



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

Институт образования
Центр психометрики и измерений в образовании

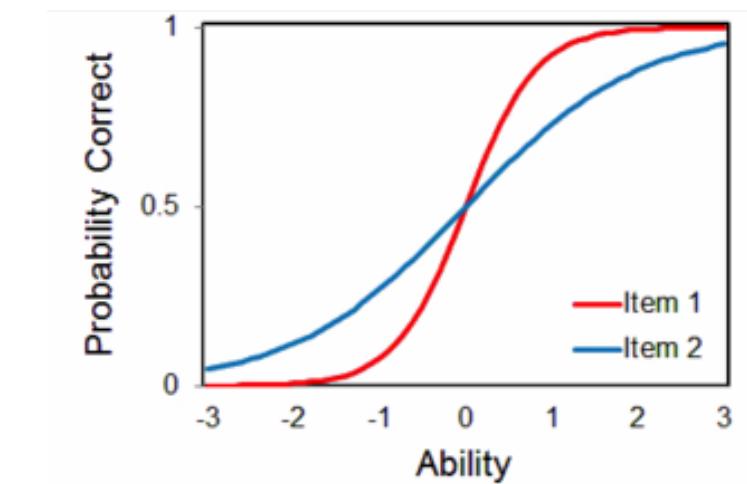
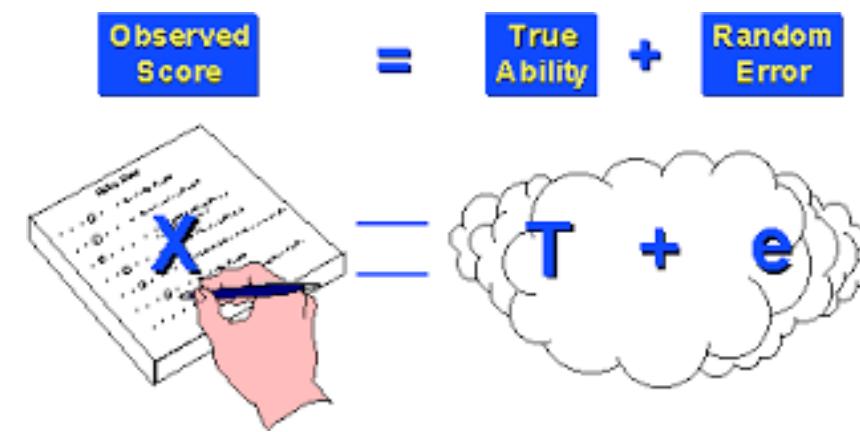
Анализ взаимодействий в психометрике

Скрябин Максим Александрович
ведущий аналитик, к.ф.-м.н.

