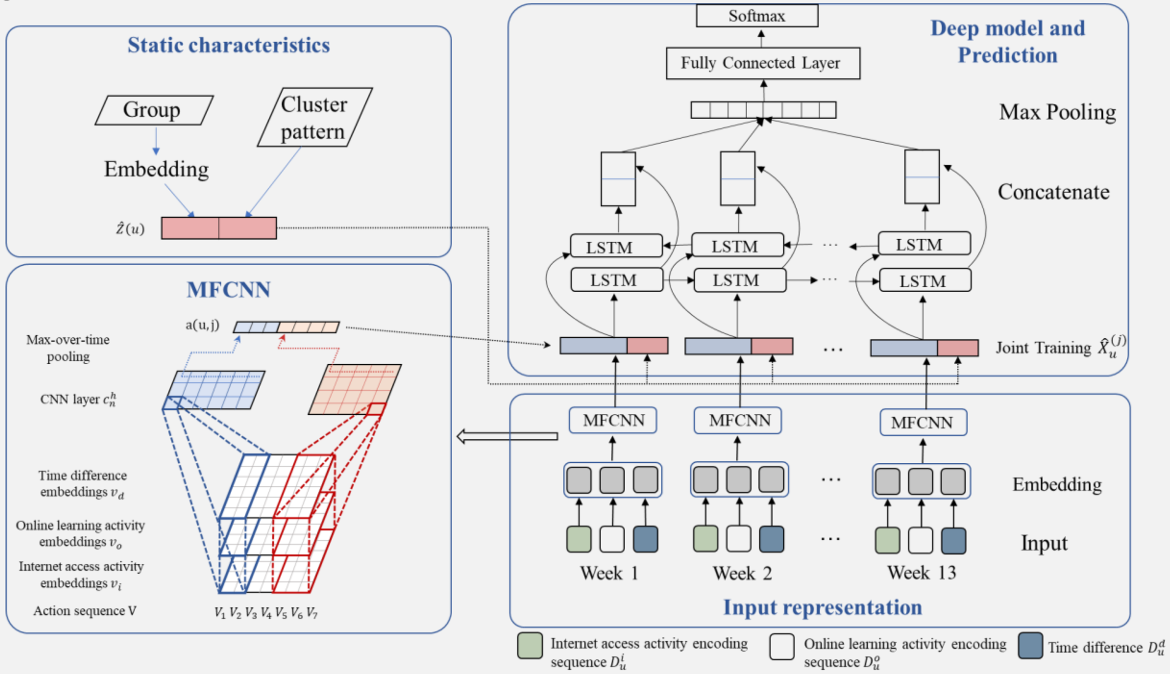


北京郵電大學
Beijing University of Posts and Telecommunications

Framework

Sequential prediction based on deep network (SPDN)

- SPDN uses four components
  1. Construct input behavior sequence
  2. Fuse multi-source data
  3. Join static information
  4. Predict students performance



Framework

SPDN —— Input Representation

- Online learning activity (O)

$$\hat{O}(u) = o_{1:M} = [o_1, o_2, \dots, o_M]$$

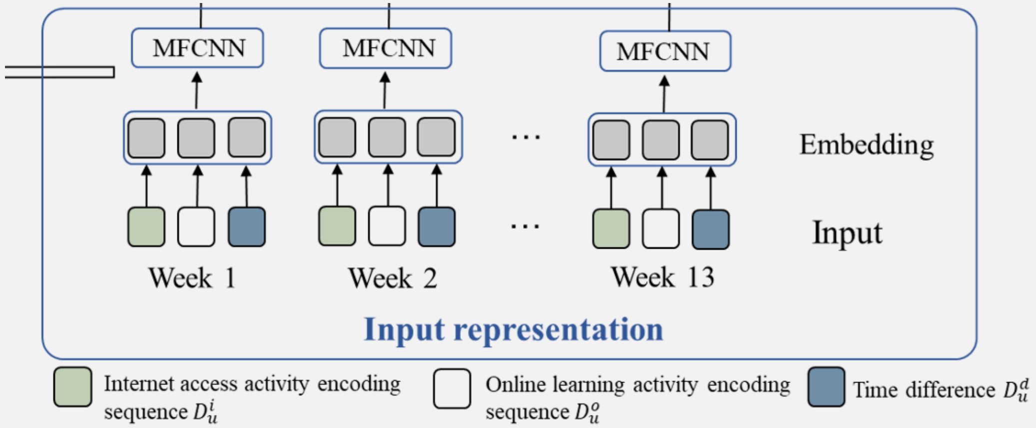
- Internet access activity (I)

