# Applying Recurrent Neural Network Models to the Assessment of Problem-solving Skills

Karen Wang, Max Arseneault {kdwang, marsenea} @stanford.edu

## Predicting

This project explores how deep learning algorithms could be applied to assess college students' scientific problem-solving skills using log data generated in an interactive circuit simulation.

Specifically, **we investigated the performance of different deep learning algorithms using sequences of students' interactions as features** to predict their problem-solving performance as measured by the solution scores.

The Recurrent Neural Network (RNN) model achieved relatively high performance as measured by accuracy, with improvements needed in the recall/precision metrics for the positive class.

## Data

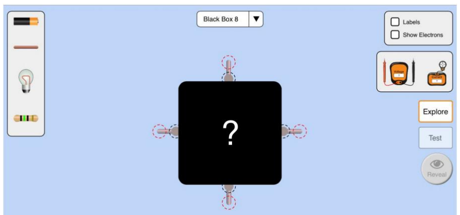



*Figure 1. Illustration of the Black Box Simulation*

- A group of college students participated in the study using the PhET Circuit Construction Kit Black Box (*Figure 1*), yielding a valid dataset of 178 samples
- Log files of individual participants' interactions were parsed into a sequence of time-stamped *events*
- The predicted variable is participants' problem-solving performance as measured by the solution score (0 - low performing, 1 - high performing)

| Feature | Description |
|---|---|
| add/delete wire | add a wire to / delete it from the circuit |
| add/delete lightbulb | add a lightbulb to / delete it from the circuit |
| add/delete resistor | add a resistor to / delete it from the circuit |
| add/delete battery | add a battery to / delete it from the circuit |
| add/delete circuit loop | connect multiple components to form a circuit loop / cut the nodes to disconnect |
| voltage measurement | use voltmeter / battery info to get voltage |
| current measuresment | use ammeter to get current |

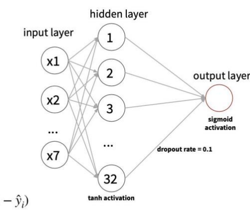
## Models

- **Baseline Neural Network**

**Input layer -** 7 features
**Hidden layer** - 32 units
**Output layer -** 1 unit
- 0 - low problem-solving skill
- 1 - high problem-solving skill

**Cost function:** binary cross entropy

$$BCE = -\frac{1}{N}\sum_{i=0}^{N} y_i \cdot log(\hat{y}_i) + (1-y_i)\cdot log(1-\hat{y}_i)$$

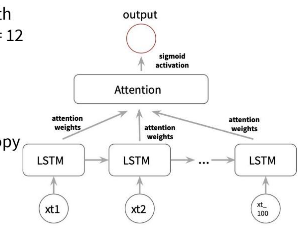



- **Recurrent Neural Network with LSTM & Attention**

**Input layer -** max sequence length
Tx = 100, event_dict_size = 12
**LSTM layer** - 16 units
dropout rate = .1

**Post-LSTM Attention layer**
**Cost function:** binary cross entropy

**Train-Dev split:** 80-20




## Results

| *Model* | Baseline Neural Net | RNN + LSTM | RNN + LSTM + Attention |
|---|---|---|---|
| Training accuracy | 0.98 | 0.94 | 1.0 |
| Test accuracy | 0.78 | 0.64 | 0.72 |
| Precision (class 0/1) | 0.83/0.50 | 0.69/0.25 | 0.74/0.60 |
| Recall (class 0/1) | 0.89/0.38 | 0.88/0.09 | 0.92/0.27 |
| F1 score (class 0/1) | 0.86/0.43 | 0.77/0.13 | 0.82/0.37 |

$Precision = \frac{TP}{TP+FP}$ $Recall = \frac{TP}{TP+FN}$ $F1 = \frac{2*Precision*Recall}{Precision + Recall}$

## Discussion & Future

- Deep learning algorithms (e.g., RNN) hold promise to model students' problem-solving skills based on large-scale log data. However, the performance of the algorithm is constrained by the small data set size
- More work is needed to improve the precision and recall of the algorithms, especially for the prediction of positive (successful) cases

## Reference & Acknowledgements

[1] Piech, Chris, et al. "Deep knowledge tracing." *Advances in neural information processing systems*. 2015.

The project received advice and guidance from Professor Carl Wieman at the Graduate School of Education and Department of Physics, and Ana-Maria Istrate, a research associate student in the Wieman Group.