CogEA: A Multi-Agent System for Cognitive Ability Annotation of Exercises by Simulating Human Behaviors

Demonstration Track

Peiran Zhang
Tsinghua University
Beijing, China
zhangpr23@mails.tsinghua.edu.cn

Nan He
Beijing University of Technology
Beijing, China
henancs@bjut.edu.cn

Binbin Qi
South China Normal University
Guangzhou, China
qidoublebins@163.com

Lifeng Sun
Tsinghua University
Beijing, China
sunlf@tsinghua.edu.cn

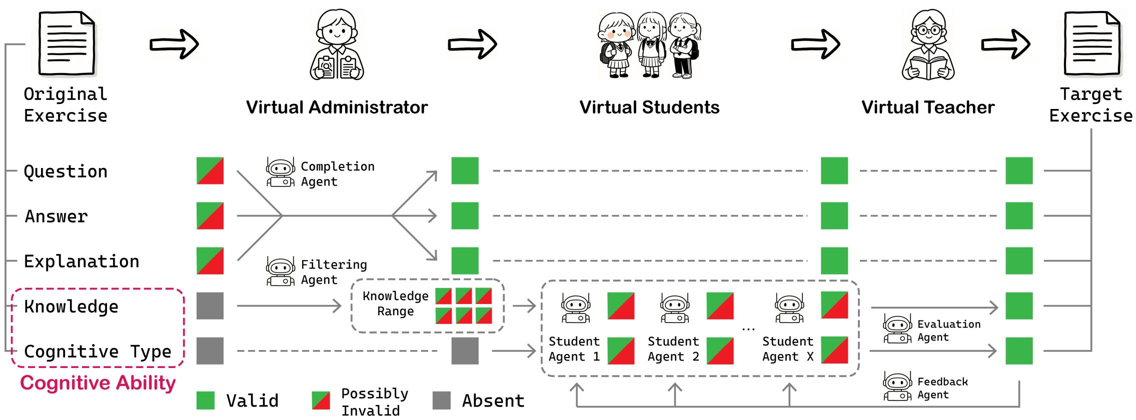

Figure 1: The exercise data structure and the multi-agent framework of the CogEA system

ABSTRACT

The rapid advancement of the online learning platform has created a demand for high-quality annotated exercise data, yet manual exercise annotation struggle to meet current exercise data requirements. This demonstration introduces the Cognitive ability-based Exercise Agent(CogEA), a Multi-Agent System(MAS) aimed at automatically annotating exercises. We propose "cognitive ability" to reflect both knowledge selection and knowledge application in exercises. In addition, using agents driven by large language models(LLM) that simulate human behaviors of the administrator, student, and teacher, the CogEA system ensures the quality of exercise annotation. The demonstration video can be found at: https://youtu.be/y3XLdWIuxNE

KEYWORDS

multi-agent; large language model; exercise annotation; cognitive ability

Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), C. Amato, L. Dennis, V. Mascardi, J. Thangarajah (eds.), May 25 – 29, 2026, Paphos, Cyprus. © 2026 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). https://doi.org/10.65109/EDGL6988

ACM Reference Format:
Peiran Zhang, Nan He, Binbin Qi, and Lifeng Sun. 2026. CogEA: A Multi-Agent System for Cognitive Ability Annotation of Exercises by Simulating Human Behaviors: Demonstration Track. In *Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), Paphos, Cyprus, May 25 – 29, 2026*, IFAAMAS, 3 pages. https://doi.org/10.65109/EDGL6988

1 INTRODUCTION

With the rapid advancement of the public and private online learning platform, the role of high-quality educational data is becoming increasingly prominent. Exercise data constitute a key component of educational data, performing the crucial function of facilitating students' knowledge acquisition and assessing their level of knowledge mastery[2, 3, 9]. However, the process of migrating exercise data to online learning platforms presents the following limitations:

(1) Limited scope of exercise annotation. Exercises serve the function of evaluating students' mastery of knowledge; therefore, the exercise should be annotated with the corresponding knowledge concept. Current exercise annotations only include the selection of knowledge concepts and do not include the application of knowledge concepts[8]. The latter plays an equally important role as the former in evaluating students' mastery of knowledge[1, 5]. Figure

Exercise	Knowledge Selection	Cognitive Ability Knowledge - Cognitive Type
determine: $(x-1)^2-1=0$ is a quadratic equation or not	quadratic equation	comprehend quadratic equation
solve the quadratic equation: $(x-1)^2-1=0$		analyse quadratic equation

Cognitive Type: retrieval / comprehension / analysis / utilization

Figure 2: Illustration of cognitive ability, which includes knowledge selection (blue) and knowledge application (red)

2 illustrates an example of knowledge selection and knowledge application included in exercises.

(2) The deficiencies and costs due to manual operation. Due to the oversights of manual operation, the original exercise on online learning platforms possibly contains deficiencies in basic exercise indexes. In addition, exercises are normally annotated manually[8], which incurs substantial costs and therefore goes against the large-scale deployment of exercise data.

The ability of large language models (LLMs) to simulate human behavior offers a promising solution to the limitations mentioned above[4]. This demonstration presents a multi-agent system CogEA. The system employs "cognitive ability" to reflect both knowledge selection and knowledge application in exercises. In addition, the agents within the system simulate the human behaviors of administrators, students, and teachers, collaborating with each other to automatically annotate high-quality exercise data. Therefore, our system is capable of supporting applications such as learning state assessment and exercise recommendation on online learning platforms[6, 7].

2 THE PROPOSED MULTI-AGENT SYSTEM

2.1 Framework

Figure 1 illustrates the exercise data structure and the multi-agent framework of the CogEA system. The exercise data structure of CogEA includes five indexes: question, answer, explanation, knowledge, and cognitive type. The multi-agent framework includes three types of collaborative agents: virtual administrator, virtual students, and virtual teacher.

2.2 Cognitive Ability-based Exercise Data Structure

The exercise data structure of the CogEA system comprises basic indexes and evaluation indexes. Basic indexes include three indexes: question, answer, and explanation. Evaluation indexes are used to evaluate students' mastery of knowledge. To reflect both knowledge selection and knowledge application in exercises, we build on Marzano's Taxonomy of Educational Objectives[5] to propose "cognitive ability" as the evaluation indexes. Cognitive ability includes two indexes: knowledge (to reflect knowledge selection) and cognitive type (to reflect knowledge application), while cognitive type classifies knowledge application into four categories: retrieval, comprehension, analysis, and utilization[5]. Figure 2 illustrates that cognitive ability is more comprehensive than the index which includes only knowledge selection.

2.3 Virtual Administrator

The virtual administrator maintains the online exercise dataset. It includes two types of agents: the completion agent and the filtering agent. Whenever users edit the exercise dataset, the completion agent checks whether the question, answer, or explanation of the edited exercise is incomplete. If the question is incomplete, the agent asks users to complete it; otherwise, the incomplete exercise is forbidden. If the answer or explanation is incomplete, the agent automatically generates the corresponding content to complete it. In addition, for each exercise, the filtering agent selects a knowledge set containing no more than 30 knowledge concepts from the knowledge base, which is prepared for subsequent annotation.

2.4 Virtual Students

The virtual student agents annotate the knowledge and cognitive categories, which reflect the cognitive ability of exercises. For each exercise, the virtual student agent simulates the problem-solving behavior of the students. Based on simulated behavior, the agent selects one knowledge concept from the knowledge set and selects one cognitive type from four available categories. To mitigate the impact of incorrect annotations, each exercise is annotated by five or more virtual student agents. These agents differ in their LLM APIs and contexts to simulate the variations among different students.