Москва, 2020

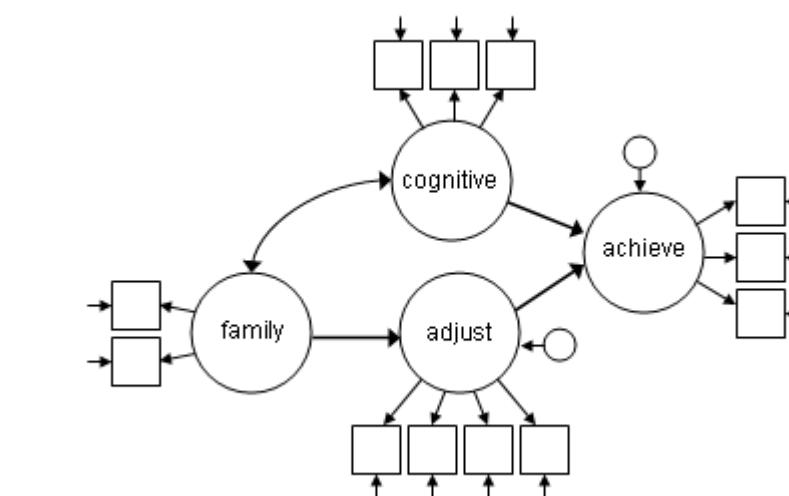


Психометрическое моделирование

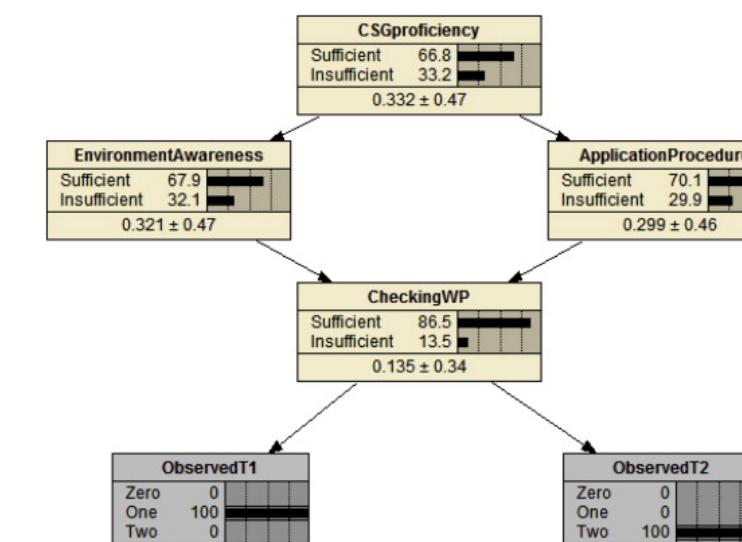


Классическая
теория
тестирования

Item Response
Theory



Структурное
моделирование,
в частности, КФА

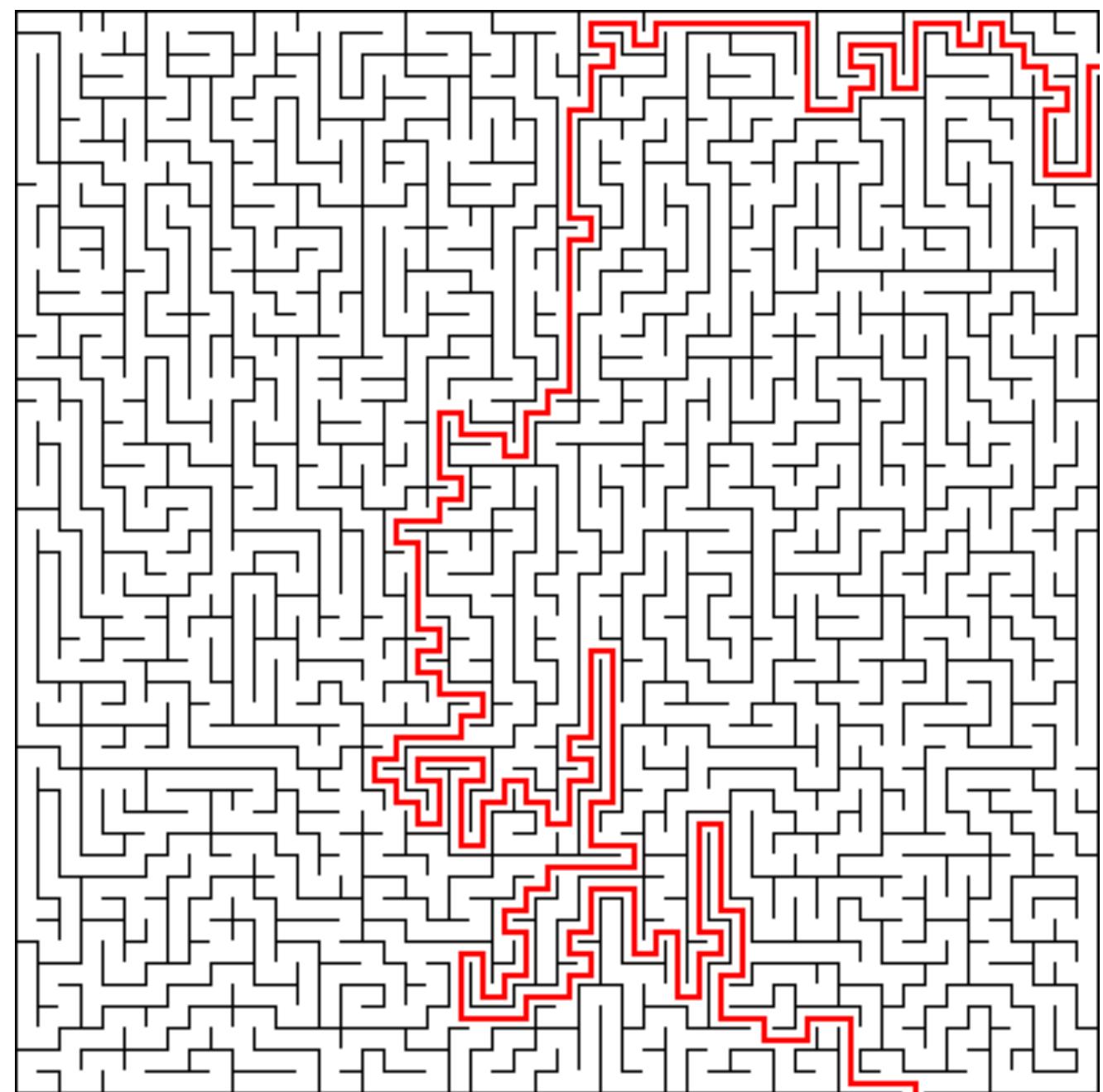


Байесовские
сети



Про что сегодня пойдет речь?

- Интеллектуальные агенты
- Анализ сценариев
- Анализ алгоритмов
- Динамические модели



Теория принятия решений и интеллектуальные агенты



Принятие решений в игре

Теория игр

Как выбрать
оптимальную
стратегию?

Поведенческая
теория игр

Какие стратегии на
самом деле
выбираются в игре?

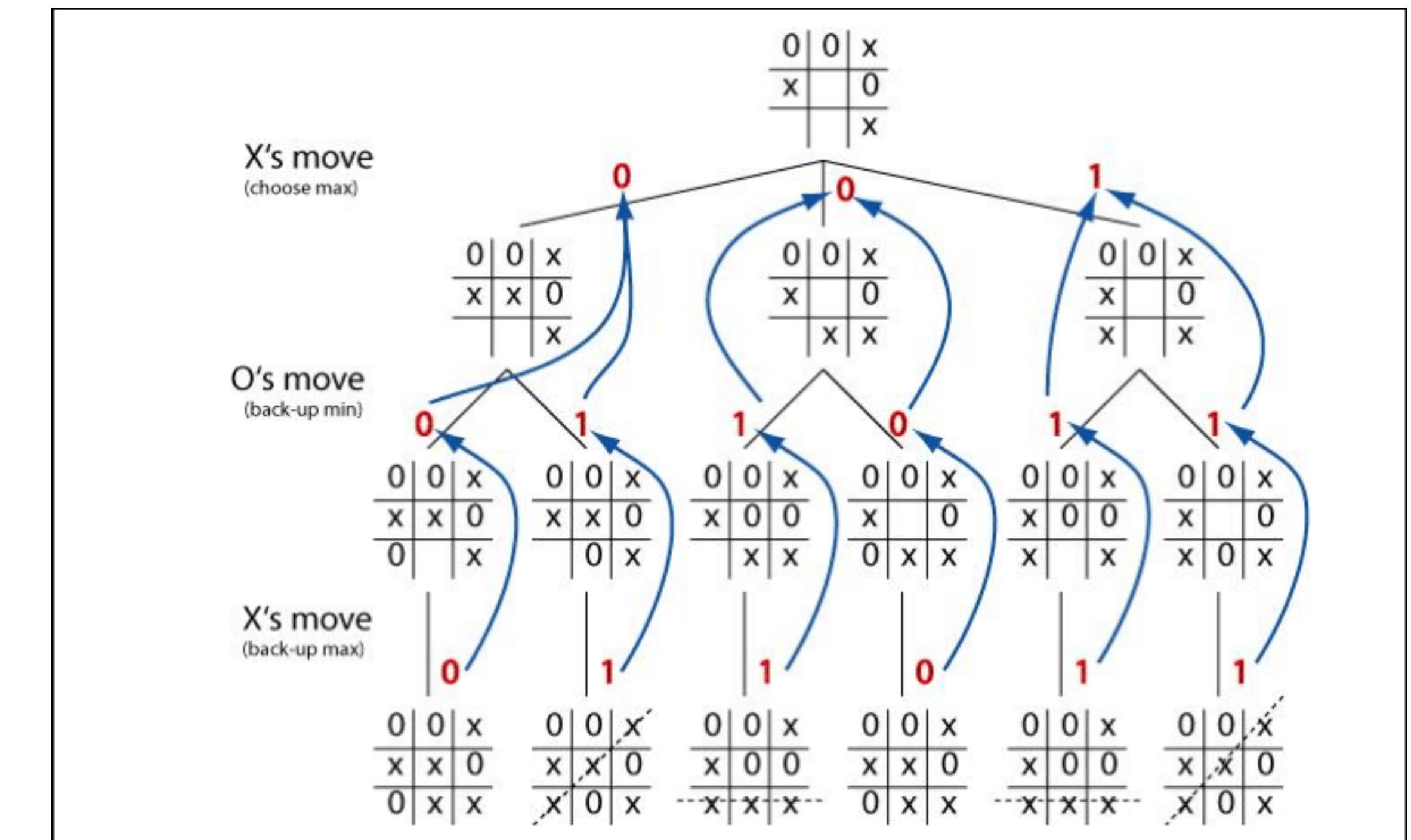
Оценивание
на основе игр (game-
based assessment)

