

ФГБОУ ВО «Московский государственный психолого-педагогический
университет»

На правах рукописи

Думин Павел Николаевич

**МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ И
ИДЕНТИФИКАЦИЯ ПАРАМЕТРОВ АДАПТИВНОГО
ТЕСТИРОВАНИЯ С УЧЕТОМ ВРЕМЕННОЙ ДИНАМИКИ
ВЫПОЛНЕНИЯ ЗАДАНИЙ**

05.13.18 – математическое моделирование, численные методы и
комплексы программ

05.13.01 – системный анализ, управление и обработка информации
(авиационная и ракетно-космическая техника)

диссертация на соискание ученой степени
кандидата физико-математических наук

Научный руководитель:
доктор технических наук, профессор
Куравский Лев Семёнович

Научный консультант:
кандидат физико-математических наук, доцент
Иванов Сергей Валерьевич

Москва, 2018

Оглавление

Введение	4
Общая характеристика работы	4
Современное состояние проблемы	9
Адаптивное тестирование с использованием байесовских сетей	13
Тестирование с использованием марковских цепей	14
Тестирование с использованием нейронных сетей	15
Глава 1 Способы адаптивного тестирования	20
1.1 Модель тестирования	20
1.2 Тестирование операторов	31
Глава 2 Методы оптимизации порядка предъявления тестов	34
2.1 Метод оптимизации последовательности предъявления тестов	34
2.2 Метод коррекции наблюдением	41
2.3 Программный комплекс	43
Глава 3 Вероятностные модели классического и игрового тестирования	46
3.1 Моделирование процесса выполнения тестовых заданий	46
3.1.1 Идентификация параметров рекуррентных зависимостей	50
3.1.2 Использование результатов моделирования	52
3.2 Процесс тестирования в игровой форме	53
3.2.1 Идентификация параметров	57
3.3 Пример практического применения	57
3.3.1 Моделирование процесса выполнения заданий теста Равена	57
3.4 Использование результатов моделирования	61
Глава 4 Разработка и сравнение численных методов идентификации моделей	62

4.1	Градиентный метод: алгоритм вычислений	63
4.2	Метод полной дискретизации: алгоритм вычислений	64
4.3	Метод дискретизации значимых параметров: алгоритм вы- числений	65
4.4	Вычислительный эксперимент	67
4.5	Параметры процедур идентификации	70
4.6	Результаты вычислительного эксперимента	73
	Заключение	85

Введение

Общая характеристика работы

Компьютерное тестирование в настоящее время широко используется с целью диагностики навыков и способностей, включая выявление степени пригодности обучающихся для выполнения тех или иных функций. Качество тестирования и достоверность его результатов в значительной степени зависят от технологий проведения тестов, которые в последние десятилетия стали предметом активных научных исследований.

Различные подходы к построению процедур тестирования рассмотрены в работах Г. Раша [90], Б. Райта [96, 97], М. Стоуна [97], Ф. Бейкера [60], Д.Р. Бергана [61], П.Ф. Лазарсфельда [24, 82], Л.С. Куравского [19–23, 27, 72–80], Г.А. Юрьева [19–23, 27, 57, 74, 77–80], А.В. Наумова [31–33], А.И. Кибзуна [16], А.О. Иноземцева [16], С.И. Панарина [15], А.Г. Шмелева и В.В. Столина [56], В.С. Аванесова [2], Чельшковой М.Б. [54] и др., послужив теоретической и методологической основой представляемой работы. В частности, в работах Кибзуна А.И., Иноземцева А.О. и Панарина С.И. подробно рассматриваются задачи построения оценок уровней подготовки испытуемых по результатам ответов на серии тестов, задачи получения статистических оценок трудностей заданий и особенности применения подхода в системе дистанционного обучения для студентов. В работах Наумова А.В. рассмотрены случаи ограниченного по времени тестирования, а также способы детекции артефактов тестирования, направленных на искажение результатов (угадывание, использование сторонних источников и т.д.). В работах Юрьева Г.А. также рассмотрены модели тестирования, позволяющие корректировать результаты тестирования с использованием специально адаптированных методов оптимальной линейной фильтрации.

Особенный интерес представляет область диагностики практических навыков и когнитивных способностей операторов сложных технических систем и игровая диагностика. Это позволяет говорить об актуальности поиска и разработки новых принципов построения технологий тестирова-

ния. Одним из наиболее перспективных результатов в этой области стал новый подход [20, 21], построенный на использовании обучаемых структур в форме марковских моделей с дискретным и непрерывным временем. Его преимуществами перед аналогичными способами тестирования являются:

- выявление и использование при построении расчётных оценок временной динамики изменения способности справляться с заданиями теста;
- возможность учёта при построении расчётных оценок времени, затрачиваемого на решение тестовых задач;
- эффективная реализация адаптивных процедур, обеспечивающих меньшее по сравнению с другими подходами число заданий, которое следует предъявлять испытуемому для получения оценок навыков или способностей с заданной точностью;
- развитая техника идентификации параметров моделей.

На основе этой технологии разработаны адаптивные процедуры тестирования и системы поддержки принятия решений (СППР), которые ускоряют процесс диагностики, оптимизируя предъявление тестовых заданий с участием эксперта. Эксперт, получая от системы необходимую информацию для анализа, предоставляет, на основе своих критериев, рекомендации по выбору следующего теста, обладающего, по сравнению с прочими, наибольшей для данного испытуемого дифференцирующей способностью. Диагностика строится на основе уточняющихся в процессе тестирования вероятностных оценок принадлежности испытуемых к различным категориям. Однако итоговые оценки при этом, как правило, определяются только по формальным результатам прохождения тестовых заданий, предъявляемых испытуемому, без учёта изменений в процессе их выполнения когнитивных способностей и психофизиологического состояния человека, что существенно ограничивает возможности измерительной процедуры, особенно в случае игровой диагностики.

Указанные особенности современных тестовых процедур позволяют говорить о необходимости:

- разработки новых подходов к моделированию процесса прохождения отдельно взятых тестовых заданий, позволяющих учесть динамику способностей испытуемых и трудностей этих заданий непосредственно во время их выполнения;
- разработки моделей процессов тестирования, проводящихся в игровой форме;
- формализации действий эксперта, обеспечивающего адаптивный выбор заданий для СППР;
- адаптации разработанных моделей для оценки уровня подготовки операторов сложных технических систем.

Таким образом, **актуальность темы диссертации** обусловлена необходимостью создания эффективных подходов к оценке навыков и способностей с помощью адаптивных компьютерных процедур, учитывающих временную динамику выполнения тестовых заданий и изменения в состоянии испытуемых, а также особенности использования этих процедур в задачах игровой диагностики и оценки уровня подготовки операторов сложных технических систем.

Цель работы: разработка математических моделей и методов адаптивного компьютерного тестирования навыков и способностей с учётом динамики изменения оцениваемых показателей и трудности заданий во времени.

Для достижения цели в процессе исследования решены следующие задачи:

- создания математической модели адаптивного тестирования в рамках СППР, позволяющей интерпретировать динамику выполнения

тестовых заданий и упражнений для оценки уровня навыков и способностей при психологическом и педагогическом тестировании, а также при управлении сложными техническими системами;

- создания вероятностной модели процесса выполнения отдельно взятого тестового задания, учитывающей изменение во времени способностей испытуемого и трудности задания;
- разработки численных методов для идентификации применяемых вероятностных моделей;
- разработки комплексов программ, реализующих разработанные методы и модели тестирования;
- оптимизации параметров процедуры компьютерного адаптивного тестирования.

Методологические основы и методы исследования: для решения поставленных задач использовались методы математического моделирования, теории оптимизации, системного анализа, статистического анализа и численные методы.

На защиту выносятся следующие научные результаты:

- математические модели адаптивного тестирования, лежащие в основе разработанной СППР;
- вероятностная модель процесса выполнения отдельно взятого тестового задания, учитывающая динамику трудностей заданий и способностей испытуемых;
- численные методы идентификации вероятностных моделей;
- комплексы программ, реализующие СППР и процедуру адаптивного тестирования, а также идентификации параметров применяемых моделей;
- методы оптимизации параметров компьютерного тестирования.

Научная новизна заключается:

- в математических моделях интерпретации результатов выполнения тестовых заданий в условиях наличия и отсутствия эксперта и адаптивности предъявляемых заданий с учетом временных ограничений;
- в вероятностной модели прохождения отдельно взятого тестового задания на базе модифицированной функции Раша;
- в методах оптимизации порядка предъявления тестов;
- в численных методах решения задачи идентификации вероятностных моделей.

Практическая значимость диссертационной работы заключается в возможности создания принципиально новых инструментов для оценки навыков и способностей, эффективность которых обусловлена настройкой используемых математических моделей по реальным результатам рассматриваемых категорий обучающихся, адаптивностью измерительной процедуры, учётом временной динамики выполнения тестовых заданий и результатом измерения, представленным в виде вероятностных оценок.

Достоверность результатов исследования подтверждается:

- оценкой адекватности полученных результатов наблюдениям с помощью статистических критериев согласия,
- практической реализацией и успешным применением системы поддержки принятия решений, созданной на основе разработанного подхода,
- вычислительными экспериментами, подтвердившими эффективность и преимущества созданных численных методов идентификации вероятностных моделей.

Апробация: теоретические основы и практические результаты работы представлены на следующих научных мероприятиях: Всероссийских научных конференциях «Нейрокомпьютеры и их применение» -2012, -2013,

-2014, -2015, 2016; Международных студенческих школах-семинарах «Новые информационные технологии» –2012, –2013; Всероссийских выставках научно-технического творчества молодежи «НТТМ» -2012, -2013, -2014, -2015». Результаты работы отмечены премией Президента РФ для поддержки талантливой молодёжи («НТТМ-2013»), медалью «За успехи в научно-техническом творчестве» («НТТМ-2013», «НТТМ-2014»), дипломом «НТТМ-2015», дипломом за лучшую работу на Международной студенческой школе-семинаре «Новые информационные технологии – 2012», дипломом за лучшую научную работу, представленную на конференции «Нейрокомпьютеры и их применение – 2013»; почетным дипломом РАЕН за лучшую научную работу (2013).

Современное состояние проблемы

В последние годы использование тестирования как способа измерения знаний, способностей и прочих характеристик стало неотъемлемой частью образовательного процесса. Под тестированием обычно понимается экспериментальный метод в психологии и педагогике, стандартизированные задания, позволяющие измерить психофизиологические и личностные характеристики, а также знания, умения и навыки испытуемого [2, 30, 49]. Среди существующих форм тестирования можно выделить компьютерное тестирование, в последнее время все больше используемое для оценки навыков и способностей испытуемых. Системы тестирования в таком случае часто используют механизм адаптивности при предъявлении тестовых заданий или, в случае использования батарей тестов, целых тестов. Адаптивное тестирование позволяет сократить время тестирования, а также повысить надежность оценок измеряемых характеристик (использование различных моделей тестирования позволяет повышать эффективность различного спектра характеристик).

Адаптивное тестирование можно разделить на классы в зависимости от выбранного критерия: цель тестирования, способ построения траектории тестирования и т.д. В работе [10] рассматривается классификация ви-

дов адаптивного тестирования. Одним из критериев для классификации видов адаптивного тестирования служит критерий «Способ построения траектории тестирования»:

- Способ с помощью теории IRT.
- Способ на основе Байесовских сетей.
- Способ на основе цепей Маркова.
- Способ на основе нейронных сетей.
- Прочие способы.

Теория IRT

Источником возникновения подхода с использованием теории является распространение логистической функции в различных гуманитарных науках. Логистическая функция вида $Y = \frac{e^x}{1 + e^x}$ широко применялась в биологии для моделирования прироста растительной массы или роста организмов. В середине XX века эту функцию предложили использовать в качестве модели психологического и педагогического тестирования [63]. Позднее ряд ученых предлагали различные сходные модели для оценки способностей испытуемых и трудностей заданий [62, 81, 84]. В ходе формирования модели IRT в ее основу был положен более развитый подход – теория латентно-структурного анализа (LSA), основная идея которого связана с выявлением латентных качеств (факторов) поведения посредством математико-статистических моделей измерения [83]. В работе [66] модель измерения определяется как «структурное построение, позволяющее соединить латентные переменные с одним или большим числом наблюдаемых переменных».

Предположения модели IRT состоят в следующем:

- 1) Существуют латентные параметры личности, не поддающиеся для непосредственного наблюдения.

- 2) Существуют переменные, связанные с латентными переменными, доступные для непосредственного измерения.
- 3) Оцениваемый латентный параметр должен быть одномерным.

Сходными исследованиями занимался Г. Раш [90]. Его модель, во многом схожая с моделью IRT, состоит в следующем. В модели IRT устанавливается соответствие между наблюдаемыми переменными и латентными параметрами (решениями конкретных тестовых задач и способностями соответственно). В качестве множества латентных параметров рассматриваются значения уровня подготовленности испытуемых θ_i , где i — номер испытуемого. Второе множество состоит из значений латентной переменной, характеризующей трудность задания β_j , где j — номер задания. Предположение Раша заключалось в том, что уровень сложности и уровень подготовленности испытуемого могут быть измерены в одних логарифмических единицах, которые Раш назвал «логит уровня знаний» и «логит уровня трудности задания» [90]. По Рашу:

- 1) Логит уровня знаний: $\theta = \ln \left(\frac{p_i}{q_i} \right)$, где p_i — доля правильных ответов испытуемого (подсчитанная как отношение доли правильных ответов к общему количеству заданий в тесте), q_i — доля неправильных ответов испытуемого ($q_i = 1 - p_i$).
- 2) Логит уровня трудности задания: $\beta = \ln \left(\frac{v_j}{u_j} \right)$, где v_j — доля правильных ответов на j -е задание теста, u_j — доля неправильных ответов на j -е задание теста.

С учетом предположения Раша об одинаковых единицах измерения двух приведенных величин, в качестве соотношения между этими параметрами была использована разность. Таким образом, модуль разности определяет расстояние, на котором находится испытуемый с конкретным уровнем подготовки θ_i от задания с определенным уровнем сложности β_j . Большое значение разности интерпретировалось как показатель низкой дифференцирующей силы задания: для испытуемого он либо слишком

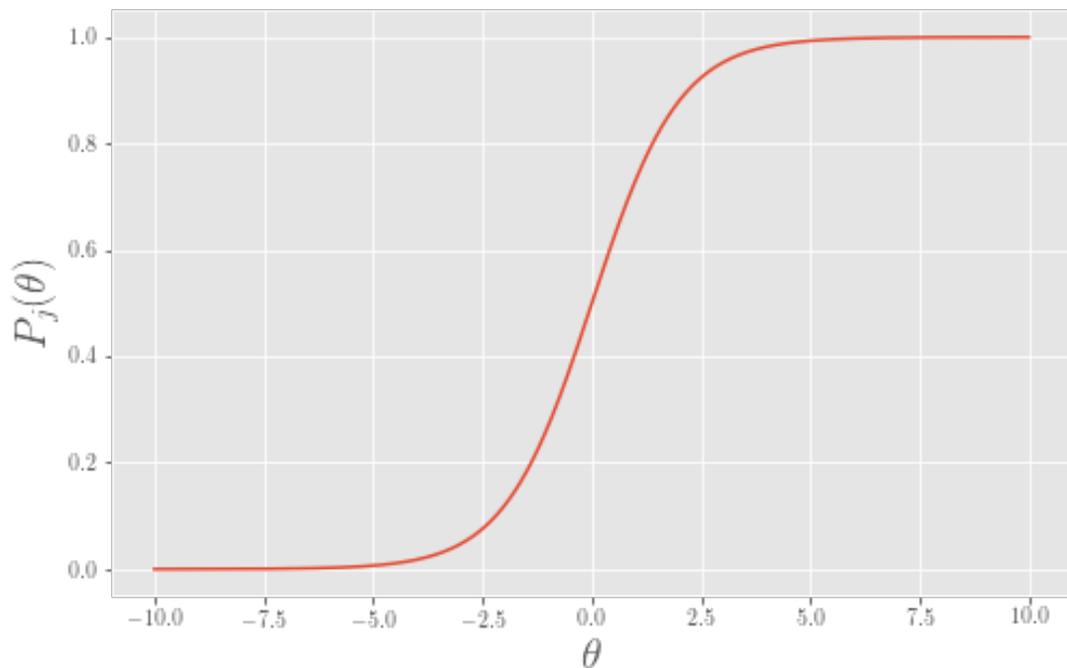


Рис. 1. График функции, иллюстрирующий зависимость вероятности правильного ответа $P_j(\theta)$ от значения θ логита уровня знаний.

сложно, в этом случае он не сможет его решить, либо слишком просто, что означает, что испытуемый подготовлен по исследуемой характеристике.

В более поздних исследованиях [2, 61, 71, 83, 84, 87] предложенная модель была усложнена и дополнена различными параметрами, позволяющими повысить информативность оценок тестирования и адаптацию под различные тестовые методики.

Применение технологии IRT приводит к следующим проблемам:

- «статичности» оценок: игнорированию изменения состояния испытуемых, оказывающего влияние на их способности;
- невозможности учёта времени, затрачиваемого на решение тестовых задач, при построении расчётных оценок, которые определяются только по формальным результатам прохождения теста;

- необходимости выполнения достаточно большого числа заданий для получения оценок с приемлемой точностью.

Адаптивное тестирование с использованием байесовских сетей

Байесовская сеть — это ациклический граф, в котором каждая вершина представляет n -значную переменную, дуги обозначают существование причинно-следственных зависимостей между переменными, сила зависимостей выражается в виде условных вероятностей, соответствующих каждой из переменных [88]. Формальное описание байесовских сетей приведено во множестве работ [34, 88]. Одним из распространенных подходов в создании адаптивных алгоритмов тестирования с помощью байесовских сетей является подход, в котором выдвигается предположение об одинаковой структуре знаний испытуемого и информационной обучающей системы [53]. В случае рассмотрения процесса тестирования по определенной дисциплине, предполагается, что структура изучаемого предмета состоит из глав, каждой из которых соответствует некоторый набор понятий. Последовательность тестовых заданий требует от испытуемого некоторого уровня владения набором понятий, которые могут понадобиться для успешного выполнения заданий теста. Помимо ситуаций, в которых испытуемый успешно выполняет задание, либо не выполняет его по причине отсутствия компетенции, с которой связан пункт теста, существуют ситуации неопределенности, когда испытуемый может случайно правильно ответить на вопрос, ответа на который он не знал, а также ситуация, в которой испытуемый неправильно отвечает на вопрос, ответ на который ему известен. Подобные случаи необходимо учитывать при построении байесовских сетей для адаптивного тестирования.

В зависимости от выбранного подхода, сопоставление конкретного объекта вершинам может меняться: каждой вершине может соответствовать элемент семантической модели знаний испытуемого [42], элементарным компетенциям [53], заданиям и т.д. Часто в разрабатываемую структуру сети закладывают ряд вспомогательных переменных, например, вершин, означающих наличие или отсутствие какой-либо внешней информации об

объектах.

Алгоритмы построения байесовских сетей для конкретной области могут различаться, однако для них характерно сохранение иерархического представления графа, что часто соответствует формату изучаемого материала в случае педагогического процесса и оценки знаний испытуемых. Вероятностный подход при построении выводов относительно оценок способностей испытуемых просто интерпретируется: как правило, система тестирования оценивает условную вероятность того, что фиксированный испытуемый, при условии правильного ответа на конкретный вопрос, обладает некоторой компетенцией (либо усвоил материал, знание которого проверяется текущим тестом). Решение об уровне способности испытуемого принимается после определения интервала, в который попала оценка условной вероятности.

Среди недостатков байесовских сетей можно выделить относительно низкую универсальность ввиду сложности составления моделей испытуемого и предметной области.

Тестирование с использованием марковских цепей

В данном подходе предполагается, что ответы на задания теста — независимые величины, а состояниями цепи являются меры сложности заданий [43]. Каждому из состояний цепи соответствует некоторый уровень сложности β_i .

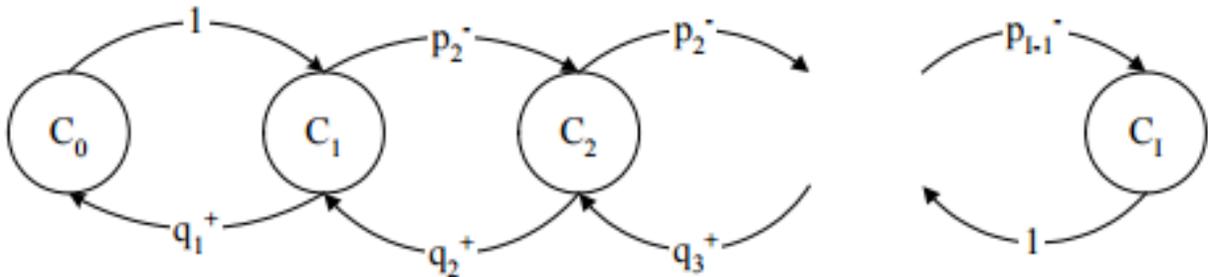


Рис. 2. Схема марковской цепи

Вероятности переходов между состояниями определяются на основа-

нии логистической функции (с использованием параметров θ и β_i):

$$p_i = \frac{e^{\left(\frac{\alpha}{\sigma}\right)(\theta - \beta_i)}}{1 + e^{\left(\frac{\alpha}{\sigma}\right)(\theta - \beta_i)}}$$

$$q_i = \frac{1}{1 + e^{\left(\frac{\alpha}{\sigma}\right)(\theta - \beta_i)}}$$

Далее задача сводится к поиску стационарных распределений цепи, не зависящих от начального состояния [43]. Механизм адаптивного тестирования задается процедурой последовательного предъявления тестовых заданий, трудность которых определяется как фиксированное состояние марковской цепи в конкретный момент времени. Если испытуемый, находясь в состоянии цепи C_i , правильно выполнил задание теста, то он переходит в состояние C_{i+1} , иначе – в состояние C_{i-1} . После завершения тестирования испытуемый оказывается в одном из состояний, наилучшим образом соответствующем его уровню подготовки [55].

Тестирование с использованием нейронных сетей

Процесс адаптивного тестирования с использованием нейросетевого конфигуратора предложен в работе [39]. Рассматриваются факторы X_1 — номер этапа тестирования, X_2 — уровень сложности тестового задания, X_3 — количество правильных ответов испытуемого. В качестве выхода: Y_i — выход нейронной сети с данными о направлении изменения сложности следующего вопроса. L — итоговая оценка испытуемого. На рисунке 3 изображена схема процесса тестирования.