$$\hat{I}(u) = i_{1:M} = [i_1, i_2, \dots, i_M]$$

- Time difference(T)

$$\hat{T}(u) = [\Delta d_1, \Delta d_2, \dots, \Delta d_M]$$

$$\Delta d_t = d_{t+1} - d_t$$



Framework

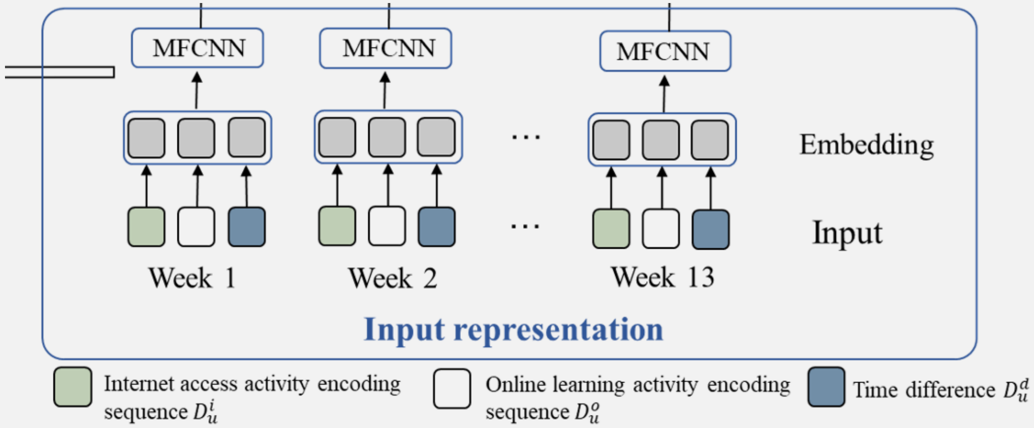
SPDN —— Input Representation

- One-hot encoding

$$\mathrm{l}(a_t^o) \in \{0,1\}^{L_o}$$

$$\mathrm{l}(a_t^i) \in \{0,1\}^{L_i}$$

$$\mathrm{l}(\Delta d_t) \in \{0,1\}^{L_d}$$



- Embedding layer

$$v_o = E_o \cdot \mathrm{l}(a_t^o)$$
$$v_i = E_i \cdot \mathrm{l}(a_t^i)$$
$$v_d = E_d \cdot \mathrm{l}(\Delta d_t)$$

Framework

SPDN ——Multi-source Fusion CNN (MFCNN)

- Compress three embedding sequences
- Input $\mathrm{V} =$

$$[[v_{o1}\ v_{i1}\ v_{d1}][v_{o2}\ v_{i2}\ v_{d2}]\ldots[v_{oM}\ v_{iM}\ v_{dM}]] \in R^{e\times M\times 3}$$

- Convolution

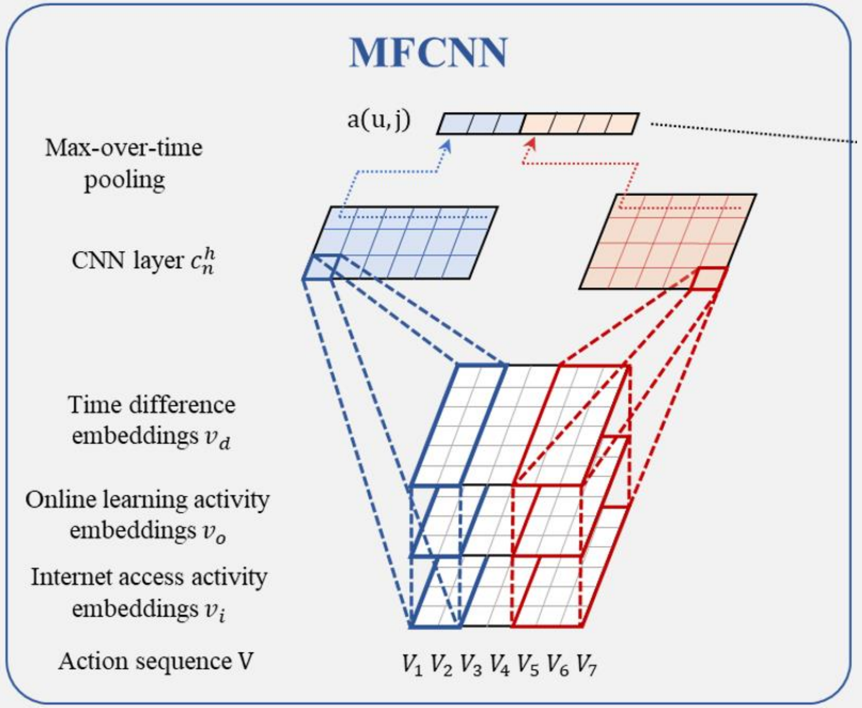
$$c_n^h = f(h * V_{n:n+k-1} + b)\ \ (0 \le n \le\ M-k+1)$$