За счет каких латентных
характеристик
выбираются такие
стратегии?



Игра “Крестики-нолики”

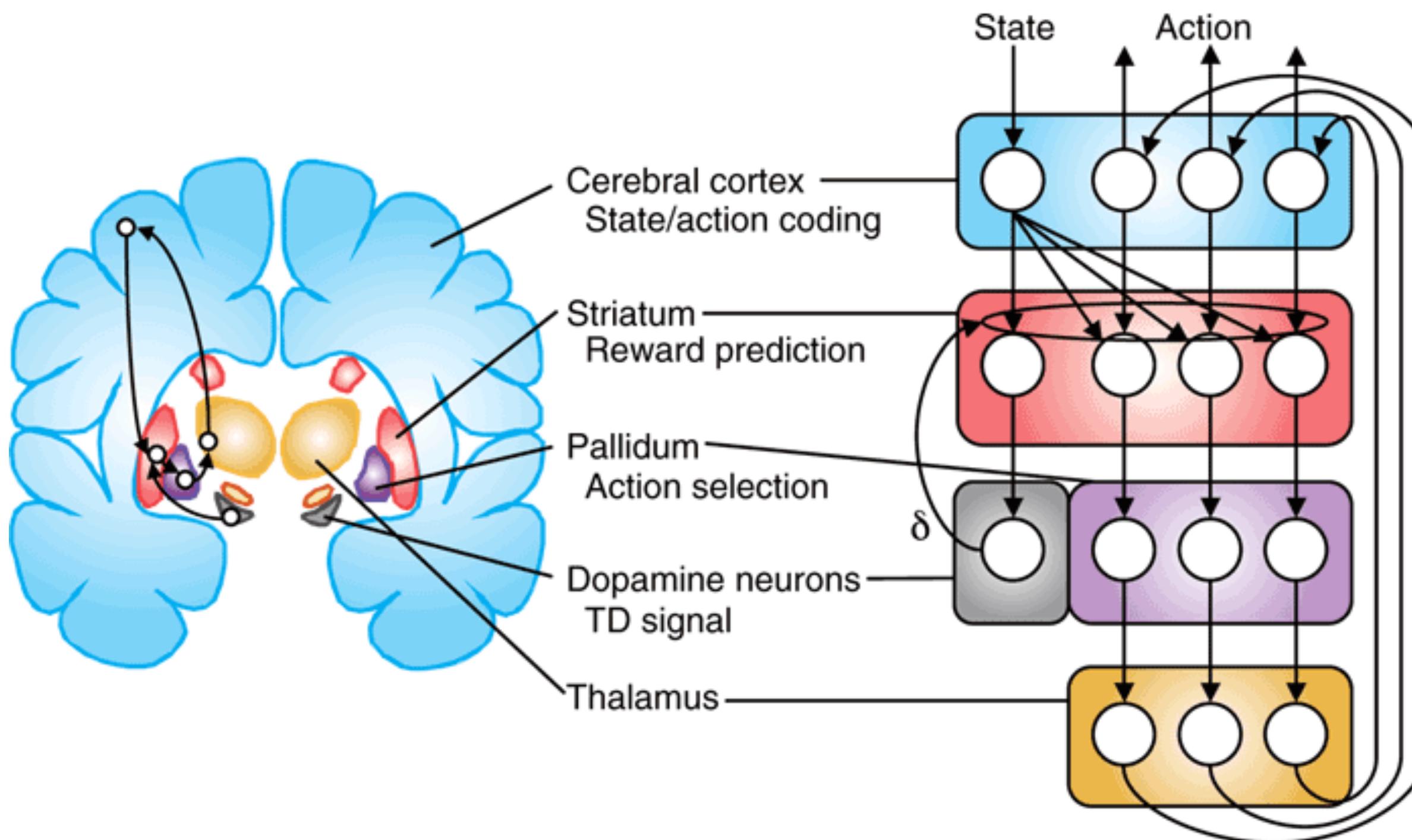
- Игра с нулевой суммой
- 5478 состояний, а с учетом симметрии и отражения - 765
- 255168 возможных игр, а с учетом симметрии и отражения - 26830
- Оптимальные стратегии (“ничейная смерть”):
 - Алгоритм Minimax
 - Reinforcement learning (обучение с подкреплением)



Алгоритм Minimax



Reinforcement Learning (обучение с подкреплением)