Примерная архитектура нейронной сети для использования в таком виде тестирования, приведена на рисунке 4.

Предложенную нейронную сети можно обучать алгоритмом обратного распространения ошибки с настройкой сигмоидальной функции. Коррекция синаптического веса, соединяющему нейроны j и i , происходит согласно правилу:

$$\nabla w_{ij}(n) = \eta \delta_j(n) y_i(n),$$

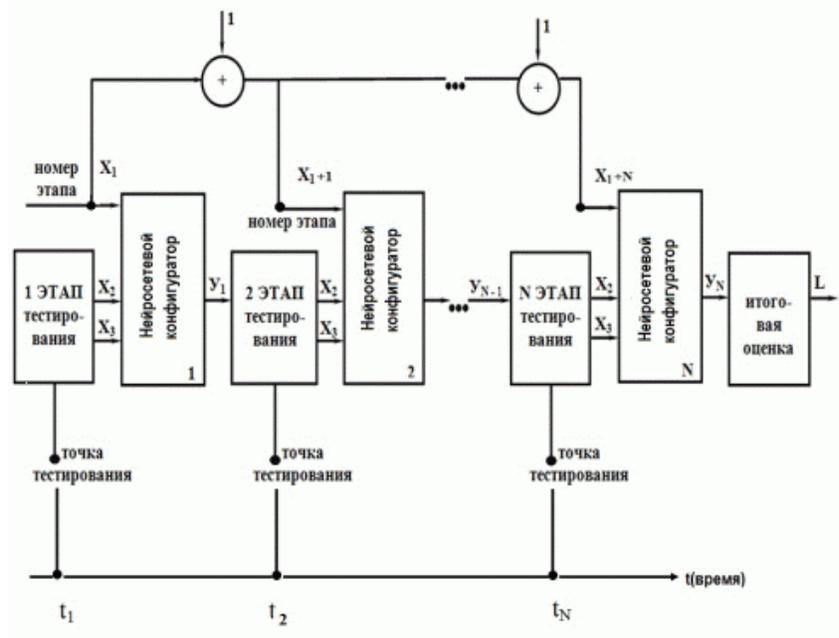


Рис. 3. Схема процесса тестирования

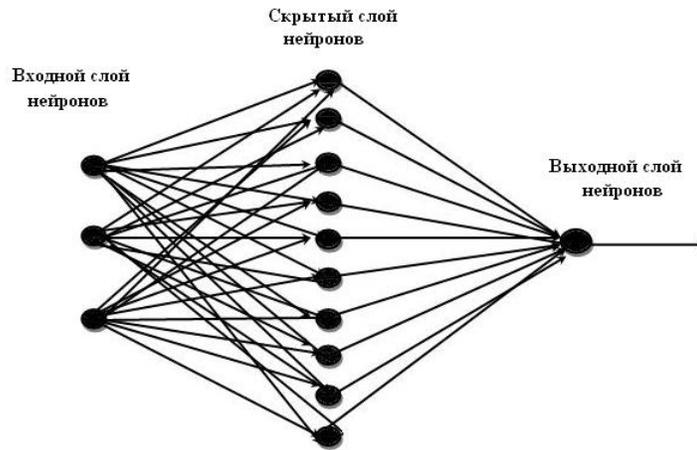


Рис. 4. Схема процесса тестирования

где $\nabla w_{ij}(n)$ – коррекция веса; n – параметр скорости обучения; $\eta \delta_j(n)$ – локальный градиент; $y_i(n)$ – входной сигнал нейрона j [34].

Локальный градиент нейрона выходного слоя выражается как:

$$\delta_j(n) = e_j(n) f'(\alpha_j(n)) = v_j(n) [d_j(n) - o_j(n)] o_j(n) [1 - o_j(n)],$$

где $o_j(n)$ – функциональный сигнал на выходе нейрона j ; $d_j(n)$ – его же-

лаемый сигнал.

Для скрытого нейрона локальный градиент равен:

$$\delta_j(n) = f'_j(n) \sum_{k=1}^M \delta_k(n) w_{kj}(n) = v_j(n) y_j(n) [1 - y_j(n)] \sum_{k=1}^M \delta_k(n) w_{kj}(n),$$

где M – количество нейронов в выходном слое.

Использование подобных подходов к реализации систем тестирования позволяет адаптировать тестовые задания к конкретному уровню подготовленности испытуемого. Выбор архитектуры нейронной сети и алгоритма ее обучения может варьироваться в зависимости от объемов выборки, временных ограничений, а также параметров, касающихся степени дискретизации пула заданий по уровням сложности.

Подробно подходы к реализации адаптивного тестирования на базе нейронных сетей изложены в работах [7, 8].

Среди недостатков применения нейронных сетей можно выделить следующие:

- 1) архитектура нейронных сетей требует адаптации под конкретные случаи использования;
- 2) обучение сети в ряде случаев приводит к тупиковым ситуациям; в таких случаях необходимо выбирать иные методы обучения сети, которые не всегда хорошо интерпретируемы.

Среди перечисленных подходов решаются сходные задачи:

- 1) адаптация процедуры измерения навыков или способностей испытуемого путем подбора заданий, наилучшим образом соответствующих его уровню способности;

- 2) повышение степени надежности оценок результатов тестирования путем использования математических моделей (модели знаний, модели изучаемого материала и т.д. [42]);
- 3) ускорение процедуры тестирования: выбор заданий с уровнем трудности, соответствующим уровню способности испытуемого, позволяет сократить количество измерений (предъявляемых заданий) для построения оценки уровня подготовки испытуемого [50].

Перечисленные подходы обладают рядом недостатков, делающих актуальными поиск и разработку новых подходов к адаптивному тестированию.

- 1) в процессе тестирования важными характеристиками являются не только трудности заданий и уровни способностей испытуемого. Одну из ключевых ролей в оценке результатов тестирования играет показатель времени: динамики прохождения испытуемыми тестовых заданий;
- 2) в главе 2 настоящей работы показано, что динамика способностей испытуемого и трудности фиксированного задания не являются неизменными величинами: с течением времени их значения могут изменяться. Описанные выше подходы игнорируют данный фактор;
- 3) в части описанных подходов для создания системы тестирования по конкретным дисциплинам необходима детализация этой дисциплины по определенным правилам: выбор темы, раздела, подраздела, понятий и т.д. Такой подход существенно снижает универсальность разрабатываемых систем;
- 4) подходы, упомянутые во введении, разработаны для проведения процедур тестирования, связанных исключительно с выполнением тестов на компьютере. В случае же, когда в качестве инструмента выполнения задания служит некоторая техническая система, описанные подходы требуют существенной переработки;

- 5) описанные подходы не рассматривают случаи, когда процедура оценки навыков или способностей испытуемых заключается не в выполнении тестовых заданий, а в некоторой игре, сложность которой с учетом успешности выполнения игровых заданий может меняться.

Глава 1. Способы адаптивного тестирования

1.1 Модель тестирования

Тестирование как процесс представляет собой некоторый набор заданий, которые предъявляются испытуемому в определенном порядке. Порядок заданий (или тестов в случае комплексного измерения способностей) может быть жестко задан в случае использования подхода, основанного на классической теории тестирования, или может изменяться согласно некоторым правилам с учетом текущих характеристик процесса. Наличие или отсутствие эксперта в общем случае не меняет содержательной части процедуры тестирования, однако может влиять на результаты и протяженность процесса тестирования в случае получения экспертом некоторой дополнительной информации о текущем состоянии испытуемого.

Предполагается, что тест состоит из определённого количества *политомических* заданий, каждое из которых может быть выполнено с различными исходами. Процедура прохождения теста описывается марковским процессом с дискретными состояниями и непрерывным временем. Модели описания динамики переходов от задания к заданию могут быть изображены в виде ориентированного графа, вершины которого представляют состояния и соответствуют заданиям теста, а дуги – переходам между ними.

В общем виде динамика изменения вероятностей пребывания в состояниях марковского процесса определяется системой уравнений Колмогорова:

$$\frac{d\mathbf{p}(t)}{dt} = \mathbf{\Lambda}\mathbf{p}(t)$$

где $0 \leq t \leq T$; $\mathbf{p}(t) = (p_1(t), \dots, p_r(t))^T$; $\mathbf{\Lambda}$ – матрица размера $r \times r$ интенсивностей переходов между состояниями модели; T – конечный момент времени.

Для описания процедуры прохождения теста, состоящего из m зада-

ний, каждое из которых допускает l различных исходов выполнения, используются процессы, представленные на рисунке 5.

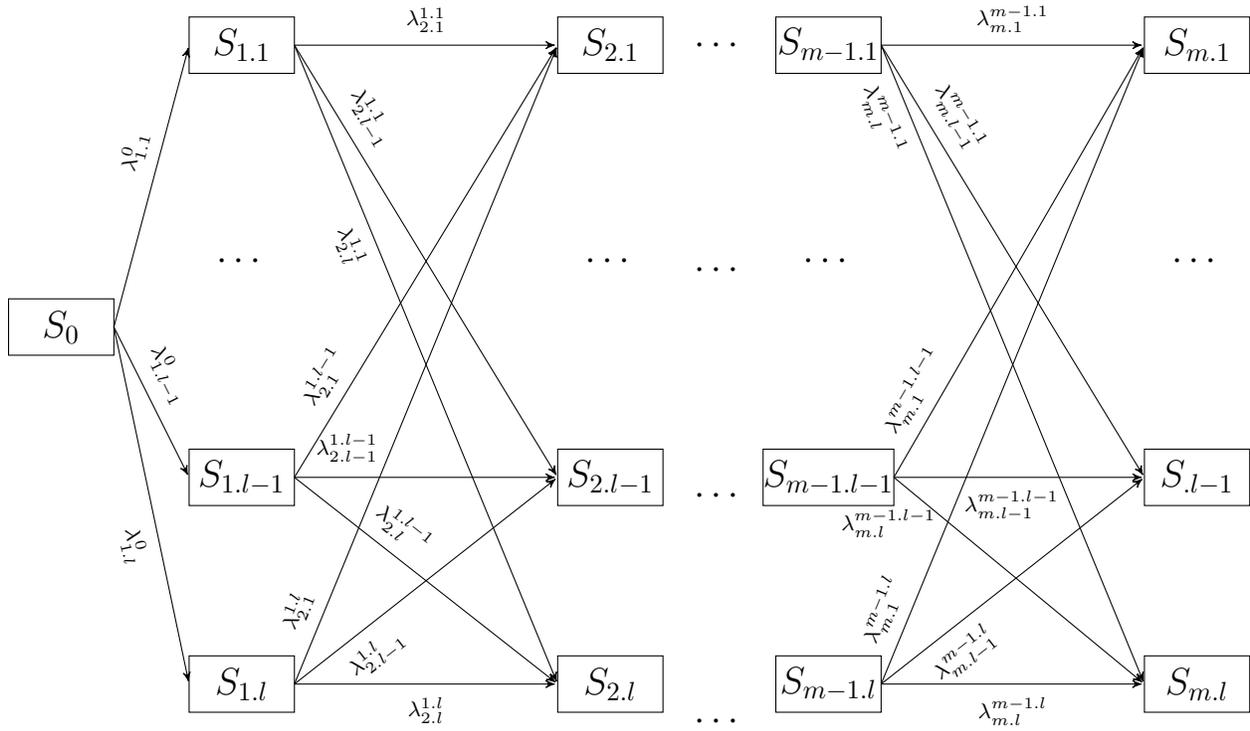


Рис. 5. Марковский процесс в случае теста, состоящего из m заданий, каждое из которых допускает l различных исходов выполнения

В этом случае система дифференциальных уравнений Колмогорова принимает следующий вид:

$$\left\{ \begin{array}{l}
 \frac{dp_{s_0}}{dt} = - \sum_{i=1}^l \lambda_{1,i}^0 * p_{s_0} \\
 \frac{dp_{s_{1,1}}}{dt} = \lambda_{1,1}^0 * p_{s_0} - \sum_{i=1}^l \lambda_{2,i}^{1,1} * p_{s_{1,1}} \\
 \vdots \\
 \frac{dp_{s_{1,l-1}}}{dt} = \lambda_{1,l-1}^0 * p_{s_0} - \sum_{i=1}^l \lambda_{2,i}^{1,l-1} * p_{s_{1,l-1}} \\
 \frac{dp_{s_{1,l}}}{dt} = \lambda_{1,l}^0 * p_{s_0} - \sum_{i=1}^l \lambda_{2,i}^{1,l} * p_{s_{1,l}} \\
 \frac{dp_{s_{2,l}}}{dt} = \sum_{i=1}^l \lambda_{2,i}^{1,i} * p_{s_{1,l}} - p_{s_{2,l}} * (\dots) \\
 \vdots \\
 \frac{dp_{s_{2,l-1}}}{dt} = \sum_{i=1}^l \lambda_{2,l-1}^{1,i} * p_{s_{1,i}} - p_{s_{2,l-1}} * (\dots) \\
 \frac{dp_{s_{2,l}}}{dt} = \sum_{i=1}^l \lambda_{2,l}^{1,i} * p_{s_{1,i}} - p_{s_{2,l}} * (\dots) \\
 \vdots \\
 \frac{dp_{s_{m-1,1}}}{dt} = - \sum_{i=1}^l \lambda_{m,i}^{m-1,1} * p_{s_{m-1,1}} - (\dots) \\
 \vdots \\
 \frac{dp_{s_{m-1,l-1}}}{dt} = - \sum_{i=1}^l \lambda_{m,i}^{m-1,l-1} * p_{s_{m-1,l-1}} - (\dots) \\
 \frac{dp_{s_{m-1,l}}}{dt} = - \sum_{i=1}^l \lambda_{m,i}^{m-1,l} * p_{s_{m-1,l}} - (\dots) \\
 \frac{dp_{s_{m,1}}}{dt} = \sum_{i=1}^l \lambda_{m,1}^{m-1,i} * p_{s_{m-1,i}} \\
 \vdots \\
 \frac{dp_{s_{m,l-1}}}{dt} = \sum_{i=1}^l \lambda_{m,l-1}^{m-1,i} * p_{s_{m-1,i}} \\
 \frac{dp_{s_{m,l}}}{dt} = \sum_{i=1}^l \lambda_{m,l}^{m-1,i} * p_{s_{m-1,i}}
 \end{array} \right. \quad (1)$$

В момент начала тестирования вероятность пребывания в состоянии S_0 задана $p_1 = 1$, а остальные вероятности пребывания в состояниях процесса равны нулю. Для любого момента времени t также выполняется нормализующее условие:

$$\sum_{i=0}^m \sum_{j=0}^l p_{s_{i,j}}(t) = 1 \quad (2)$$

Для проведения идентификации разработанных моделей необходимо провести эксперимент, в котором выборка разделена по уровню исследуемой способности на диагностируемые группы. Для реализации этого разделения можно использовать оценки экспертов. Далее необходимо, фиксируя время выполнения каждого задания испытуемыми, получить набор временных точек и состояний, в которых находятся испытуемые в фиксированные моменты времени – наблюдаемые частоты пребывания в состояниях. Эта информация используется для идентификации свободных параметров моделей для каждой категории испытуемых.

Значения свободных параметров идентифицируются путем сравнения наблюдаемых и гипотетических вероятностей. На основе ожидаемых и наблюдаемых частот пребывания в фиксированных состояниях модели в заданные моменты времени вычисляются такие значения параметров, которые доставляют минимум статистике Пирсона:

$$\chi^2 = \sum_{d=0}^{D-1} \sum_{i=1}^r \frac{(p_i(t_d)N - F_{id})^2}{p_i(t_d)N},$$

где $\{F_{id}\}_{i=1,\dots,r}$ – число элементов, попавших в i -ое состояние модели в моменты времени $\{t_d\}_{d=0,\dots,D-1}$; $N = \sum_{i=1}^r F_{id}$; D – количество моментов времени, в которые фиксировались наблюдения F_{id} ; $0 \leq t_d \leq T$. Эта статистика используется как мера соответствия модели наблюдениям. Прогнозируемые вероятности пребывания в состояниях модели вычисляются методами численного интегрирования системы дифференциальных уравнений. Используемый метод идентификации рассмотрен в главе 4.

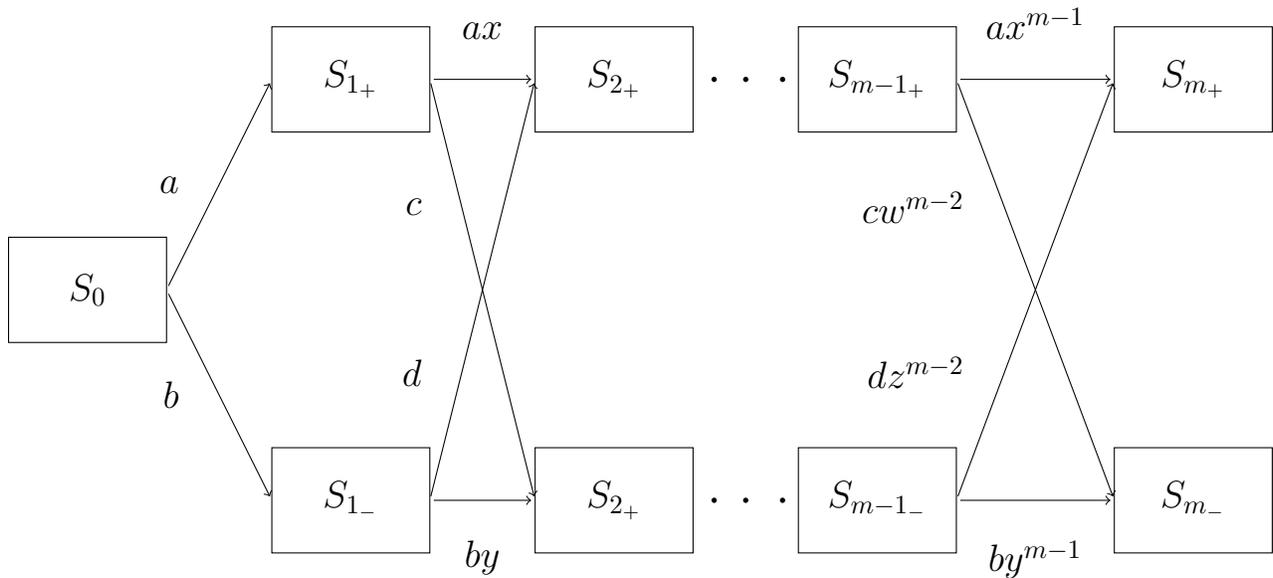


Рис. 6. Пример зависимости интенсивностей переходов между состояниями.

Зависимости между интенсивностями переходов

Для упрощения задачи идентификации параметров моделей допустимо вводить зависимости между интенсивностями переходов между состояниями. Пример такой зависимости приведен на рис. 6. Данная схема соответствует дихотомическому закрытому типу теста с количеством заданий равным m . Состояния S_{n+} , где $n = 1, \dots, m$, соответствуют успешному выполнению заданий, состояния S_{n-} , где $n = 1, \dots, m$ – неуспешному выполнению заданий. С течением времени значения интенсивностей переходов между состояниями снижаются, что соответствует идее повышения сложности заданий и накопленной усталости испытуемого. Тип зависимости интенсивностей переходов необходимо выбирать индивидуально: структура модели (типы и количество вопросов теста), качественные характеристики выборки, цель тестирования могут варьироваться. Оптимальный выбор подобных зависимостей опирается на технику проверки статистических гипотез.

При выполнении ряда общих условий, значения статистики χ^2 , получаемые при подстановке истинных решений уравнений, асимптотически описываются распределением χ^2 с $r - s - 1$ степенями свободы, где r – чис-

ло уравнений, s – число идентифицируемых параметров, причем вычисленные значения свободных параметров при увеличении объема выборки испытуемых сходятся по вероятности к искомому решению. Это позволяет использовать приведенную статистику для проверки гипотезы о том, что полученный прогноз согласуется с результатами наблюдений. Данный способ идентификации свободных параметров даёт решения, близкие к полученным методом максимального правдоподобия.

Согласно теореме Г. Крамера, при выполнении условий:

- 1) $\sum_{i=1}^r p_i(\lambda_1, \dots, \lambda_s) = 1.$

- 2) $p_i(\lambda_1, \dots, \lambda_s) > c^2 > 0$ при всех i .

- 3) Все p_i имеют непрерывные производные $\frac{\partial p_i}{\partial \lambda_j}$ и $\frac{\partial^2 p_i}{\partial \lambda_j \partial \lambda_k}$.

- 4) Матрица $U = \left\{ \frac{\partial p_i}{\partial \lambda_j} \right\}$, где $i = 1, \dots, r$ и $j = 1, \dots, s$, имеет ранг S

рассмотренная задача идентификации параметров имеет единственное решение, которое при увеличении объема наблюдений N сходится по вероятности к искомому решению.

Пусть возможные результаты некоторого случайного эксперимента ϑ разбиты на r непересекающихся групп, и предположим, что вероятность получения результата, принадлежащего к i -ой группе, равна

$$p_i^0 = p_i(\lambda_1^0, \dots, \lambda_s^0), \text{ где } \lambda_0 = (\lambda_1^0, \dots, \lambda_s^0) - \text{внутренняя точка интервала } A.$$

Пусть F_i обозначает число результатов, принадлежащей к i -ой группе, которое было получено в последовательности n повторений эксперимента ϑ , так что $\sum_{i=1}^r F_i = n$.

Тогда уравнение $\sum_{i=1}^s \frac{F_i - np_i}{p_i} \frac{\partial p_i}{\partial \lambda_j} = 0$, где $j = 1, 2, \dots, s$ имеет в точности одну систему решений $\lambda = (\lambda_1, \dots, \lambda_s)$, такую, что λ сходится по распределению к случайной величине λ_0 при $n \rightarrow \infty$. Значение χ^2 , получаемое при подстановке этих значений λ_j в уравнение $\chi^2 = \sum_{i=1}^r \frac{[F_i - np_i(\lambda_1, \dots, \lambda_s)]^2}{np_i(\lambda_1, \dots, \lambda_s)}$, в пределе при $n \rightarrow \infty$ имеет распределение χ^2 с $r - s - 1$ степенями свободы.

Для решения рассмотренной системы обыкновенных дифференциальных уравнений Колмогорова достаточные условия теоремы Крамера выполняются для любого сечения процесса при $t > 0$, поэтому выводы указанной теоремы применимы к рассматриваемой задаче.