2.5 Virtual Teacher

The virtual teacher comprises two types of agents: the evaluation agent and the feedback agent. The evaluation agent evaluates and integrates the annotations from each virtual student agent and determines the final annotation. For virtual students who made annotation errors, the feedback agent generates feedback and adds it to that virtual student's interaction memory. Consequently, the virtual student agent embeds the previous feedback into the prompt via a RAG module, thus avoiding previous mistakes in subsequent annotations.

2.6 Demonstration Setup

The automated annotation and maintenance function of the CogEA system is deployed on an exercise dataset from a middle school that contained 5100 exercises. The LLMs-based agents invoke the APIs provided by GPT, Gemini, and DeepSeek.

3 THE ACCURACY OF ANNOTATION

To validate the accuracy of cognitive ability annotation performed by our MAS, we randomly selected the annotation of 100 exercises from our dataset and invited four experts to evaluate the annotation. Expert evaluations indicated that 96% of the extraction is accurate, providing preliminary validation of the reliability of the annotation.

4 CONCLUSION

This demonstration presents an innovative application of the CogEA system for automated exercise annotation and maintenance in online learning platforms. The exercised processed by the CogEA system show promising quality and usability, due to the novel cognitive ability-based exercise data structure, LLM-based human behavior simulation, and multi-agent collaboration.

ACKNOWLEDGMENTS

This work is supported by National Key Research and Development Program of China No.2022ZD0115903.

REFERENCES

[1] Benjamin Bloom. 1956. *Taxonomy of Educational Objectives: The Classification of Educational Goals, Handbook I: Cognitive Domain.* Longman, New York.

[2] Xinghe Cheng, Xufang Zhou, Liangda Fang, Chaobo He, Yuyu Zhou, Weiqi Luo, Zhiguo Gong, and Quanlong Guan. 2025. NR4DER: Neural Re-ranking for Diversified Exercise Recommendation. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '25).* Association for Computing Machinery, New York, NY, USA, 1738–1747.

[3] Quanlong Guan, Fang Xiao, Xinghe Cheng, Liangda Fang, Ziliang Chen, Guanliang Chen, and Weiqi Luo. 2023. KG4Ex: An Explainable Knowledge Graph-Based Approach for Exercise Recommendation. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (CIKM '23).* Association for Computing Machinery, New York, NY, USA, 597–607.

[4] Kelvin Guu, Kenton Lee, Zora Tung, Panupong Pasupat, and Ming-Wei Chang. 2020. REALM: Retrieval-Augmented Language Model Pre-Training. In *INTERNATIONAL CONFERENCE ON MACHINE LEARNING, VOL 119 (Proceedings of Machine Learning Research, Vol. 119).*

[5] RJ Marzano and JS Kendall. 2007. *The New Taxonomy of Educational Objectives (2nd ed.).* Corwin Pres, Thousand Oaks CA.

[6] Hiromi Nakagawa, Yusuke Iwasawa, and Yutaka Matsuo. 2019. Graph-based Knowledge Tracing: Modeling Student Proficiency Using Graph Neural Network. In *IEEE/WIC/ACM International Conference on Web Intelligence (ICLR).* ACM, Thessaloniki Greece, 156–163. https://doi.org/10.1145/3350546.3352513

[7] Chris Piech, Jonathan Bassen, Jonathan Huang, Surya Ganguli, Mehran Sahami, Leonidas Guibas, and Jascha Sohl-Dickstein. 2015. Deep Knowledge Tracing. In *ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 28 (NIPS 2015) (Advances in Neural Information Processing Systems, Vol. 28).*

[8] Shuanghong Shen, Qi Liu, Zhenya Huang, Yonghe Zheng, Minghao Yin, Minjuan Wang, and Enhong Chen. 2024. A Survey of Knowledge Tracing: Models, Variants, and Applications. *IEEE Transactions on Learning Technologies* 17 (2024), 1858–1879.

[9] Zhengyang Wu, Ming Li, Yong Tang, and Qingyu Liang. 2020. Exercise recommendation based on knowledge concept prediction. *KNOWLEDGE-BASED SYSTEMS* 210 (2020).