EXTRACTING PREREQUISITE RELATIONSHIPS WITH DKT

Sida Gao

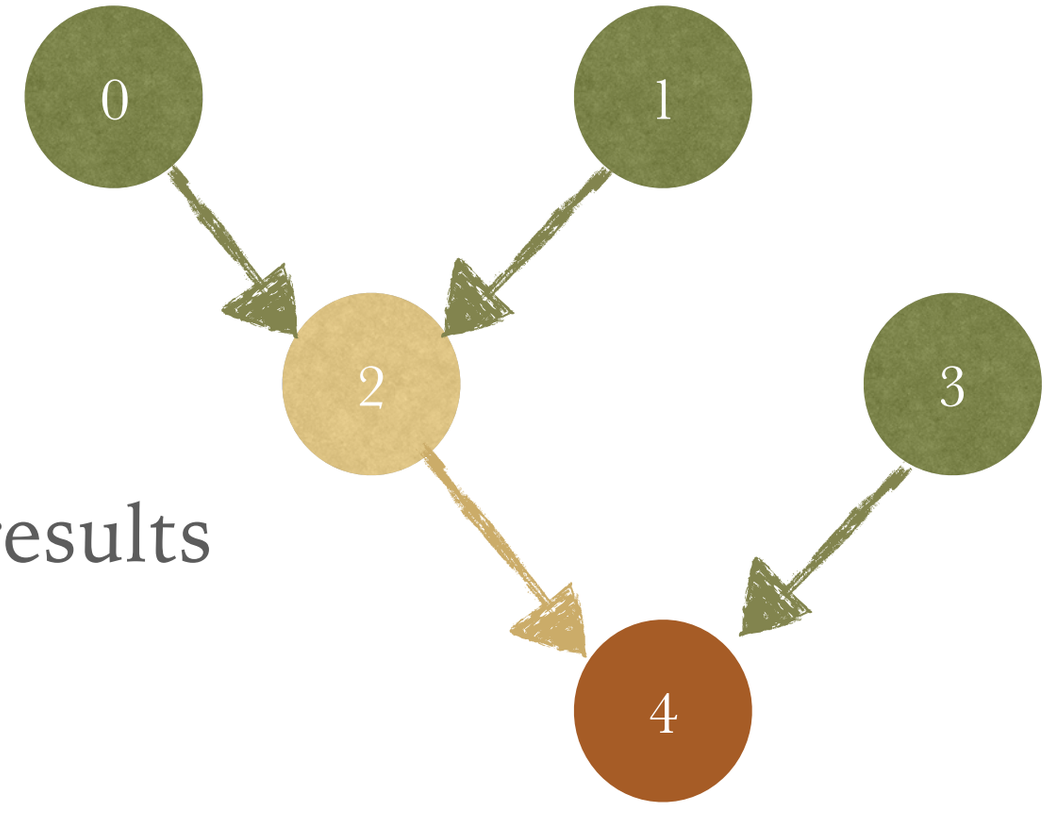
2016/8/26

DEEP KNOWLEDGE TRACING

- LSTMs, with four gates, are like BKTs
- Inner states: model the correlations between skills
- Better to implement from scratch
- Trainable and readable initial states

SANITY CHECK

- BKT Simulation: with prerequisite relationships
- Oracle BKT inference model
- Visualization of the tracing results



