Deep Learning to Predict Student Outcomes

Hak Kim | AI Team

UDACITY

Backgrounds

The return of the MOOC

- Education is no longer a one-time event but a lifelong experience
- Working lives are now so lengthy and fast-changing
- People must be able to acquire new skill throughout their careers
- MOOC is broadening access to cutting-edge vocational subjects

In the World of MOOCs

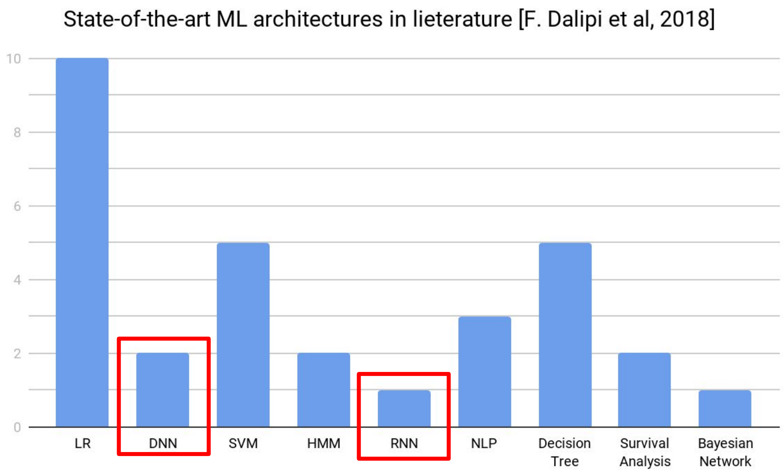
Challenges

- Significant increase in student numbers
- Impracticality to assess each student's level of engagement by *human* instructors
- Increasingly fast dev and update cycle of course contents
- Diverse demographics of students in each online classroom

Student Outcome Prediction

- Challenging problem to forecast the future performance of students as they interact with online coursework
- Reliable *early-stage* prediction of a student's future outcome is critical to facilitate timely (educational) interventions during a course

Only a few prior works have explored the problem from a deep learning perspective!



Student Interaction Trace

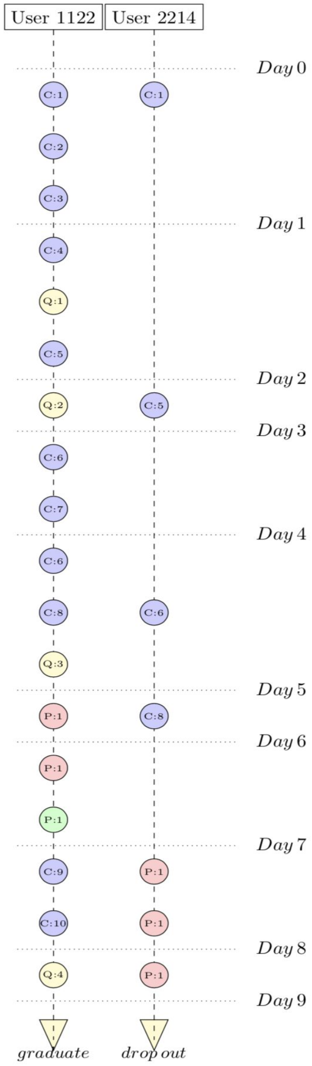
Graduate (*user 1122*)

- First views four lecture videos, then gets a quiz correct
- In the subsequent 13 events, watches a series of videos, answers quizzes and submits first project three times and finally get it passed

Drop-out (*user 2214*)

- Views four lectures sparsely and fails the first project three times in a row

*With this raw student activity data, **GritNet** can make a prediction as to whether or not student would graduate (or any other student outcomes)*



UDACITY

GritNet

Input

- Feed time-stamped student raw events as a sequence (w/o engineered features)
- Encode them by one-hot encoding of possible event tuple *{action, timestamp}* - to capture student's learning speed