- Max pooling

$$\tilde{c}^h = \max\{c_1^h, c_2^h, \ldots, c_{M-k+1}^h\}$$

- Output the student u's online behavior representation in $j^{th}(0\ \le j \le 13)$ week

$$\mathrm{a(u,j)} = [\tilde{c}^{h_1}\ \tilde{c}^{h_2}\ \ldots\ \tilde{c}^{h_m}]$$

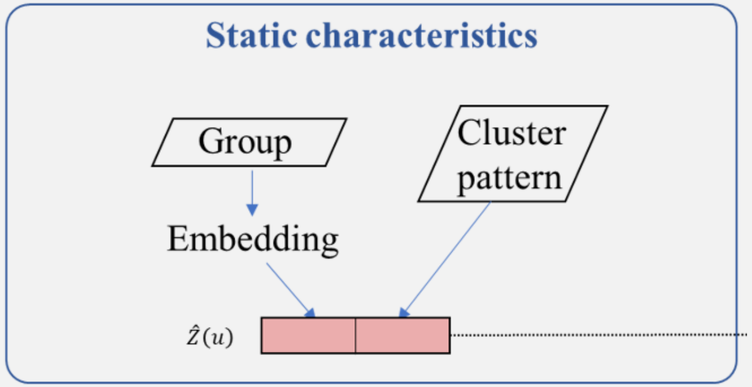


Framework

SPDN —— Static Characteristics Component

- Group id embedding vector v_{Z_g}
- Cluster pattern one-hot vector $\mathbb{1}(Z_p)$
- Static feature vectors

$$\hat{Z}(u) = \left[v_{Z_g} \oplus \mathbb{1}(Z_p)\right]$$

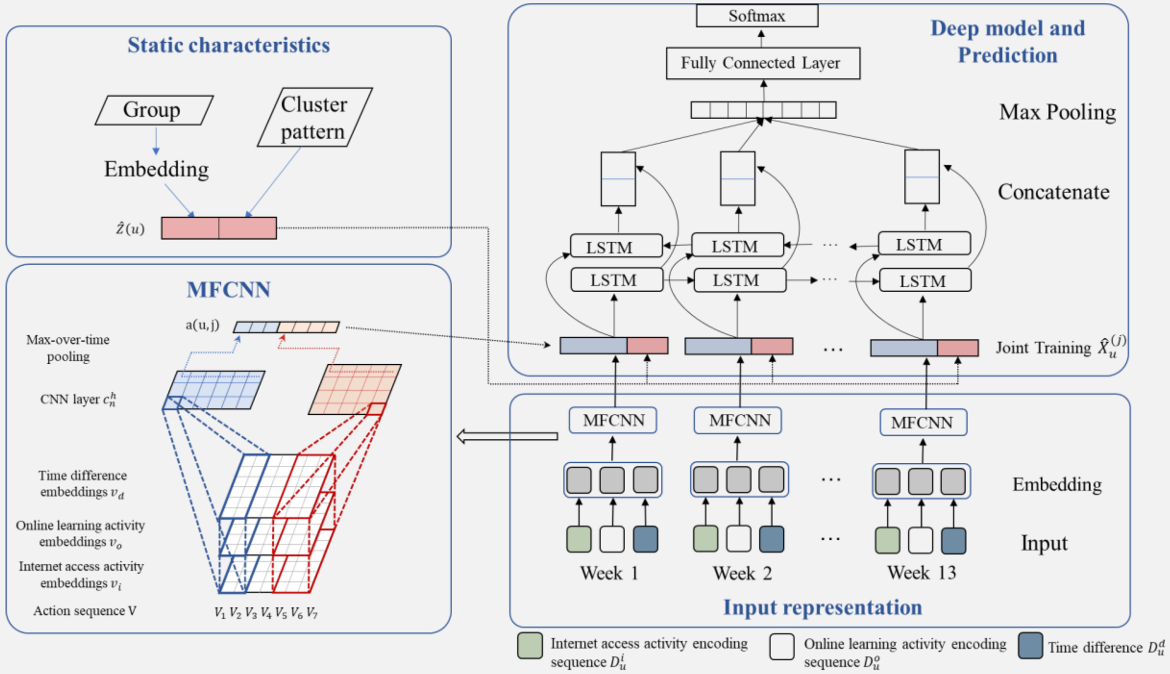


北京郵電大學
Beijing University of Posts and Telecommunications

Framework

SPDN —— Bi-LSTM and Prediction

- Join static feature vectors with weekly behavior representation
- Bi-LSTM layer
- Max pooling layer
- SoftMax layer
- Output students' at-risk probability $\hat{y}(u) \in [0,1]$.
- Training objective function:
  - Cross entropy



$$L(\theta) = -\sum_{u \in U}[y(u)\log(\hat{y}(u)) + (1 - y(u))\log(1 - \hat{y}(u))]$$