	q1|0	q4|0	q3|0	q0|1	q4|1	q3|1	q1|0	q3|0	q1|0	q3|1	q1|1	q1|1	q4|0	q4|0	q4|1	q4|0	q1|1	q1|1	q1|1	q2|0	q2|0	q4|1	q2|1	q2|1	q1|1	q4|0	q3|1	q4|0	q0|1	q1|1	q1|1	q1|1	q4|0	q4|0	q1|1	q3|1	q2|1	q4|1	q4|1	q0|1	q2|1	q2|0	q4|1	q2|1	q2|1	q2|0	q2|1	q2|1	q1|1	q1|1
0	0.38	0.43	0.43	0.43	0.83	0.83	0.85	0.83	0.83	0.83	0.85	0.83	0.86	0.87	0.87	0.85	0.87	0.85	0.87	0.86	0.84	0.8	0.82	0.89	0.93	0.9	0.89	0.89	0.9	0.88	0.92	0.89	0.91	0.9	0.9	0.89	0.91	0.93	0.91	0.92	0.91	0.93	0.91	0.92	0.92	0.93	0.91	0.93	0.93	0.92
1	0.25	0.4	0.47	0.46	0.42	0.41	0.41	0.46	0.49	0.47	0.45	0.74	0.88	0.9	0.89	0.88	0.89	0.89	0.92	0.9	0.92	0.9	0.9	0.91	0.92	0.9	0.91	0.87	0.93	0.86	0.93	0.88	0.92	0.9	0.9	0.89	0.9	0.91	0.9	0.91	0.89	0.91	0.91	0.92	0.9	0.92	0.91	0.92	0.9	0.91
2	0.21	0.16	0.28	0.23	0.25	0.27	0.26	0.22	0.24	0.22	0.21	0.24	0.25	0.26	0.27	0.32	0.29	0.3	0.29	0.3	0.32	0.35	0.39	0.63	0.81	0.82	0.8	0.75	0.8	0.77	0.85	0.8	0.83	0.79	0.79	0.79	0.77	0.85	0.87	0.89	0.85	0.87	0.85	0.88	0.86	0.88	0.88	0.88	0.87	0.88
3	0.33	0.31	0.41	0.58	0.57	0.57	0.81	0.81	0.83	0.79	0.89	0.91	0.92	0.92	0.91	0.89	0.9	0.9	0.91	0.89	0.86	0.85	0.88	0.9	0.89	0.88	0.89	0.89	0.92	0.88	0.93	0.89	0.92	0.89	0.91	0.91	0.91	0.91	0.9	0.93	0.92	0.92	0.88	0.93	0.91	0.92	0.88	0.92	0.9	0.91
4	0.21	0.18	0.19	0.19	0.19	0.22	0.23	0.22	0.21	0.22	0.2	0.21	0.21	0.23	0.22	0.28	0.24	0.23	0.21	0.21	0.17	0.19	0.24	0.24	0.26	0.26	0.28	0.3	0.29	0.32	0.3	0.31	0.31	0.31	0.3	0.29	0.32	0.31	0.52	0.75	0.72	0.71	0.66	0.8	0.77	0.77	0.74	0.78	0.74	0.74

(a) DKT result on student 9999.

	q1|0	q4|0	q3|0	q0|1	q4|1	q3|1	q1|0	q3|0	q1|0	q3|1	q1|1	q1|1	q4|0	q4|0	q4|1	q4|0	q1|1	q1|1	q1|1	q2|0	q2|0	q4|1	q2|1	q2|1	q1|1	q4|0	q3|1	q4|0	q0|1	q1|1	q1|1	q1|1	q4|0	q4|0	q1|1	q3|1	q2|1	q4|1	q4|1	q0|1	q2|1	q2|0	q4|1	q2|1	q2|1	q2|0	q2|1	q2|1	q1|1	q1|1
0	0.41	0.41	0.41	0.41	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9
1	0.27	0.25	0.25	0.25	0.25	0.25	0.25	0.24	0.24	0.24	0.24	0.69	0.86	0.86	0.86	0.86	0.86	0.89	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9
2	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.24	0.24	0.24	0.61	0.79	0.79	0.79	0.79	0.79	0.79	0.79	0.79	0.79	0.79	0.79	0.79	0.84	0.84	0.84	0.84	0.85	0.84	0.84	0.85	0.85	0.85	0.85	0.85	0.85
3	0.34	0.34	0.34	0.31	0.31	0.31	0.8	0.8	0.64	0.64	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9
4	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.22	0.21	0.21	0.21	0.21	0.21	0.21	0.25	0.25	0.25	0.25	0.23	0.23	0.22	0.22	0.22	0.22	0.22	0.22	0.22	0.22	0.22	0.22	0.51	0.75	0.75	0.75	0.75	0.83	0.83	0.83	0.83	0.83	0.83	0.83

(b) Oracle BKT result on student 9999.

