A Unified Adaptive Testing System Enabled by Hierarchical Structure Search

Junhao Yu[1], Yan Zhuang[1], Zhenya Huang[1,2], Qi Liu[1,2], Xin Li[1], Rui Li[3], Enhong Chen[1]
[1]State Key Laboratory of Cognitive Intelligence, University of Science and Technology of China
[2]Institute of Artificial Intelligence, Hefei Comprehensive National Science Center
[3]School of Computer Science and Technology, Xidian University.

➢Motivation

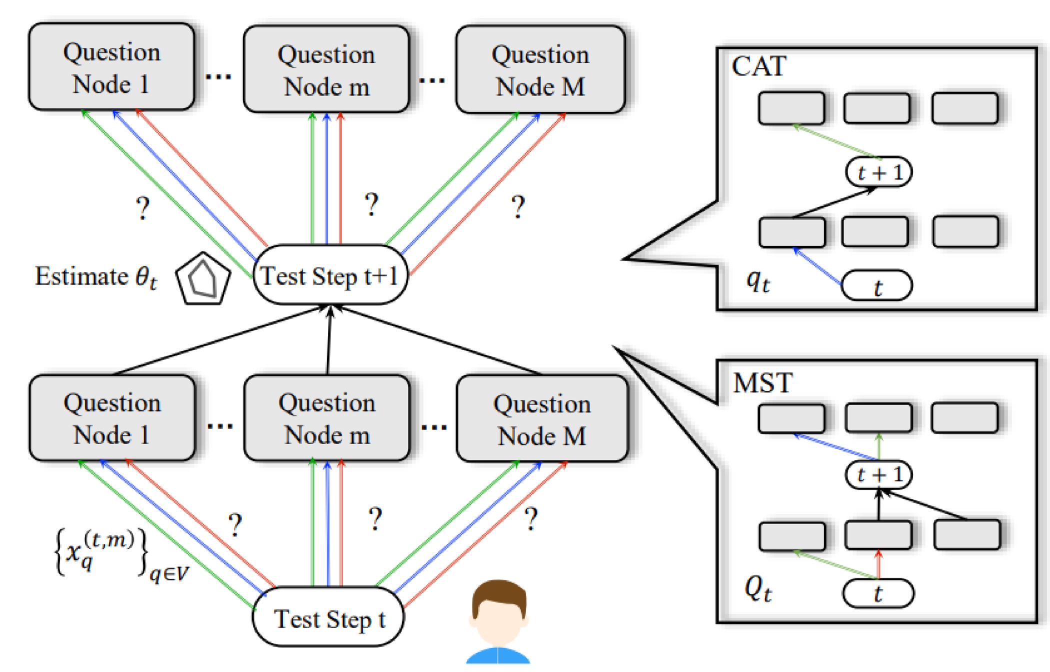
Adaptive Testing System (ATS) is an intelligent educational application that dynamically adjusts questions to match individual skill levels, providing personalized assessments. Widely used in standardized tests such as GRE and TOEFL, ATS enhances accuracy in skill assessment and reduces the number of required questions. Despite the diverse existing forms of ATS, there lacks a unified framework for modeling different testing formats. The contributions of our paper are as follows:

- For the first time, various forms of Adaptive Testing Systems (ATS) are redefined as a unified structured search problem, leveraging data-driven dynamic navigation of question selection and eliminating the need for manual expert design.
- A differentiable hierarchical optimization algorithm is designed to select appropriate questions, with theoretical guarantees on convergence and gradient estimation error.
- Across various real-world datasets, this framework demonstrates superior accuracy and efficiency compared to other ATS methods, with more stable training convergence.

➢Our Proposed UATS Framework

A. Framework Overview

The UATS (**U**nified **A**daptive **T**est **S**ystems) framework realizes automatic problem selection by unifying various forms of adaptive test systems into a test structure search problem. It uses a data-driven approach with differentiable hierarchical optimization algorithms that dynamically adjust question selection according to students' abilities, thereby avoiding the process of designing tests manually



The illustration of UATS framework.

