By Marshall Grimmett, Ayse Arslan, David Chen

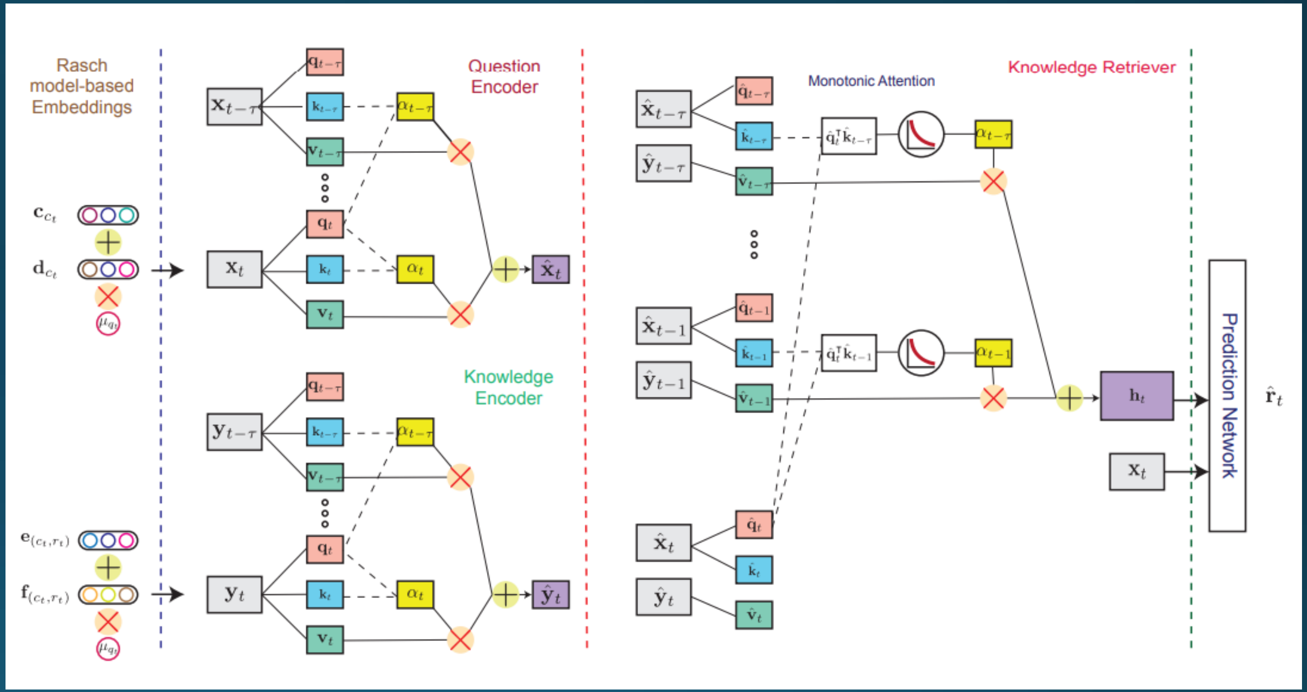
Knowledge Tracing with Neural Nets

Context-Aware Knowledge Tracing (AKT)

- A proposed solution for the limited interpretability problem with deep neural network-based models in KT.
- Use of a novel monotonic attention mechanism
- Use of exponential decay, context aware relative distance measure and similarity measure.
- Use of Rasch model

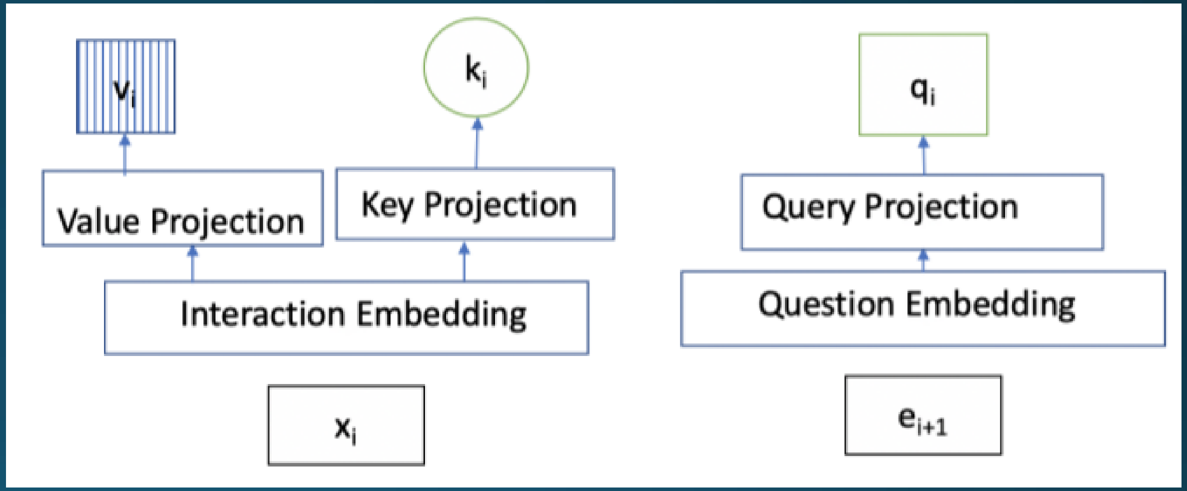
Four important components:

- **Two self-attentive encoders: questions and knowledge acquisition**
- **A single attention-based knowledge retriever**
- **A feed forward response prediction model**



Self-Attentive Knowledge Tracing (SAKT)

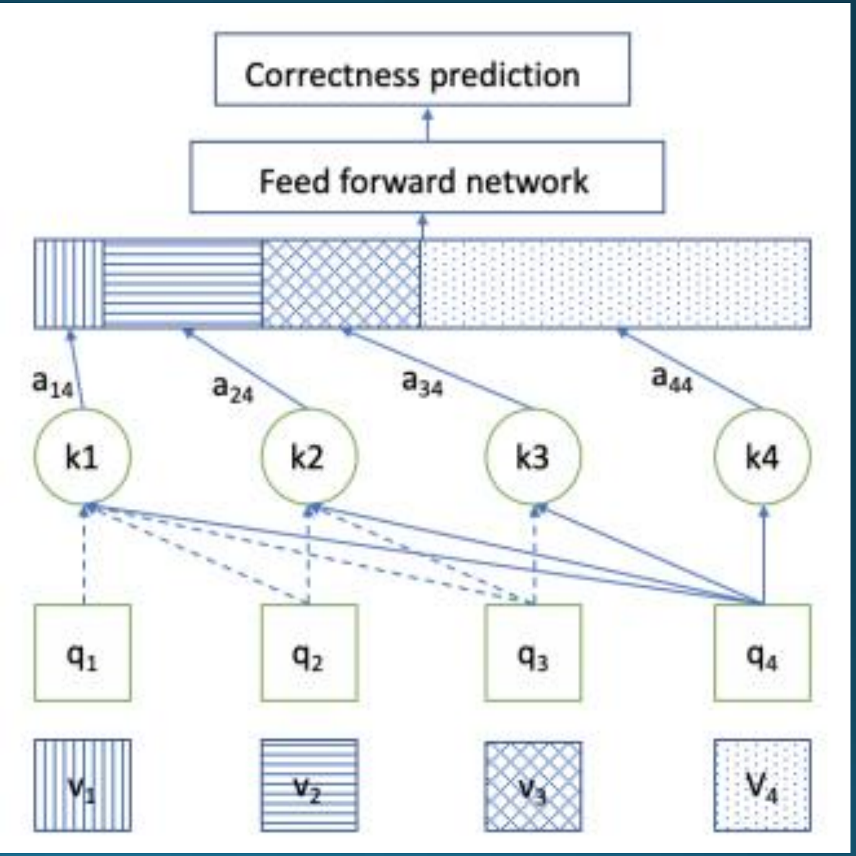
- SAKT is a method where we choose KC and find all other KC that has a relationship with that KC. With all these KC, we would be weighed each KC differently depending on their importance to KC we choose. Then we would calculate the student's performance bas on the student prior experience
- Comparing to other KT, SAKT is running on a parallel base model compare to the recurrent neural network (RNN) where other models run on



Self-Attentive Knowledge Tracing

Sakt has different layer during computation

- Embedding layer- where the all-student knowledge is stored
  during computation.
- Self-attention layer: due to each KC has different weight. Self-attention layer would calculate the weight of each data on the result
- Feed Forward layer -this layer would take the multi-layer result from self- attention layer and combine in into one layer
- Prediction layer: this would take the result from Feed Forward layer and calculate the performance of the student



