

Клонирование заданий сценарного типа

Грачева Дарья Александровна
аспирант, стажер-исследователь

Тарасова Ксения
к.п.н., зам. заведующего

Лаборатория измерения новых конструктов
и дизайна тестов ЦПИО НИУ ВШЭ



Клонирование – это подход разработки заданий, предполагающий создание максимально похожих заданий с едиными психометрическими характеристиками. В классической терминологии мы можем говорить о клонах как о параллельных вариантах заданий.

Задания сценарного типа фокусируются как на самом решении проблемы, так и на процессе решения, и представляют собой среду, в которой тестируемые демонстрируют уровень владения навыками в ответ на проблему.

Зачем нужны клоны заданий сценарного типа?

- Проведение мониторинговых исследований
- Снижение частоты списывания
- Возможность продемонстрировать навыки в ситуациях с разным содержанием
- Снижение затрат на разработку по сравнению с созданием нового сценарного задания; и сохранение качества уже валидизированного задания
- «Продление жизни» инструменту
- Справедливое оценивание независимо от задания

Art or Science?

Общие
подходы к
разработке
вариантов
теста

Существует несколько подходов к созданию вариантов теста

- «Традиционный подход» - создание вариантов по спецификации тестов и использование экспертной оценки для их валидации – слишком **субъективный**
- Случайный отбор заданий из банка с использованием статистических методов – слишком **технический**

Традиционный подход ориентируется на отдельные задания, является затратным как по ресурсам, так и по времени разработки.

Предполагает активное участие содержательных экспертов, несколько итераций обсуждений и правок

Часто не проходит проверку психометрического качества и параллельности

Подход, основанный на случайном отборе, требует наличие банка заданий, отсутствие локальной зависимости между заданиями, часто игнорирует содержательную сторону теста

Методы машинного обучения – слабо интерпретируемы, сложно выдержать сюжетную структуру заданий сложного формата

Развитие технологий **автоматической генерации заданий** (AIG)

- Автогенерация на основе **форм (шаблонов) заданий** (*template-based AIG*)
- Подход выявления **обязательных и вариативных элементов заданий**. В отличие от вариативных, обязательные элементы влияют на трудность задания. [Irvine, 2002]
- **«Слабая»** теория создания шаблонов заданий (Weak Theory) или Item Generation With 1-Layer Models with key characteristics

«Задание-родитель» задает структуру задания. На его основе создается новый вариант с помощью изменения элементов заданий для достижения вариантов с едиными психометрическими характеристиками. Такие задания называются изоморфами или заданиями-клонами [Clause et al, 2018; Bejar, 2003].

- **«Сильная»** теория создания шаблонов заданий (Strong Theory), Модели логических структур

Отделение когнитивной операции от контекста. Создание логической структуры, лежащей в основе шаблона задания. Обычно используется для точных наук

Yesterday, a veterinarian treated [x1] birds, [x2] cats, [x3] dogs. What was the ratio of the number of cats treated to the total number of animals treated by the veterinarian?

Правильный ответ:
 $x2 / (x1 + x2 + x3)$

Модели с 1 слоем (Weak Theory)

*Манипулируем обязательными элементами задания,
подбирая максимально похожие для клонов*

A 25-year-old female sees her doctor and reports that she's been experiencing a [Cough Type] cough. Upon examination, she presents with an oral temperature of [Temperature]. What is the most likely diagnosis?

- 1: Hay Fever
- 2: Ear Infection
- 3: Common cold**
- 4: Acute Sinusitis
- 5: Seasonal Influenza

Для элемента
"Temperature"
подбираются значения
с заданными
ограничениями.
Например, t от 37 до 39

*Возможно изменение правильного ответа и/или дистракторов

Манипулируем вариативными элементами задания.

A [Age]-year-old [male] sees her doctor and [described] that she's been experiencing a mild cough. Upon examination, she presents with an oral temperature of 37.8. What is the most likely diagnosis?