Experiments

Prediction Performance

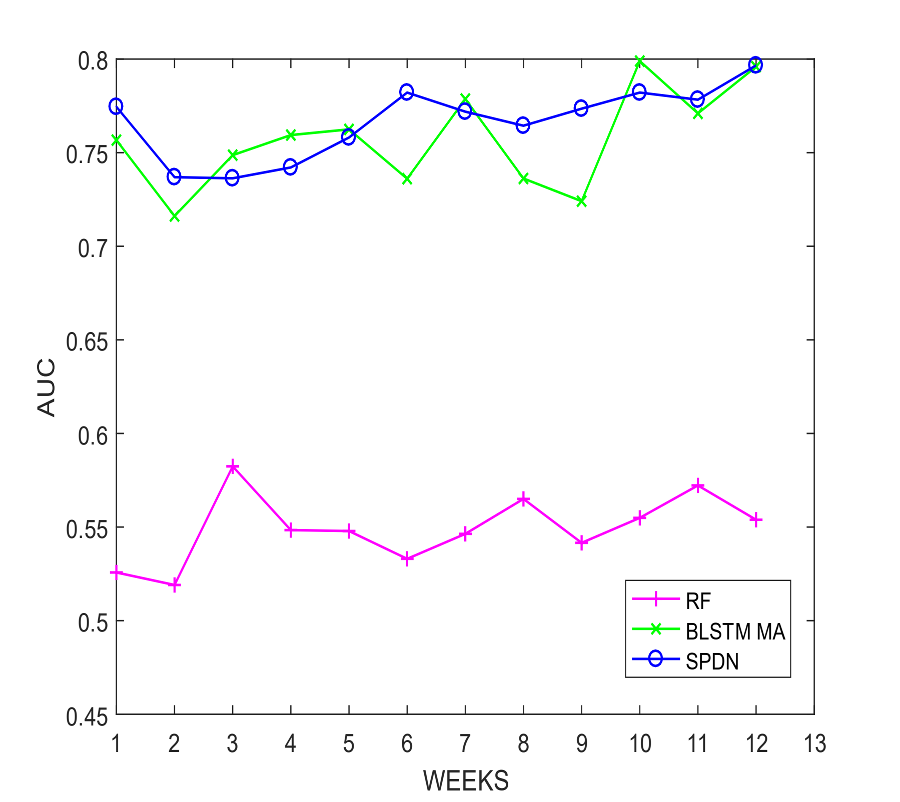
Table 1. Overall results

Approaches	Accuracy (%)	AUC (%)	F1	
			Positive	Negative
SPDN	**73.51**	**79.67**	**0.65**	**0.78**
SPDN_without_MFCNN	70.30	76.31	0.57	0.76
Logistic Regression	52.48	52.20	0.41	0.59
Naive Bayesian	53.27	58.09	0.49	0.57
random forest	61.78	54.64.	0.32	0.73
Decision Tree	65.15	60.09	0.51	0.72

Experiments

Early prediction

- Comparisons of the SPDN and baseline models in terms of mean AUC for early prediction



*BLSTM_MA is the same as SPDN without MFCNN

北京郵電大學
Beijing University of Posts and Telecommunications

Experiments

Factors importance

Table 2. Contribution analysis for different engagement activities

Removed feature	Accuracy	AUC	F1	
			Positive	Negative
Total	0.7129	0.7911	0.66	0.74
Online learning activity	0.7364	0.7831	0.654	0.78
Internet access activity	0.6908	0.7771	0.644	0.728
Static characteristic	0.7128	0.7419	0.62	0.77

Summary

Conclusion

- SPDN makes full use of two datasets and joins the static information to predict students' performance.
- SPDN has considerable improvement over baseline and gets close to the best performance in the first few weeks.
- Internet access activities have a greater impact on students' academic performance prediction .

北京郵電大學
Beijing University of Posts and Telecommunications

THANKS!
Q&A