Обозначим: $0 \leq t \leq T$; $p(t) = (p_1(t), \dots, p_r(t))$

Утверждение 1. Пусть матрица $U = \left\{ \frac{\partial p_i}{\partial \lambda_j} \right\}$, где $i = 1, \dots, r$ и $j = 1, \dots, s$ имеет ранг s . Пусть \mathbf{p} – решение системы уравнений:

$$\frac{d\mathbf{p}}{dt} = \Lambda \mathbf{p}$$

Тогда справедливо, что уравнение $\sum_{i=1}^s \frac{F_i - np_i}{p_i} \frac{\partial p_i}{\partial \lambda_j} = 0$, где $j = 1, \dots, s$, имеет в точности одно решение $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_s)$. При этом статистика $\chi^2 = \sum_{i=1}^r \frac{[F_i - np_i(\lambda_1, \dots, \lambda_s)]^2}{np_i(\lambda_1, \dots, \lambda_s)}$ в пределе при $n \rightarrow \infty$ имеет распределение χ^2 с $r - s - 1$ степенями свободы, где r – число уравнений в решаемой системе.

Доказательство. Проверим условия теоремы Крамера:

$$1) \sum_{i=1}^r p_i(\lambda_1, \dots, \lambda_s) = 1.$$

Для любого момента времени t сумма вероятностей пребывания в состояниях равна 1. Выполнение данного условия обусловлено наличием нормализующего условия $\boxed{2}$ для системы уравнений $\boxed{1}$.

$$2) p_i(\lambda_1, \dots, \lambda_s) > c^2 > 0 \text{ при всех } i.$$

Все состояния марковской сети достижимы в некоторый интервал времени $\exists t p_{i,j} > 0$.

Покажем, что существует точка, в которой все вероятности $p_i > 0$ при некотором наборе $\{\lambda_{1.1}^0, \dots, \lambda_{m.l}^{m-1.l}\}$. Предположим, что это неверно. Тогда при любом наборе $\{\lambda_{1.1}^0, \dots, \lambda_{m.l}^{m-1.l}\} \exists p = 0$. В силу нормализующего условия $\sum_{i=0}^m \sum_{j=0}^l p_{s_i,j}(t) = 1$ существует хотя бы одно зна-

чение p для g -го состояния $p_g > 0$. Рассмотрим смежное к g -му состоянию состояние f , для которого $p_f = 0$. Осуществляем малое приращение Δt , причем настолько малое, чтобы значение p_g оставалось положительным. В данный интервал времени $[t; \Delta t]$ значение p_f станет положительным. Применяя данный алгоритм к состояниям, вероятность пребывания в которых $p = 0$, получаем противоречие первоначальному утверждению о существовании точки, для которой все вероятности $p_i = 0$.

- 3) Все p_i имеют непрерывные производные $\frac{\partial p_i}{\partial \lambda_j}$ и $\frac{\partial^2 p_i}{\partial \lambda_j \partial \lambda_k}$.

Общий алгоритм решения подобных систем обыкновенных дифференциальных уравнений $\mathbf{X}' = A\mathbf{X}(t)$ с помощью матричной экспоненты имеет вид

$$\mathbf{X}(t) = \exp(tA)C,$$

где C в данном случае – n -мерный вектор, выражающий начальные условия задачи Коши $\mathbf{p} = (p_0, p_1, \dots, p_n)^T$. Для рассматриваемой системы $\frac{d\mathbf{p}}{dt} = \mathbf{M}(\boldsymbol{\lambda})\mathbf{p}$ решение будет иметь вид бесконечно дифференцируемого выражения $\mathbf{p}(t) = \exp(M(\boldsymbol{\lambda})t)\mathbf{p}_0$, где $\mathbf{p}_0 = \mathbf{p}(t = t_0)$.

- 4) Матрица $U = \left\{ \frac{\partial p_i}{\partial \lambda_j} \right\}$, где $i = 1, \dots, r$ и $j = 1, \dots, s$, имеет ранг s .

Введение зависимостей между интенсивностями состояний (описанных в п. [1.1](#), рис. [6](#)) обеспечивает выполнимость данного условия.

Процедура адаптивного тестирования заключается в последовательном предъявлении экспертом испытываемому тестов из заданного набора. Тесты не повторяются. Процедура назначения тестов завершается по достижении заданного условия, накладываемого на вероятность наиболее

правдоподобного варианта категории испытуемого. Если указанное условие не выполняется в течение всей процедуры тестирования, то испытуемому предъявляется полный набор из m заданий.

Оценка возможности отнесения испытуемого к одной из категорий производится по формуле Байеса. Оценки рассчитываются для каждой из рассматриваемых категорий в контрольный момент времени для наблюдаемого состояния процесса, в котором испытуемый оказывается после выполнения последнего задания:

$$P(C_i|S(t_d)) = \frac{P(C_i)P(S(t_d)|C_i)}{\sum_{k=1}^I P(C_k)P(S(t_d)|C_k)},$$

где C_i – событие, связанное с наличием у испытуемого i -й категории ($i = 1, \dots, I$); $S(t_d)$ – событие, связанное с нахождением в наблюдаемом состоянии сети в контрольный момент времени t_d ; $P(C_i)$ – априорная вероятность наличия у испытуемого i -й категории; $P(S(t_d)|C_i)$ – вероятность нахождения в наблюдаемом состоянии сети в контрольный момент t_d времени при наличии i -й категории; $P(C_i|S(t_d))$ – вероятность i -й категории при условии нахождения в наблюдаемом состоянии сети в контрольный момент времени t_d . Вероятности $P(S(t_d)|C_i)$ вычисляются путём численного интегрирования приведённых выше уравнений Колмогорова.

Наиболее вероятной категорией $i^* = \arg \max_i P(C_i|S)$ является та, для которой достигается наибольшая условная вероятность $\max_i \{P(C_i|S(t_d))\}_{i=1, \dots, I}$, вычисленная по результатам прохождения последнего выполненного задания. Распределение вероятностей $\{P(C_i|S(t_d))\}_{i=1, \dots, I}$ позволяет оценить степень дифференциации этой оценки.

В ситуации, когда процесс тестирования находится под контролем эксперта, испытуемый и эксперт работают за разными мониторами [11]. Монитор испытуемого служит для предъявления тестов и заданий или ссы-

лок на них, а монитор эксперта – для отображения результатов тестирования и информации, необходимой для выбора очередного теста, а также выставления оценок. После выбора очередного теста экспертом на мониторе испытуемого появляется ссылка на тест, который следует выполнить (рисунок 7). Испытуемый не должен видеть данные, отображаемые на мониторе эксперта. В зависимости от типа тестовой процедуры, испытуемый, получив задание, пытается его выполнить или находясь за монитором компьютера (рисунок 8а), или в «полевых» условиях (рисунок 8б).

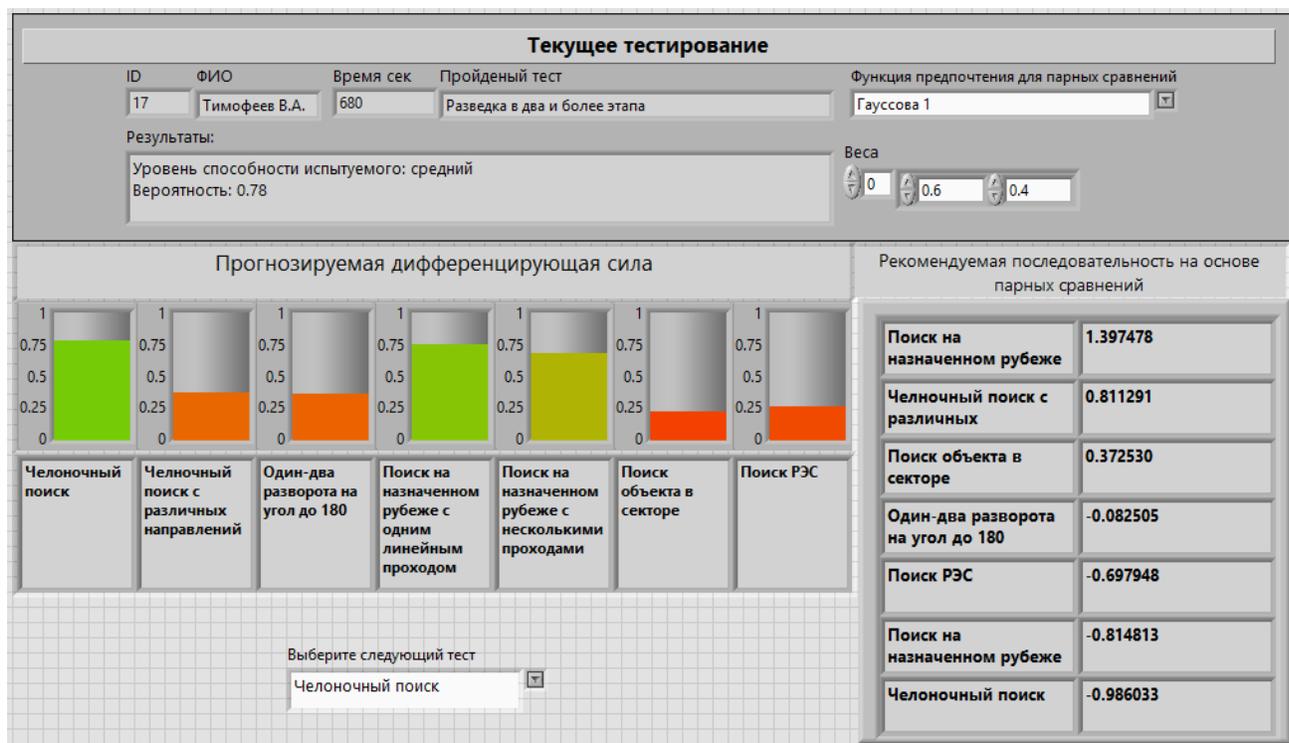


Рис. 7. Окно результатов и выбора очередного при тестировании операторов сложных технических систем

После попытки выполнения каждого теста вычисляются:

- оценки вероятностей различных вариантов диагнозов испытуемых при условиях их принадлежности к каждой из диагностируемых групп и нахождения в контрольный момент времени в наблюдаемом состоянии сети;
- вероятности различных вариантов диагнозов при условиях принад-



Рис. 8. Выполнение теста в обычных (а) и полевых условиях (б)

лежности испытуемого к каждой из диагностируемых групп и нахождения по истечении заданного фиксированного интервала времени в конечном состоянии каждого из непредъявленных к данному моменту тестов.

Стратегия эксперта, принимающего решение о последовательности предъявления тестов, обусловлена необходимостью обеспечения за наименьшее время наибольшей дифференциации результата тестирования, представленной условием, накладываемым на вероятность наиболее правдоподобной категории. При отсутствии дополнительных аргументов (чаще всего полученных на основе наблюдений за испытуемым), эксперту следует выбирать в качестве очередного тот тест, который приводит к наибольшей дифференциации самой вероятной категории от остальных вариантов при условии пребывания испытуемого по истечении заранее фиксированного интервала времени в соответствующем конечном состоянии. Такой подход обеспечивает близкую к оптимальной последовательность предъявления тестов, которая, тем не менее, может быть изменена экспертом в любой момент.

1.2 Тестирование операторов сложных технических систем

В случае, если задания теста представлены не в классическом виде, а в форме задания, предполагающего управление некоторой технической системой, модель может быть модифицирована. Один из классических вариантов диагностики операторов — отработка схем управления системой в полевых условиях. Предполагается, что испытуемый, воздействуя на систему, должен выполнить ряд задач за определенное время. Динамика выполнения заданий в таких условиях является одним из ключевых факторов. Типичные задачи такого вида объезд или облет участка по определенной траектории дистанционно управляемой системой. В таком случае модель, соответствующая заданию, может приобретать вид, приведенный на рисунке [9](#)



Рис. 9. Модель, отражающая структуру задания для оператора технических систем

При использовании такого вида моделей начальное и конечное состояния соответствуют старту и окончанию выполнения задания, а промежуточные состояния соответствуют некоторым элементам процесса управления системой: осуществление разворота, изменение высоты и т.д. На рисунке [10](#) приведен график динамики вероятностей пребывания в различных состояниях для модели из 8 состояний. Отклонения от нормативного времени выполнения упражнения в целом и конкретных элементов в частности, будет отражено в модели. В табл. [1](#) приведены некоторые ошибки и соответствующие им изменения в модели.

Таблица 1. Таблица ошибок и соответствующих изменений характеристик процесса

Описание	Отражение в модели
Некорректный старт	Превышение времени пребывания в первом состоянии модели
Выход за границы территории	Оценка интенсивности переходов между состояниями текущей траектории ниже, чем аналогичная оценка для нормативной траектории
Изменение масштаба выполняемого маневра	Оценка интенсивности переходов между состояниями текущей траектории выше, чем аналогичная оценка для нормативной траектории
Некорректное завершение упражнения	Превышение времени пребывания в состоянии посадки (возврата и посадки)

В целом замедление либо ускорение по сравнению с нормативным временем будет явно отражаться на динамике пребывания испытуемых в различных состояниях. На рисунке 10 приведена динамика пребывания в состояниях для модели из 8 состояний, разные линии соответствуют разным временным показателям выполнения задания: медленного, нормального и высокого. В данной задаче рассматривается случай двух классов: испытуемых с высоким уровнем способностей, выполняющих облет территории за нормативное время (синяя линия на рисунке 10), и испытуемых с низким уровнем способностей, выполняющий упражнение либо медленно, либо слишком быстро (пунктирные линии на рисунке 10).

Значения оценок вероятностей пребывания в фиксированные моменты времени в различных состояниях характеризует принадлежность испытуемого к определенной диагностируемой группе (рисунок 9). Зная оценки вероятностей для различных диагностируемых групп (высокой квалифи-

кации, низкой квалификации и т.д.), можно рассчитать оценку вероятности принадлежности конкретного испытуемого к каждой из групп с помощью Байесовского подхода.

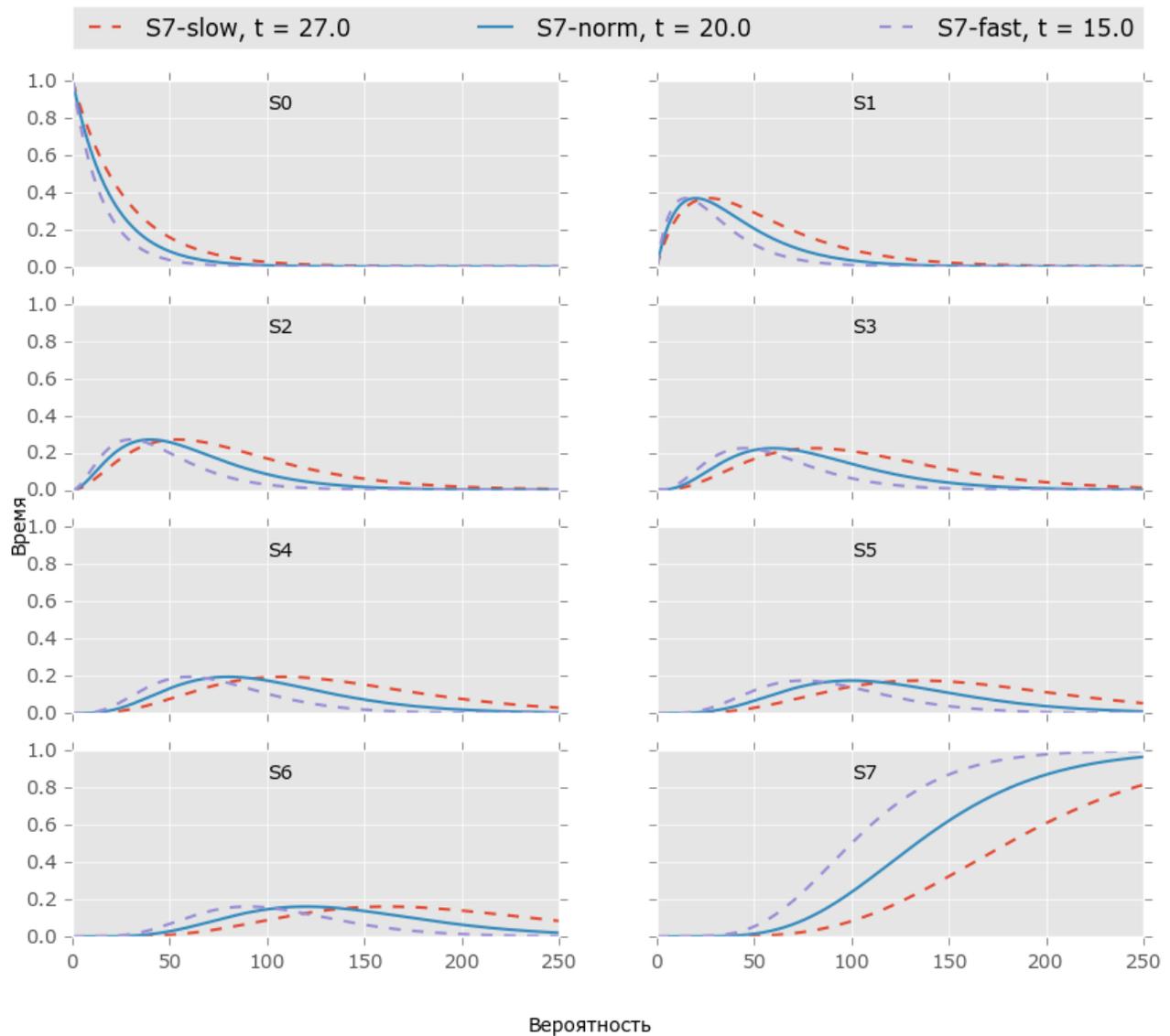


Рис. 10. Динамика пребывания в различных состояниях модели в случае медленного (красная линия), нормального (синяя линия) и ускоренного (сиреневая линия) выполнения задания

Глава 2. Методы оптимизации порядка предъявления тестов

Эффективность процедуры оценки уровня навыков или способностей обусловлена возможностью поиска наиболее вероятного диагноза за наименьшее время. Рассматривая оценки, характеризующие дифференцирующую силу каждого не предъявленного теста, эксперт выбирает вариант, обеспечивающий наибольшую дифференциацию при отнесении испытуемого к одной из категорий. Для этого разработаны два подхода: *метод оптимизации последовательности предъявления тестов* и *метод коррекции наблюдением*.

2.1 Метод оптимизации последовательности предъявления тестов

Метод оптимизации последовательности предъявления тестов относится к семейству методов принятия решений, основанных на парных сравнениях альтернатив по набору критериев. Его преимуществом является возможность построения ранжированной последовательности альтернатив [13, 26, 41]. Эксперту при использовании данного метода нет необходимости в определении величин различных параметров, влияющих на результат.

В качестве альтернатив рассматриваются непредъявленные тесты: $G = \{g_1, g_2, \dots, g_m\}$, где m – количество оставшихся в наборе тестов, в качестве критериев – $K = \{k_1, k_2, \dots, k_n\}$, n – число категорий, к которым возможно отнести испытуемого в процессе тестирования («высокий уровень способностей», «средний уровень способностей», «низкий уровень способностей» и т.д.).

Введем величину $d_{u,v}$ – разность между двумя альтернативами u и v из множества G по одному из критериев k из множества K . Оцен-

ка среднеквадратического отклонения $\hat{\sigma}$ вычисляется по формуле: $\hat{\sigma} = \sqrt{\frac{1}{m} \sum_{i=1}^m (x_i - \bar{x})^2}$, где x_i – значение альтернативы g_i по фиксированному критерию k , \bar{x} – среднее значение всех альтернатив по фиксированному критерию k , m – количество альтернатив. Введем величины w_1, \dots, w_k , являющиеся весами критериев k_1, \dots, k_n . На веса накладывается ограничивающее условие $\sum_{k=1}^n w_k = 1$, w_k – вес критерия k . В случае отсутствия предпочтений эксперта относительно весов альтернатив, значения весов w_k могут полагаться равными друг другу $w_1 = \dots = w_k$.

Введем функцию предпочтения, выражающую степень предпочтения одного теста другому по конкретному критерию:

$$H(d_{u,v}) = \begin{cases} 0, & d_{u,v} \leq 0 \\ \frac{1 - \exp\left(-\frac{d_{u,v}^2}{\hat{\sigma}_k}\right)}{1 + \exp\left(-\frac{d_{u,v}^2}{\hat{\sigma}_k}\right)} & d_{u,v} > 0 \end{cases}$$

где $d_{u,v}$ – разность значений объектов u и v по выбранному критерию, $\hat{\sigma}_k$ – оценка среднеквадратического отклонения по текущему критерию, u – индекс альтернативы. Оценка среднеквадратического отклонения по критерию σ_k вычисляется на основе вероятностей $P(C_i | S_{m.l}(t))$ принадлежать к определенной категории при условии нахождения по истечении фиксированного интервала времени в конечном состоянии $S_{m.l}$. Значение данной функции предпочтения помимо $d_{u,v}$ зависит также от стандартного отклонения по выбранному критерию. График значений функции предпочтения расстояния и параметра σ_k изображен на рисунке [11](#)

Как видно на графике, с увеличением среднеквадратического отклонения значение функции предпочтения для конкретной разности $d_{u,v}$ существенно меняется. В случае, когда альтернативы имеют близкие значения по критериям (а следовательно небольшое среднеквадратическое отклоне-

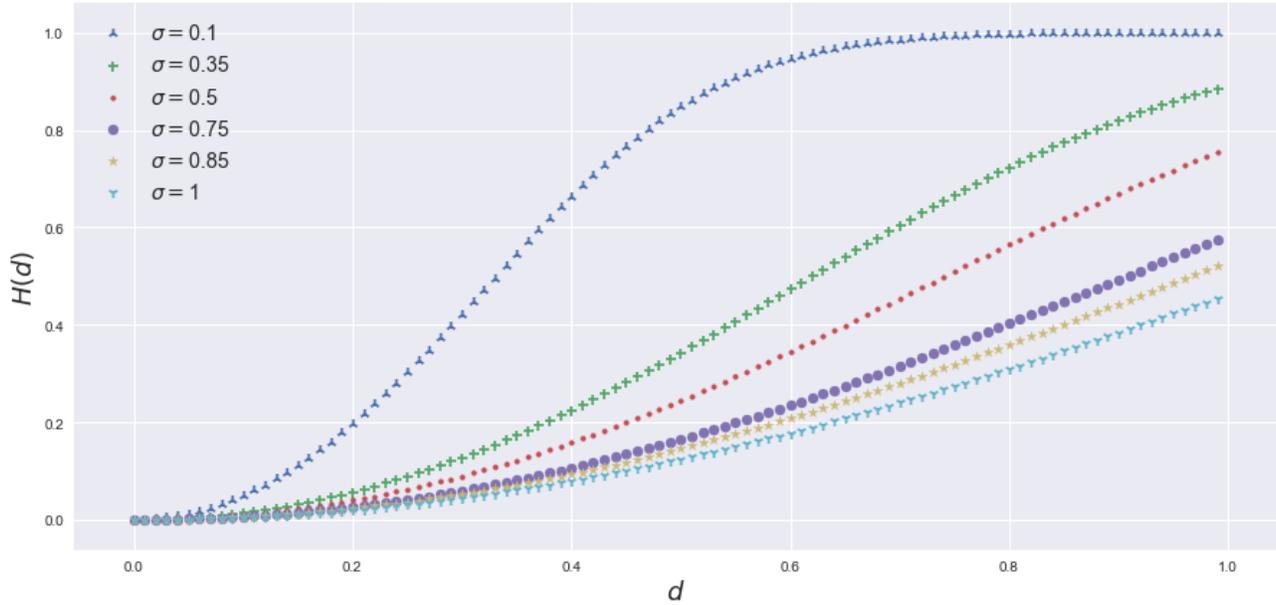


Рис. 11. График зависимости значения функции предпочтения от $d_{u,v}$ и σ_k .