Для элемента "Age"
подбираются значения
с заданными
ограничениями.
*Например, возраст в
диапазоне от 20 до 35.*

Манипулируем структурой предложений (перефразировать)

A 25-year-old male experienced temperature of 37.8. He also had a mild cough. He told his doctor about the symptoms during regular examination. What the doctor should diagnosis?

В этом примере другие
элементы не
варьируются

Пример шаблона задания

Table 3.2 1-Layer Key Features Medical Item Model

| Parent Item: | |
|----------------|---|
| | <p>A 22-year-old female sees her doctor and reports that she's been experiencing a mild cough and slight body aches that have developed over a few days. Upon examination, she presents with an oral temperature of 37°C. What is the most likely diagnosis?</p> <p>1: Hay Fever 2: Ear Infection 3: Common cold 4: Acute Sinusitis 5: Seasonal Influenza</p> |
| Item Model: | |
| <i>Stem</i> | <p>A [Age]-year-old female sees her doctor and reports that she's been experiencing a [Cough Type] cough and [Body Aches] that have developed [Onset]. Upon examination, she presents with an oral temperature of [Temperature]. What is the most likely diagnosis?</p> |
| <i>Element</i> | <p>Age: 18–30, by 1 Cough Type: 1. mild, 2. hacking, 3. severe Body Aches: 1. slight body aches, 2. slight body pains, 3. severe body aches, 4. severe body pains Onset: 1. over a few days, 2. within 3-6 hours, 3. suddenly Temperature: 1. 37°C; 2. 37.8°C; 3. 39°C; 4. 39.5°C</p> |
| <i>Key</i> | Common cold; Seasonal flu |

«Банк» **дистракторов**

Правильный ответ
тоже может меняться
как из-за контекста, так
и из-за изменения
условий задачи

Задания-клоны (изоморфы) – это задания, созданные по одному шаблону и имеющие одинаковые психометрические характеристики

Table 1
Possible Combinations of Structurally and Psychometrically Isomorphic Tasks

| Psychometrically | Structurally | |
|------------------|---|--|
| | Not Isomorphic | Isomorphic |
| Not Isomorphic | Example: “Different” items with different psychometric properties | Examples: Instances from the same task template with different psychometric properties (e.g., a family of items created to assess learning on the same construct, but the items do not have similar item difficulty values; an item and itself on a later form if item drift has occurred) |
| Isomorphic | Example: “Different” items with the same psychometric properties | Instances from the same task template with the same psychometric properties (e.g., a family of items created to assess learning on the same construct and the items also provide similar item difficulty values, an item and itself on another form) |

- Позволяет упростить процесс разработки
 - Фокус смещается с конкретного задания на создание похожих заданий по шаблону
 - Не исключает использование экспертов. Однако эксперты могут быть заменяемы
 - Увеличивает вероятность создать задания с похожими психометрическими характеристиками и содержанием
-

Классические методы автогенерации включают три этапа:

1. Создание когнитивной модели ~ описание содержательной области, для которой будут генерироваться задания и их когнитивная трудность
2. Создание шаблона (модели) задания – шаблон, позволяющий понять, какие элементы заданий мы можем менять
3. Генерация задания (часто с использованием компьютерных методов)

Задания сценарного типа (Scenario-based task) – это совокупность заданий, объединенных общим контекстом. SBT предлагают ситуацию, где тестируемый последовательно решает ряд проблем, демонстрируя уровень владения навыком.

Задания сценарного типа удобно разрабатывать в логике метода доказательной аргументации (ECD).