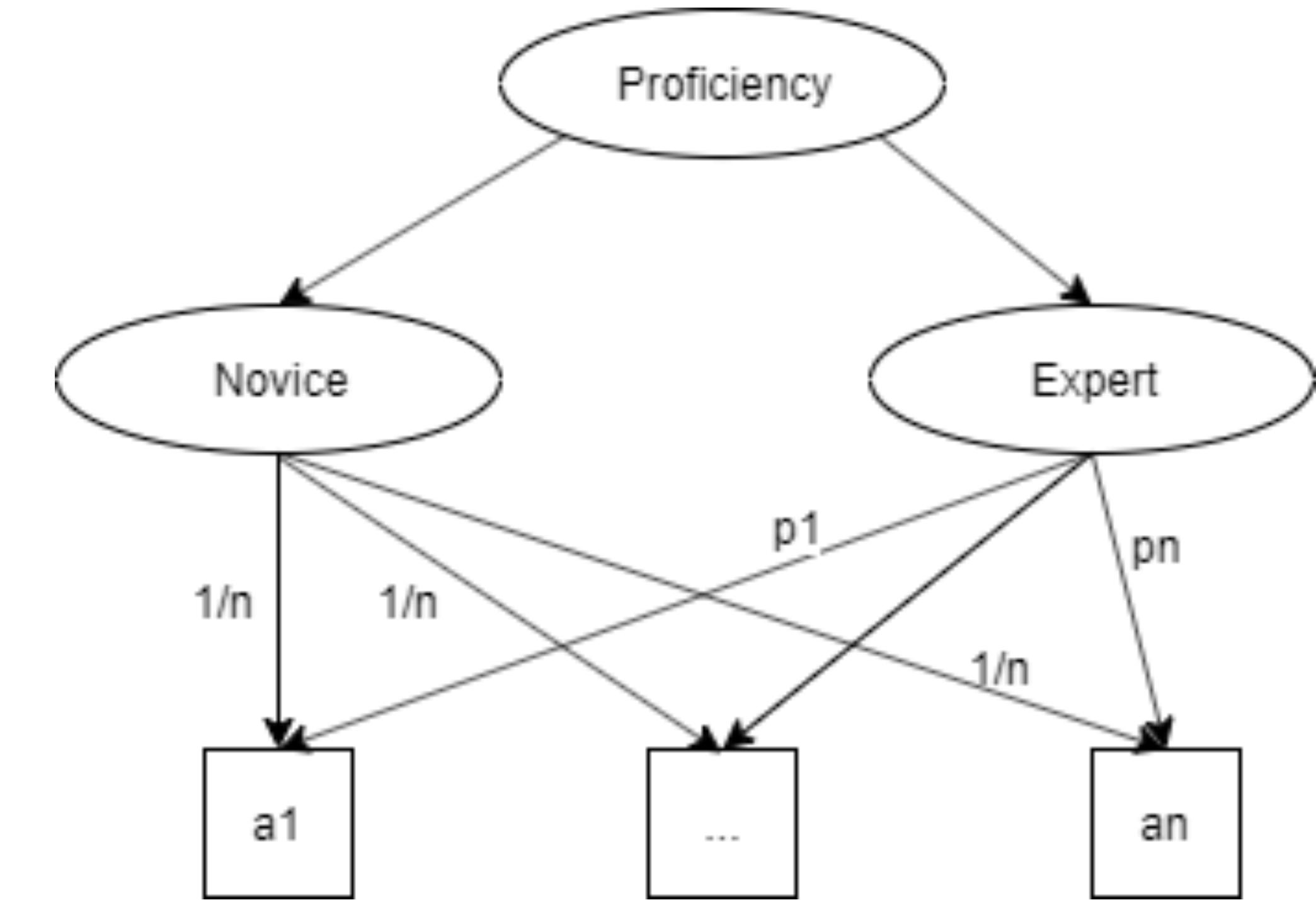




Смешивание стратегий

- **Новичок (novice):**
 - Выбирает стратегию наугад
- **Эксперт (mastery):**
 - Выбирает оптимальную стратегию

Для лучшего оценивания мы смешиваем выбор стратегии интеллектуальных агентов (в играх так часто задается уровень сложности)





Сложности и перспективы

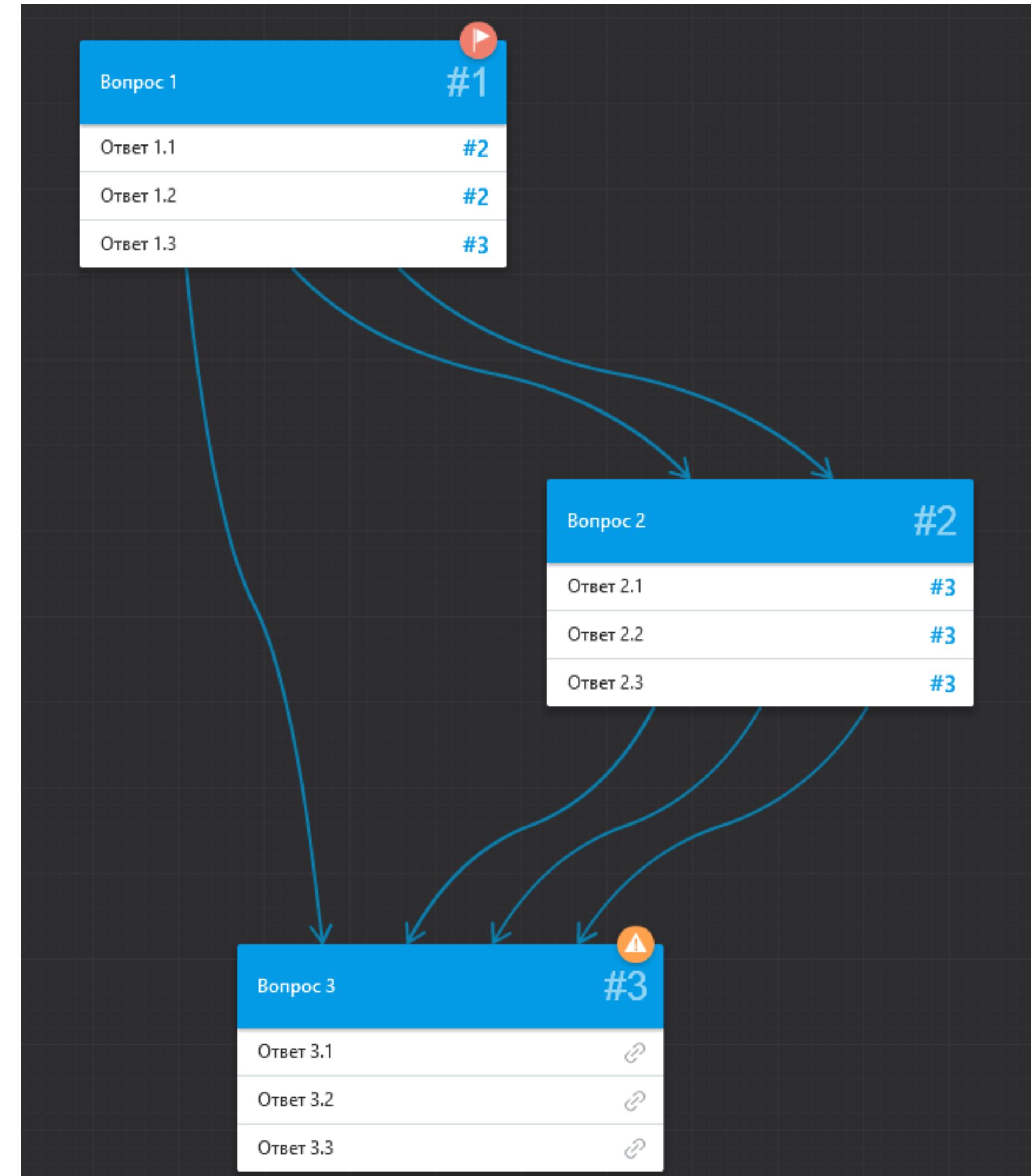
- Хотелось бы игру посодержательнее с интересным латентным конструктом.
«Крестики-нолики» - это модельный пример
- Нужны валидные данные о том, как люди играют в игры. На данных имитационного моделирования «банк заданий» не откалибровать