ние σ_k) малая разность $d_{u,v}$ будет возвращать большее значение функции, чем для случая большего среднеквадратического отклонения.

Построение ранжированной последовательности тестов включает следующие этапы:

- 1) вычисление значения функции $H(d_{u,v})$ предпочтения между сравниваемыми тестами u и v по каждому критерию;
- 2) выбор n весов w_k для каждого критерия;
- 3) расчет индексов предпочтения по формуле:

$$Q_{u,v} = \sum_{k=1}^n w_k H(d_{u,v}^k),$$

где $H(d_{u,v}^k)$ – значение функции предпочтения для между u -й и v -й альтернативами, n – число критериев, w_k – вес k -го критерия.

- 4) расчет коэффициентов предпочтения Φ_p и Φ_n производится по фор-

мулам:

$$\Phi_p = \sum_{v=1, u \neq v}^m Q_{u,v}$$

$$\Phi_n = \sum_{u=1, v \neq u}^m Q_{v,u}$$

где u и v – индексы альтернатив, m – количество рассматриваемых альтернатив.

- 5) построение последовательности предпочтительных вариантов; при этом для определения оценки критерия, влияющего на позицию альтернативы в последовательности, необходимо рассчитать итоговый коэффициент Φ :

$$\Phi = \Phi_p - \Phi_n,$$

где u и v – индексы альтернатив, m – количество рассматриваемых альтернатив. Упорядочение элементов списка альтернатив по соответствующему им значению итогового коэффициента Φ позволит определить предпочтительную последовательность предъявления тестов испытуемому.

Пример построения предпочтительной последовательности тестов

После прохождения каждого теста, помимо вероятностей принадлежности испытуемого к каждой категории $P(C_i|S(t_d))$, вычисляются вероятности различных вариантов категорий при условиях принадлежности испытуемого к каждой из категорий и нахождения по истечении заданного фиксированного интервала времени в конечном состоянии $S_{m.l}$ для каждого из непредъявленных к данному моменту тестов:

$$P(C_i|S_{m.l}(t)) = \frac{P(C_i)P(S_{m.l}|C_i)}{\sum_{k=1}^I P(C_k)P(S_{m.l}|C_k)}, l = 1, \dots, l,$$

где C_i – событие, связанное с наличием у испытуемого i -го варианта диагноза ($i = 1, \dots, I$); $S_{m.l}(t)$ – событие, связанное с нахождением в состоянии $S_{m.l}(t)$ по истечении заданного интервала времени; $P(C_i)$ – априорная вероятность принадлежности испытуемого к i -ой категории; $P(S_{m.l}|C_i)$ – вероятность нахождения в состоянии $S_{m.l}$ по истечении заданного интервала времени в случае принадлежности испытуемого к i -ой категории; $P(C_i|S_{m.l})$ – вероятность принадлежности к i -ой категории при условии нахождения в состоянии $S_{m.l}$ по истечении заданного интервала времени. Вероятности $P(S_{m.l}|C_i)$ вычисляются с помощью численного интегрирования уравнений Колмогорова.

В рассмотренной ниже задаче 2 категории испытуемых: «готовые к школе» (1) и «не готовые к школе» (2).

По формуле полной вероятности вычисляем вероятность правильной классификации:

$$P(A) = P(S_{m.1}|C_1) * P(C_1) + P(S_{m.2}|C_2) * P(C_2),$$

где $P(S_{m.1}|C_1)$ – вероятность принадлежности к категории «готовые к школе» при условии нахождения в состоянии $S_{m.1}$;

$P(C_1)$ – априорная вероятность принадлежать к категории «готовые к школе»;

$P(S_{m.2}|C_2)$ – вероятность принадлежности к категории «не готовые к школе» при условии нахождения в состоянии $S_{m.2}$;

$P(C_1)$ – априорная вероятность принадлежать к категории «не готовые к школе».

В качестве априорных вероятностей могут быть использованы экспертные оценки, либо оценки, полученные на предыдущих этапах тестирования.

Значениями критериев в данной задаче являются оценки вероятностей:

$$k_1 = P(C_1|S_{m.1})$$

$$k_2 = P(C_2|S_{m.2})$$

Высокие значения критериев позволяют оценить дифференцирующую силу непредъявленного теста: эксперту, чтобы за наименьшее время отнести испытуемого к одной из категорий, достаточно предъявлять тесты, имеющие высокий показатель дифференцирующей силы.

- 1) Вычислим значение функции предпочтения $H(d_{u,v})$ для всех тестов попарно (табл. 2).

Таблица 2. Значения альтернатив по каждому критерию.

	Класс «готовый к школе»	Класс «не готовый к школе»
Тест 1	0,73	0,56
Тест 2	0,31	0,46
Тест 3	0,2	0,3
Тест 4	0,6	0,29

- 2) Полагаем равными веса критериев $w_1 = w_2 = 0.5$.
- 3) Вычисляем индексы предпочтения попарно для всех тестов и двух критериев – класса «готовый к школе» и класса «неготовый к школе» (табл. 3 и табл. 4 соответственно). На пересечении строки и столбца расположено значение, характеризующее степень предпочтения теста (по строке) над тестом (по столбцу) по фиксированному критерию.

Значения функции $Q_{u,v}$ для всех тестов приведены в табл. 5

- 4) Вычисленные значения коэффициентов Φ_p и Φ_n приведены в табл. 6. Данные коэффициенты являются списками, содержащими построчную сумму значений таблицы 5 (для Φ_p), и сумму по столбцам (для Φ_n). Вычисленный коэффициент итогового предпочтения является условием сортировки альтернатив для получения предпочтительной последовательности предъявления тестов.

Таблица 3. Значения индексов предпочтений тестов для первого критерия (класса «готовый к школе»).

	Тест 1	Тест 2	Тест 3	Тест 4
Тест 1	0	0,958	0,995	0,183
Тест 2	0	0	0,131	0
Тест 3	0	0	0	0
Тест 4	0	0,726	0,941	0

Таблица 4. Значения индексов предпочтений тестов для второго критерия (класса «неготовый к школе»).

	Тест 1	Тест 2	Тест 3	Тест 4
Тест 1	0	0,371	0,989	0,993
Тест 2	0	0	0,760	0,81
Тест 3	0	0	0	0,003
Тест 4	0	0	0	0

Таблица 5. Итоговые индексы предпочтения.

	Тест 1	Тест 2	Тест 3	Тест 4
Тест 1	0	1,330	1,985	1,176
Тест 2	0	0	0,892	0,810
Тест 3	0	0	0	0,003
Тест 4	0	0,726	0,941	0

- 5) Построенная предпочтительная последовательность приведена в табл. 7.

Таблица 6. Коэффициенты предпочтения Φ_p , коэффициенты обратного предпочтения Φ_n , итоговые коэффициенты Φ .

	Φ_p	Φ_n	Φ
Тест 1	4,492	0	0,492
Тест 2	1,702	2,056	-0,353
Тест 3	3,900	3,819	-3,816
Тест 4	1,668	1,990	-0,322

Таблица 7. Предпочтительный порядок предъявления тестов и соответствующие значения коэффициента Φ .

№ теста	Тест	Φ
1	Тест 1	4,492
2	Тест 4	-0,322
3	Тест 2	-0,353
4	Тест 3	-3,816

2.2 Метод коррекции наблюдением

Предположим, что в рамках тестирования могут применяться тесты из множества G , каждый из которых состоит из определённого числа заданий. Пусть $P_g(C_i|y(t))$ – вероятность i -й категории испытуемого при условии нахождения в состоянии сети $y \in Y$ в момент времени t при прохождении теста g , а T_{max} – наибольшее допустимое время прохождения одного теста. Обозначим как $d_g(y(t))$ разность между наибольшей условной вероятностью $\max_{i \in I} P_g(C_i|y(t))$ и вероятностью следующей по порядку выбора категории с теми же самыми параметрами, которую далее будем называть *дифференциацией категорий испытуемых*. Введём наименьшую допустимую дифференциацию d_{min} , которую далее будем использовать как критерий завершения вычислений. Будем говорить, что задания теста выполняются *адекватно*, если в каждый момент времени байесовские оценки,

описанные в главе 1, обеспечивают корректный выбор категории испытуемого. Будем называть марковские процессы, представляющие прохождение теста, *согласованными*, если они позволяют адекватное выполнение заданий.

Метод коррекции наблюдением обеспечивает построение близкой к оптимальной последовательности тестов и опирается на следующий алгоритм:

- 1) положить параметр $t := 0$, параметр d_{min} определяется экспертом;
- 2) Определить тест $g_* \in G$ и момент времени $t_* \in [0; T_{max}]$, которые обеспечивают наибольшую дифференциацию категорий испытуемых среди наименьших дифференциаций в данный момент времени по всем состояниям:

$$(g_*, t_*) = \arg \max_{g \in G, t \in [0; T_{max}]} \min_{y(t) \in Y} d_g(y(t));$$

- 3) провести тестирование с использованием выбранного теста g_* до момента времени $t + t_*$, определив состояние y_* , в котором в момент времени $t + t_*$ находится наблюдаемый испытуемый;
- 4) используя формулы Байеса (см. главу 1), определить наиболее вероятную категорию испытуемого i_* в момент времени $t + t_*$. Если $d_{g_*}(y_*(t + t_*)) \geq d_{min}$, то завершить вычисления ;
- 5) положить параметр $t := t + t_*$;
- 6) определить тест $g_* \in G$ и момент времени $t_* \in [t; t + T_{max}]$, которые обеспечивают наибольшую дифференциацию в случае наиболее вероятной категории i_* .

$$(g_*, t_*) = \arg \max_{g \in G, t \in [t; t + T_{max}]} d_g(y(t));$$

- 7) Перейти к шагу 3.

Утверждение 2. Если:

- задания теста выполняются адекватно,
- марковские процессы, представляющие прохождение теста, согласованы, и
- тесты обеспечивают монотонное возрастание дифференциаций,

то дифференциация, полученная в результате применения метода коррекции наблюдением, хуже наибольшей дифференциации, достижимой в данный момент времени, не более чем на разность наибольшей дифференциации первого выбранного теста и дифференциации, полученной на шаге 2 приведённого выше алгоритма.

Доказательство. Предположим, что существует другая последовательность с большей дифференциацией. Найдём 1-ый шаг, где эти последовательности различаются, и получим, что после этого шага дифференциация меньше. Согласно сделанному предположению, дифференциация, полученная в результате применения метода коррекции наблюдением, меньше. Поскольку тесты обеспечивают монотонное возрастание дифференциаций, то и итоговая дифференциация меньше. Это противоречит шагу № 6 алгоритма.

2.3 Программный комплекс

Структура системы поддержки принятия решений представлена на рисунке 12. Комплекс программ разработан с использованием графического языка G и системы программирования *LabVIEW* [3, 44, 46, 49]. Процесс создания модуля идентификации моделей сопровождался вычислительным экспериментом, анализ результатов которого производился с использованием языка R и системы программирования *RStudio* [28, 85, 94, 95]. Система реализована в составе нескольких модулей, часть из которых отвечает за реализацию процедуры тестирования (предъявления тестов, управления

порядком предъявления тестов и т.д.), часть – за хранение данных, необходимых для функционирования предложенных алгоритмов:

- 1) Модуль «Тестовые задания» содержит в себе перечни вопросов и допустимых вариантов ответов для каждого вопроса. Информация хранится в текстовых файлах.
- 2) Модуль «Характеристики тестов» содержит в себе описания тестов: назначение, целевая группа испытуемых, описание, инструкции к выполнению, инструкции к обработке. Информация хранится в текстовых файлах.
- 3) Модуль «Выборочные данные» содержит в себе исходные выборочные данные для идентификации параметров моделей: метки временных точек, указатель категории испытуемого, название теста, сведения о правильности выполнения заданий теста и т.д. Информация хранится в текстовых файлах с разделителями.
- 4) Модуль «Данные моделей» содержит в себе вычисленные параметры моделей, используемые далее в процессе исполнения компонентов модуля тестирования.
- 5) «Модуль идентификации моделей» содержит в себе алгоритмы идентификации параметров моделей, описанных в главе 1. Описание алгоритма дискретизации значимых параметров приведено в главе 4.
- 6) «Модуль тестирования» является основным модулем. Основные функции: предъявление тестовых заданий, подсчет вероятностей отнесения испытуемого к категориям, фиксация временных показателей исполнения теста. Компонентом данного модуля является
- 7) «Модуль оптимизации предъявления тестов», выполняющий функцию рекомендательного алгоритма, описанного в главе 2, позволяющего выбрать следующий тест с учетом вычисленных дифференциаций на текущем этапе тестирования.

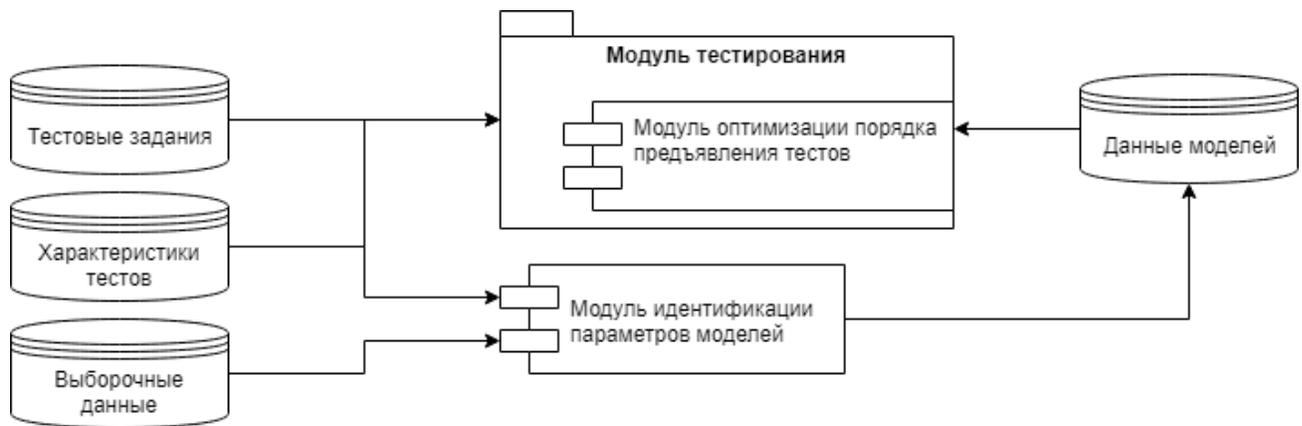


Рис. 12. Структура программного комплекса

Для каждого теста, реализованного в системе, хранится набор текстовых файлов, содержащий данные о количестве тестов в методике, количестве вопросов, возможных исходов каждого задания (модуль «Характеристики тестов»). В модуле «Данные моделей» хранятся оценки параметров моделей для разных групп испытуемых и разных тестах, содержащихся в системе. «Модуль идентификации моделей» реализует предложенные в главе 4 методы идентификации параметров моделей: метод полной дискретизации, методы дискретизации значимых параметров.

Модуль тестирования реализует процедуру предъявления и проверки результатов тестирования, также в нем содержится реализованные методы оптимизации порядка предъявления тестов: метод парных сравнений и метод коррекции наблюдением.

В качестве выборочных данных используются наблюдения, полученные при проведении тестирования с диагностическими группами, уровень способности которых известен заранее.

Глава 3. Вероятностные модели классического и игрового тестирования

3.1 Моделирование процесса выполнения тестовых заданий

Определим величины, необходимые для представления процесса выполнения задания. Для этого рассмотрим конечные генеральные совокупности заданий I и испытуемых J , состоящие, соответственно, из N и M элементов. Пусть $p_{ij}(t)$ есть вероятность того, что j -й испытуемый выполнит i -е задание до момента времени t включительно; $v_i(t)$ есть вероятность того, что произвольно выбранный испытуемый не выполнит i -е задание до момента времени t включительно при условии, что все испытуемые из генеральной совокупности J имеют равную вероятность быть выбранными; $u_j(t)$ есть вероятность того, что j -й испытуемый выполнит произвольно выбранное задание до момента времени t включительно при условии, что все задания из генеральной совокупности I имеют равную вероятность быть выбранными. Время решения задач рассматривается относительно каждой конкретной задачи, а не относительно всего теста. Таким образом, решение каждой задачи начинается с нулевого момента времени.

Величины $v_i(t)$ и $u_j(t)$ рассматриваются, соответственно, как меры трудности i -го задания и способностей j -го испытуемого [80].

Пусть τ есть продолжительность малого минимального интервала времени, в течение которого возможно выполнение задания. Эта величина является параметром исследуемой системы тестовых заданий. Тогда, рассматривая полную группу из двух событий A_j и \bar{A}_j , где A_j есть наличие выполненного j -го задания в момент времени t , и, используя формулу полной вероятности, вероятность $p_{ij}(t)$ того, что j -й испытуемый выполнит i -е

задание до момента времени $t + \tau$ включительно, выражается следующим образом:

$$p_{ij}(t + \tau) = p_{ij}(t)r_p + (1 - p_{ij}(t)) L(u_j(t) - v_i(t), a_\tau, b_\tau),$$

где $p_{ij}(0) = 0$; r_p есть вероятность исправить данное правильное решение на неправильное в течение интервала времени $[t; t + \tau]$;

функция $L(u_j(t) - v_i(t), a_\tau, b_\tau)$ выражает вероятность выполнения задания в течение интервала времени $[t; t + \tau]$ при условии, что в момент времени t задание не выполнено; a_τ, b_τ – параметры, идентифицируемые по результатам наблюдений. В случае простейших тестовых заданий, предполагающих выбор одного из предложенных вариантов ответа, параметр r_p равен единице. Для более сложных заданий, требующих определённых логических рассуждений или аналитических выкладок, значение r_p может быть меньше единицы.

Учитывая накопленный эмпирический опыт [37, 56], можно полагать, что функция L зависит от разности мер способностей испытуемого $u_j(t)$ и трудности задания $v_i(t)$, а также некоторых параметров, обусловленных величиной τ . В качестве аппроксимации этой функции целесообразно использовать модифицированную функцию Раша [90] следующего вида:

$$L(u_j(t) - v_i(t), a_\tau, b_\tau) = \frac{e^{a_\tau(u_j(t) - v_i(t) - b_\tau)}}{1 + e^{a_\tau(u_j(t) - v_i(t) - b_\tau)}} = \frac{1}{1 + e^{a_\tau(v_i(t) - u_j(t) + b_\tau)}}$$

Выбор указанной аппроксимации обусловлен её асимптотическими свойствами: при существенном превышении уровня трудности уровнем способностей вероятность выполнения задания стремится к единице, а в случае обратного соотношения – к нулю.

Пусть $v_i^*(t) = 1 - v_i(t)$ есть вероятность того, что произвольно выбранный испытуемый выполнит i -е задание до момента времени t включительно при условии, что все испытуемые из генеральной совокупности J имеют

равную вероятность быть выбранными. Тогда, рассматривая полную систему из двух событий B_i и \bar{B}_i , где B_i есть наличие выполненного i -го задания в момент времени t , и используя формулу полной вероятности, величина $v_i^*(t + \tau)$ выражается следующим образом:

$$v_i^*(t + \tau) = v_i^*(t)r_v + (1 - v_i^*(t)) P_{v,i}(a_\tau, b_\tau, t),$$

где величина r_v аналогична r_p и представляет вероятность не найти ошибку в представленном ранее к моменту t результате выполнения соответствующего задания в течение интервала времени $[t; t + \tau]$; $P_{v,i}(a_\tau, b_\tau, t)$ есть вероятность выполнения произвольно выбранным испытуемым i -го задания в течение интервала времени $[t; t + \tau]$.

