ВЕРОЯТНОСТНЫЙ МЕТОД ФИЛЬТРАЦИИ АРТЕФАКТОВ ПРИ АДАПТИВНОМ ТЕСТИРОВАНИИ

КУРАВСКИЙ Л. С., *Московский городской психолого-педагогический университет, Москва*
ЮРЬЕВ Г. А., *Московский городской психолого-педагогический университет, Москва*

В статье представлен метод фильтрации результатов адаптивного тестирования, построенного на использовании обучаемых структур в форме марковских моделей с непрерывным временем. Устранение артефактов, обусловленных различными формами некорректного целенаправленного вмешательства в процедуру испытаний, выполняется на основе сравнения наблюдаемых и прогнозируемых результатов ответов на вопросы с помощью фильтра Калмана, адаптированного для решения рассматриваемой задачи.

Ключевые слова: адаптивное тестирование, марковские модели, фильтр Калмана.

1. Введение

Компьютерное тестирование в настоящее время широко используется в медицине, психологии и образовании с целью диагностики, определения уровня компетенций и пригодности испытуемых для выполнения тех или иных функций, включая контроль качества обучения. Качество тестирования и достоверность его результатов в значительной степени зависят от технологий проведения тестов, которые в последние десятилетия стали предметом активных научных исследований.

В первое время тесты строились на основе классической модели тестирования (Карданова, 2008; Тюменева, 2007; Gregory, 2007; Gulliksen, 1950), в основе которой лежит теория погрешности измерений, заимствованная из физики: полагалось, что измеряемые характеристики имеют некоторые «истинные» значения, искажаемые случайными и систематическими погрешностями. Этот подход получил определенное распространение, однако его практическому применению препятствует ряд существенных недостатков:

- возникают проблемы при сравнении сходных особенностей тестируемых, выявленных с помощью разных методик;
- не решается проблема валидности;
- тестовые баллы становятся недостаточно надежными в областях экстремальных значений;
- технология в целом недостаточно надежна и универсальна.

Для преодоления указанных проблем была разработана новая технология тестирования, основанная на латентно-структурном анализе и названная *теорией ответов на вопросы* (Item Response Theory – IRT)[1] (Тюменева, 2007; Baker, 2001). В ней реализована концепция *адаптивного тестирования*, согласно которой тестируемому с определенной текущей расчетной оценкой уровня знаний или способностей на каждом шаге тестирования вы-

[1] В русскоязычной литературе также используются и другие варианты ее названия: *стохастическая теория тестов, математическая теория измерений, современная теория тестирования, теория латентных черт, теория характеристических кривых заданий, теория моделирования и параметризации педагогических тестов* и т. д.

числяются и предлагаются задания определенной сложности. Основная концепция новой теории, предложенная Г. Рашем в 1960 году (Rasch, 1980), предполагает, что вероятность правильного ответа на задание определяется разностью уровня знаний или способностей и трудности теста. В зависимости от условий прикладной задачи на практике используются и другие, более сложные модели, построенные на базе данной концепции (Аванесов, 2003; Rasch, 1980; Wright, Stone, 1979; Wright, Masters, 1982).

При применении технологии IRT возникают следующие проблемы:

- «статичность» оценок: игнорирование того факта, что результат тестирования вследствие усталости испытуемых и других факторов может, вообще говоря, существенно изменяться со временем, принимая различные значения в процессе сеанса тестирования;
- невозможность учета времени, затрачиваемого на решение тестовых задач, при построении расчетных оценок;
- необходимость выполнения достаточно большого числа заданий для получения оценок с приемлемой точностью;
- сложность вычисления распределения вероятностей возможных результатов теста, что необходимо для оценки их надежности;
- сравнительно сложная для практической реализации процедура оценки точности результата, связанная с применением метода максимального правдоподобия и расчетом доверительных интервалов.