Шаблон проектирования (Pattern Design) может являться примером менее структурированного шаблона для генерации

А что с сценарными заданиями?

| Название | Описание | Как это в дальнейшем ляжет в CAF |
|--|--|----------------------------------|
| Обоснование | Объяснение почему этот шаблон является важным, позволяет собрать необходимые свидетельства о ключевых компетенциях | |
| Ключевые компетенции | Описание КК, являющиеся целью этого паттерна | Student Model |
| Дополнительные компетенции | Другие знания / навыки/умения, которые могут потребоваться | Student Model |
| Потенциальное наблюдаемое поведение | То наблюдаемое поведение респондентов при взаимодействии с заданием, которое будет свидетельствовать о проявлении КК Касаются конкретных аспектов результатов деятельности. Описывают качества, сильные стороны или степени характеристик продуктов деятельности - свидетельства, которые продукты труда передают о КК (например, " количество...", " качество...", " уровень...", "вид..."). | Evidence Model |
| Потенциальный результат деятельности (work products) | Результат (например, устный ответ, письменное описание), в котором мы можем обнаружить проявление (свидетельства) КК | Task Model |
| Характерные особенности | Обязательные элементы ситуаций, в которых можно получить свидетельства | Task Model |
| Вариативные элементы | Аспекты ситуаций, которые можно варьировать для того, чтобы сместить сложность или фокус | Task Model |

Особенностью сценарных задания является контекст сценария. То есть SBT не отделимы от контекста.

Что такое контекст? Единого мнения на этот счет нет.

- Контекст в сценарных заданиях ассоциируют с «историей»
- Элемент задания, который не относится к предметному содержанию и является дополнением к заданию. Так, к контексту может относиться и *сложность текстов, «ресурсы» задания (например, картинки), роль и тип контекста, степень абстрактности, наличие нерелевантной основной задаче информации* и пр. [Ruiz-Primo, M. A., & Li, M.,2015]

Контекст определяет основную проблемную **ситуацию** задания и развитие ситуации (сюжета) – последовательность действий, отношения между этапами задания, персонажами и пр.

В целом, контекстом могут считаться любые элементы задания, характеризующую проблемную ситуацию.

Исследования показывают, что **контекст заданий может сказываться на результатах тестирования**

- Знакомый контекст дает преимущество даже при решении математических задач
- Для заданий PISA по естествознанию задания с информационными текстами оказались труднее, чем задания, где тексты положены в повседневный контекст.
- Для тех же заданий описательная роль контекста (контекст необходим для решения задачи) связана с более низкими результатами, чем сценарная роль контекста (контекст для создания истории) [Ruiz-Primo, M. A., & Li, M., 2015]
- В исследовании креативности контекст «виртуального мира», показанного через VR-шлемы, сказывался на рисунках детей [Nelson J., Guegan J., 2019]

Это связано с тем, что для заданий с контекстом характерно появление большой доли **нерелевантной дисперсии в результатах**.


- Например, исследование performance tasks по естествознанию показало, что большая доля дисперсии результатов связано с взаимодействием тестируемого и задания ~ 50%


Поэтому когда мы пытаемся клонировать задания с контекстом – мы с большой вероятностью можем только увеличить долю необъясненной дисперсии☹

Задание «Аквариум» - «Террариум»

Измерение критического мышления

Применение на
практике






Мы с мамой пришли в магазин, и оказалось, что пресноводные крабы закончились! Было очень обидно, но мы всё-таки купили других крабов — морских. Теперь придётся немного изменить аквариум! Я нашла несколько ссылок с информацией об устройстве аквариума для морского краба. Выбери одну ссылку, которая заслуживает доверия, и прочитай, что там написано.

Обустройство аквариума для морских крабов.
Основные правила.
zooacademic.ru

А я на море! Как быстро построить дом для морского краба - личный опыт.
animalblog.ru

[Выйти к списку заданий](#)

← C10 → 

Задание «Аквариум» - «Террариум»

Контекст: Помоги мне выбрать крабов и построить для них аквариум.

1. Подбор нового содержания сценария – гекконы и террариум
2. Выделение обязательных и вариативных элементов оригинального задания с упором на максимальную похожесть

Выделение **обязательных элементов** задания: ситуация, свидетельства проявления критического мышления, механики, симуляторы, последовательность действий внутри сценария.