The picture above shows a general sketch of the frame. In each selection, the selection parameters (i.e., edge weights) are updated according to the students' ability. Then the combination of questions is selected according to the edge weights. After answering, the ability is updated, which is an adaptive test process

B. Theoretical Analysis

First, let α represent the edge weight for question selection. θ_t represents the student's ability at step t. The testing problem can be formulated as a bilevel optimization problem, as shown in the right expression.

$$\min_{\alpha} L(\theta_t^*(\alpha), D_u)$$
$$\text{s.t.}\ \ \theta_t^*(\alpha) = \arg\min_{\theta} L(\theta, D_t(\alpha))$$

Unified Definition of ATS

However, optimizing it directly can be slow. To address this, we use a K-step approximation solution instead of the exact solution for the inner function, while ensuring bounded convergence error, as shown in the left expression.

$$\|\nabla_\alpha L(\theta_t^K(\alpha_t), D_u) - \nabla_\alpha L(\theta_t^*(\alpha_t), D_u)\| \le \frac{B}{\mu}\left(L^2(1-\gamma\mu)^K - \frac{M(\tau\mu + L\rho)}{\mu}(1-\gamma\mu)^{\frac{K+1}{2}}\right) + \frac{ML(1-\gamma\mu)^K}{\mu} + \frac{BM(\tau\mu + L\rho)}{\mu^2}(1-\gamma\mu)$$

Gradient Estimation Error Bound

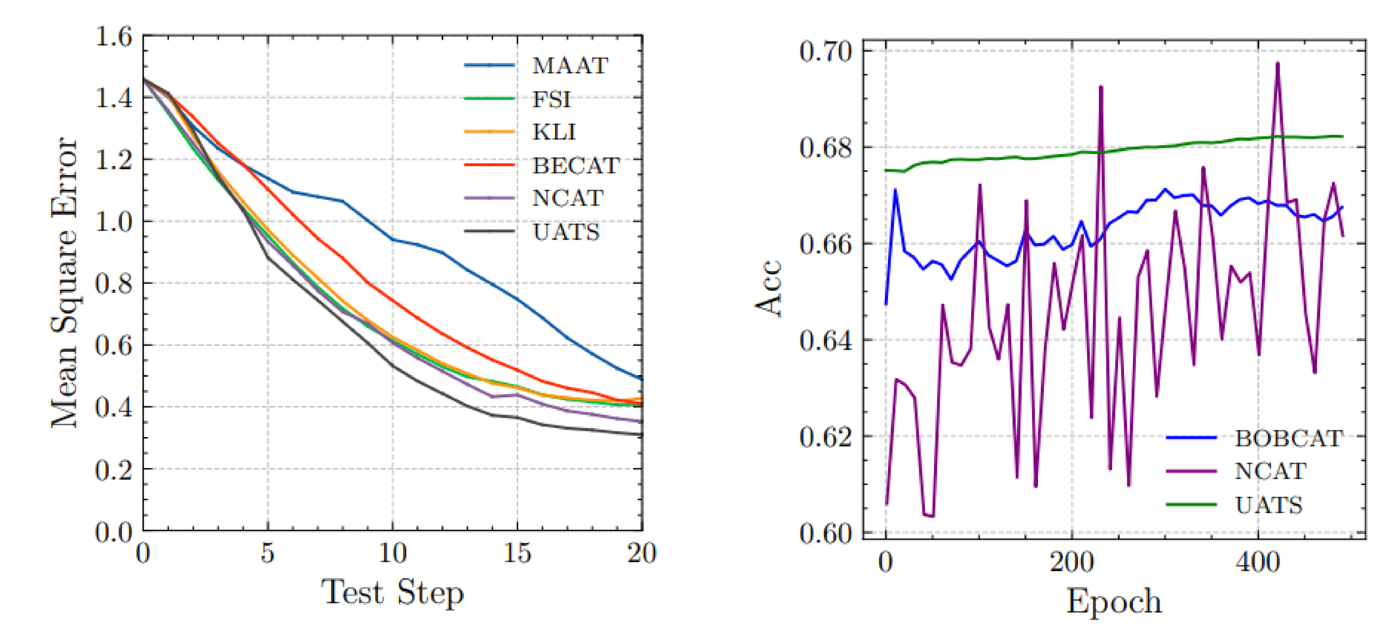
$$\frac{1}{T}\sum_{t=0}^{T-1}\|\nabla L(\theta_t^*(\alpha_t), D_u)\|^2 \le \frac{16W}{T}(L(\theta_0^*(\alpha_0), D_u) - L(\theta_T^*(\alpha_T), D_u)) + C$$

Convergence Analyze

According to the error theorem obtained above, we further prove the convergence of the UATS method and ensure the stability of the method

C. Experimental Verification

The following figure shows the capability estimation simulation and convergence analysis experiment. As you can see, there is less MSE error between UATS estimate ability and true ability when answering the same question. Not only that, the results of UATS are more stable during training.



Demonstration of experimental effect

➢Conclusion

We introduced the UATS framework, unifying CAT and MST methodologies for adaptive testing. It automates question selection, optimizing accuracy and efficiency. UATS demonstrated superior performance, dynamically adjusting to student responses for precise ability estimation and stable training convergence in real-world educational settings.

Contacts:
Junhao Yu
yujunhao@mail.ustc.edu.cn