Указанные проблемы делают актуальной разработку новых технологий тестирования. В этой работе рассматриваются новые аспекты применения разработанного ранее авторами подхода к адаптивному тестированию (Куравский, Баранов, 2001, 2002; Куравский и др., 2005; Куравский и др., 2003; Куравский и др., 2010; Куравский и др., 2011; Куравский, Юрьев, 2011 а; Куравский, Юрьев, 2011 б; Kuravsky, Malykh, 2004; Kuravsky, Baranov, 2003, 2004, 2005; Kuravsky et al., 2010; Kuravsky et al., 2011), построенного на использовании *обучаемых структур в форме марковских моделей* с дискретным и непрерывным временем. Его особенностями, обеспечивающими преимущества перед аналогичными способами тестирования, являются:

- выявление и использование при построении расчетных оценок временной динамики изменения способности справляться с заданиями теста;
- возможность учета при построении расчетных оценок времени, затрачиваемого на решение тестовых задач;
- возможность исследования временной динамики знаний или способностей как в дискретной, так и в непрерывной временной шкале;
- меньшее по сравнению с другими подходами количество заданий, которые следует предъявлять испытуемому для получения оценок знаний или способностей с заданной точностью, что ускоряет процесс тестирования;
- получение распределения вероятностей возможных результатов теста в качестве конечного результата;
- развитая техника идентификации параметров моделей.

Одной из наиболее серьезных проблем, возникающих в процессе тестирования, является появление в истории ответов испытуемого искажающих результаты *артефактов*, обусловленных подсказками, угадыванием и другими формами некорректного целенаправленного вмешательства в процедуру испытаний. Представленная выше технология адаптивного тестирования позволяет бороться с этими явлениями, устраняя артефакты на основе сравнения наблюдаемых и прогнозируемых результатов ответов на вопросы для разных уровней способностей испытуемых. В качестве инструмента для сопоставления в данной работе предлагается использовать *фильтр Калмана* (Тихонов и др., 2009; Шахтарин, 2010) – нестационарную систему с обратной связью, включающую в себя как составную часть формирующий фильтр, воспроизводящий идеализированную модель поведения.

Выбор фильтра Калмана для устранения артефактов тестирования среди близких по содержанию подходов является оптимальным решением, поскольку он наилучшим образом согласуется с принятой концепцией адаптивного тестирования и контекстом ее использования. В частности, этот фильтр:

- в отличие от фильтра Винера способен обрабатывать текущую информацию об ответах испытуемого в реальном времени, формируя свои оценки сразу же после получения очередного ответа и не требуя полного протокола тестирования, который недоступен до завершения всей процедуры ответов на вопросы;
- в отличие от фильтра Стратоновича использует только линейные методы оценки, наилучшим образом согласующиеся с применяемой линейной дифференциальной моделью адаптивного тестирования, и не приводит к неоправданному усложнению процесса решения;
- в отличие от фильтра Льюинбергера учитывает ошибки наблюдений и обеспечивает оптимальные оценки.

Далее кратко представлен новый подход к адаптивному тестированию, основанный на использовании марковских моделей, поставлена задача фильтрации артефактов с помощью фильтра Калмана и рассмотрены особенности ее решения.

2. Марковские модели адаптивного тестирования

2.1. *Структура и математическое описание применяемых марковских моделей с непрерывным временем. Процедура оценки знаний или способностей*

Оценка вероятностей различных уровней способностей проводится по результатам тестирования с использованием параметрических математических моделей, описывающихся *марковскими случайными процессами с дискретными состояниями и непрерывным и дискретным временем* (Овчаров, 1969; Саати, 2010). Дальнейшее изложение относится только к моделям с непрерывным временем. Непосредственно наблюдаемой величиной является трудность выполняемого теста, измеряемая в логитах. Допустимый диапазон значений этой величины делится на несколько интервалов, каждый из которых рассматривается как отдельное состояние x_i, i=0,1,...,n, в котором тестируемый может находиться с некоторой вероятностью, переходя из одного состояния в другое по определенным правилам. Длина указанных интервалов определяет разрешающую способность оценок, получаемых в процессе тестирования. В свою очередь, число состояний определяется желаемой разрешающей способностью оценок и доступным объемом выборки [2].

---

[2] Рассматривая непрерывно изменяющуюся характеристику как дискретную величину, мы теряем часть информации (это имеет место при любой идеализации). Однако эти потери несущественны в случае достаточно больших выборок, когда мы имеем возможность устанавливать длину интервалов состояний так, чтобы она не превышала ошибок измерений.