Выделение **вариативных элементов задания**: сюжетные элементы (элементы контента)

Задание: Выдели предложения с важной информацией о том, что точно понадобится в аквариуме для крабов / террариуме для гекконов

Добавить в блокнот

Обустройство аквариума начинается с укладки грунта. На дно аквариума для крабов нужно положить слой песка. А вот рыбкам подойдет более крупный грунт.

Добавить в блокнот

Обустройство террариума начинается с укладки грунта. На дно террариума для гекконов нужно положить слой мягкой земли. А вот черепахам подойдут крупные опилки и щепки.

Задание «Динозавр»: оценка степени достоверности источника информации

Реальный пример 4К

Искать книгу в библиотеке / ссылку в инт

Энциклопедия динозавров

dino.wiki.org

Описание **массоспондила** и его место среди других динозавров

Обсуждение динозавров

dino.forum.org

Сайт для обмена мнениями о **массоспондиле** и других травоядных динозаврах

Сказки о динозаврах

dino.tales.org

Как поссорился **Массоспондил** Петрович с Птеродактилем Ивановичем

Энциклопедия динозавров – то что определяет достоверность источника (обязательный элемент)

Энциклопедия **ежей** – вариативный контекстный элемент

Учебник про динозавров – изменение обязательного элемента

Обязательные элементы регулируют трудность задания, а значит, связаны с конструктом, который вы измеряете

Вариативные элементы обычно служат контекстной цели или являются частью оформления

Применение на
практике

Что можно изменять в задании? Вариантов масса

- Изменение в «стеме»
- Дизайн экранов
- Изменение в вариантах ответа (ключ + дистракторы)
- Изменение характеристик персонажей
- Порядок заданий в тексте
- Порядок вариантов ответа в тесте
- Замена отдельных слов, перефразирование ...

На что стоит обратить внимание

- Релевантность ситуации
- Соответствие лексики целевой аудитории (сложность терминов)
- Степень знакомства (может играть как в плюс, так и в минус)
- Однозначность интерпретации (элемент не ведет к усложнению или перенасыщению контекста, не вызывает дополнительных интерпретаций)

Вариативность клонов может достигаться за счет дизайна экранов



Кажется этой **книги** нет в библиотеке! Что делать? Выбрать и прочитать другую **книгу** или не читать, а сделать так, как мне кажется правильно?

Прочитай другую книгу - верно

Можешь не читать, справимся без книги

Кажется, эта **ссылка** не открывается! Что делать? Найти другую **ссылку** или не читать, а сделать так, как мне кажется правильно?

Поищи информацию по другим ссылкам - верно

Можешь не искать, справимся и так

Первый вариант задания оказался значительно **труднее**.

Как понять, какие элементы являются обязательными, а какие – вариативными?

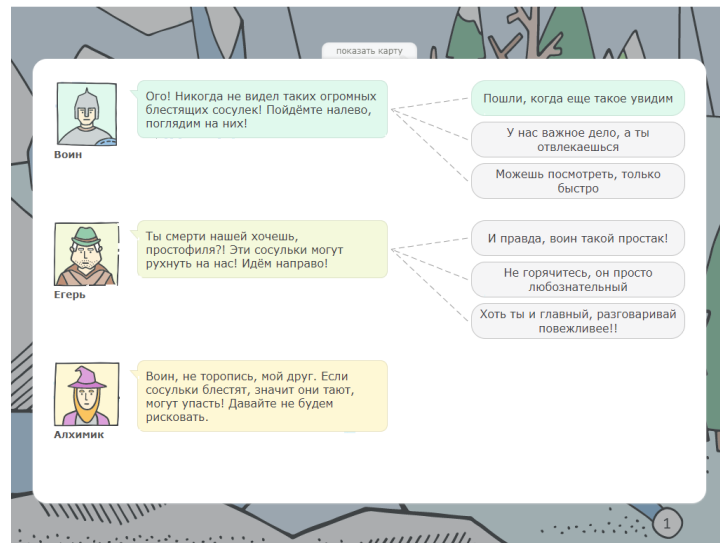
Конечно, здесь не обойтись без **мнения эксперта или разработчика**

Другой путь – **эмпирический**. Какие элементы задания действительно сказываются на трудности?