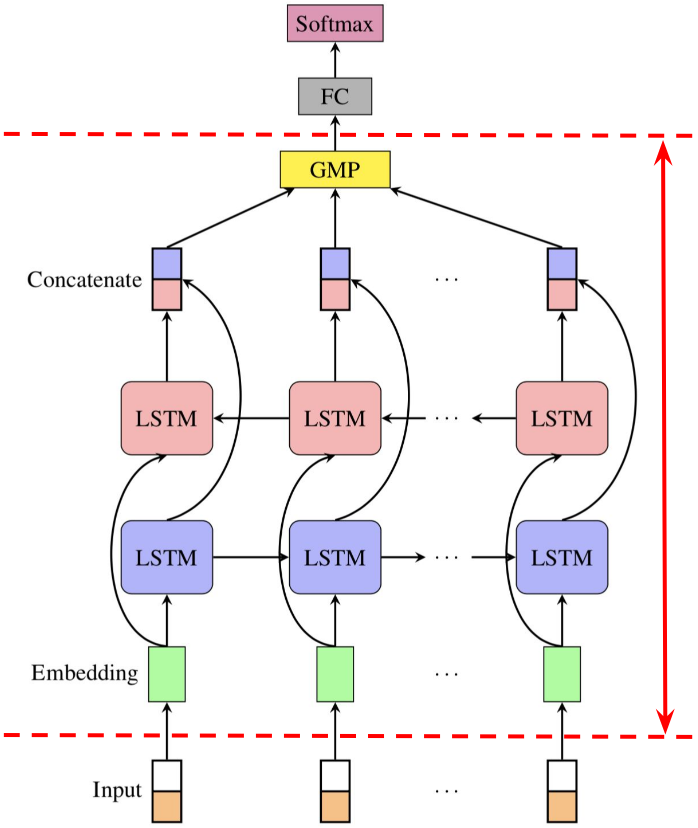
Output

- Probability of graduation or any other student outcomes

Model obs.

- **GMP layer output** captures *generalizable* sequential representation of input event seq. (i.e., most relevant part even effective to deal with imbalanced nature of data)
- **FC layer** learns the course-specific features

***GritNet** can be viewed as a sequence-level embedding extractor plus a classifier*



Real Time Prediction

In model dev.

- GritNet can be trained and tested on the same course data
- **Question**: GritNet models transfer well to different courses or could be deployed for *real-time* prediction with *on-going* courses (before the course finishes)?

Unsupervised domain adaptation

- GritNet trained from previous (source) course but to be deployed on another (target) course
- E.g., v1 → v2, existing → newly-launched ND program

Domain Adaptation

Pseudo-Labels generation

- No target labels from the target course
- Use a trained GritNet to assign hard one-hot labels (Step 4)

Algorithm 1 Domain Adaptation with GritNet

Require: Source course data $\mathcal{X}_{source}$, Source label $\mathcal{Y}_{source}$, Target data $\mathcal{X}_{target}$, Threshold θ

1: Set source training set as $\mathcal{T}_{source} = (\mathcal{X}_{source}, \mathcal{Y}_{source})$
2: Train $GritNet_{source}$ with $\mathcal{T}_{source}$
3: Evaluate on $\mathcal{X}_{target}$: $\hat{\mathcal{Y}}_{pred}$:=$GritNet_{source}(\mathcal{X}_{target})$
4: Assign pseudo-labels to $\mathcal{X}_{target}$: $\mathcal{Y}_{label}$:=$\mathbb{1}(\hat{\mathcal{Y}}_{pred} \geq \theta)$
5: Update target training set as $\mathcal{T}_{adapt} = (\mathcal{X}_{target}, \mathcal{Y}_{label})$
6: Freeze all the $GritNet_{source}$ but the last FC layer and continue training with $\mathcal{T}_{adapt}$

Domain Adaptation (cont'd)

Transfer Learning

- Continue training the last FC layer on the target course while keeping the GritNet core fixed (Step 6)
- The last FC layer limits the number of params to learn

Algorithm 1 Domain Adaptation with GritNet

Require: Source course data $\mathcal{X}_{source}$, Source label $\mathcal{Y}_{source}$, Target data $\mathcal{X}_{target}$, Threshold θ

1: Set source training set as $\mathcal{T}_{source} = (\mathcal{X}_{source}, \mathcal{Y}_{source})$
2: Train $GritNet_{source}$ with $\mathcal{T}_{source}$
3: Evaluate on $\mathcal{X}_{target}$: $\hat{\mathcal{Y}}_{pred}$:=$GritNet_{source}(\mathcal{X}_{target})$
4: Assign pseudo-labels to $\mathcal{X}_{target}$: $\mathcal{Y}_{label}$:=$\mathbb{1}(\hat{\mathcal{Y}}_{pred} \geq \theta)$
5: Update target training set as $\mathcal{T}_{adapt} = (\mathcal{X}_{target}, \mathcal{Y}_{label})$
6: Freeze all the $GritNet_{source}$ but the last FC layer and continue training with $\mathcal{T}_{adapt}$

Very practical solution for faster adaptation time and applicability for even a smaller size target course!

