Appendix of "QCCDM: A Q-Augmented Causal Cognitive Diagnosis Model for Student Learning"

Shuo Liu[a], **Hong Qian**[a;*], **Mingjia Li**[a] **and Aimin Zhou**[a]

[a]Shanghai Institute of AI for Education and School of Computer Science and Technology, East China Normal University, Shanghai 200062, China

The appendix is organized as follows:

- Appendix 1 introduces the detailed experiment setting of QCCDM.
- Appendix 2 demonstrates the validity and effectiveness of Q-augmentation in the context of cognitive diagnosis.
- Appendix 3 illustrates the detailed effect of hyperparameter λ on three real-world datasets.

1 Detailed Experiment Settings

To ensure the consistency across models, we set the embedding dimension to be the same as the number of knowledge attributes for Q-based CDMs. In addition, for Q-irrelevant CDMs, we empirically set the latent dimension to 16, which has shown to perform well [6]. For models that use neural networks as the interactive function, we use the same number of hidden layers and hidden units. Specifically, we adopt a common tower structure, where deeper layers have half the number of neurons as its next lower layer. In QCCDM, we set λ to be 0.1 for datasets with a large number of knowledge attributes (e.g., Junyi) and 0.01 for datasets with a normal number of knowledge attributes (e.g., Math1 and Math2). To initialize the model parameters, we employ the Xavier normal initialization [2], which is widely used in deep learning models. For optimization, we use the Adaptive moment estimation [3], which has shown to be effective in training deep neural networks. QCCDM is implemented with Pytorch [4]. It is usually set to a default value of 0.01. The hyperparameter of batch normalization is default and $p = 0.5$ for all the dropout layers.

2 Analysis of Q-Augmentation (Q5)

To demonstrate the validity and effectiveness of Q-augmentation, i.e., Q5 in experiment of the main paper, we conduct a regression analysis [1, 5] to validate its rationality. Specifically, we use the sum of questions' difficulty as the dependent variable and Q as the independent variable. The results present in Table 1 show that Q-augmentation outperforms manually labeled Q in terms of R-squared, Adj. R-squared, F-statistic, and Prob (F-statistic), which highlights the significance of exploring the relationship between latent attributes and exercises in describing the relationship between question difficulty and knowledge attributes.

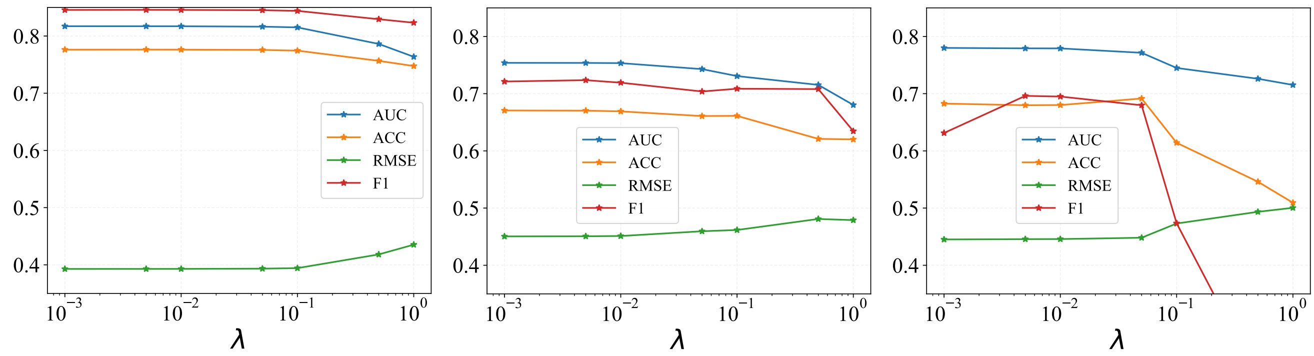
* Corresponding Author. E-mail: hqian@cs.ecnu.edu.cn.

Table 1. Validating Q-augmentation with regression analysis. In each row, an entry is bold if its mean value is the best.

Datasets	Metrics	QCCDM	NCDM
Math1	R-squared	**0.997**	0.825
	Adj. R-squared	**0.989**	0.125
	F-statistic	**121.7**	27.79
	Prob (F-statistic)	**0.00016**	0.00291
Math2	R-squared	**0.999**	0.825
	Adj. R-squared	**0.953**	0.125
	F-statistic	**27.38**	1.179
	Prob (F-statistic)	**0.0006**	0.506

3 Supplement of Hyperparameter Analysis (Q6)

We conduct a hyperparameter experiment to investigate the effect of λ on the density of $\widetilde{Q}$ across all datasets (Q6 in experiment of the main paper). The results, present in Figure 1, show that increasing the value of λ leads to sparser $\widetilde{Q}$, which can negatively impact the model's performance, as indicated by the lower F1 in Math2. On the other hand, a higher value of λ may result in each exercise being related to all knowledge attributes, which is not reasonable. Therefore, it is important to maintain a reasonable range for λ to ensure that the relationship between exercises and latent knowledge attributes is effectively captured while maintaining good prediction performance for students.



(a) On Junyi (b) On Math1 (c) On Math2

Figure 1. The performance (AUC, ACC, RMSE and F1) under different λ values on three datasets.

References

[1] Rudolf J. Freund, William J. Wilson, and Ping Sa, *Regression Analysis*, Elsevier, 2006.

[2] Xavier Glorot and Yoshua Bengio, 'Understanding the difficulty of training deep feedforward neural networks', in *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics*, pp. 249–256, Sardinia, Italy, (2010).

[3] Diederik P. Kingma and Jimmy Ba, 'Adam: A method for stochastic optimization', in *Proceedings of the 3rd International Conference on Learning Representations*, San Diego, CA, (2015).

[4] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Z. Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala, 'Pytorch: An imperative style, high-performance deep learning library', in *Advances in Neural Information Processing Systems 32*, pp. 8024–8035, British Columbia, Canada, (2019).

[5] Dongbo Tu, Yan Cai, Liangxu Gao, and Xunda Wang, *Advanced cognitive diagnosis*, Beijing Normal University Publishing Group, 2019.

[6] Fei Wang, Qi Liu, Enhong Chen, Zhenya Huang, Yuying Chen, Yu Yin, Zai Huang, and Shijin Wang, 'Neural cognitive diagnosis for intelligent education systems', in *Proceedings of the 34th AAAI Conference on Artificial Intelligence*, New York, NY, (2020).