Простые методы: корреляция между трудностью задания и наличием определенного элемента

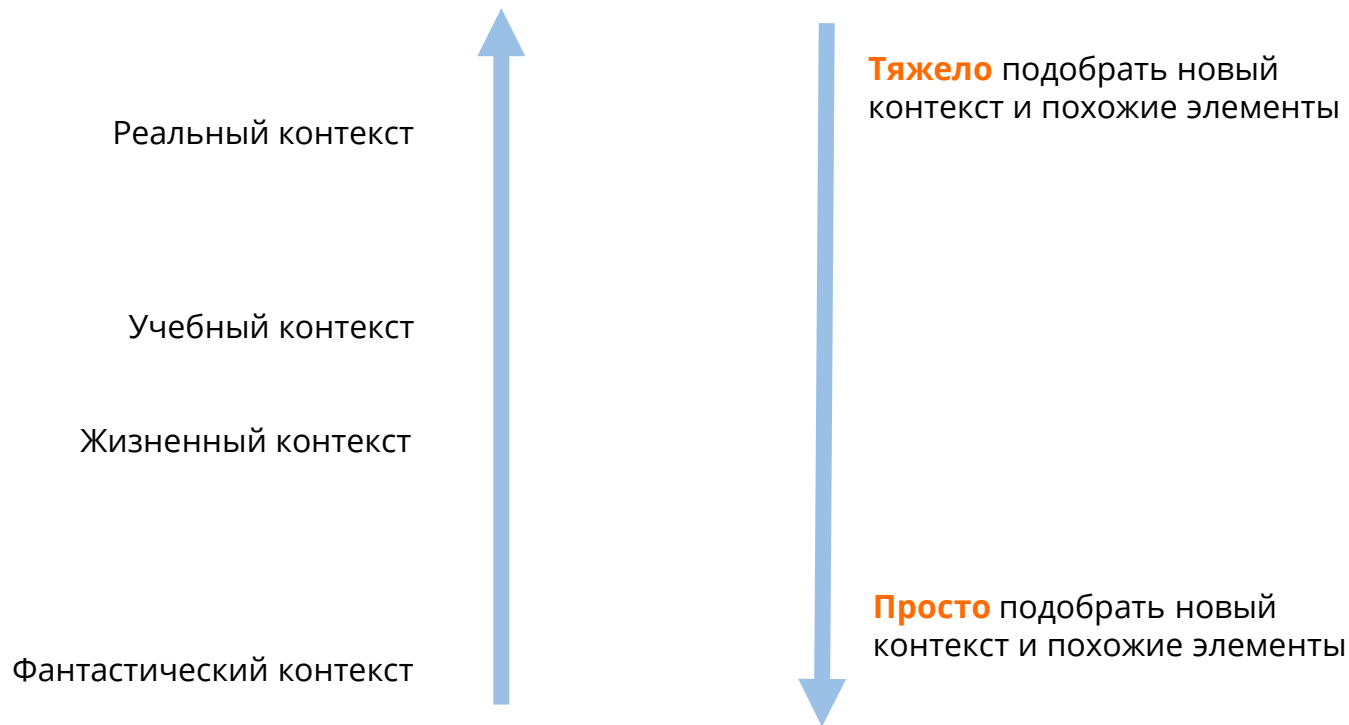
Более продвинутый метод: экспланаторные модели (linear logistic test model, LLTM)

На примере заданий по коммуникации и кооперации **текстовые характеристики (сложность текстов), количество дистракторов не имеют значимого вклада в трудность задания, в отличие от положения правильного ответа** (если правильный ответ первый – задание будет проще, чем если он последний или в середине)



**Обязательные
и вариативные
элементы**

Свобода в выборе элементов напрямую связана с контекстом задания



Packet Tracer Skills Assessments (PTSAs)

Задача – создание
изоморфов.