DATA QUALITY ISSUES IN ASSISTMENTS

- Duplicated records

- Non-chronological order

- Items associated with multiple skills

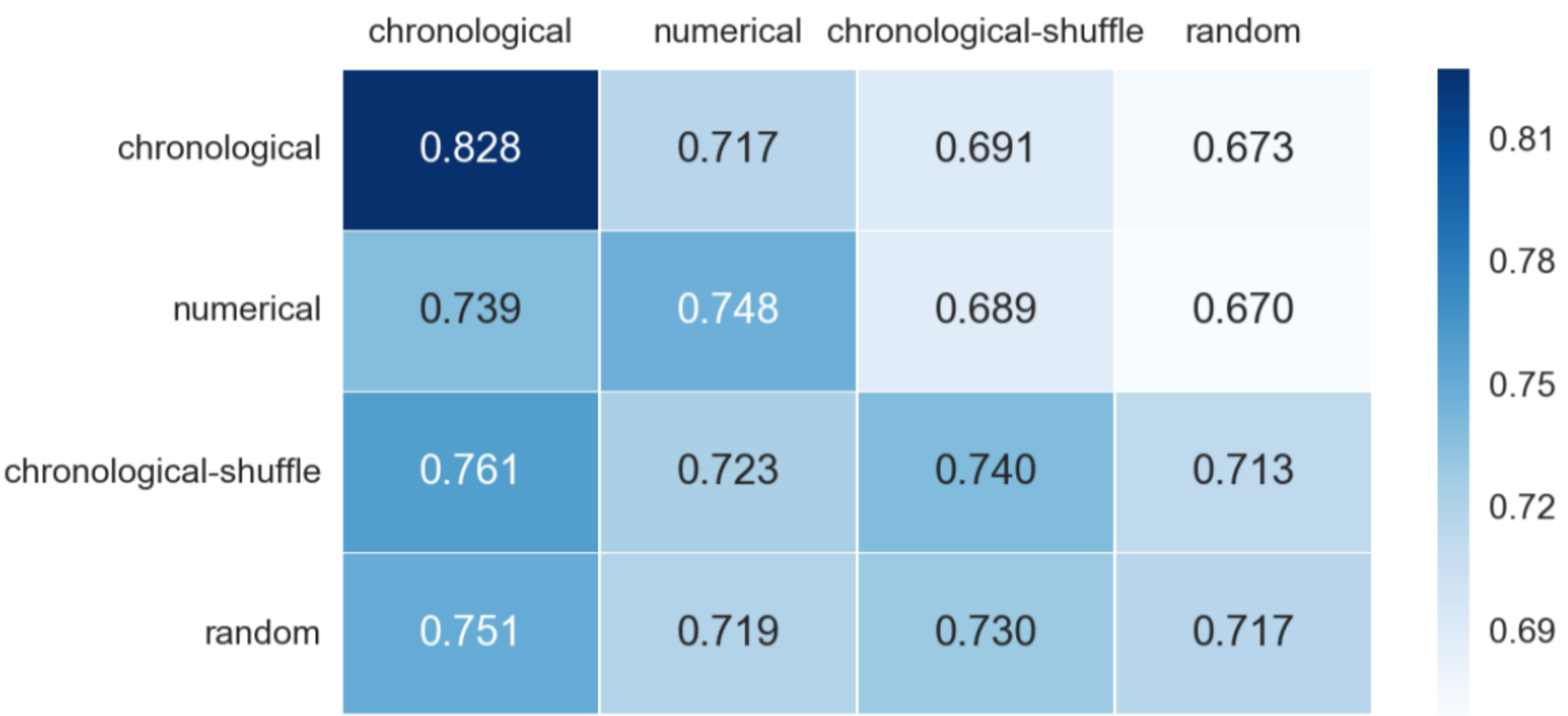
ORDER SCHEMES

- Chronological

- Numerical

- Chronological-Shuffle

- Random


chronological
numerical
chronological-shuffle
random
chronological
numerical
chronological-shuffle
random
0.828
0.717
0.691
0.673
0.739
0.748
0.689
0.670
0.761
0.723
0.740
0.713
0.751
0.719
0.730
0.717
0.81
0.78
0.75
0.72
0.69