Experimental Setup

Udacity Data

Dataset	Enrolled From	Students	Contents (i)	Quizzes (j)	Projects (k)	AvgSeq. Length	Graduates	Grad. Rates
ND-A v1	2/3/2015	5,626	471	168	4	421	1,202	21.4%
ND-A v2	4/1/2015	2,230	471	168	4	881	453	20.3%
ND-B	3/22/2017	4,377	568	84	10	675	1,726	39.4%
ND-C	1/14/2017	4,761	346	50	5	430	2,198	46.2%

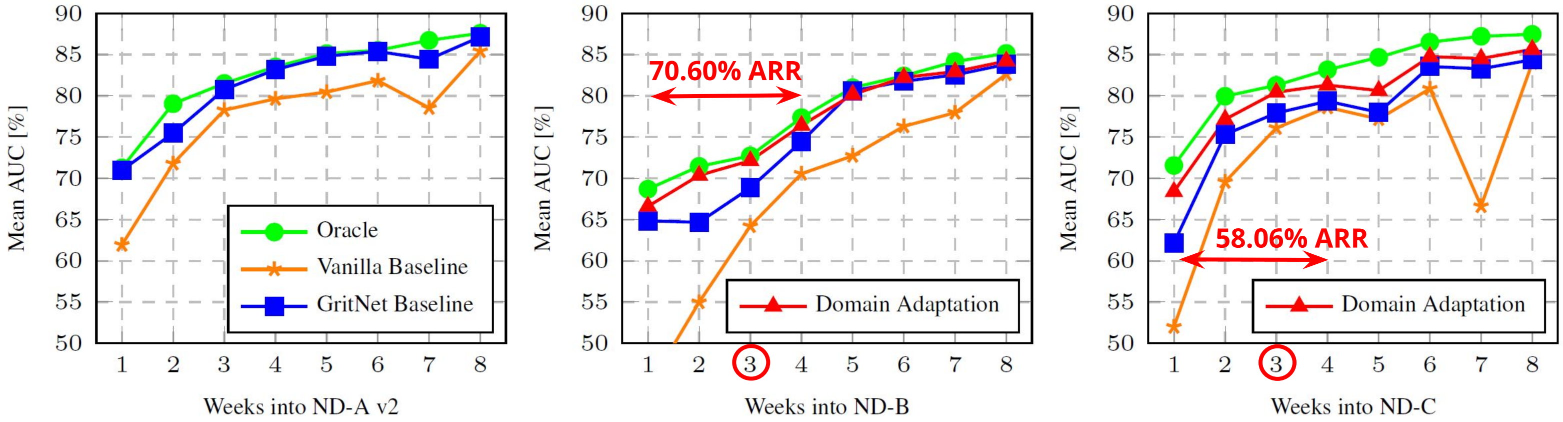
Evaluation Scenarios

- Vanilla baseline: LR model trained on *source* course, evaluated on *target* course
- GritNet baseline: GritNet trained on *source* course, evaluated on *target* course
  - BLSTM layers of 256 cell dimensions per direction and embedding layer of 512 dimension
- Domain adaptation
- Oracle: use oracle target labels instead of pseudo-labels

Generalization Performance

Results

- Mean AUC Recovery Rate (ARR): AUC reduction that unsupervised domain adaptation provides divided by the absolute reduction that oracle training produces
- Clear wins of GritNet baseline vs. vanilla baseline
- Domain adaptation enhances real-time predictions substantially in the first **four weeks**!



Up to **75.07%** ARR at week 3

Up to **84.88%** ARR at week 3

(a) ND-A v1 to ND-A v2

(b) ND-A v1 to ND-B

(c) ND-A v1 to ND-C

Summary

GritNet *is the current state of the art for student outcomes prediction*

- **P1:** Does not need any feature engineering (learns from raw input)
- **P2:** Can operate on any student event data associated with a timestamp (even when highly imbalanced)
- **P3:** Can be transferred to new courses without any (students' outcome) label!

Full papers available online (ICLR'19, EDM'18)

Wrap Up

"This is not the end. It is not even the beginning of the end. But it is, perhaps, the end of the beginning."

#KDD2019 #DL4Ed Workshop

2019 KDD Workshop on Deep Learning for Education (DL4Ed)

CFP at *ml4ed.cc/2019-kdd-workshop*

We are happy to announce that we will be organizing a half-day workshop at KDD 2019, as part of the Deep Learning Day. Details on the workshop will be updated soon. KDD 2019 will be held in Anchorage, Alaska, USA during August 4-8, 2019. See here.

Thank You!