Анализ сценариев



Паттерн-дизайн и его анализ

- Студент: Дудинов Тимур Батырович
(«Прикладная математика»)
- Графовидная модель оценивания латентных характеристик респондентов с помощью диалоговых тренажеров /
A Graph-based Model for Assessment of Respondents' Latent Characteristics with Dialogue Simulations





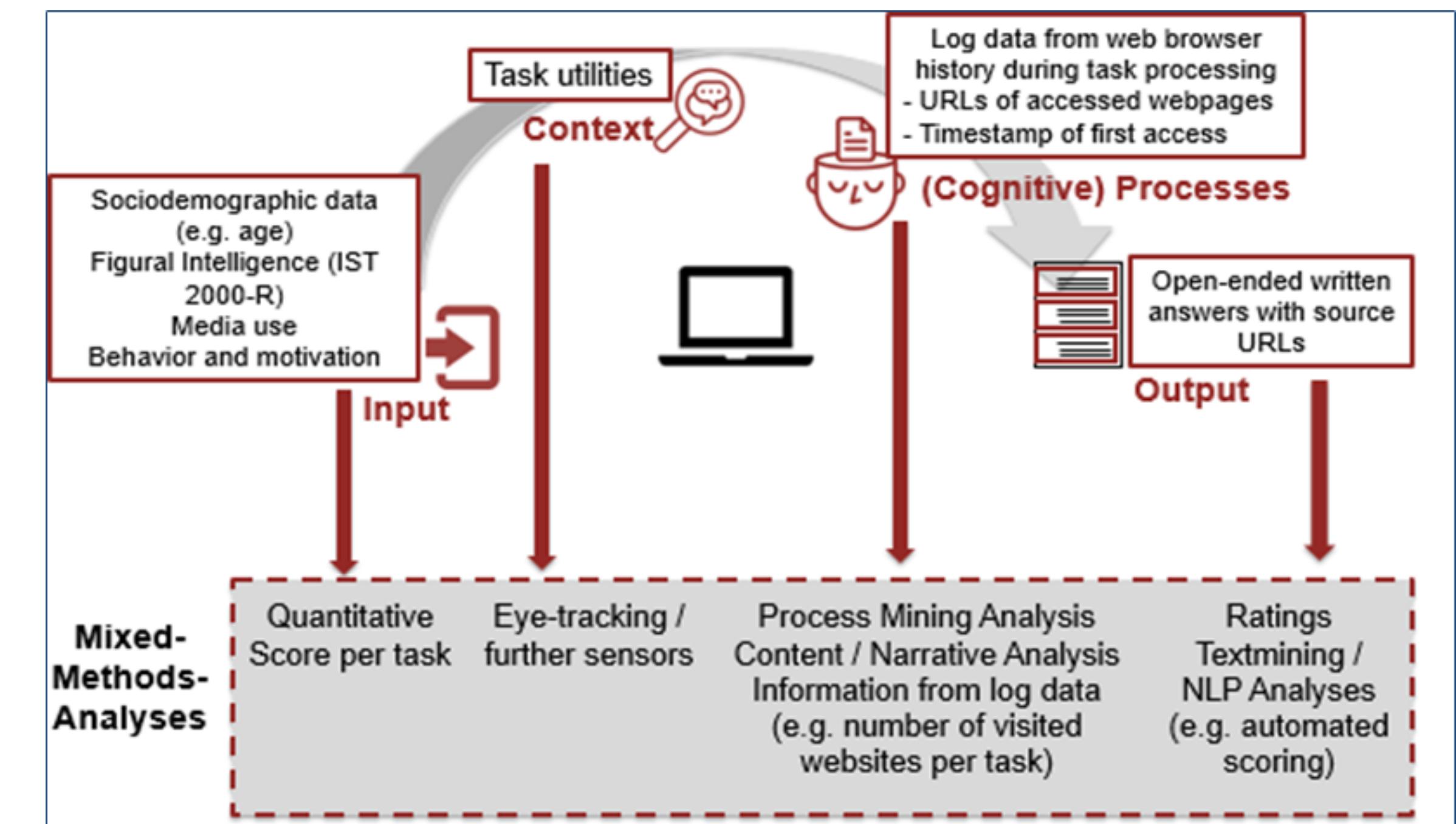
Сложности и перспективы

- Проверить модель на эмпирических данных, например, на игре из инструмента 4К
- Связать модель с анализом логов: ICL Test и CORA



Тест на критическое мышление для студентов вузов

- Прототипы: HElghten® Critical Thinking Assessment (ETS) и CORA (Critical Online Reasoning Skills, Университет Майнца)
- **Свободное использование интернет-ресурсов (open-book exam)**
- Методология разработки теста – Evidence Centered Design (ECD)
- Выявление паттернов поведения респондентов, автоматическое оценивание