MULTI-SKILL SCHEMES

【A, B, C】 correct

- Repeated scheme

A correct

B correct

C correct

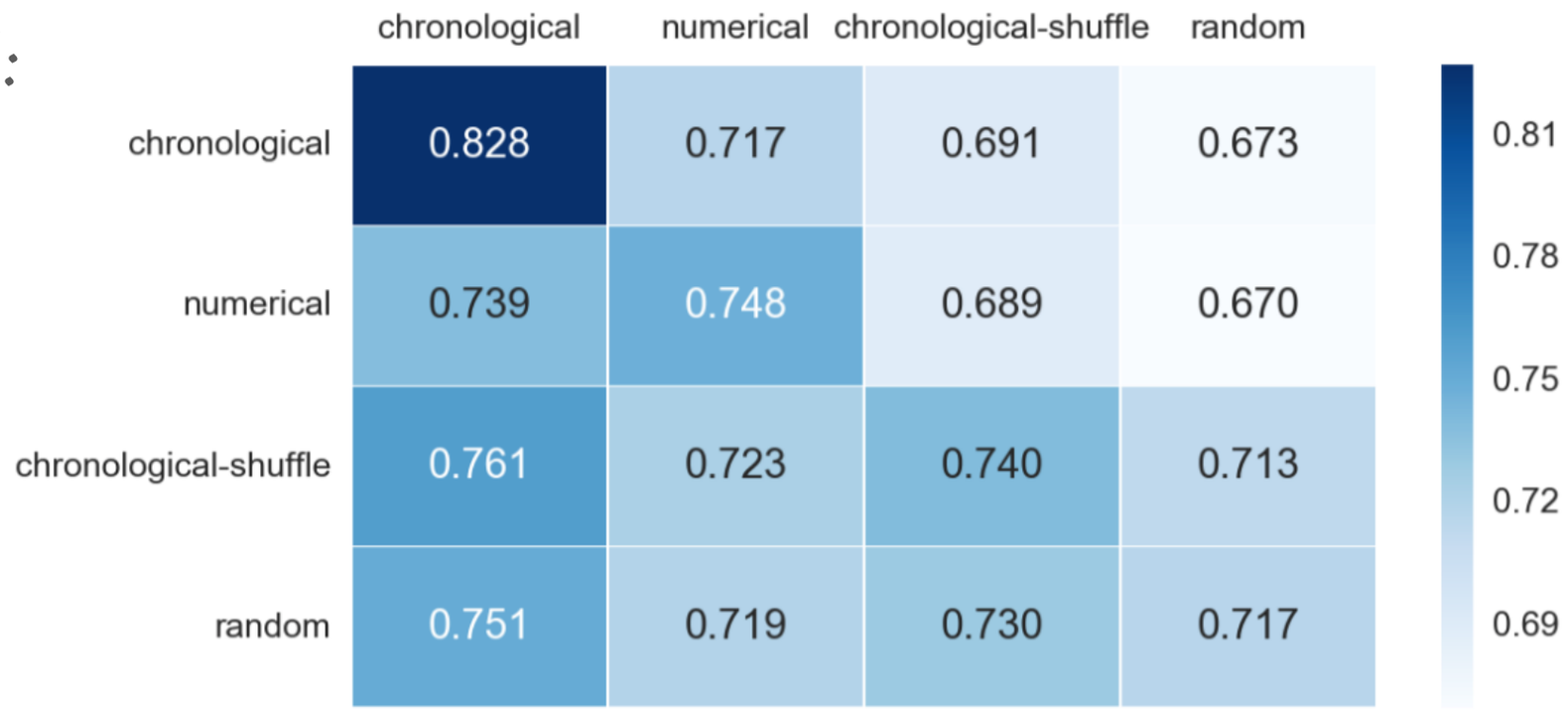
- Joint scheme

D correct

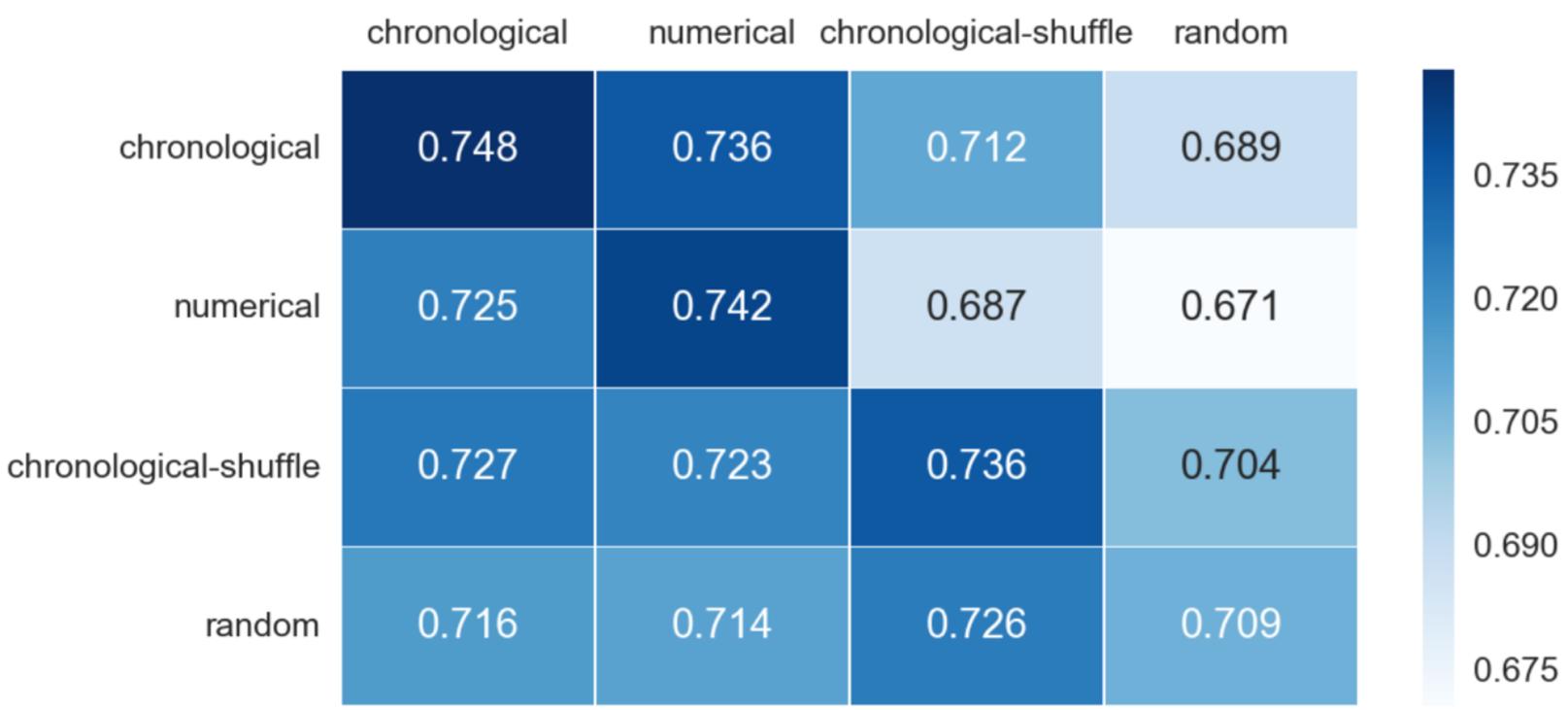
- Multi-hot scheme

【A, B, C】 correct

Repeated:

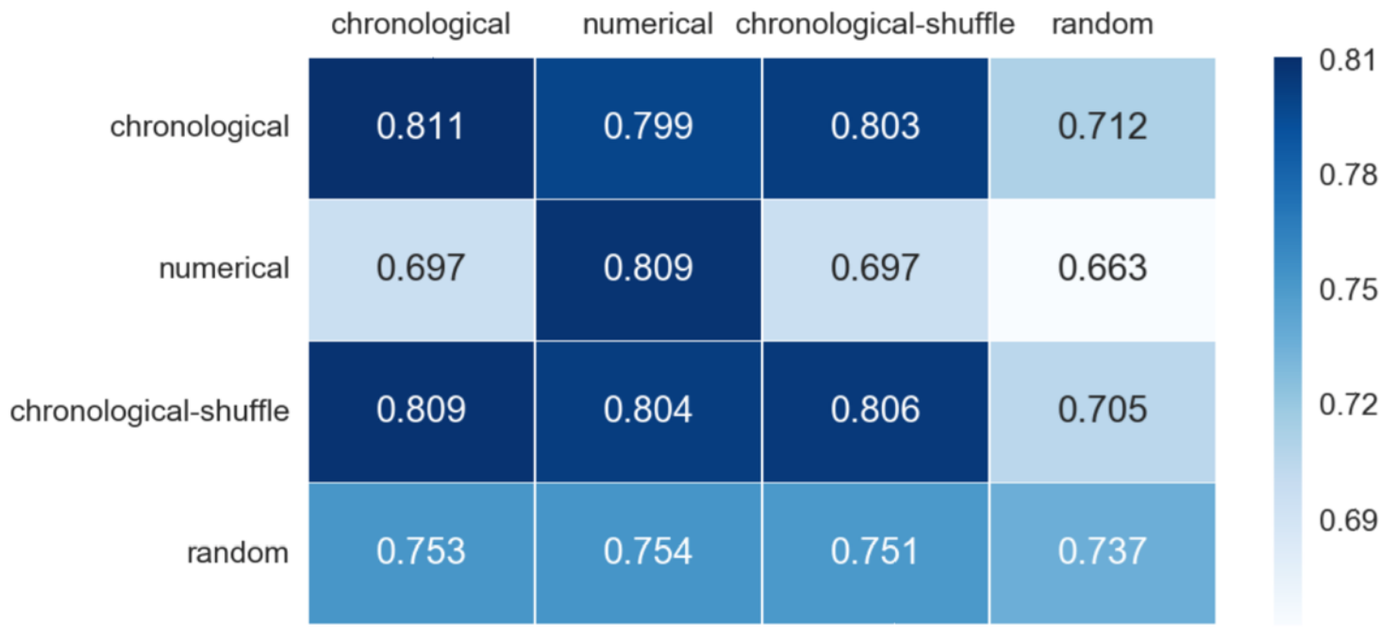


Joint:

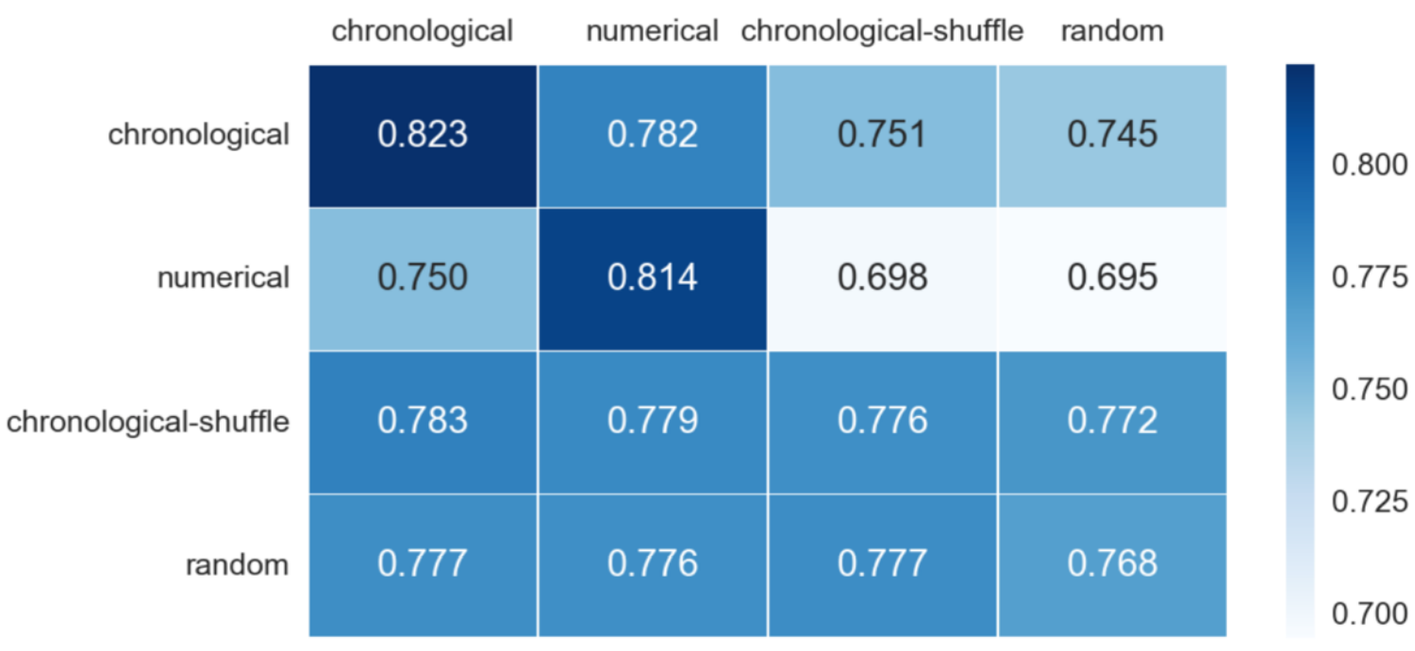


Joint scheme is somehow equivalent to multi-hot scheme?