Как трудности заданий, так и способности тестируемых измеряются в единой безразмерной *шкале логитов*, выражающей соотношение долей правильных и неправильных ответов. Перевод в шкалу логитов осуществляется по формуле:

$$C = ln\frac{r}{1-r}\ ,$$

где C – значение в шкале логитов, r – вероятность правильного выполнения задания. В случае оценки трудности этот параметр характеризует возможность выполнения определенного задания для всего множества тестируемых, а в случае оценки способностей – результаты определенного тестируемого для всего множества допустимых заданий. Статистические приближения указанных величин получаются после замены в приведенной формуле вероятности r на ее выборочные оценки.

Если обозначить верхнюю и нижнюю границы диапазона возможных значений трудности тестов как D_{bot} и D_{top}, состояние x_0 будет соответствовать интервалу от D_{bot} до $D_{bot}+(D_{top}-D_{bot})/(n+1)$, состояние x_1 – интервалу от $D_{bot}+(D_{top}-D_{bot})/(n+1)$ до $D_{bot}+2(D_{top}-D_{bot})/(n+1)$ и т. д.

Модели для описания динамики этих переходов представляются ориентированными графами, в которых вершины[3] соответствуют состояниям, а дуги[4] соответствуют переходам.

В случае моделей с непрерывным временем процесс тестирования может рассматриваться как случайное блуждание по графу с переходами из одного состояния в другое согласно направлениям дуг. Эти переходы мгновенны и происходят в случайные моменты времени.

Предполагается, что для них выполняются следующие два свойства *пуассоновских потоков событий*:

- *ординарность* (поток называется ординарным, если вероятность появления двух и более событий в течение малого интервала времени намного меньше, чем вероятность появления за это же время одного события);

- *независимость приращений* (это свойство означает, что количества событий, попадающих в два непересекающихся интервала, не зависят друг от друга).

Можно показать, что в рассматриваемых потоках число событий X, попадающих в любой временной интервал длины τ, начинающийся в момент t, распределено согласно *закону Пуассона*:

$$P_{t,\tau}(X=m)=\frac{a(t,\tau)^m}{m!}e^{-a(t,\tau)}\ ,$$

где $P_{t,\tau}(X=m)$ – вероятность появления m событий в течение рассматриваемого интервала, $a(t,\tau)$ – среднее число событий, попадающих в интервал длины τ, начинающийся в момент времени t. Далее будут рассматриваться только *стационарные потоки* (в которых $a(t,\tau)=\eta\tau$, $\eta=const$). Параметр η называется *интенсивностью стационарного потока*. Он равен среднему числу событий в единицу времени. Средняя продолжительность времени между двумя смежными событиями в этом случае равна $1/\eta$.

---

[3] Обозначаются как прямоугольники.

[4] Обозначаются как стрелки.

Упомянутые выше предположения о свойствах потоков событий обычны для прикладных задач, так как эти потоки (или потоки, близкие к ним по свойствам) часто встречаются на практике благодаря предельным теоремам для потоков событий (Овчаров, 1969; Саати, 2010).

Для моделей с непрерывным временем неизвестными (свободными) параметрами модели являются интенсивности потоков событий. Их значения определяются путем сравнения наблюдаемых и прогнозируемых гистограмм, описывающих распределения частот пребывания в состояниях модели, а именно: вычисляются значения, обеспечивающие наилучшее соответствие наблюдаемых и ожидаемых частот попадания в определенное состояние системы в заданные моменты времени. Прогнозируемые вероятности нахождения в состояниях получаются путем численного интегрирования систем уравнений Колмогорова.

Марковские модели с непрерывным временем и свободными параметрами, которые идентифицируются по данным наблюдений, называются *сетями Маркова* (Куравский, Баранов, 2002; Куравский и др., 2003; Kuravsky, Baranov, 2003, 2004, 2005).