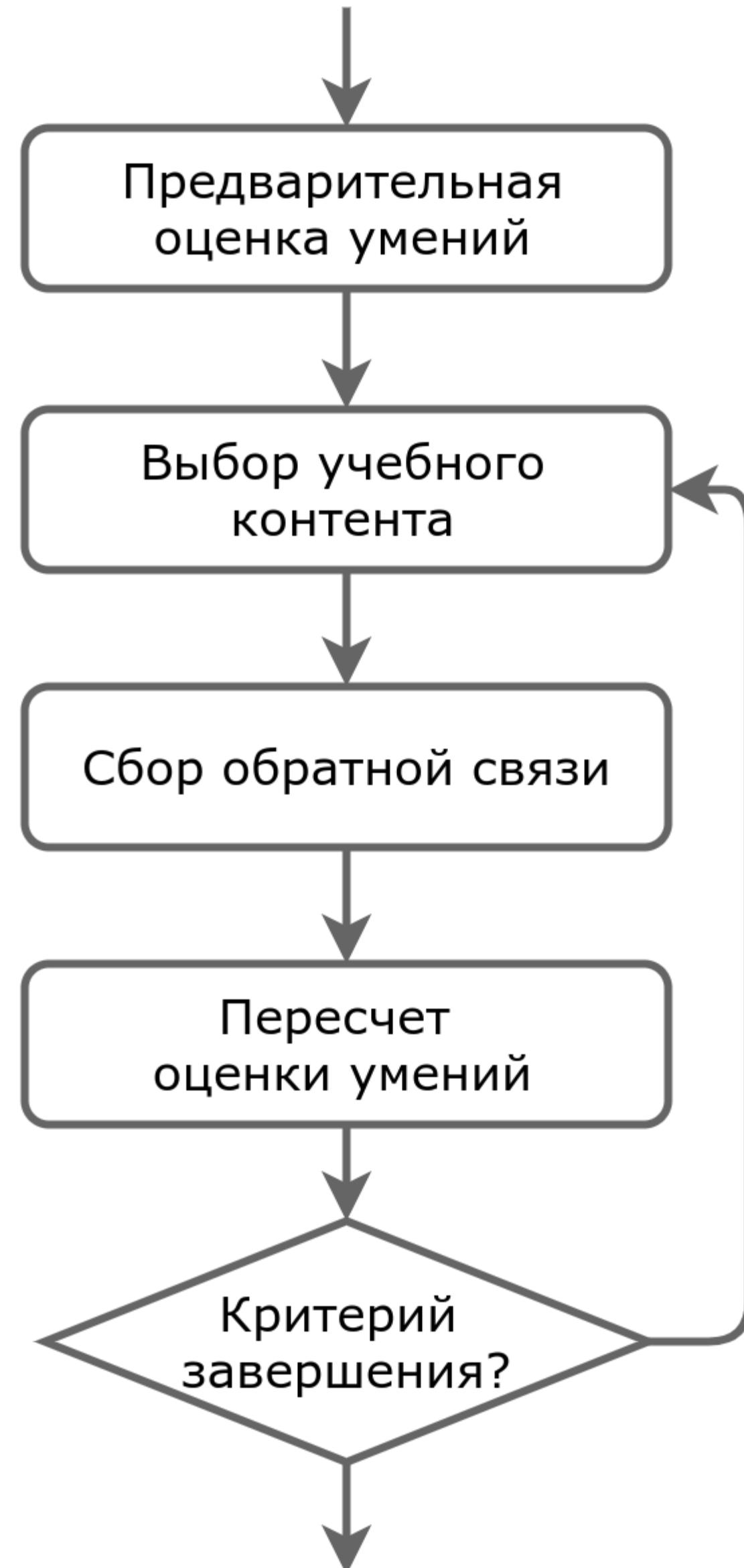
Источник: CORA

Анализ алгоритмов



Адаптивное тестирование и обучение

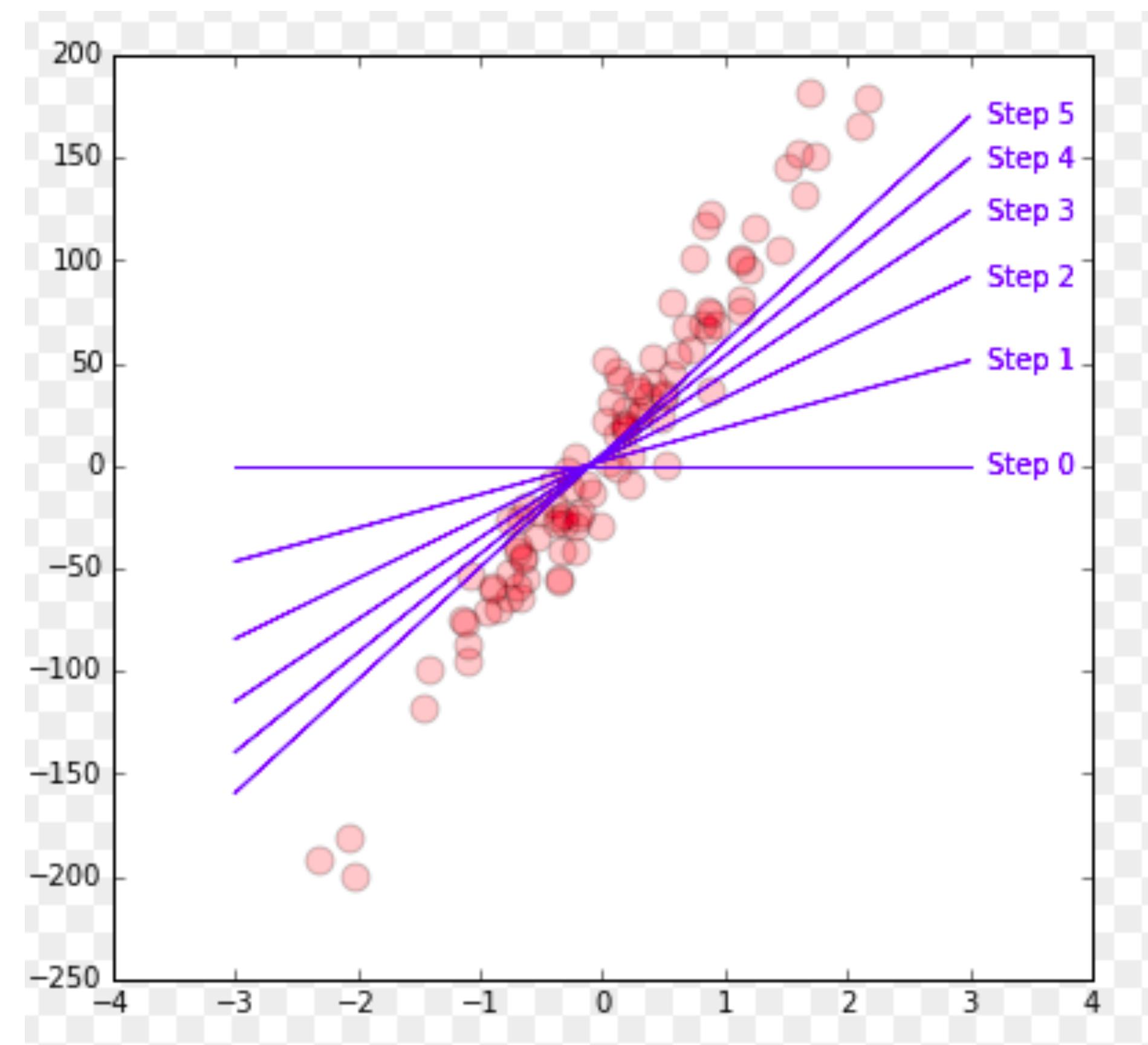
Адаптивное тестирование	Адаптивное обучение
Измерение характеристик	Развитие учащихся и обратная связь
Более стрессовое	Менее стрессовое
Обучающий эффект не желателен	Обучающий эффект необходим
Наиболее точное измерение характеристик	Наиболее сильный обучающий эффект