Используя формулу полной вероятности, величину $P_{v,i}$ можно выразить через функцию L :

$$P_{v,i}(a_\tau, b_\tau, t) = \sum_{j=1}^M p_{v,j} L(u_j(t) - v_i, a_\tau, b_\tau),$$

где $p_{v,j}$ есть вероятность выбора j -го испытуемого. Поскольку все испытуемые имеют равную вероятность быть выбранными, то $p_{v,j} = \frac{1}{M}$, а

$$P_{v,i}(a_\tau, b_\tau, t) = \frac{1}{M} \sum_{j=1}^M p_{v,j} L(u_j(t) - v_i, a_\tau, b_\tau).$$

Рассуждая аналогично, несложно выразить $u_j(t + \tau)$:

$$u_j(t + \tau) = u_j(t)r_u + (1 - u_j(t)) P_{u,j}(a_\tau, b_\tau, t),$$

где величина r_u аналогична r_p и r_v , $P_{u,j}$ есть вероятность выполнения j -м испытуемым произвольно выбранного задания в течение интервала времени $[t; t + \tau]$. Как и в случае $P_{v,i}$, величину $P_{u,j}$ можно выразить через

функцию L :

$$P_{u,j}(a_\tau, b_\tau, t) = \sum_{i=1}^N p_{u,i} L(u_j(t) - v_i(t), a_\tau, b_\tau),$$

где $p_{u,i}$ есть вероятность выбора i -го задания. Поскольку все задания имеют равную вероятность быть выбранными, то $p_{u,i} = \frac{1}{N}$, а

$$P_{u,j}(a_\tau, b_\tau, t) = \frac{1}{N} \sum_{i=1}^N p_{u,i} L(u_j(t) - v_i(t), a_\tau, b_\tau).$$

Таким образом, предположение об адекватности модифицированной функции Раша наблюдениям определяет следующие рекуррентные формулы для вычисления вероятностей $v_i^*(t)$, $u_j(t)$ и $p_{ij}(t)$ при $i = 1, \dots, N$ и $j = 1, \dots, M$ в дискретные моменты времени $\{k\tau\}_{k=1}^\infty$:

$$\begin{aligned} p_{ij}(t + \tau) &= p_{ij}(t)r_p + (1 - p_{ij}(t)) L(u_j(t) - v_i(t), a_\tau, b_\tau), \\ v_i^*(t + \tau) &= v_i^*(t)r_v + \frac{1}{M} (1 - v_i^*(t)) \sum_{j=1}^M L(u_j(t) - v_i(t), a_\tau, b_\tau), \\ u_j(t + \tau) &= u_j(t)r_u + \frac{1}{N} (1 - u_j(t)) \sum_{i=1}^N L(u_j(t) - v_i(t), a_\tau, b_\tau), \end{aligned}$$

с начальными условиями $p_{ij}(\tau)$, $v_i^*(\tau)$ и $u_j(\tau)$ и идентифицируемыми по результатам наблюдений параметрами a_τ, b_τ, r_p, r_v и r_u . Параметр τ задаётся заранее и является частью постановки задачи.

Вычисление значений функций $p_{i_0 j_0}(t)$ при фиксированных i_0 и j_0 в дискретные моменты времени $\{k\tau\}_{k=1}^\infty$ требует одновременного использования всех представленных выше рекуррентных формул для $v_i^*(t)$, $u_j(t)$ при $i = 1, \dots, N$ и $j = 1, \dots, M$, что является преодолимой проблемой при

решении практических задач, включая прогнозирование и идентификацию. Однако, для упрощения расчётов, в определённые моменты времени целесообразно использовать предположение об идеальности генеральных совокупностей заданий I и испытуемых J . Назовём указанные генеральные совокупности идеальными в момент времени t , если в них в этот момент времени различные уровни способностей испытуемых и трудностей заданий представлены так, что $v_i(t) = \frac{i}{N}$, $u_j(t) = \frac{j}{M}$. В этом случае расчётные формулы для $p_{i_0j_0}(t)$ используют только функции $v_{i_0}^*(t)$ и $u_{j_0}(t)$, приобретая простой вид автономной системы из трёх уравнений:

$$\begin{aligned} p_{i_0j_0}(t + \tau) &= p_{i_0j_0}(t)r_p + (1 - p_{i_0j_0}(t)) L(u_{j_0}(t) - v_{i_0}(t), a_\tau, b_\tau), \\ v_{i_0}^*(t + \tau) &= v_{i_0}^*(t)r_v + \frac{1}{M} (1 - v_{i_0}^*(t)) \sum_{j=1}^M L\left(\frac{j}{M} - v_{i_0}(t), a_\tau, b_\tau\right), \\ u_{j_0}(t + \tau) &= u_{j_0}(t)r_u + \frac{1}{N} (1 - u_{j_0}(t)) \sum_{i=1}^N L\left(u_{j_0}(t) - \frac{i}{N}, a_\tau, b_\tau\right). \end{aligned}$$

3.1.1 Идентификация параметров рекуррентных зависимостей

Данными наблюдений, по которым идентифицируются параметры рекуррентных зависимостей a_τ, b_τ, r_p, r_v и r_u , являются матрицы ответов, определяемые для заданных контрольных моментов времени по результатам выполнения корректно подготовленных наборов тестовых заданий репрезентативными группами испытуемых. Каждая из этих матриц представляют собой прямоугольную таблицу, на пересечении k -й строки и l -го столбца которой находится единица, если k -й испытуемый выполнил l -е задание к моменту времени t , которому соответствует рассматриваемая таблица, и нуль – в противном случае. Вычисление для каждой строки и столбца отношения числа содержащихся в них единиц к общему количеству столбцов или строк в матрице ответов определяет выборочные оценки вероятностей $u_j(t)$ и $v_i^*(t)$, соответственно.

Эти оценки вероятностей позволяют идентифицировать параметры

a_τ, b_τ, r_p, r_v и r_u (как правило, для упрощения задачи можно полагать, что $r_p = r_v = r_u = r$). Используя представленное выше упрощающее предположение об идеальности генеральных совокупностей заданий I и испытуемых J , параметры a_τ, b_τ и r целесообразно идентифицировать для каждой пары вероятностей $u_j(t)$ и $v_i^*(t)$, где $i \in 1, \dots, N$ и $j \in 1, \dots, M$, автономно. Идентифицированные параметры, которые можно обозначить как $a_{\tau,ij}, b_{\tau,ij}$ и r_{ij} , используются для вычисления функций $u_j(t), v_i^*(t)$ и $p_{ij}(t)$ в заданные моменты времени с помощью определённых выше рекуррентных зависимостей. Очевидно, что идентификация без привязки к i и j даёт менее точные результаты.

Оценки идентифицируемых параметров могут быть определены, опираясь на критерий, характеризующий меру соответствия прогнозируемым значениям функций $v_i^*(t_k)$ и $u_j(t_k)$ в контрольные моменты времени $\{t_k\}_{k=1}^K$ наблюдаемых оценок, представляющих распределения по заданному подмножеству генеральной совокупности заданий $I_0 \in I$ объёмом N_0 количеств испытуемых $F_i(t_k)$, выполнивших i -е задания, и распределения по заданному подмножеству генеральной совокупности испытуемых $J_0 \in J$ объёмом M_0 количеств заданий $Q_j(t_k)$, выполненных (-ми) испытуемыми. В качестве такого критерия используется статистика Пирсона:

$$\chi^2 = \sum_{k=1}^K \left[\sum_{i=1}^{N_0} \frac{(F_i(t_k) - v_i^*(t_k)N_0)^2}{v_i^*(t_k)N_0} + \sum_{j=1}^{M_0} \frac{(Q_j(t_k) - u_j(t_k)M_0)^2}{u_j(t_k)M_0} \right]$$

Величина χ^2 является мерой соответствия в том смысле, что ее большие значения означают плохое согласование прогнозируемых и наблюдаемых данных, а малые значения – хорошее согласование. Таким образом, решение задачи идентификации сводится к нахождению таких параметров a_τ, b_τ и r , которые обеспечивают минимальное значение указанного критерия, что обеспечивается выполнением численной процедуры многомерной нелинейной оптимизации [73].

Приведённый выше способ идентификации называется методом минимума χ^2 [18, 19]. Согласно теореме Крамера [18] при выполнении ря-

да общих условий, в случае адекватности прогнозируемых процессов наблюдениям, рассмотренная задача имеет единственное решение, которое сходится по вероятности к искомому решению, а значения статистики χ^2 асимптотически описываются распределением χ^2 с $K(N_0 + M_0) - \gamma - 1$ степенями свободы, где γ – число идентифицируемых параметров. Это позволяет использовать приведённую выше статистику для проверки гипотезы о том, что прогнозируемые оценки согласуются с результатами наблюдений.

Использование статистики Пирсона корректно только при выполнении условий указанной выше теоремы. Если эти условия не выполнены, то вычисление оценок идентифицируемых параметров с использованием критерия χ^2 остаётся возможным, однако полученное решение может быть не единственным, а значения соответствующей статистики не обязаны быть распределены как χ^2 . Следует отметить, что одним из признаков нарушения условий теоремы является невыполнение неравенства $K(N_0 + M_0) > \gamma + 1$.

В случае нарушения условий теоремы для оценки степени соответствия модели наблюдениям следует подобрать критерий, отличный от указанного выше. Один из таких критериев строится на основе представленного в работе [79] анализа регрессионной зависимости между наблюдаемыми и прогнозируемыми оценками сравниваемых величин в контрольные моменты времени.

3.1.2 Использование результатов моделирования

Идентификация параметров рекуррентных зависимостей обеспечивает возможность прогнозирования вероятностных функций $p_{ij}(t)$ на любом временном интервале, что позволяет:

- оценивать пригодность тестовых заданий для измерения заявленных характеристик испытуемых;
- решать различные типы диагностических задач с учётом динамики

выполнения тестовых заданий;

- оценивать временную динамику и шансы выполнения определённых заданий в зависимости от их трудностей и категории испытуемых;
- определять полезное время выполнения заданий, под которым понимается наименьший интервал времени, по истечении которого вероятность их выполнения остаётся неизменной в пределах заданной точности.

3.2 Процесс тестирования в игровой форме

Определим величины, необходимые для представления процесса игры. Для этого рассмотрим конечные генеральные совокупности игровых ситуаций I и игроков J , состоящие, соответственно, из N и M элементов. Пусть $v(t)$ есть вероятность того, что произвольно выбранный из генеральной совокупности J игрок не завершит приемлемым для себя образом заданную игру до момента времени t включительно при условии, что все игроки из J имеют равную вероятность быть выбранными; $u(t)$ есть вероятность того, что произвольно выбранный игрок определённой категории $J_0 \subseteq J$ завершит заданную игру приемлемым для себя образом до момента времени t включительно при условии, что все допустимые игровые ситуации из подмножества генеральной совокупности $I_0 \subseteq I$ на каждом ходе игры имеют равную вероятность быть созданными. В зависимости от типа игры, под её завершением приемлемым образом понимается набор заданного количества баллов, достижение определённого счёта или просто победа в игре [11].

Величины $v(t)$ и $u(t)$ рассматриваются, соответственно, как меры трудности игры и способностей игрока.

Пусть τ есть продолжительность минимального интервала времени, в течение которого оцениваются характеристики игры и игроков. Эта величина является параметром исследуемой игры и может включать один или серию ходов, сделанных игроком. Выбор τ задаёт масштаб времени.

Функция $L(u(t) - v(t), a_\tau(t), b_\tau(t))$ выражает вероятность завершения игры приемлемым образом в течение интервала времени $[t; t + \tau]$ при условии, что в момент времени t игра таким образом не завершена; $a_\tau(t)$ и $b_\tau(t)$ – функции, идентифицируемые по результатам наблюдений.

Функция L зависит от разности мер способностей игрока $u(t)$ и трудности игры $v(t)$, а также некоторых параметров, обусловленных величиной τ . В качестве аппроксимации этой функции рассматривается модифицированная функция Раша [11] следующего вида:

$$L(u(t) - v(t), a_\tau, b_\tau) = \frac{e^{a_\tau(u(t)-v(t))-b_\tau}}{1 + e^{a_\tau(u(t)-v(t))-b_\tau}} = \frac{1}{1 + e^{a_\tau(u(t)-v(t))+b_\tau}}$$

Выбор указанной аппроксимации обусловлен её асимптотическими свойствами: при существенном превышении уровня трудности уровнем способностей вероятность выполнения задания стремится к единице, а в случае обратного соотношения – к нулю.

Пусть $v^*(t) = 1 - v(t)$ есть вероятность того, что произвольно выбранный из генеральной совокупности J игрок завершит приемлемым для себя образом заданную игру до момента времени t включительно при условии, что все игроки из J имеют равную вероятность быть выбранными. Тогда, рассматривая полную систему из двух событий B и \bar{B} , где B есть наличие завершённой игры в момент времени t , и используя формулу полной вероятности, величина $v^*(t + \tau)$ выражается следующим образом:

$$v^*(t + \tau) = v^*(t) + (1 - v^*(t)) * P_v(a_\tau, b_\tau, t),$$

где $P_v(a_\tau, b_\tau, t)$ есть вероятность завершения приемлемым образом заданной игры произвольно выбранным из генеральной совокупности J игроком в течение интервала времени $[t; t + \tau]$.

Используя формулу полной вероятности, величину P_v можно выразить

через функцию L :

$$P_v(a_\tau, b_\tau, t) = \sum_{j=1}^M p_{v,j} L(u_j(t) - v_t, a_\tau, b_\tau),$$

где $p_{v,j}$ есть вероятность выбора j -го игрока категории $J_0 \subseteq J$, $u_j(t)$ есть вероятность того, что j -й игрок завершит заданную игру приемлемым для себя образом до момента времени t включительно при условии, что все допустимые игровые ситуации из подмножества генеральной совокупности $I_0 \subseteq I$ на каждом ходе игры имеют равную вероятность быть созданными. Поскольку все игроки имеют равную вероятность быть выбранными, то $p_{v,j} = \frac{1}{M}$, а

$$P_v(a_\tau, b_\tau, t) = \frac{1}{M} \sum_{j=1}^M L(u_j(t) - v(t), a_\tau, b_\tau).$$

Аналогично выражается величина $u(t + \tau)$:

$$u(t + \tau) = u(t) + (1 - u(t))P_u(a_\tau, b_\tau, t),$$

где $P_u(a_\tau, b_\tau, t)$ есть вероятность завершения приемлемым образом заданной игры игроком определённой категории $J_0 \subseteq J$ в течение интервала времени $[t; t + \tau]$. Как и в случае P_v , величину P_u можно выразить через функцию L :

$$P_u(a_\tau, b_\tau, t) = \sum_{i=1}^{N_0} p_{u,i} L(u(t) - v_i(i) a_\tau, b_\tau),$$

где $p_{u,i}$ есть вероятность создания i -й игровой ситуации из подмножества генеральной совокупности $I_0 \subseteq I$, $v_i(t)$ есть вероятность того, что произвольно выбранный из генеральной совокупности J игрок не завершит приемлемым для себя образом заданную игру до момента времени t вклю-

чительно в случае i -й игровой ситуации при условии, что все игроки из J имеют равную вероятность быть выбранными, а N_0 — среднее число игровых ситуаций из подмножества генеральной совокупности $I_0 \subseteq I$, возникающих после каждого хода игры. Поскольку все игровые ситуации имеют равную вероятность быть созданными, то $p_{u,i} = \frac{1}{N_0}$, а

$$P_u(a_\tau, b_\tau, t) = \frac{1}{N_0} \sum_{i=1}^{N_0} L(u(t) - v_i(i)a_\tau, b_\tau)$$

Таким образом, предположение об адекватности модифицированной функции Раша наблюдениям определяет следующие рекуррентные формулы для вычисления вероятностей $v^*(t)$ и $u(t)$ в дискретные моменты времени $\{k\tau\}_{k=1}^\infty$:

$$v^*(t + \tau) = v^*(t) + \frac{1}{M}(1 - v^*(t)) \sum_{j=1}^M L(u_j(t) - v(t), a_\tau, b_\tau),$$

$$u(t + \tau) = u(t) + \frac{1}{N_0}(1 - u(t)) \sum_{i=1}^{N_0} L(u(t) - v_i(t), a_\tau, b_\tau),$$

с начальными условиями $v^*(0) = 0$ и $u(0) = 0$ и идентифицируемыми по результатам наблюдений функциями $a_\tau(t)$ и $b_\tau(t)$. Параметр задаётся заранее и является частью постановки задачи.

Для упрощения расчётов целесообразно использовать предположение об идеальности совокупностей игровых ситуаций I_0 и игроков J . Назовём указанные генеральные совокупности идеальными в момент времени t , если в них в этот момент времени различные уровни способностей игроков и трудностей игровых ситуаций представлены так, что $v_i(t) = \frac{1}{N_0}$, $u_j(t) = \frac{j}{M}$. В этом случае расчётные формулы используют только функции $v^*(t)$

и $u(t)$, приобретая вид автономной системы из двух уравнений:

$$v^*(t + \tau) = v^*(t) \frac{1}{M} (1 - v^*(t)) \sum_{j=1}^M L\left(\frac{j}{M} - v(t), a_\tau, b_\tau\right),$$

$$u(t + \tau) = u(t) + \frac{1}{N_0} (1 - u(t)) \sum_{i=1}^{N_0} L(u(t)) - \frac{i}{N_0}, a_\tau, b_\tau).$$

3.2.1 Идентификация параметров

Данными наблюдений, по которым идентифицируются параметры функций $a_\tau(t)$ и $b_\tau(t)$, являются выборочные оценки вероятностей $u_j(t)$.

Оценки идентифицируемых параметров могут быть определены, опираясь на критерий, характеризующий меру соответствия прогнозируемым значениям функции $u(t_k)$ в контрольные моменты времени $\{t_k\}_{k=1}^k$ наблюдаемых оценок, представляющих распределения количеств игроков $F(t_k)$ заданного подмножества генеральной совокупности испытуемых $J_0 \subseteq J$ объёмом M_0 , завершивших игру до момента времени t_k приемлемым образом. В качестве такого критерия используется статистика Пирсона:

$$\chi^2 = \sum_{k=1}^K \left[\frac{(F(t_k) - u(t_k)M_0)^2}{u(t_k)M_0} \right]$$

3.3 Пример практического применения

3.3.1 Моделирование процесса выполнения заданий теста Равена

В качестве наблюдаемых зависимостей трудностей и способностей использовались выборочные оценки, полученные по результатам тестирования группы испытуемых, состоящей из 90 человек в возрасте от 15 до 20 лет ($M = 16,7$ лет, $SD = 0,8$ лет) с нормальным или скорректированным до нормального зрением. Тестирование производилось с использованием компьютерной реализации матричного теста продуктивного интеллекта

Равена (продвинутая прогрессивная версия для взрослых) [37]. Время и результат решения каждого задания, каждым испытуемым, фиксировалось в базе данных, в процессе выполнении заданий.

Результаты тестирования, использованные для расчёта выборочных зависимостей, необходимых для дальнейшей идентификации моделей, были представлены матрицей ответов испытуемых, включающей в себя метку верного/неверного ответа, а также матрицей, содержащей время решения каждого задания.

По матрице, содержащей время решения каждого задания, были оценены выборочные зависимости трудностей всех заданий и способностей всех испытуемых от времени. Оценка выборочной зависимости трудности задания от времени выполнялась следующим образом: по матрице времён выполнения заданий для каждого из значений параметра времени, выбранных с единичным шагом в диапазоне от 0 до 200 секунд, рассчитывалась доля испытуемых в выборке, которые не выполнили задание к соответствующему моменту. Оценка выборочной зависимости способности испытуемого от времени выполнялась аналогично: по матрице времён выполнения заданий для каждого из указанных выше значений параметра времени рассчитывалась доля заданий в выборке, которые были выполнены испытуемым к соответствующему моменту. Полученные зависимости вместе с их усреднёнными аналогами приведены на рисунке 13 (усреднение долей проводилось в каждый момент времени отдельно).

Формы выборочных зависимостей трудностей заданий и способностей испытуемых от времени явно демонстрируют выраженную связь с величиной трудностей заданий и способностей испытуемых в классическом понимании. Однако можно наблюдать, что вариация кривых зависимости способности испытуемых во времени содержит также необъяснённые компоненты, которые, вероятно, связаны с такими характеристиками испытуемых, как скорость принятия решений, уверенность в выбранном ответе и т.п.

Параметры описанной динамической модели были идентифицированы в среде программирования RStudio [28, 85], с использованием встроенной

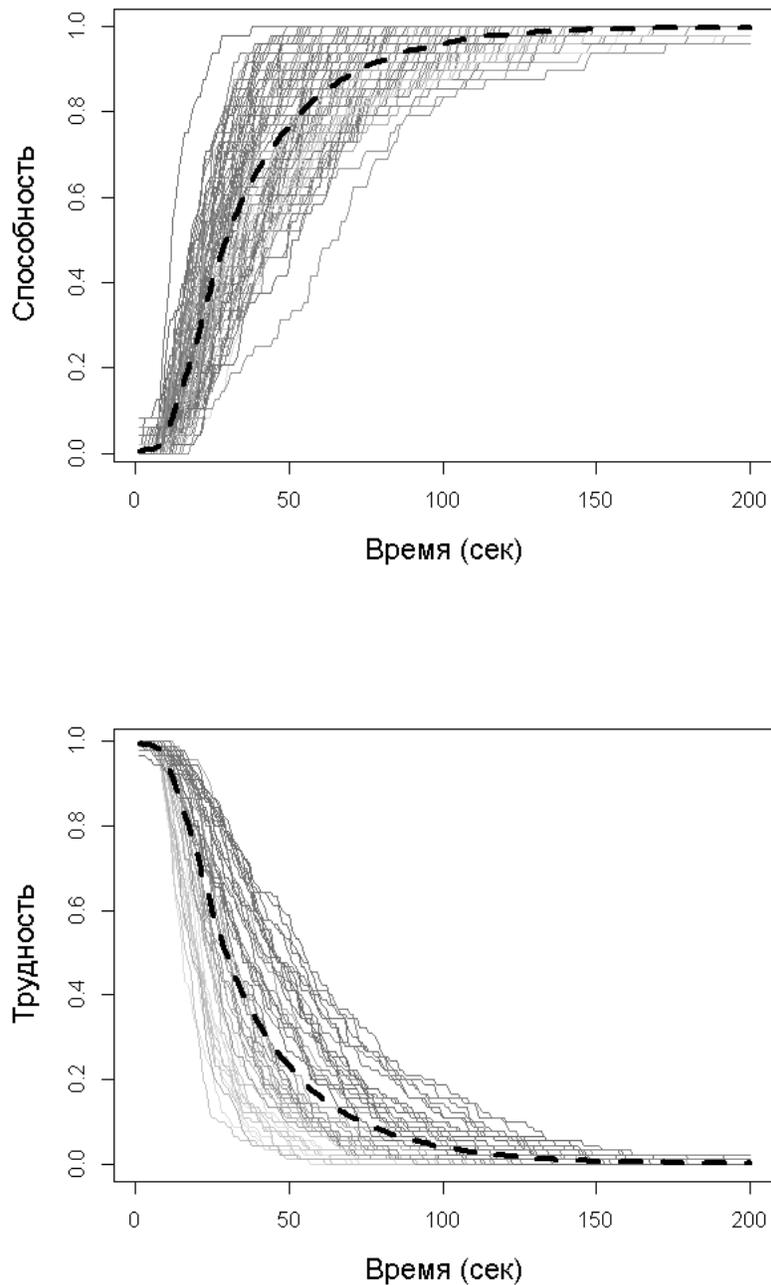


Рис. 13. Тонкими линиями показаны выборочные зависимости способностей испытуемых (верхний график) и трудностей заданий (нижний график) от времени (каждый отдельный график соответствует испытуемому/заданию). Жирными чёрными линиями показаны результаты усреднения соответствующего набора выборочных зависимостей.