Для описания того, как вероятности нахождения в заданных состояниях изменяются со временем, применяются сети Маркова, организованные по так называемой схеме «гибели и размножения»[5] (рис. 1). Эта схема представляет собой конечную цепь из $n+1$ состояния, в которой переходы из состояния x_k ($k\neq 0$, $k\neq n$) возможны только в предшествующее состояние x_{k-1} или в следующее по порядку состояние x_{k+1}. Из состояний x_0 и x_n доступны только состояния x_1 и x_{n-1}, соответственно.

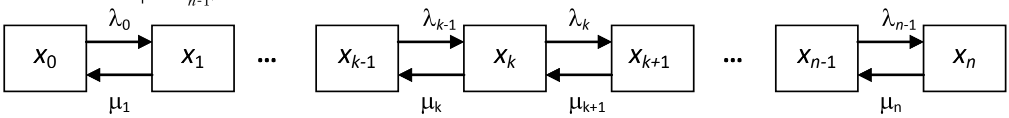

Рис. 1. Сеть Маркова, представляющая процесс тестирования с непрерывным временем: x_i ($i=0,1,...,n$) – состояния, λ_i ($i=0,1,...,n-1$) и μ_i ($i=1,2,...,n$) – интенсивности переходов

Динамика вероятностей нахождения в различных состояниях указанной схемы описывается следующей системой обыкновенных дифференциальных *уравнений Колмогорова*:

$$\frac{dp_0(t)}{dt} = -\lambda_0 p_0(t) + \mu_1 p_1(t),$$

$$\dots\dots\dots\dots\dots\dots\dots\dots$$

$$\frac{dp_k(t)}{dt} = -(\lambda_k + \mu_k) p_k(t) + \lambda_{k-1} p_{k-1}(t) + \mu_{k+1} p_{k+1}(t)$$

$$(k = 1,2,\dots,n-1),$$

$$\dots\dots\dots\dots\dots\dots\dots\dots$$

$$\frac{dp_n(t)}{dt} = -\mu_n p_n(t) + \lambda_{n-1} p_{n-1}(t),$$

где $p_*(t)$ есть вероятность нахождения в состоянии x_* в момент времени t; $*$ – номер состояния; λ_i ($i=0,1,...,n-1$) и μ_i ($i=1,2,...,n$) – интенсивности переходов между состояниями, которые определяются отдельно для каждого из рассматриваемых уровней способностей. Для интегрирования указанной системы необходимо задать начальные условия: $p_0(0)$, $p_1(0)$,..., $p_n(0)$. Нормализующее условие $\sum_{k=0}^{n} p_k(t) = 1$ выполняется в любой момент времени.

[5] Она была впервые применена в биологии для анализа динамики роста популяций.

Для упрощения задачи, а также для обеспечения приемлемой процедуры идентификации интенсивности потоков часто полагаются зависящими от индекса i по определенным правилам, включая тривиальный вариант: $\lambda_0=\lambda_1=...=\lambda_{n-1}=\lambda$ и $\mu_1=\mu_2=...=\mu_n=\mu$. Оптимальный выбор подобных зависимостей опирается на технику проверки статистических гипотез. В случае моделей с дискретным временем аналогичные зависимости исследуются для вероятностей переходов.

Процедура адаптивного тестирования заключается в последовательном предъявлении испытуемому задач, трудность которых определяется состоянием сети или цепи Маркова, в котором он находится в данный момент. Если испытуемый, находясь в состоянии x_i, решает задачу, он переходит в состояние x_{i+1}, в противном случае – в состояние x_{i-1}. По завершении тестирования он оказывается в одном из состояний $x*$, наилучшим образом соответствующих его уровню способностей. Принцип выбора очередного теста заключается в выборе задачи, трудность которой примерно соответствует уровню способностей испытуемого. Согласно проведенным наблюдениям и результатам современной теории тестирования, это обеспечивает наилучшую *дифференциацию* испытуемых по уровню их способностей.

2.2. *Идентификация марковских моделей с непрерывным временем*

Идентификации марковских моделей проводятся по выборкам испытуемых, отдельно для каждого из рассматриваемых уровней способностей. Каждому уровню способностей C_i, $i=1,...,I$ при этом ставится в соответствие свой уникальный набор оценок параметров модели, что позволяет в дальнейшем выявлять значение этого показателя, наилучшим образом согласующегося с наблюдениями. Таким образом, вероятности и интенсивности переходов являются функциями двух характеристик: уровня способностей и трудности задачи. Число уровней способностей – это дискретный параметр, который задает разрешающую способность оценки данной характеристики и устанавливается при решении каждой прикладной задачи в зависимости от объема выборки испытуемых, имеющейся у исследователя при решении задачи идентификации, и желаемой точности результата.