Алгоритмы реального времени (онлайн-алгоритмы)

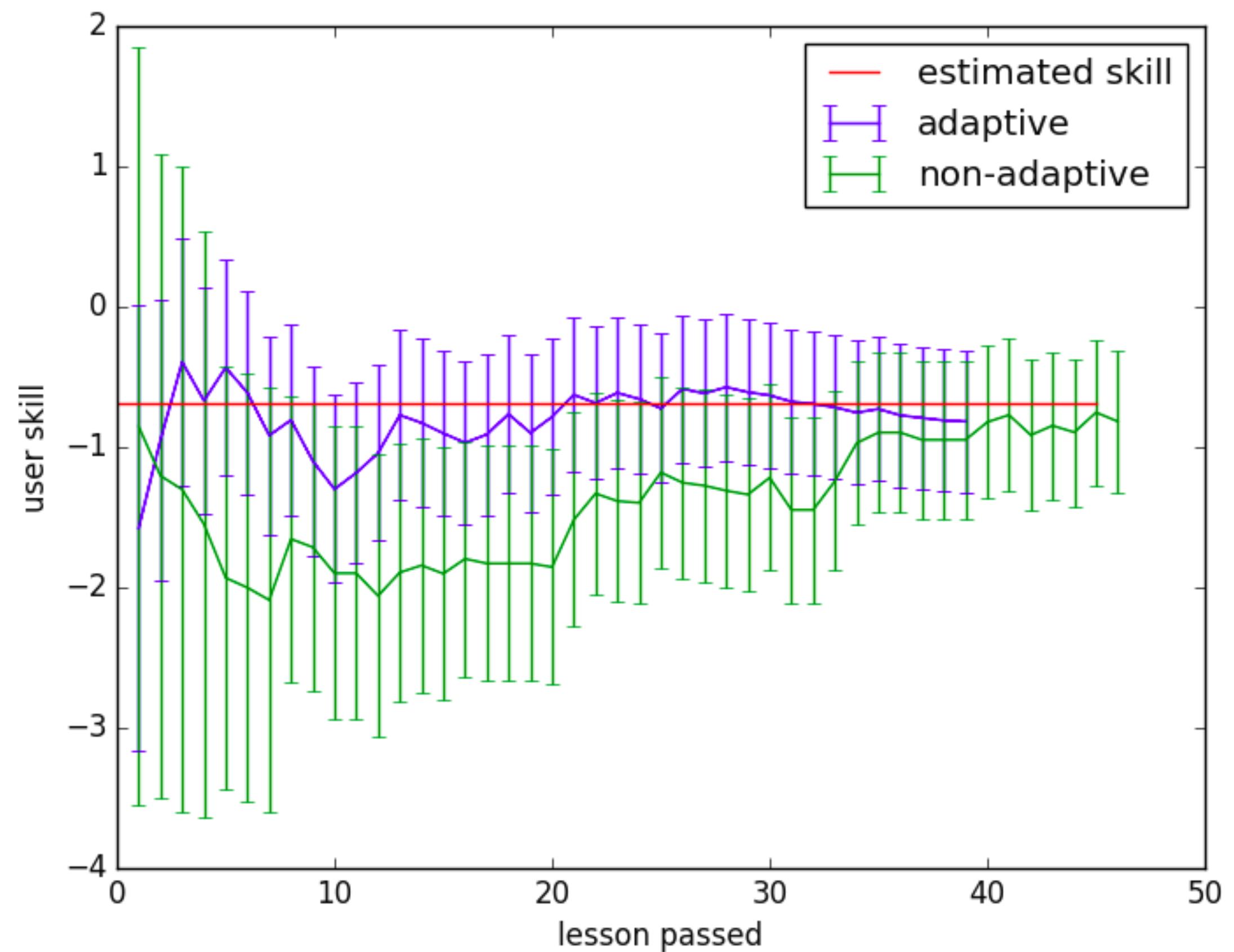
- Оптимизация параметров модели происходит при поступлении новых данных (т.н. **стохастический градиентный спуск/подъем**)
- В адаптивном тестировании и обучении используется для пересчета умений





Банк заданий и алгоритм рекомендации

- Определены психометрические характеристики заданий (банк заданий откалиброван)
- Можно ли использовать адаптивный алгоритм для банка заданий? Нужно проводить исследование с помощью имитационного моделирования:
 - выбор задания
 - пересчет умений
 - условие завершения





Адаптивный алгоритм в инструменте ПРОГРЕСС

- Критерий завершения на основе правила: три ошибки подряд или четыре ошибки в блоке заданий
- Предварительные результаты: точность оценивания зависит от порядка предъявления заданий, так как они по разному каллибруются

The Adaptive Algorithm with a Rule-Based Termination Criterion for Reading Literacy Assessment of Primary School Students

Maxim Skryabin

Centre for Psychometrics and Educational Measurement,
National Research University Higher School of Economics, Russia

Computerized adaptive testing (CAT) is a measurement approach where tailored tests for respondents are generated in real-time on the basis of their responses to previous items (Lord, 1980). CATs require six main components: an item bank of calibrated items, an entry or starting rule, an item selection rule, a response model, a method for scoring ability, and a termination criterion or stopping rule (Weiss & Kingsbury, 1984).