Dataset Processing

- Column selection and data cleaning
- K fold cross validation and train/test/split
- AKT/SAKT/DKT/DKVMN data format
  - Student Id
  - Questions
  - KCs
  - Responses
- DKT+ data format
  - Student Id
  - Questions
  - Responses

Dataset	Learners	Problems	Knowledge Concepts	Responses
ASSISTments2005-2006	2833	1187	85	323,388

Table 1: Dataset statistics


```
1402
355,355,355,355,355,2556,2556,2556,2556,333,754,626,226,1384,1384,1384,1444,1444,1444,1444,1444,1444,1444,
45,79,45,79,79,16,16,16,16,6,15,9,36,14,14,14,65,65,53,53,53,53,65,76,76,13,13,13,13,13,76,76,76,76,21,9,9,
1,1,0,1,0,0,0,1,0,1,0,0,0,1,0,0,1,0,1,1,1,0,0,1,0,1,1,1,0,0,1,1,1,0,1,1,1,1,0,1,0,0,0,0,1,1,1,1,0,1,0,1,0,
1369
326,920,2650,2650,2650,5758,5758,5758,2247,2247,2247,2247,2247,2247,2247,2247,2247,2247,2247,2247,2247,
50,2,57,26,26,7,7,7,21,21,21,21,21,21,21,21,21,21,21,21,21,21,21,45,45,46,43,28,28,28,28,28,28,28,28,
0,1,1,0,0,1,1,0,1,1,1,1,0,1,0,1,0,1,0,0,0,0,1,0,0,1,0,0,1,1,1,1,1,1,1,0,1,0,0,1,1,0,1,1,1,1,0,1,0,1,1,1,0,
38
3146,3146,3146,3146,3146,1439,3693,2449,626,626,626,626
22,22,22,22,22,48,55,16,9,9,9,9
1,0,1,0,0,0,0,0,1,1,1,0
1504
3146,2642,5738,467,49,49,49,49,49
22,29,1,6,6,6,44,44,44
0,1,0,0,1,0,1,0,0
410
2556,2556,2556,2556,2556,3146,2642,5738,5738,5738,2556,2556,2556,2556,2556,2556,2556,2556,2556,2556,2556,
16,16,16,16,16,22,29,1,1,1,16,16,16,16,16,16,16,16,16,16,16,16,16,16,16,16,16,16,16,16,16,16,16,16,16,16,
0,0,1,0,1,0,0,1,1,0,1,1,1,1,0,1,1,1,0,1,1,1,1,1,1,0,1,1,1,1,1,1,0,1,1,1,1,1,1,1,0,1,0,0,0,1,1,0,1,0,0,1,1,
665
31,5513,355,2624,2635,4731,4731,4731,4731,4731,4731,615,615,615,615,615,615,615,615,615,615,615,615,2600,
45,38,79,26,16,57,75,57,57,57,57,21,21,80,80,80,80,21,27,27,27,27,27,59,16,21,66,68,68,68,68,68,68,12,12,
```

Experiments and Results

- K Fold Cross Validation with 60/20/20 train/validate/test split
- Evaluation metric (AUC)
- Low performance with SAKT
- DKT+ outperformed the others
- AKT performed second best
- Rasch embeddings improved results
- SAKT ablation study showed small changes between parameters.

Dataset	Learners	Problems	Knowledge Concepts	Responses
ASSISTments2005-2006	2833	1187	85	323,388

Table 1: Dataset statistics

(Table 2) Main results

DKT	DKT+	DKVMN	SAKT	AKT-NR	AKT-R
0.6099	0.8214	0.6555	0.5692	0.6742	0.6928

(Table 3) With and without Rasch embeddings

DKT	DKT-R	DKT+	DKVMN	DKVMN-R	SAKT	SAKT-R	AKT-NR	AKT-R
0.6077	0.6099	0.8214	0.6326	0.6555	0.5698	0.5698	0.6742	0.6928

(paper 1)

(Table 4) SAKT ablation study

Architecture	SAKT
Default	0.5698
No Dropout	0.5697
Single head	0.5696
0 blocks	0.5664
2 blocks	0.5696

Conclusion

- Compared the performance of different Neural Network models
- AKT performed better than all models, except DKT+
- SAKT performed poorly
- Rasch embeddings improved the models
- Future work:
  - Run DKT+ with Rasch embeddings and K-fold cross validation
  - SAKT needs improvement