OTHER DATASETS



(c) BKT Simulation



(d) Fraction, 105 skills

1) Blocking vs Mixing

2) Skill distribution over an episode

3) Skill distribution in train and test set

EXTRACTING PREREQUISITES

- One proposed method for DKT

$$J_{ij} = \frac{y\left(j|i\right)}{\sum_k y\left(j|k\right)}$$

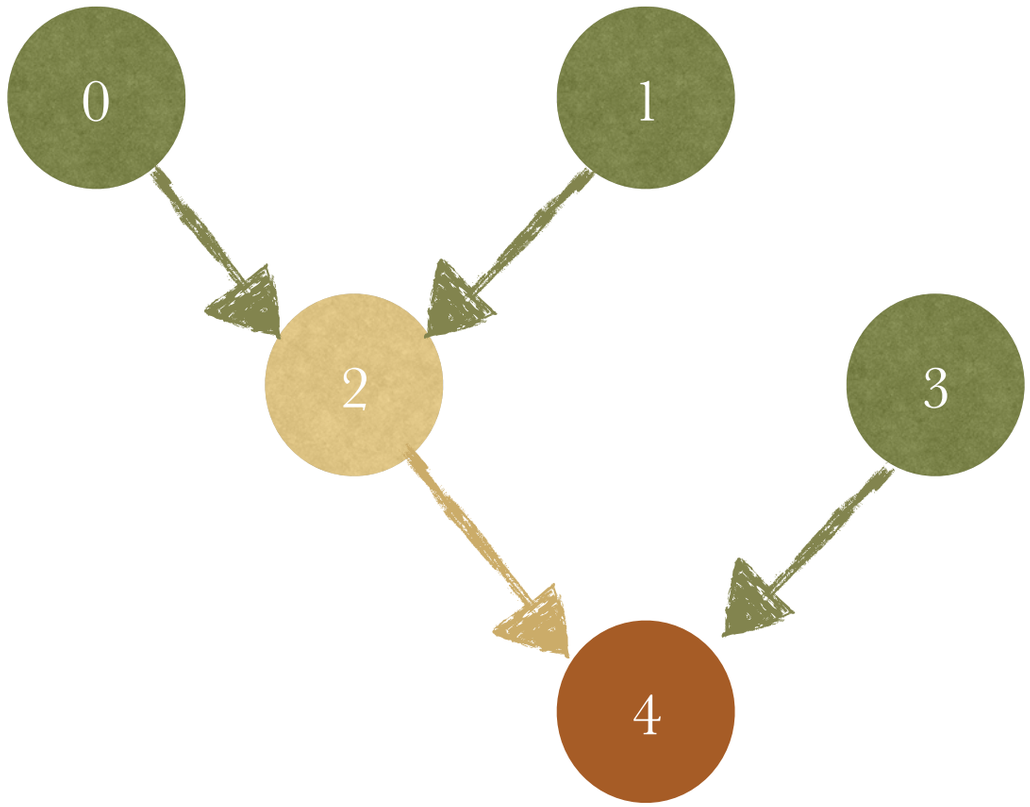
- Impure items

	q0|1	q0|1	q0|1	q0|1	q0|1	q0|1	q0|1	q0|1	q0|1	q0|1	q0|1	q0|1	q0|1	q0|1	q0|1	q0|1	q0|1	q0|1	q0|1	q0|1
0	0.41	0.82	0.86	0.89	0.9	0.91	0.92	0.92	0.92	0.92	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93
1	0.29	0.25	0.25	0.22	0.21	0.21	0.21	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.19	0.19	0.19	0.19	0.19
2	0.22	0.22	0.22	0.21	0.2	0.19	0.19	0.18	0.18	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.16	0.16	0.16
3	0.39	0.34	0.19	0.12	0.079	0.059	0.049	0.043	0.04	0.038	0.037	0.037	0.037	0.036	0.036	0.036	0.036	0.036	0.036	0.036
4	0.18	0.18	0.16	0.15	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16

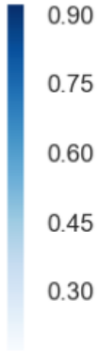
0.8
0.6
0.4
0.2

(a) Skill 0

1) Independent skills' beliefs are changing.

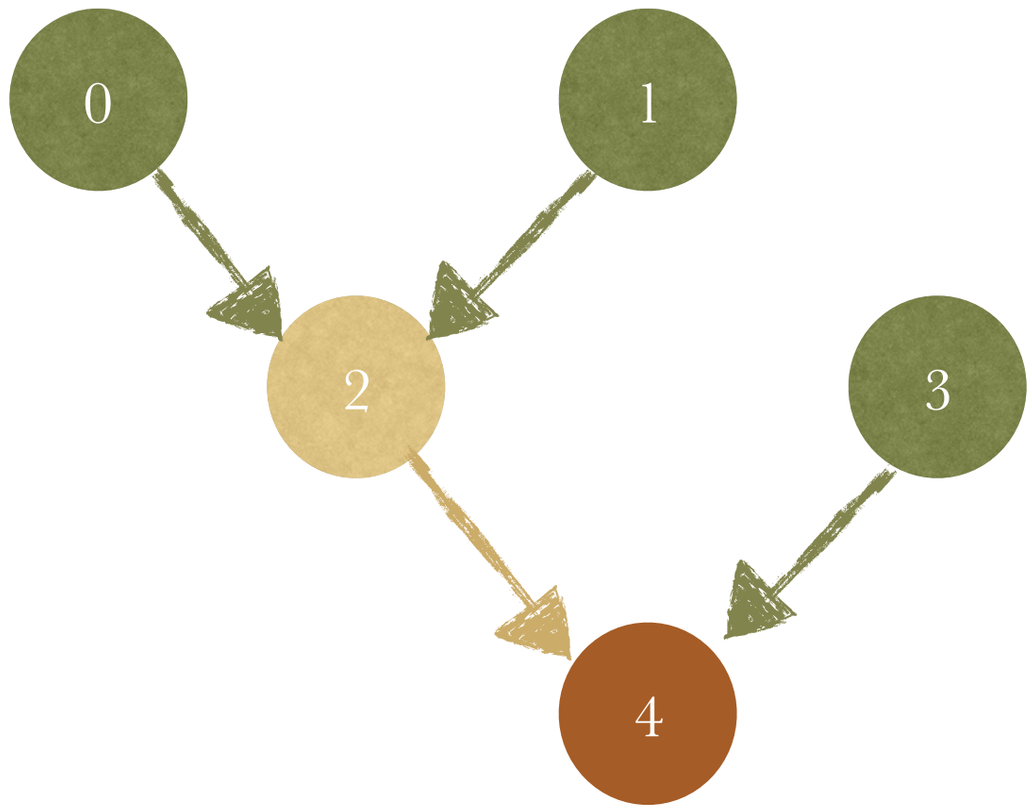


	q4|1	q4|1	q4|1	q4|1	q4|1	q4|1	q4|1	q4|1	q4|1	q4|1	q4|1	q4|1	q4|1	q4|1	q4|1	q4|1	q4|1	q4|1	q4|1	q4|1
0	0.41	0.42	0.44	0.49	0.53	0.57	0.62	0.66	0.72	0.79	0.85	0.88	0.89	0.9	0.91	0.92	0.92	0.92	0.92	0.92
1	0.29	0.29	0.28	0.32	0.35	0.4	0.47	0.55	0.64	0.76	0.86	0.89	0.88	0.88	0.89	0.89	0.89	0.89	0.89	0.89
2	0.22	0.19	0.18	0.2	0.23	0.26	0.32	0.41	0.54	0.68	0.76	0.78	0.77	0.78	0.8	0.81	0.82	0.82	0.82	0.83
3	0.39	0.29	0.23	0.3	0.37	0.48	0.62	0.74	0.83	0.88	0.91	0.9	0.88	0.88	0.89	0.9	0.9	0.9	0.9	0.9
4	0.18	0.19	0.22	0.27	0.33	0.43	0.56	0.71	0.83	0.9	0.92	0.9	0.86	0.85	0.86	0.87	0.88	0.88	0.88	0.88



(e) Skill 4

2) Takes more than one step to show correlations.