The precision of ability estimation, when the adaptive algorithm is used, depends on all six components included in CAT. But in this study, we only focus on a termination criterion, or stopping rule, that can be divided into fixed-length or variable-length rules, which refers to whether the test length is equal or varied across respondents. When fixed-length termination criteria are used, a predetermined number of items are administered. However, the precision of ability estimation is not consistent (Leroux & Dodd, 2014).

Variable-length termination criteria include the standard error, minimum-information, absolute-change-in-theta stopping rules, etc. (van der Linden & Glas, 2000; Babcock & Weiss, 2012). These types of stopping rules are often combined, i.e., a variable-length termination criterion is used until a maximum number of items is administered.

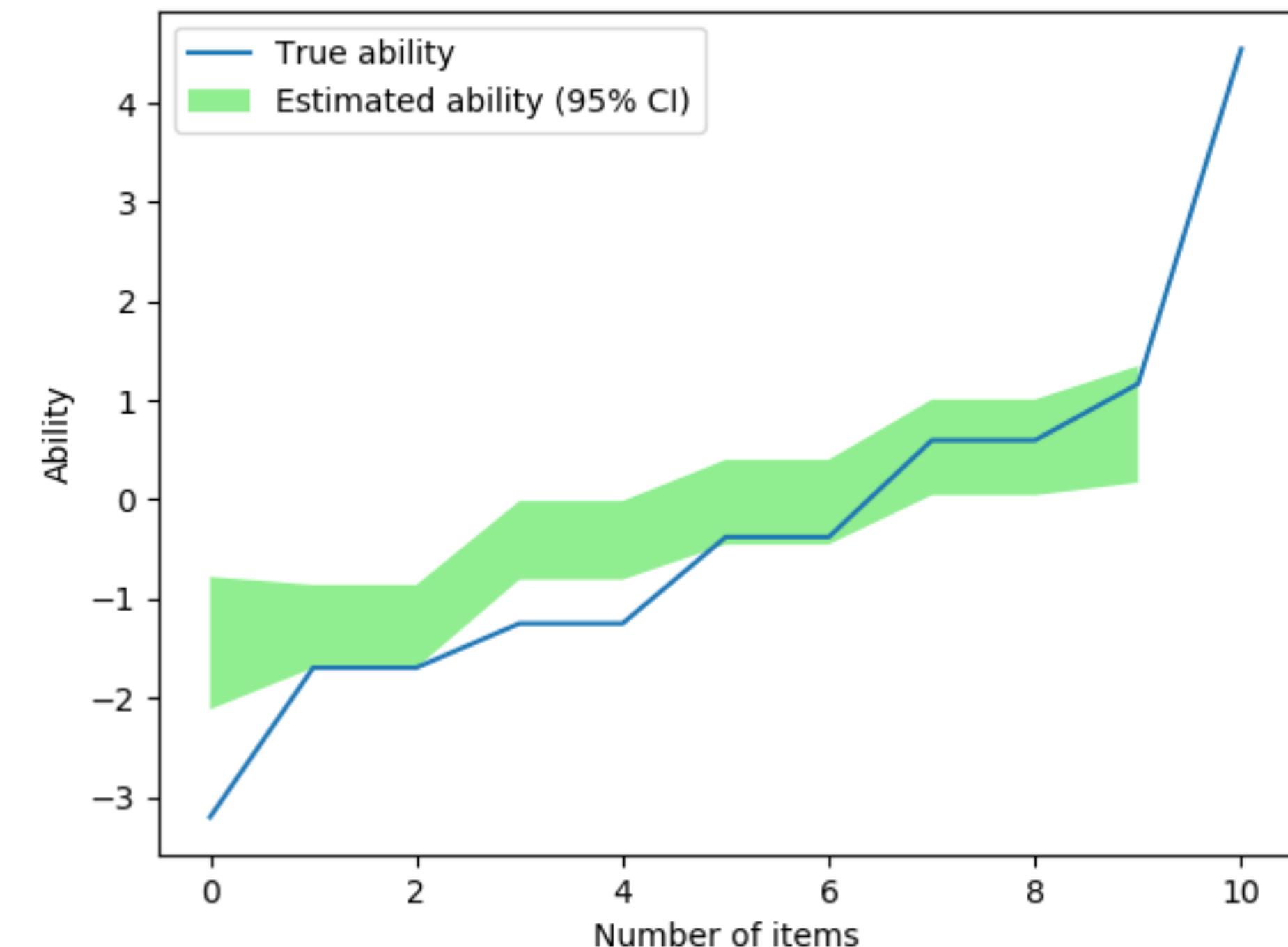
The performance of the termination criterion can depend on item bank characteristics or whether items are dichotomous or polytomous (Boyd et al., 2010). Therefore, it is important

Динамические модели



Динамические модели

- Assessment as learning: измеряемая характеристика меняется со временем
- Характеристики заданий - трудность и **обучающий эффект**
- Важно **наличие обратной связи**
- Применяется в формирующем оценивании и обучающих системах





Что за кадром?

- Интеллектуальные технологии: автоматический скринг, автоматическая генерация заданий
- Имитационное моделирование (simulation study)
- Учебная аналитика: отчеты обратной связи учителям и др.



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

Скрябин Максим Александрович
Ведущий аналитик, к.ф.-м.н.

E-mail: mskryabin@hse.ru