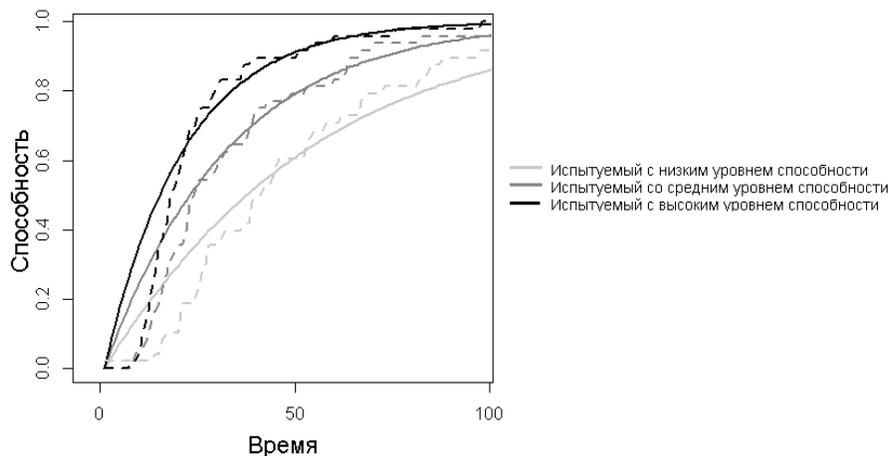


Рис. 14. Примеры результирующих модельных (непрерывная кривая) и выборочных (пунктирная кривая) зависимостей способностей испытуемых с низким ($a_\tau = 0,158; b_\tau = 24,800$), средним ($a_\tau = 0,077; b_\tau = 44,155$) и высоким ($a_\tau = 0,057, b_\tau = 51,522$) уровнями способностей, оценёнными классическим образом.

реализации алгоритма модифицированного численного метода оптимизации Бroyдена-Флетчера-Гольдфарба-Шенно, применяемого для оценки параметров моделей в задачах машинного обучения в условиях, когда параметры модели следует искать в заданных интервалах значений.

В качестве исходных данных для идентификации моделей из выборки испытуемых случайным воспроизводимым образом были отобраны 27 испытуемых с относительно низким $((0; 0,48])$, средним $((0,48; 0,6])$ и высоким $((0,6; 1])$ уровнями способностей, оцениваемыми как отношение количества решенных заданий к конечному моменту времени к общему количеству заданий, и относительно малым $((0; 30])$, средним $((30; 40])$ и большим $((40; 80])$ средним временем выполнения всех заданий теста. На каждую группу приходилось по 3 испытуемых. Для каждого испытуемого идентифицировалось 48 моделей (по одной модели на каждое задание) – всего 1296 моделей. Примеры модельных зависимостей способностей и трудностей представлены на рисунке [14](#) и рисунке [15](#).

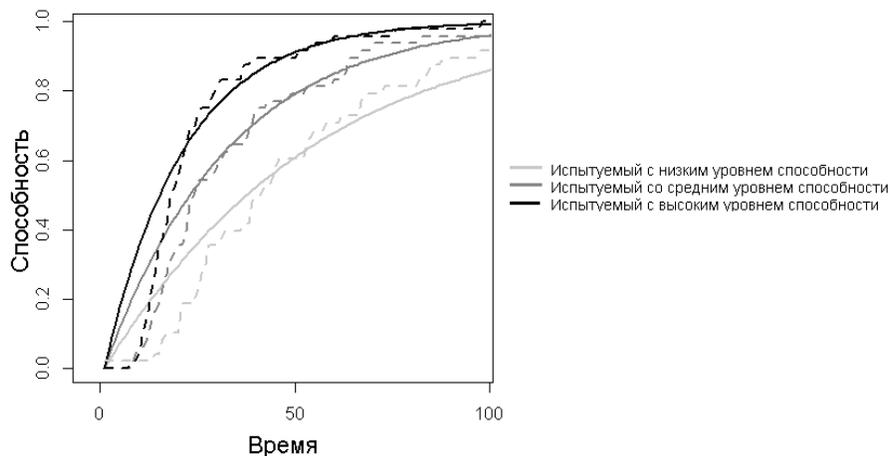


Рис. 15. Примеры результирующих модельных (непрерывная кривая) и выборочных (пунктирная кривая) зависимостей трудностей заданий с низким (задание №3, $a_\tau = 0,063$; $b_\tau = 46,582$), средним (задание №27, $a_\tau = 0,067$; $b_\tau = 50,670$) и высоким (задание №44, $a_\tau = 0,100$; $b_\tau = 40,629$) уровнями трудностей, оценёнными классическим образом.

3.4 Использование результатов моделирования

Идентификация параметров рекуррентных зависимостей обеспечивает возможность прогнозирования вероятностных функций $p_{ij}(t)$ на любом временном интервале, что позволяет:

- оценивать пригодность тестовых заданий для измерения заявленных характеристик испытуемых;
- решать различные типы диагностических задач с учётом динамики выполнения тестовых заданий;
- оценивать временную динамику и шансы выполнения определённых заданий в зависимости от их трудностей и категории испытуемых;
- определять полезное время выполнения заданий, под которым понимается наименьший интервал времени, по истечении которого вероятность их выполнения остаётся неизменной в пределах заданной точности .

Глава 4. Разработка и сравнение численных методов идентификации моделей

Для приведенных в главе 1 систем уравнений ставится задача идентификации набора параметров. Вычисляются значения, обеспечивающие наилучшее соответствие наблюдаемых и ожидаемых частот попадания в определенное состояние системы в заданные моменты времени. При этом определяется набор интенсивностей λ , обеспечивающий наименьшее значение статистики Пирсона:

$$\chi^2(\lambda) = \sum_{d=0}^{D-1} \sum_{i=0}^n \frac{(p_i(\lambda, t_d)N - F_i)^2}{p_i(\lambda, t_d)N},$$

где $N = \sum_{i=0}^n F_{id}$. Эта статистика используется как мера соответствия модели наблюдениям.

Доказано, что, при выполнении ряда общих условий, значения указанной статистики χ^2 , получаемые при подстановке истинных решений, асимптотически описываются распределением χ^2 с $n - m - 1$ степенями свободы, где m – число определяемых параметров, причем вычисленные значения свободных параметров при увеличении объема выборки сходятся по вероятности к искомому решению [18]. Это позволяет использовать статистику Пирсона для проверки гипотезы о том, что полученный прогноз согласуется с данными наблюдений.

Таким образом, постановка задачи предполагает решение обратной задачи, в которой коэффициенты дифференциальных уравнений определяются по фрагментам решения, заданным наблюдениями. Представленная выше постановка может формально рассматриваться как прямая задача с усложнённым вычислением критерия χ^2 , которая решается стандартными методами оптимизации.

Для её решения, в частности, допустимы алгоритмы локальной оптимизации, включая градиентные методы, и стохастические алгоритмы.

Однако опыт их практического применения позволяет говорить о неприемлемо большом времени вычислений в случае задач высокой размерности (при больших значениях параметра m), а также о проблемах, обусловленных возможной неединственностью получаемого решения.

Для решения этих проблем предложены представленные численные методы, предназначенные для идентификации набора параметров λ , определяющего матрицу M . Их особенностью является использование начальных приближений, вычисляемых по данным наблюдений, а также оценок чувствительности критерия оптимизации к малым изменениям идентифицируемых параметров. Новым элементом второго из предложенных подходов является улучшение временных характеристик применяемого комбинаторного алгоритма подбора значений идентифицируемых переменных за счёт замены на каждой итерации вычислений перебора различных сочетаний этих значений в окрестности их текущих оценок перебором значений только тех переменных, к изменению которых наиболее чувствителен критерий оптимизации. Для оценки этой чувствительности используются разностные аппроксимации соответствующих частных производных.

Рассмотренные алгоритмы могут эффективно применяться для настройки параметров обучаемых структур, включая нейронные сети.

4.1 Градиентный метод: алгоритм вычислений

- 1) Используя имеющиеся данные наблюдений, вычислить в качестве начальных оценок параметров $\lambda_i (i = 0, \dots, m)$, обозначаемых как $\lambda_i^0 (i = 0, \dots, \dots, m)$, выборочные средние значения количеств переходов между парами соответствующих состояний марковского процесса в единицу времени (или, что иногда удобнее, величины, обратные выборочным средним продолжительностям времени между двумя смежными переходами для этих же пар состояний). Полученные оценки рассматривать как начальные приближения к идентифицируемым параметрам.
- 2) Задать счётчик итераций $j := 0$, установить предельно допустимое

число итераций j_{max} .

- 3) Установить шаг градиентного метода $h := h_0$.
- 4) Если $j = j_{max} - 1$, то завершить работу.
- 5) В точке текущего приближения λ^j вычислить оценку градиента $\nabla \chi^2(\lambda)$ с использованием центрально-разностной аппроксимации, при этом значения $\mathbf{p}(\lambda t_d)$ определить путём численного решения задачи Коши для системы уравнений $\frac{d\mathbf{p}(t)}{dt} = \mathbf{M}(\lambda)\mathbf{p}$ с заданными начальными условиями.
- 6) Выполнить коррекцию текущего приближения: $\lambda^{j+1} := \lambda^j - h\nabla \chi^2(\lambda^j)$.
- 7) Если $\chi^2(\lambda^{j+1}) < \chi^2(\lambda^j)$, то положить $j := j + 1$ и перейти к шагу 3, иначе перейти к шагу 8.
- 8) Если $\chi^2(\lambda^{j+1}) \geq \chi^2(\lambda^j)$, то положить $\lambda^{j+1} := \lambda^j$, $j := j + 1$, $h := \frac{h}{2}$, и перейти к шагу 4.

4.2 Метод полной дискретизации: алгоритм вычислений

- 1) Используя имеющиеся данные наблюдений, вычислить в качестве начальных оценок параметров $\lambda_i (i = 0, \dots, m)$, обозначаемых как $\lambda_i^0 (i = 0, \dots, \dots, m)$, выборочные средние значения количеств переходов между парами соответствующих состояний марковского процесса в единицу времени (или, что иногда удобней, величины, обратные выборочным средним продолжительностям времени между двумя смежными переходами для этих же пар состояний). Полученные оценки рассматривать как начальные приближения к идентифицируемому параметрам.
- 2) Положить $j := 1$, $k := 1$, $\beta_1 := 1 + \gamma_1$, где γ_1 – параметр алгоритма.

- 3) Для каждой из текущих оценок вычислить идентифицируемых параметров $\lambda_i^j (i = 0, \dots, m)$ на j -ой итерации вычислить верхнюю λ_i^{j+} и нижнюю λ_i^{j-} границу сдвига по формулам: $\lambda_i^{j+} := \beta_j \lambda_i^j$, $\lambda_i^{j-} := \beta^{-1} \lambda_i^j$.
- 4) Выполнив полный перебор всех вариантов текущих оценок набора параметров $\lambda_j = (\lambda_0^j, \dots, \lambda_m^j)^T$, в которых каждый из его компонентов $\lambda_i^j (i = 0, \dots, m)$ принимает только три возможных значения из множества
- $$\left\{ \lambda_i^{j-1-}, \lambda_i^{j-1}, \lambda_i^{j-1+} \right\}.$$
- Выбрать из указанных вариантов одну из оценок $\lambda_{j,*} := (\lambda_{0,*}^j, \dots, \lambda_m^j)^T$ обеспечивающих минимальное значение критерия χ^2 , равное $\chi^{2j,min}$.
- 5) Если для всех компонентов выбранного набора $\lambda_{j,*}$, выполняется равенство $\lambda_{i,*}^j = \lambda_i^{j-1} (i = 0, \dots, m)$, то положить $k := k + 1$, $\gamma_k := \frac{\gamma_{k-1}}{2}$.
- 6) Положить $j = j + 1$, $\beta_j = 1 + \gamma_k$, $\lambda_j = \lambda_{j-1}$.
- 7) Если $\chi_{j-1,min}^2 > \chi_*^2$, где χ_*^2 параметр алгоритма, то перейти к шагу 3, иначе завершить вычисления.

Временная сложность одной итерации рассмотренного алгоритма $O(3^{m+1})$, что приводит к неприемлемо длительным вычислениям при больших значениях m . Для того, чтобы временная сложность стала приемлемой для практических вычислений, перебор вариантов текущих оценок набора параметров на шаге 4 приведённого выше алгоритма следует осуществлять не для всех параметров $\lambda_i^j (i = 0, \dots, m)$, а только для их подмножества, состоящего из $l < m + 1$ элементов, к изменению которых наиболее чувствителен критерий χ^2 .

4.3 Метод дискретизации значимых параметров: алгоритм вычислений

- 1) Используя имеющиеся данные наблюдений, вычислить в качестве начальных оценок параметров $\lambda_i (i = 0, \dots, m)$, обозначаемых как

$\lambda_i^0 (i = 0, \dots, m)$, выборочные средние значения количеств переходов между парами соответствующих состояний марковского процесса в единицу времени (или, что иногда удобнее, величины, обратные выборочным средним продолжительностям времени между двумя смежными переходами для этих же пар состояний). Полученные оценки рассматривать как начальные приближения к идентифицируемым параметрам.

- 2) Положить $j := 1, k := 1, \beta_1 := 1 + \gamma_1$, где γ_1 – параметр алгоритма.
- 3) Оценить чувствительность критерия χ^2 в малой окрестности каждой из текущих оценок идентифицируемых параметров $\lambda_i^j (i = 0, \dots, m)$ на j -ой итерации алгоритма, используя для этого разностные аппроксимации абсолютных значений частных производных:

$$\delta_i(\varepsilon) = \left| \frac{\chi^2(\lambda_0^j, \dots, \lambda_i^j + \varepsilon, \dots, \lambda_m^j) - \chi^2(\lambda_0^j, \dots, \lambda_i^j, \dots, \lambda_m^j)}{\varepsilon} \right|,$$

где $i = 0, \dots, m; \varepsilon$ – параметр алгоритма.

- 4) Выбрать l текущих оценок идентифицируемых параметров $\lambda_{i_q}^j$, где $i_q \in \{0, \dots, m\}; q = 1, \dots, l; l$ – параметр алгоритма, имеющих наибольшие значения оценок чувствительности $\delta_i(\varepsilon)$.
- 5) Для каждой из выбранных на шаге 4 текущих оценок идентифицируемых параметров $\lambda_{i_q}^j$ на j -ой итерации алгоритма, где $i_q \in \{0, \dots, m\}; q = 1, \dots, l$; вычислить верхнюю $\lambda_{i_q}^{j+}$ и нижнюю $\lambda_{i_q}^{j-}$ границу сдвига оценки по формулам: $\lambda_{i_q}^{j+} := \beta_j \lambda_{i_q}^j, \lambda_{i_q}^{j-} := \beta_j^{-1} \lambda_{i_q}^j$
- 6) Выполнив полный перебор всех вариантов текущих оценок набора параметров $\lambda_j = (\lambda_0^j, \dots, \lambda_m^j)^T$, в которых каждый из его компонентов $\lambda_{i_q}^j$, где $i_q \in \{0, \dots, m\}; q = 1, \dots, l$; принимает только три возможных значения из множества:

$$\left\{ \lambda_{i_q}^{j-1,-}, \lambda_{i_q}^{j-1}, \lambda_{i_q}^{j-1,+} \right\},$$

а оставшиеся компоненты – значение своей текущей оценки на j -й итерации алгоритма, выбрать из указанных вариантов одну из оценок $\lambda_{j,*} := (\lambda_{0,*}^j, \dots, \lambda_{m,*}^j)^T$, обеспечивающих минимальное значение критерия χ^2 , равное $\chi_{j,min}^2$

- 7) Если для всех компонентов выбранного набора $\lambda_{j,*}$ выполняется равенство $\lambda_{i,*}^j = \lambda_i^{j-1}$ ($i = 0, \dots, m$), то положить $k := k + 1$, $\gamma_k := \frac{\gamma_{k-1}}{2}$.
- 8) Положить $j := j + 1$, $\beta_j := 1 + \gamma_k$, $\lambda_j := \lambda_{j-1,*}$.
- 9) Если $\chi_{j-1,min}^2 > \chi_*^2$, где χ_*^2 – параметр алгоритма, то перейти к шагу 3, иначе завершить вычисления.

Для решаемых на практике типовых задач общее время вычислений зависит от параметра l монотонно, достигая наименьшего значения при $l = 1$. Временная сложность одной итерации модифицированного алгоритма при $l = 1$ есть $O(m)$, что свидетельствует об асимптотически линейной зависимости времени вычислений, необходимого для выполнения одной итерации, от числа идентифицируемых параметров.

4.4 Вычислительный эксперимент

Для оценки эффективности предложенных алгоритмических решений, был создан пул тестовых структур, близких по своим параметрам и структуре к моделям, используемым при решении реальных прикладных задач [20, 21, 79]. В качестве базовой рассматривалась марковская модель, организованная по схеме «гибели и размножения». В процессе моделирования управляемыми параметрами были число состояний такой модели и число независимых параметров, через которые выражались интенсивности переходов между состояниями.

Самым простым случаем является ситуация, при которой все интенсивности λ равны между собой, а самым сложным – случай, при котором каждый из переходов имеет собственную интенсивность. В процессе генерации моделей из заданного числового интервала случайным образом

выбирались значения переменных, сопоставленных интенсивностям переходов между состояниями [22].

Пример типовой модели представлен на рис. 16 в виде ориентированного графа.

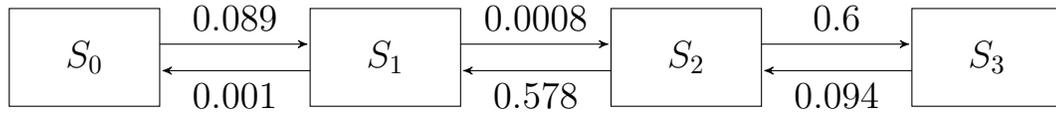


Рис. 16. Организованная по схеме «гибели и размножения» марковская модель из 4 состояний. Интенсивности переходов определяются 6 независимыми параметрами.

Структура связей подобного графа может быть описана матрицей смежности, элементы которой содержат значения интенсивностей переходов.

Таблица 8. Пример матрицы смежности, отражающей структуру связей и интенсивности переходов, для модели с 4 состояниями и 6 независимыми параметрами.

	S_0	S_1	S_2	S_3
S_0		0,89		
S_1	0,001		0,0008	
S_2		0,578		0,6
S_3			0,094	

Матрица, представленная в таблице 8 позволяет сформировать следующую систему обыкновенных дифференциальных уравнений Колмогоро-

ва:

$$\begin{cases} \frac{dp_{S_0}(t)}{dt} = p_{S_1}(t) * 0,001 - p_{S_0}(t) * 0,089 \\ \frac{dp_{S_1}(t)}{dt} = p_{S_0}(t) * 0,089 + p_{S_2}(t) * 0,578 - p_{S_1}(t) * (0,001 + 0,0008) \\ \frac{dp_{S_2}(t)}{dt} = p_{S_1}(t) * 0,0008 + p_{S_3}(t) * 0,094 - p_{S_2}(t) * (0,578 + 0,6) \\ \frac{dp_{S_3}(t)}{dt} = p_{S_3} * 0,6 - p_{S_3}(t) * 0,094 \end{cases}$$

Для всех созданных моделей в начальный момент времени t_0 вероятность пребывания в состоянии S_0 была равна единице, а во всех прочих – нулю. Для каждой возможной комбинации «Число состояний/число параметров» (до 10 состояний) было сгенерировано по 20 примеров (рисунок [17](#)).

Рис. 17. Структура сформированного пула моделей для идентификации.

Для систем уравнений Колмогорова, соответствующих сгенерированным моделям, на временном интервале $[0;1000]$ численно решалась задача Коши [\[51\]](#). Вычисленные значения вероятностей $\mathbf{P}(t) = (p_{S_0}(t), \dots, p_{S_n}(t))^T$ для контрольных точек $\{t_d\}_{d=0, \dots, D-1}$, сохранялись в файле модели. Чтобы избежать ряда вычислительных проблем, связанных с представлением малых вероятностей, вместо χ^2 в качестве минимизируемого критерия использовалась статистика:

$$Q = \sum_{d=0}^{D-1} \sum_{i=0}^n (p_i(t_d, \lambda) - \tilde{p}_i(t_d))^2,$$

где $\tilde{p}_i(t_d)$ – вероятность пребывания в i -том состоянии модели в момент времени t_d , полученная при интегрировании уравнений модели на этапе её генерации.

Для вычислительного эксперимента было сгенерировано 1800 моделей, каждая из которых представлялась XML-файлом [\[36\]](#), содержащим матрицу смежности; значения независимых параметров, определяющих ин-

тенсивности переходов; параметры, используемые при численном интегрировании уравнений модели; имена состояний; значения вероятностей $P(t) = (p_{S_0}(t), \dots, p_{S_n}(t))$ в контрольных точках $\{t_d\}_{d=0, \dots, D-1}$.

Таблицы 9 и 10 содержат число сгенерированных моделей для каждого набора параметров и каждого числа состояний.

Программное обеспечение для вычислительного эксперимента реализовано на языке G в среде графического программирования National Instruments LabVIEW. Характеристики системы, работавшей под управлением ОС Microsoft Windows 7 Professional (SP1), следующие:

- материнская плата ASRock Z77 Extreme4-M Socket LGA 1155 (1066 – 2800 МГц);
- четырёхядерный процессор Inter Core i5-3330 с частотой 3000 МГц;
- 2 модуля оперативной памяти Corsair XMS3 4 Гб DDR3 1333 МГц.

4.5 Параметры процедур идентификации

Поскольку при генерации пула данных для вычислительного эксперимента отсутствует информация, необходимая для определения начальных оценок интенсивностей переходов, оценки параметров λ_i^0 на первом шаге представленных выше алгоритмов устанавливались равными заданным константам. Значение параметра γ_1 , определяющего сдвиг от текущих оценок идентифицируемых величин, устанавливалось равным 0,9. Число контрольных моментов времени D было равно 100. По достижении значения Q , удовлетворяющего неравенству $\frac{Q}{(n+1)D} \leq 0,001$, поиск останавливался, а соответствующая модель считалась идентифицированной. Время вычислений, затрачиваемое на каждую модель, ограничивалось 200 секундами. По его истечении поиск останавливался, а соответствующая модель считалась неидентифицированной. В случаях методов

Таблица 9. Число моделей, сгенерированных для каждого числа параметров

Число параметров	Число моделей
1	180
2	180
3	160
4	160
5	140
6	140
7	120
8	120
9	100
10	100
11	80
12	80
13	60
14	60
15	40
16	40
17	20
18	20

Таблица 10. Число моделей, сгенерированных для каждого числа состояний

Число состояний	Число моделей
2	40
3	80
4	120
5	160
6	200
7	240
8	280
9	320
10	360

полной дискретизации и дискретизации двух значимых параметров идентификация проводилась только для моделей, имеющих не более шести параметров.