Изменение системы
подключений
(**обязательный**) + названий
(**вариативный**)

Примеры
исследований
для сложных
форматов и
навыков

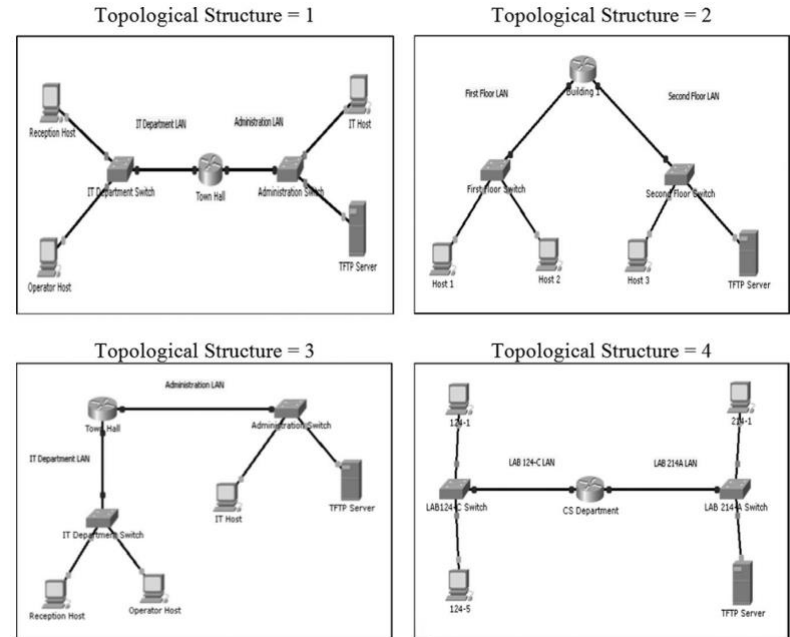
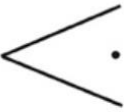




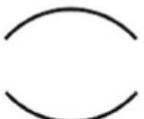




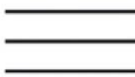





Figure 1. Four topological structures of device networks that served as structurally isomorphic tasks on the Packet Tracer Skills Assessment.

Test of Creative Imagery Abilities

Поворот элементов

| TCIA A - 1 | TCIA A - 2 | TCIA A - 3 | TCIA A - 4 | TCIA A - 5 | TCIA A - 6 | TCIA A - 7 |
|---|---|--|---|---|---|---|
|  |  |  |  |  |  |  |
| TCIA B - 1 | TCIA B - 2 | TCIA B - 3 | TCIA B - 4 | TCIA B - 5 | TCIA B - 6 | TCIA B - 7 |
|  |  |  |  |  |  |  |

[Jankowska D. M., Karwowski M. 2015.]

**Примеры
исследований
для сложных
форматов и
навыков**

Тест критического мышления в сценарном формате – не описан процесс разработки клонов

Wang D., Liu H., Hau K. T. Automated and interactive game-based assessment of critical thinking //Education and Information Technologies, 2021. pp. 1-23.

Проблемы клонирования

- Излишняя «**похожесть**» заданий, эффект запоминания – зависит от цели создания клонов. Возможно увеличения количества элементов клонирования. В АГЗ это модели с большим числом слоев (n-layer models)
 - Строгий шаблон не всегда приводит к более сопоставимым результатам [Грачева Д. А., Тарасова К. В., 2022]
 - В идеальном мире: прежде чем делать клон – **нужно разобраться с оригиналом!**
- Проверка структуры инструмента
 - Оценка психометрических характеристик задания (индикатора)
 - Анализ работы дистракторов

Иначе ошибки оригинала воспроизведутся в клоне!

Другие примеры клонирования сценарных заданий