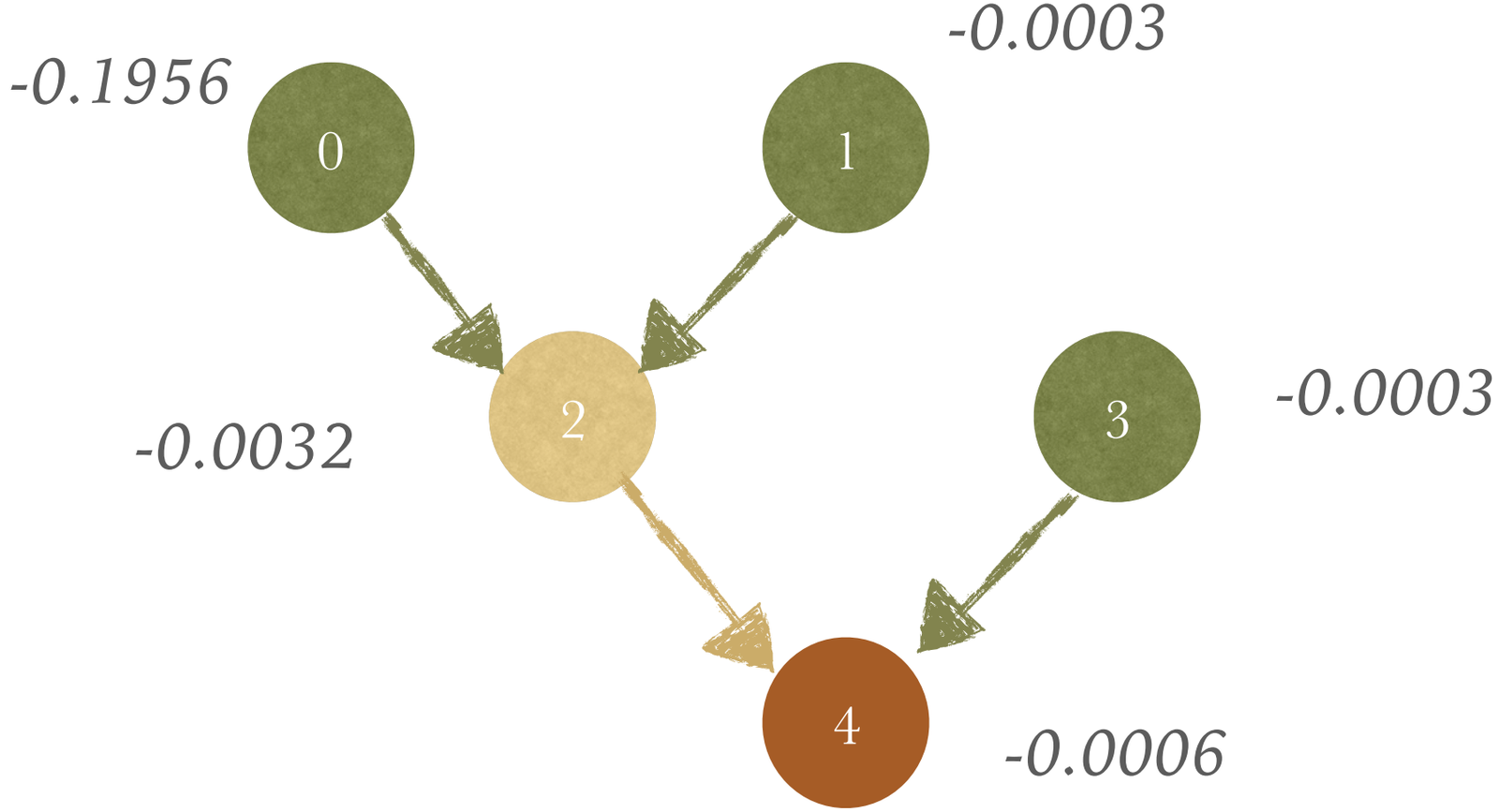


OUR METHOD

- Leverage the AUC drop after shuffling.
- Which skills to shuffle? (Single one or a group?)
- (We can not rely on the AUC drop of a shuffled skill.)
- Shuffle train set or test set?
- Set a threshold to filter the noise?

PRELIMINARY RESULTS

- On fraction, about 20 out of 105 skills suffer a significant (>0.1) AUC drop after doing chronological-shuffle for all skills.
- Shuffle skill 0 on test set (simulation):



SUMMARY

- Data preprocessing is vital and error-prone.

- DKT captures inter-skill relationships with noise.

- Do “shuffle to see AUC drop” experiments on Fraction
- Figure out what’s DKT leveraging other than the order of skills
- Explicitly support multi-skill problems in DKT.