4.6 Результаты вычислительного эксперимента

После идентификации каждой модели характеристики проведённого процесса вычислений сохранялись в виде XML-файла. Сформированный таким образом набор данных включал название метода идентификации, число состояний, число параметров модели, временные отметки для каждой итерации и другую необходимую информацию.

Созданные файлы использовались для формирования выборки результатов вычислительных экспериментов со следующими характеристиками: число состояний *States* и параметров *Params* модели, идентификационный номер модели *Trial*, евклидово расстояние *Distance* начальных оценок параметров до их значений, заложенных при генерации модели, количество итераций *Iterations*, выполненных до достижения одного из критериев останова, результирующее значение *ResFit* минимизируемого критерия, целевое значение *GoalFit* минимизируемого критерия, булев флаг *MissFitFlag*, определяющий успешность идентификации модели (0 – модель успешно идентифицирована, 1 – модель не идентифицирована), номинальный фактор *Alg* с пятью значениями, указывающими на использованный алгоритм (0 – градиентный метод, 1 – метод полной дискретизации, 2 – метод дискретизации 2-х значимых параметров, 3 – метод дискретизации 3-х значимых параметров, 4 – метод дискретизации 4-х значимых параметров). Полученные данные анализировались в программной среде для статистических вычислений *R*.

В таблице [11](#) приведено количество неидентифицированных моделей для каждого количества состояний, Таблица [12](#) содержит в себе данные о количестве неидентифицированных моделей и количестве параметров.

В таблице [13](#) представлено среднее число итераций для каждого метода и количеств состояний.

Таблица 11. Количество неидентифицированных моделей для каждого метода и количества состояний. 0 – градиентный метод, 1 – метод полной дискретизации, 2 – метод дискретизации 2-х значимых параметров, 3 – метод дискретизации 3-х значимых параметров, 4 – метод дискретизации 4-х значимых параметров.

Метод \ Кол-во состояний	0	1	2	3	4
2	1	0	0		
3	5	7	0	4	2
4	16	14	0	3	3
5	19	14	5	7	1
6	46	18	11	9	4
7	68	21	14	15	15
8	85	21	39	20	11
9	114	18	62	26	19
10	143	27	82	33	40

Таблица 12. Количество неидентифицированных моделей для каждого метода и количества параметров. 0 – градиентный метод, 1 – метод полной дискретизации, 2 – метод дискретизации 2-х значимых параметров, 3 – метод дискретизации трех значимых параметров, 4 – метод дискретизации 4-х значимых параметров.

Кол-во параметров \ Метод	0	1	2	3	4
	1	9	0	–	–
2	6	2	0		–
3	19	17	3	11	
4	37	18	9	5	11
5	43	26	10	7	5
6	39	77	18	11	8
7	39	–	22	7	0
8	41	–	15	6	7
9	47	–	22	11	5
10	42	–	19	12	8
11	33	–	17	14	10
12	32	–	17	10	8
13	24	–	14	5	6
14	32	–	14	8	8
15	22	–	10	1	4
16	18	–	15	7	6
17	7	–	1	1	3
18	7	–	7	3	6

Таблица 13. Среднее количество итераций моделей для каждого метода и количества параметров. 0 – градиентный метод, 1 – метод полной дискретизации, 2 – метод дискретизации 2-х значимых параметров, 3 – метод дискретизации трех значимых параметров, 4 – метод дискретизации 4-х значимых параметров.

Метод \ Кол-во параметров	0	1	2	3	4
1	10,58	3,62	–	–	–
2	239,23	6,22	1,38		–
3	439,68	8,26	11,89	35,80	
4	564,01	10,42	15,76	12,68	10,98
5	522,40	12,62	9,22	14,21	5,65
6	493,98	13,20	11,62	14,00	5,84
7	460,81	–	13,94	6,99	3,27
8	380,44	–	11,37	6,00	5,87
9	359,28	–	10,21	5,82	3,97
10	344,98	–	10,21	6,01	4,77
11	298,03	–	8,34	6,45	4,35
12	268,08	–	10,56	6,35	4,8
13	230,45	–	9,97	5,83	3,93
14	260,73	–	10,08	5,48	4,15
15	234,58	–	9,10	5,73	3,88
16	178,98	–	9,60	6,65	5,55
17	147,70	–	9,10	5,90	4,45
18	138,95	–	9,30	5,90	4,45

Зависимости доли моделей, которые не были идентифицированы из-за превышения заданного лимита времени, от числа параметров моделей при разных методах идентификации представлены на рисунке 18. Эти результаты свидетельствуют о том, что разработанные методы идентификации, за исключением метода полной дискретизации при более чем пяти параметрах, обеспечивают существенное ускорение вычислений по сравнению с рассмотренным классическим градиентным методом.

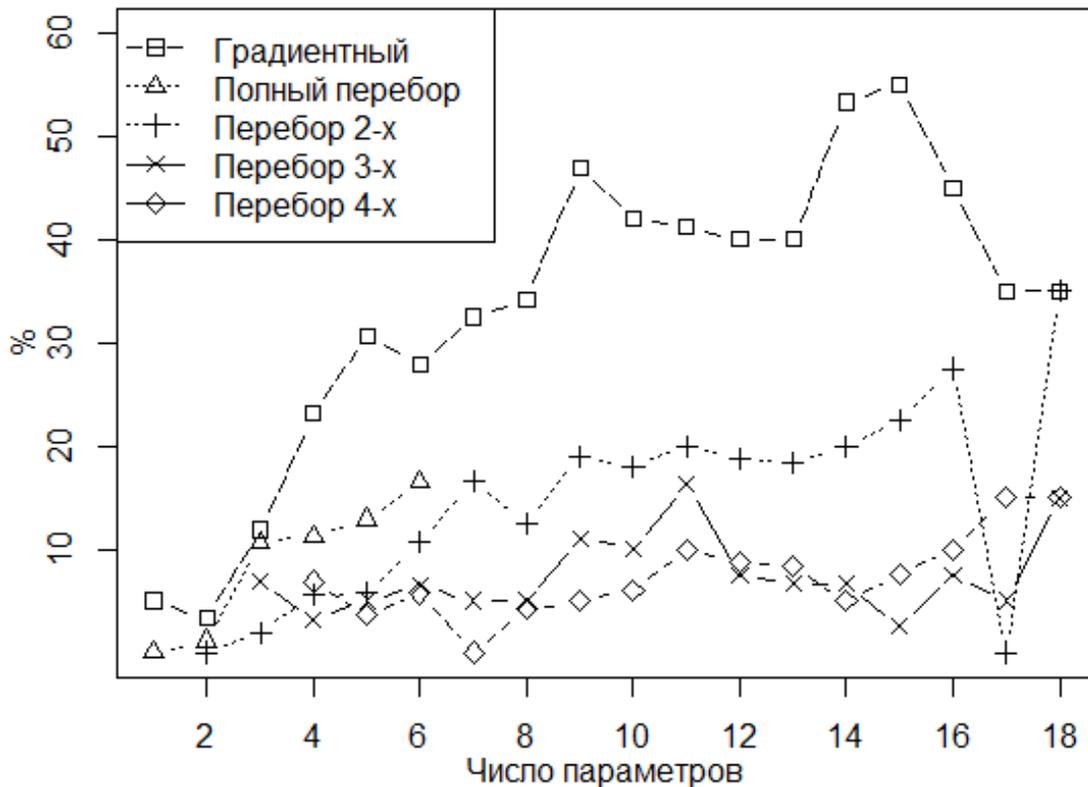


Рис. 18. Зависимости доли неидентифицированных моделей от числа параметров моделей при разных методах идентификации.

На рисунке 19 представлены диаграммы размаха, отражающие зависимости статистических характеристик времени вычислений (размаха, межквартильного интервала и медианы) от числа состояний и используемого метода идентификации (слева) и от числа параметров модели и используемого метода идентификации (справа). Вдоль горизонтальной оси на этом и последующем рисунках указаны составные показатели, первый компонент которых есть число состояний или параметров, а второй – индекс метода (1 – метод полной дискретизации, 2 – метод дискретизации

2-х значимых параметров). Приведённые графики иллюстрируют показательный характер зависимости времени вычислений от числа параметров модели для метода идентификации, основанного на полной дискретизации комбинаций оценок, и близкий к линейному характер такой зависимости в случае дискретизации двух значимых параметров.

Те же самые статистические характеристики для градиентного метода и методов дискретизации трёх и четырёх значимых параметров, обозначенных, соответственно, индексами 0, 3 и 4, представлены на рисунке 20. Динамика роста медианы времени вычислений на приведённых графиках позволяет говорить о том, что градиентный метод наименее подвержен влиянию числа идентифицируемых параметров. Однако, как было показано выше, этот метод в несколько раз чаще, чем другие подходы, не обеспечивает нахождение решения (если число параметров больше восьми, решение не определяется примерно в 34-55% случаев). Это существенно ограничивает практическую применимость градиентного метода и, по-видимому, объясняется известной чувствительностью этого способа решения к выбору начальных приближений.

Также на рисунке 21 отдельно приведены зависимости среднего времени вычислений от числа параметров моделей. При этом в качестве единицы измерения использовано среднее время вычислений в случае дискретизации трёх значимых параметров.

В таблице 14 представлены статистические оценки значимости влияний различных характеристик на время вычислений, полученные для следующей линейной регрессионной зависимости:

$$Time = b_0 + b_1 States + b_2 Params + b_3 Distance + b_4 Alg1 + b_5 Alg2 + b_6 Alg3 + b_7 Alg4 + \varepsilon,$$

где $Time$ – время вычислений, $States$ – число состояний модели, $Params$ – число параметров модели, $Distance$ – евклидово расстояние от начальных оценок параметров до их оптимальных значений, $Alg1$, $Alg2$, $Alg3$, $Alg4$ – бинарные показатели, принимающие значение 1 для результатов, полученных, соответственно, полной дискретизации, дискретизации двух

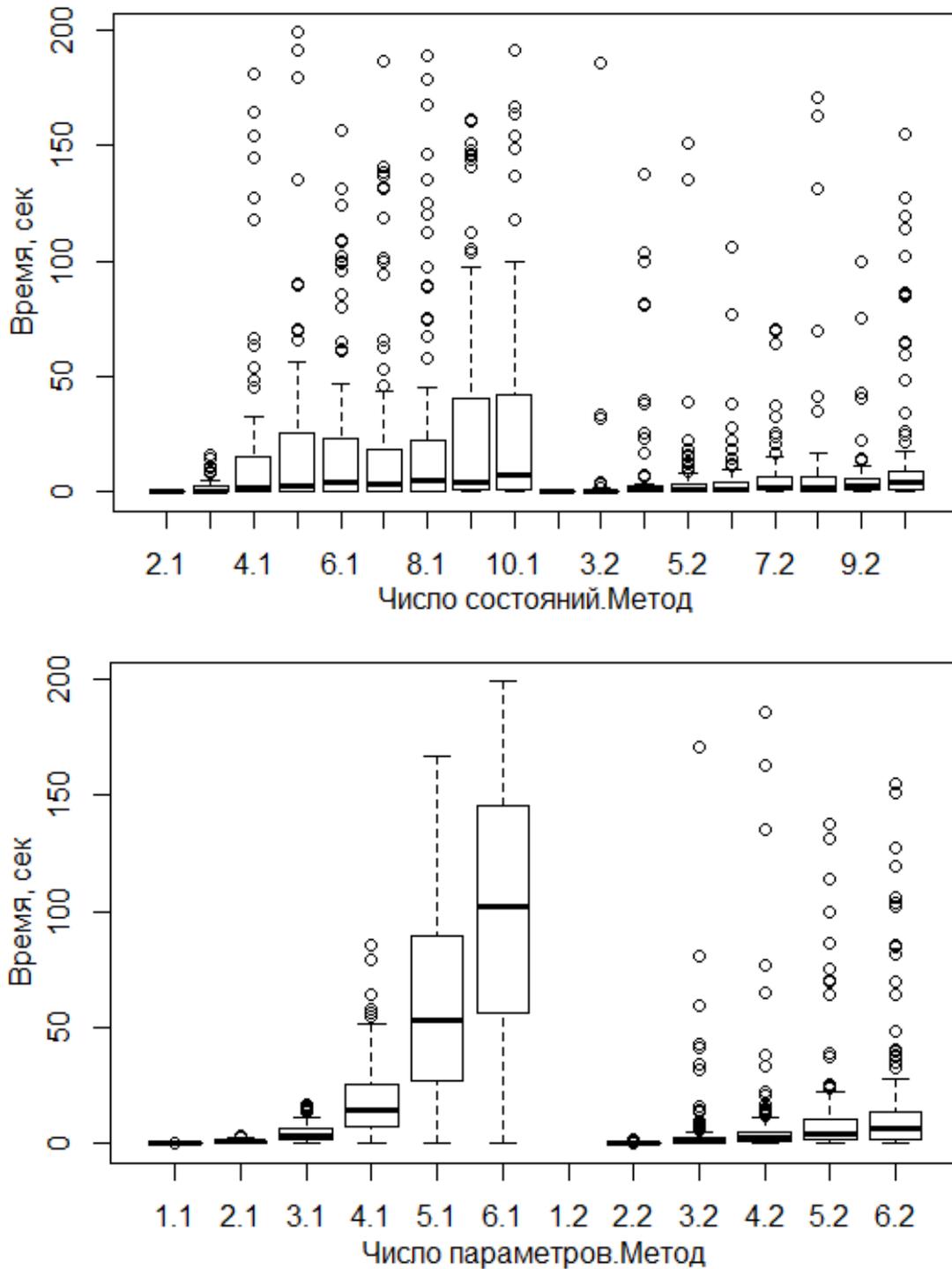


Рис. 19. Зависимости статистических характеристик времени вычислений (размаха, межквартильного интервала и медианы) от числа состояний и параметров и используемого метода идентификации. Вдоль горизонтальной оси указаны составные индексы, первый компонент которых есть число состояний (верхний график) или параметров (нижний график), а второй – указатель метода (1 – метод полной дискретизации, 2 – метод дискретизации двух значимых параметров).

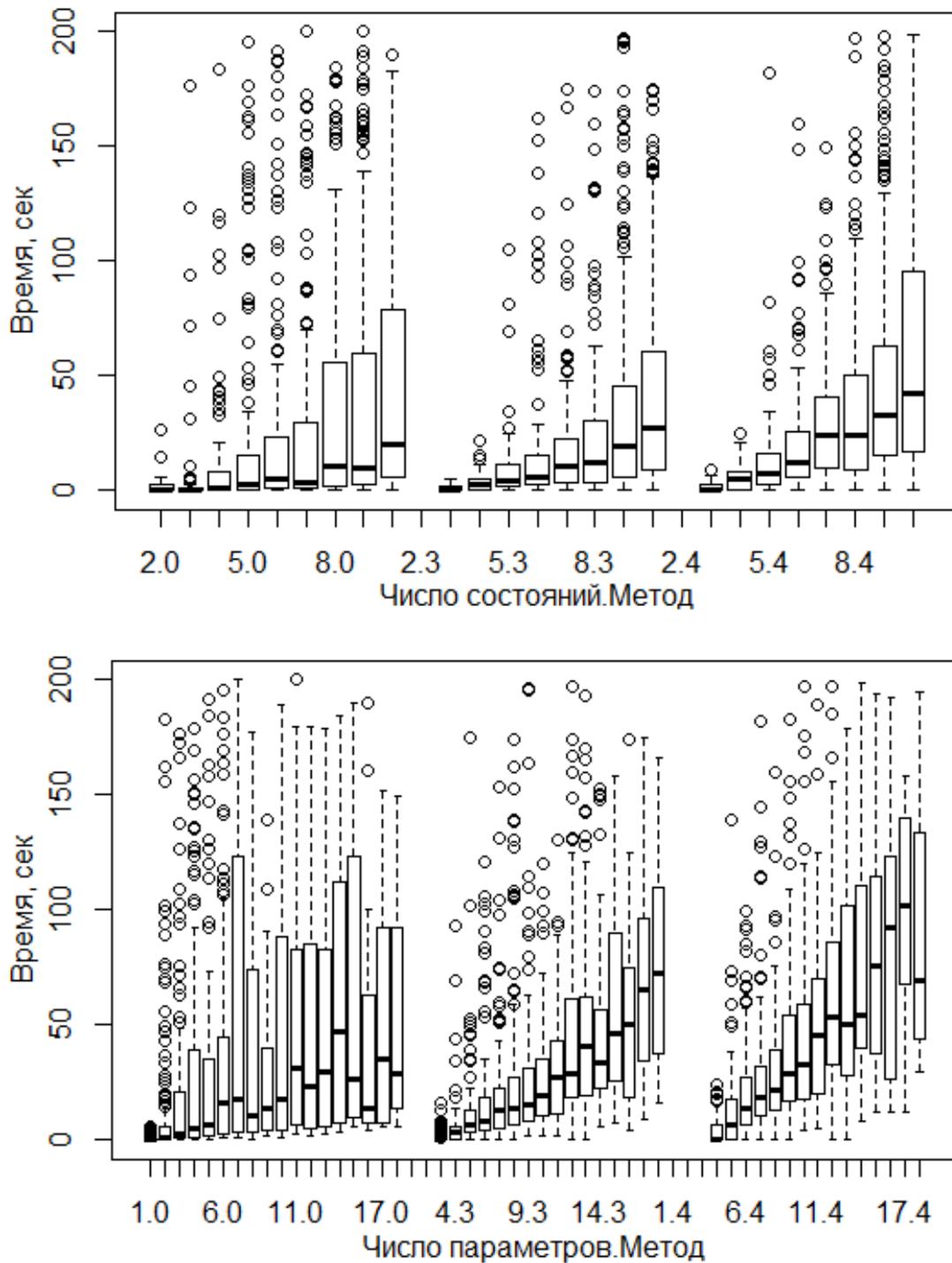


Рис. 20. Зависимости статистических характеристик времени вычислений (размаха, межквартильного интервала и медианы) от числа состояний и параметров и используемого метода идентификации. Вдоль горизонтальной оси указаны составные индексы, первый компонент которых есть число состояний (верхний график) или число параметров (нижний график), а второй – указатель метода (0 – градиентный метод, 3 – метод дискретизации трёх значимых параметров, 4 – метод дискретизации четырёх значимых параметров).

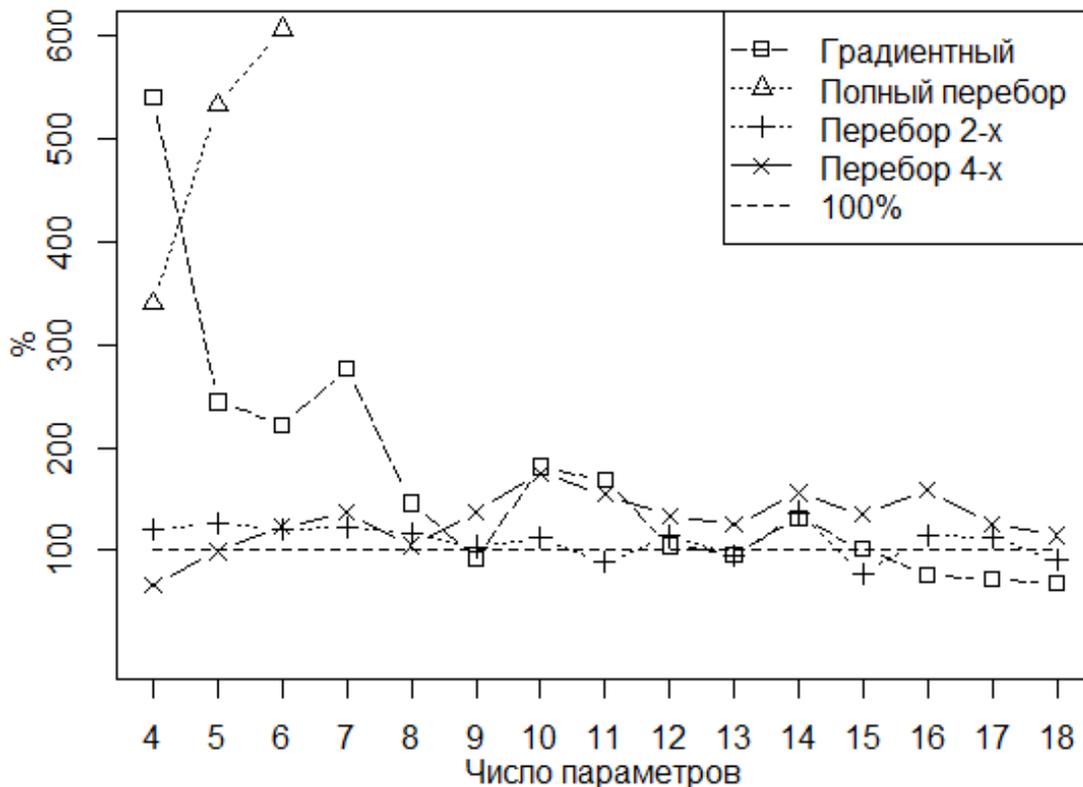


Рис. 21. Зависимости среднего времени вычислений от числа параметров моделей для различных методов идентификации. В качестве единицы измерения использовано среднее время вычислений в случае дискретизации трёх значимых параметров (100%).

значимых параметров, дискретизации трех значимых параметров и дискретизации четырёх значимых параметров, и значение 0 – в противном случае.

Статистические оценки значимости влияний различных характеристик на время вычислений, полученные путём анализа заданной линейной регрессионной зависимости, свидетельствуют о том, что как рассмотренная регрессионная модель в целом, так и все её компоненты статистически значимы. Анализ коэффициентов, соответствующих бинарным показателям, представляющим метод идентификации, позволяет говорить о существенном ускорении вычислений при использовании методов дискретизации значимых параметров по сравнению с классическим градиентным методом [27].

Статистическая проверка различий распределений времени вычисле-

Таблица 14. Статистические оценки значимости влияния различных характеристик на время вычислений, полученные путём анализа заданной линейной регрессионной зависимости.

R^2 (прив.): 0,26; F -статистика: 309,3 при 7 и 6026 ст. св.; $p < 2,2E-16$					
Параметр модели	Оценка	Стандартная ошибка	t -стат.	p -знач.	Статистич. значимость
b_0	-5,94	1,88	-3,15	0,00165	есть
States	1,65	0,24	6,86	7,95E-12	есть
Params	3,95	0,14	27,27	$< 2,0E-16$	есть
Distance	14,35	1,34	10,7	$< 2,0E-16$	есть
Alg1	1,52	1,65	0,91	0,36	нет
Alg2	-17,55	1,45	-12,1	$< 2,0E-16$	есть
Alg3	-21,43	1,5	-14,29	$< 2,0E-16$	есть
Alg4	-12,71	1,56	-8,16	4,21E-16	есть

ний для разных алгоритмов была проведена с использованием двух критериев:

- непараметрического двухвыборочного критерия Колмогорова-Смирнова;
- непараметрического U-критерия Манна-Уитни (Уилкоксона-Манна-Уитни).

Двухвыборочный тест Колмогорова-Смирнова позволяет проверить однородность двух выборок. Статистика критерия измеряет различие между двумя функциями распределения, построенными по выборкам. Подсчет значения статистики производится по формуле:

$$\lambda_{\text{эмп.}} = d_{\text{max}} * \sqrt{\frac{n_1 * n_2}{n_1 + n_2}},$$

где d_{max} – наибольший модуль разностей относительных частот. Формируются следующие гипотезы:

$$H_0 : F(x) = G(x) \forall x$$

$$H_1 : F(x) \neq G(x)$$

Критерий Манна-Уитни позволяет оценить значимость различий между двумя выборками по количественно измеренному признаку. Критерий является ранговым, U -статистика критерия вычисляется по формуле:

$$U = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - T_x,$$

где n_1, n_2 – объемы выборок, T_x – сумма рангов. Для случаев, когда $n_1 > 3$ и $n_1 + n_2 \geq 20$ нормальное приближение для критерия является достаточно точным [5]. Формируются следующие гипотезы:

$$H_0 : P(X > Y) = P(Y > X)$$

$$H_1 : P(X > Y) > P(Y > X)$$

Таблица 15. Результаты проверки однородности распределения времен вычислений и сравнения средних значений времени вычисления для разработанных методов и метода градиентного спуска.

	Среднее время (с)	λ -стат.	p -знач.	U -стат.	p -знач.
Метод градиентного спуска	55.502				
Метод дискретизации 2-х значимых параметров	29.558	0.19397	0.0000	499982.5	0.00000
Метод дискретизации 3-х значимых параметров	24.641	0.21917	0.0000	448351.0	0.00001
Метод дискретизации 4-х значимых параметров	27.551	0.20711	0.0000	381343.0	0.00916

Результаты процедуры применения тестов приведены в таблице 15.

Результаты проведенных статистических тестов позволяют говорить о том, что при выполнении процедуры идентификации с использованием методов дискретизации значимых параметров время, затрачиваемое на вычисления, статистически значимо меньше аналогичного времени для градиентного метода.

Представленные в главе 4 результаты свидетельствуют о том, что разработанные методы идентификации имеют преимущества перед классическим градиентным методом при числе идентифицируемых параметров, не превышающем 9, при этом статистически значимые различия средних времени вычислений подтверждены примененными статистическими тестами.

Заключение

ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ, ВЫНОСИМЫЕ НА ЗАЩИТУ

- 1) Методика компьютерного тестирования с участием или без участия эксперта (инструктора), особенностями которой являются:
 - а) математическая модель адаптивного тестирования, представленная марковским процессом с дискретными состояниями и непрерывным временем, учитывающая временную динамику выполнения заданий испытуемым;
 - б) механизм адаптивного предъявления тестовых заданий, базирующийся на вычислении показателей дифференцирующей силы непредъявленных тестов;
- 2) Вероятностная модель процесса прохождения отдельного тестового задания, позволяющая учесть динамику способностей испытуемых и трудностей заданий во время их выполнения и предполагающая построение количественных оценок приведенных характеристик с помощью модифицированной функции Раша.
- 3) Численные методы идентификации вероятностных моделей. Эффективность методов обоснована проведенной серией вычислительных экспериментов.
- 4) Комплекс программ, обеспечивающий проведение эффективной процедуры оценки навыков и способностей как с использованием классических тестов, так и при управлении сложными техническими системами.
- 5) Метод оптимизации последовательности предъявления тестовых заданий. Метод основан на анализе показателей дифференцирующей силы различных тестов, позволяя на основе идентифицированных

по выборке параметров процедур тестирования строить близкую к оптимальной последовательность предъявления тестовых заданий, обеспечивающую наибольшую условную вероятность принадлежности испытуемых к предусмотренным категориям .

Список литературы

1. Аббасов М. Э. Методы оптимизации: Учебное пособие, — СПб.: Издательство «ВВМ», 2014. — 64 с.
2. Аванесов В.С. Математические модели педагогического измерения. М.: Исследовательский центр проблем качества подготовки специалистов, 1994. — 26 с.
3. Блюм П. LabVIEW. Стиль программирования. Москва, ДМК Пресс, 2009, 400 с.
4. Вадзинский Р. Статистические вычисления в среде Excel. Библиотека пользователя. — СПб.: Питер, 2016. — 608 с.: ил.
5. Ван дер Варден Б.Л. Математическая статистика. — М.: Издательство иностранной литературы, 1960. 435 с.: ил.
6. Васильев Ф. П. Методы оптимизации. — М.: Факториал Пресс, 2002. 824 с.: ил.
7. Глущенко А. Ю. Разработка метода адаптивного управления обучением по индивидуальной образовательной траектории: дис. ... канд. техн. наук. ФГОУ ВПО «Государственный технологический университет «Московский институт стали и сплавов» Москва, 2009.
8. Горюшкин Е.И. Использование нейросетевых технологий в адаптивном тестировании по информатике в вузе: дис. ... канд. пед. наук, ГОУ ВПО «Курский государственный университет», Москва, 2009.
9. Гублер Е. В., Генкин А. А. Применение непараметрических критериев статистики в медико-биологических исследованиях. — Л., 1973.
10. Данг Х. Ф., Камаев В. А., Шабалина О. А. Метод разработки алгоритмов адаптивного тестирования // Известия ВолгГТУ. 2012. №13. URL: <http://cyberleninka.ru/article/n/metod-razrabotki-algoritmov-adaptivnogo-testirovaniya> (дата обращения: 11.10.2016).

11. Думин П.Н., Куравский Л.С. Анализ психологических аспектов игры на основе модифицированной функции Раша. // Нейрокомпьютеры: разработка, применение. 2016. №4. С. 43-47.
12. Думин П.Н. Концепция системы поддержки принятия решений // Нейрокомпьютеры: разработка, применение. 2016. №9. С. 14-21.
13. Думин П.Н. Концепция системы поддержки принятия решений // Нейрокомпьютеры: разработка, применение. 2016. №9. С. 14-21.
14. Думин П.Н. Решение многокритериальной задачи выбора тестового задания в системе поддержки принятия решений. // Нейрокомпьютеры: разработка, применение. 2014. № 10. С. 47 – 51.
15. Кибзун А. И., Панарин С. И. Стохастическая модель модифицируемости системы дистанционного обучения // Вестник Московского авиационного института, 2009, Т. 16, №7, с. 76-79.
16. Кибзун А. И., Иноземцев А. О., «Оценивание уровней сложности тестов на основе метода максимального правдоподобия» // Автоматика и телемеханика, 2014, № 4, с. 20–37
17. Кирий В.Г., Ульянов Д.А., Применение марковской модели для оценки уровня знаний при адаптивном тестировании // Вестник Ир-ГТУ. Иркутск. 2004, №3 (19), 21-26.
18. Крамер Г. Математические методы статистики. – М.: Мир, 1976. – 648 с.
19. Куравский Л.С., Баранов С.Н., Юрьев Г.А. Синтез и идентификация скрытых марковских моделей для диагностики усталостного разрушения. - Нейрокомпьютеры: разработка и применение, №12, 2010, с. 20-36.
20. Куравский Л.С., Марголис А.А., Мармалюк П.А., Юрьев Г.А., Думин П.Н. Обучаемые марковские модели в задачах оптимизации поряд-

ка предъявления психологических тестов. – Нейрокомпьютеры: разработка и применение. 2013. №4. С. 28–38.

21. Куравский Л.С., Мармалюк П.А., Алхимов В.И., Юрьев Г.А. Новый подход к построению интеллектуальных и компетентностных тестов. // Моделирование и анализ данных. 2013. №1. С. 4–28.
22. Куравский Л.С., Мармалюк П.А., Юрьев Г.А., Думин П.Н. Идентификация марковских процессов по статистическим данным. Нейрокомпьютеры: разработка, применение. 2015. № 5. С. 42 – 47.
23. Куравский Л.С., Мармалюк П.А., Юрьев Г.А., Думин П.Н., Панфилова А.С. Вероятностное моделирование процесса выполнения тестовых заданий на основе модифицированной функции Раша. Вопросы психологии. 2015. № 4. С. 109-118.
24. Лазарсфельд П.Ф. Измерение в социологии // Американская социология. М.: Прогресс, 1972.
25. Ларин Р.М., Плясунов А.В., Пяткин А.В. Методы оптимизации. Примеры и задачи: Учеб. Пособие, Новосибирский университет, Новосибирск, 2003, 115 с.
26. Лотов А. В., Поспелова И. И. Многокритериальные задачи принятия решений: Учебное пособие. – М.: МАКС Пресс, 2008. – 197 с.
27. Мармалюк П.А., Юрьев Г.А., Куравский Л.С., Думин П.Н. Результаты вычислительного эксперимента по сравнению методов идентификации марковских процессов. // Нейрокомпьютеры: разработка, применение. 2015. № 9. С. 44-50.
28. Мастицкий С.Э., Шитиков В.К. Статистический анализ и визуализация данных с помощью R. 2014. – Электронный ресурс: <http://r-analytics.blogspot.com> (дата обращения: 05.09.2016).
29. Мельникова М.Б. Теория и практика конструирования педагогических тестов. М.: Логос, 2002. 410 с.

30. Михеев О.В. Математические модели педагогических измерений // Педагогические измерения, 2004, №2. -С.75-88.
31. Наумов А.В., Иноземцев А.О. Алгоритм формирования индивидуальных заданий в системах дистанционного обучения // Вестник компьютерных и информационных технологий. – 2013. – Т. 1, № 6. – С. 35-42.
32. Наумов, А.В. , Джумурат А.С., Иноземцев А.О. Система дистанционного обучения математическим дисциплинам CLASS.NET // Вестник компьютерных и информационных технологий. – 2014. - Т. 1, № 10. – С. 36-40.
33. Наумов А.В. , Мхитарян Г.А. О задаче вероятностной оптимизации для ограниченного по времени тестирования / А.В. Наумов, Г.А. Мхитарян // Автоматика и телемеханика. – 2016. – № 9. – С. 124-135.
34. Николенко С., Тулупьев А., Сироткин А. Байесовские сети. Логико-вероятностный подход. М.: Наука. 2006. 600 с.
35. Овчаров Л.А. Прикладные задачи теории массового обслуживания. – М.: Машиностроение, 1969. – 324 с.
36. Одиночкина С.В. Основы технологий XML — СПб: НИУ ИТМО, 2013. — 56 с.
37. Равен Д. Педагогическое тестирование: проблемы, заблуждения, перспективы // Пер.с англ. М.: «Когито-Центр», 2001. 144 с.
38. Саати Т.Л. Элементы теории массового обслуживания и её приложения. – М.: ЛИБРОКОМ, 2010. – 520 с.
39. Савченко Е.Ю. Применение модифицированных алгоритмов обучения нейронных сетей в задачах адаптивного тестирования. // Научный аспект. 2012. № 4.

40. Сидоренко Е.В. Методы математической обработки в психологии. – СПб.: ООО «Речь», 2000. – 350 с., ил.
41. Соловьев В. И. Методы оптимальных решений: Учебное пособие. М.: Финансовый университет, 2012. 364 с.
42. Сологуб Г. Б. Построение и использование байесовской сети для моделирования знаний студента в интеллектуальной системе тестирования. // Компьютерные инструменты в образовании. 2012. № 2.
43. Строганов Д. В., Свободин В. Ю., Ягудаев Г. Г., Сычева Н. В. Оценка эффективности процедур адаптивного тестового контроля. // Наука и образование. 2012. № 11. 40-48 сс.
44. Суранов А. Я. LabVIEW 8.20. Справочник по функциям. Санкт-Петербург, ДМК Пресс, 2007 г. – 536 с.
45. Тихонов В.И., Шахтарин Б.И., Сизых В.В. Случайные процессы. Примеры и задачи. /Т.5. Оценка сигналов, их параметров и спектров. Основы теории информации. – М.: Горячая линия–Телеком, 2009. – 400 с.
46. Тревис Дж., Кринг Дж. LabVIEW для всех. М.: Изд. ПриборКомплект, 2008. – 880 с.
47. Трусов П.В. Введение в математическое моделирование. М.: Логос, 2007. – 440 с.
48. Тюменева Ю.А. Психологическое измерение. – М.: Аспект-Пресс, 2007.
49. Федосов В. П. Цифровая обработка звуковых и вибросигналов в LabVIEW (+ DVD-ROM). Санкт-Петербург, ДМК Пресс, 2009. – 1296 с.

50. Фоминова Н.С. Разработка и анализ стохастической и аппроксимационной моделей адаптивного тестирования для информационно-управляющих систем: дис. ... канд. тех. наук. Московский государственный институт электронной техники, Москва, 2010.
51. Формалев В. Ф., Ревизников Д. Л. Численные методы. – М.: ФИЗМАТЛИТ, 2004. – 400 с.
52. Хайкин С. Нейронные сети. Полный курс. 2-ое изд. 2006. — М., СПб., Киев: Вильямс.
53. Хлопотов М.В. Применение байесовской сети при построении моделей обучающихся для оценки уровня сформированности компетенций // Наукоедение. 2014. № 5 (24).
54. Челышкова М.Б. Теория и практика конструирования педагогических тестов. М.: Логос, 2002. – 432 с.
55. Шан М. А. Автоматизированная информационная система адаптивного обучения на основе компетентностного подхода: дис. ... канд. техн. наук. Кабардино-Балкарский государственный университет им. Х. М. Бербекова, Нальчик, 2014.
56. Шмелёв А. Г., Столин В. В. Практикум по психодиагностике. Дифференциальная психометрика. М.: Изд-во Моск. ун-та, 1984. – 151 с.
57. Юрьев Г.А. Математическая модель интерпретации результатов компьютерного тестирования с использованием марковских сетей: дис. ... канд. ф.-м. наук, ГБОУ ВПО МГППУ, Москва, 2013.
58. Amponsah S.K., Darkwah K. F., Inusah A. Logistic preference function for preference ranking organization method for enrichment evaluation (PROMETHEE) decision analysis. // African Journal of Mathematics and Computer Science Research Vol. 5(6), pp. 112-119, 15 March, 2012.

59. Andreychenko A., Mikeev L., Spieler D., Wolf V. Parameter Identification for Markov Models of Biochemical Reactions // Computer Aided Verification. Vol. 6806 of the series Lecture Notes in Computer Science, pp 83-98.
60. Baker F.B. The Basics of Item Response Theory. ERIC Clearinghouse on Assessment and Evaluation, University of Maryland, College Park, MD, 2001.
61. Bergan J.R. Rasch Versus Birnbaum: New Arguments in an Old Debate // Assessment Technology, Incorporated. [Электронный ресурс] URL: <http://www.ati-online.com/pdfs/researchK12/RaschVsBirnbaum.pdf> (дата обращения: 02.10.2014).
62. Billingsley P. Probability and Measure. Wiley series in probability and mathematical statistics, 1995, 519 p.
63. Binet AND., Simon T.H.E. The Development of Intelligence in Young Children. Vineland, NJ: The Training School, 1916.
64. Birnbaum A. Some Latent Trait Models and Their Use in Inferring an Examinee's Ability / In: F.M. Lord and M.R. Novick. Statistical Theories of Mental Test Scores. Reading, Mass: Addison - Wesley, 1968. - 568p.
65. Boatemaa B. Multi Criteria Ranking of Telecommunication Carriers using Promethee Method. Thesis submitted to the department of mathematics, Kwame Nkrumah University of science and technology, 2012, 120 p.
66. Bollen K.A. Structural Equations with Latent Variables. P. 182., N - Y, Wiley & Sons, 1989. - 514p.
67. Downey A. B. Think Python. — O'Reilly Media, 2012, 300 p.
68. Fedyanin D. On Parameter Identification Methods for Markov Models Applied to Social Networks.
69. Fulop J. Introduction to Decision Making Methods. 2014.

70. Goodman, L. A. Exploratory latent structure analysis using both identifiable and unidentifiable models. // *Biometrika*, № 61, 1974, 167-171 pp.
71. Gregory R.J. *Psychological testing: History, principles, and applications* (5th edition). - New York: Pearson, 2007.
72. Kuravsky L. S., Malykh S. B. Application of Markov models for analysis of development of psychological characteristics. - *Australian Journal of Educational & Developmental Psychology*, 2004, Vol 2, pp 29-40.
73. Kuravsky L. S., P. A. Marmalyuk, G. A. Yuryev, P. N. Dumin. A numerical technique for the identification of discrete-state continuous-time Markov models. *Applied Mathematical Sciences*, Vol. 9, 2015, no. 8, 379-391.
74. Kuravsky L.S. and Baranov S.N. Condition monitoring of the structures suffered acoustic fatigue failure and forecasting their service life. - *Proc. Condition Monitoring 2003*, Oxford, United Kingdom, pp. 256-279, July 2003.
75. Kuravsky L.S. and Baranov S.N. Neural networks in fatigue damage recognition: diagnostics and statistical analysis. - *Proc. 11th International Congress on Sound and Vibration*, St.-Petersburg, Russia, pp. 2929-2944, July 2004.
76. Kuravsky L.S. and Baranov S.N. The concept of multifactor Markov networks and its application to forecasting and diagnostics of technical systems. – In: *Proc. Condition Monitoring 2005*, Cambridge, United Kingdom, pp. 111-117, July 2005.
77. Kuravsky L.S., Baranov S.N. and Yuryev G.A. Synthesis and identification of hidden Markov models based on a novel statistical technique in condition monitoring. – In: *Proc. 7th International Conference on Condition Monitoring & Machinery Failure Prevention Technologies*, Stratford-upon-Avon, England, June 2010. – 23 pp.

78. Kuravsky L.S., Margolis A.A., Marmalyuk P.A., Panfilova A.S., Yuryev G.A., Dumin P.N. A Probabilistic Model of Adaptive Training. // Applied Mathematical Sciences. 2016. Vol. 10, 2016, no. 48, 2369 – 2380.
79. Kuravsky L.S., Marmalyuk P.A., Baranov S.N., Alkhimov V.I., Yuryev G.A. and Artyukhina S.V. A New Technique for Testing Professional Skills and Competencies and Examples of its Practical Applications. – Applied Mathematical Sciences. Vol. 9. 2015. No 21. Pp. 1003–1026.
80. Kuravsky L.S., Marmalyuk P.A., Yuryev G.A., Dumin P.N., Panfilova A.S. Probabilistic modeling of a testing procedure. // Applied Mathematical Sciences. 2015. T. 9. № 81-84. C. 4053-4066.
81. Lawley D.N. On Problems Connected with Item Selection and Test Construction // Proceedings of the Royal Society of Edinburgh. Section A Mathematical and Physical Sciences. 43 v. LXI, part III, p. 273 - 287, 1943.
82. Lazarsfeld P.F. The interpretation and computation of some latent structures. // Measurement and Prediction, N-Y, John Wiley and Sons, 1950.- p.415-472, 1950. Linacre J. M., Wright B. D. (1998). BIGSTEPS: Rasch-model computer programs. Chicago, IL: Mesa Press.
83. Linacre J. M., Wright B. D. (1998). BIGSTEPS: Rasch-model computer programs. Chicago, IL: Mesa Press.
84. Lord F.M. A Theory of Test Scores. Psychometric Monographs, 1952, 7. Richmond. Publ. by Psychometric Society. – 84 pp.
85. Matloff N. Art of R Programming - A Tour of Statistical Software Design, 2011. Mills C. N., Stocking M. L. (1996). Practical issues in large-scale computerized adaptive testing. Applied Measurement in Education, 9(4), 287-394.
86. Mills C. N., Stocking M. L. (1996). Practical issues in large-scale

- computerized adaptive testing. *Applied Measurement in Education*, 9(4), 287-394.
87. Milovanovic I. *Python Data Visualization Cookbook*. – Packt Publishing, 2015, 302 p.
 88. Pearl J. *Probabilistic Reasoning in Expert Systems: Networks of Plausible Inference*. San Francisco: Morgan Kaufmann, 1988.
 89. Podvezko V., Podviezko A. Dependence of multi-criteria evaluation result on choice of preference functions and their parameters. // *Baltic Journal on Sustainability*, No. 16(1), 2011, 143–158 pp.
 90. Rasch, G. *Probabilistic models for some intelligence and attainment tests*. // Copenhagen, Danish Institute for Educational Research, expanded edition (1980) with foreword and afterword by B.D. Wright. Chicago: The University of Chicago Press. 1960/1980).
 91. Rossant C. *IPython Interactive Computing and Visualization Cookbook*. – Packt Publishing, 2014, 520 p.
 92. Snijders Tom A.V. *Statistical Models for Social Networks* // *Annual Review of Sociology* Vol. 37, 131-153, 2011.
 93. Tomić V., Marinković Z., Janošević D. PROMETHEE method implementation with multi-criteria decisions. // *Mechanical Engineering*, Vol. 9, No 2, 2011, pp. 193 – 202 pp.
 94. Verzani J. *Getting Started with RStudio*. 2011.
 95. Wickham H. *Ggplot2 - Elegant Graphics for Data Analysis*. 2009.
 96. Wright B.D., Masters G.N. *Rating scale analysis. Rasch measurements*. – Chicago: MESA Press, 1982.
 97. Wright B.D., Stone M.N. *Best Test Design*. - Chicago: MESA Press, 1979.