С каждой изменяющейся со временем гистограммой пребывания в состояниях модели связывается марковский процесс с дискретными состояниями. *Статистика Пирсона*

$$X^2=\sum_{k=0}^{n}\frac{(F_k-p_kN)^2}{p_kN} ,$$

где N – число элементов в выборке, p_k – прогнозируемая вероятность попадания в k-е состояние модели, а F_k – наблюдаемая частота нахождения в k-м состоянии модели, используется как мера соответствия в том смысле, что ее большие значения означают плохое согласование прогнозируемых и наблюдаемых результатов, а малые значения – хорошее согласование. Для идентификации модели минимизируется сумма указанных статистик в те моменты времени, для которых имеются результаты наблюдений. Наблюдаемые количества попаданий в различные интервалы трудностей задач определяются по результатам тестирования группы испытуемых. В качестве искомых оценок свободных параметров моделей используются значения, обеспечивающие наилучшее соответствие наблюдаемых и прогнозируемых частот попадания в определенное состояние системы в заданные моменты времени.

Доказано, что при выполнении ряда общих условий значения статистики Пирсона X^2, получаемые при подстановке истинных решений, асимптотически описываются распределением χ^2 с $n-l$ степенями свободы, где l – число определяемых параметров, причем вычисленные значения свободных параметров при увеличении объема выборки сходятся по вероятности к искомому решению (Крамер, 1976). Это позволяет использовать приведенную статистику для проверки гипотезы о том, что полученный прогноз согласуется с результатами наблюдений. Данный способ идентификации свободных параметров называется *методом минимума* χ^2 (Крамер, 1976) и дает решения, близкие к полученным, методом максимального правдоподобия.

Используемая процедура вычисления оцениваемых параметров состоит из двух этапов. На подготовительном этапе с помощью электронной таблицы для указанной системы дифференциальных уравнений кодируется численная схема интегрирования, позволяющая вычислять вероятностные функции p_k (Куравский, Баранов, 2002; Куравский и др., 2003; Kuravsky, Baranov, 2003). Эти функции вычисляются с некоторым заданным временным шагом. Для вычисления решения с приемлемой точностью оказались достаточными *методы Рунге-Кутта* или их эквиваленты.

На заключительном этапе запускается численная процедура многомерной нелинейной оптимизации[6] (Куравский, Баранов, 2002; Куравский и др., 2003; Kuravsky, Baranov, 2003), позволяющая получать искомые значения свободных параметров. Полученные оценки свободных параметров рассматриваются как характеристики модели, выявленные в результате наблюдений. Рассмотренный критерий также позволяет сравнивать между собой различные варианты марковских моделей, выбирая среди них оптимальные.

2.3. *Поиск оптимального решения*

Зная состояние модели, в котором оказался тестируемый после решения последней предложенной ему задачи, и рассчитав вероятность нахождения в этом состоянии в заданный момент времени для каждого из рассматриваемых уровней способностей с помощью дифференциальных зависимостей (см. раздел 2.1), можно оценить вероятности пребывания в указанном конечном состоянии по формуле Байеса:

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

где C_i – событие, связанное с наличием у тестируемого i-го уровня способностей (i=1,...,I), S – событие, связанное с нахождением в заданном конечном состоянии модели в заданный момент времени, $P(C_i)$ – априорная вероятность появления i-го уровня способностей у тестируемого, $P(S|C_i)$ – вероятность нахождения в заданном конечном состоянии модели в заданный момент времени при наличии i-го уровня способностей, $P(C_i|S)$ – вероятность i-го уровня способностей при условии нахождения в заданном конечном состоянии модели в заданный момент времени.

Уровень способностей, при котором достигается наибольшая условная вероятность

$$P(C_{max}|S) = \max_i \{P(C_i|S)\}_{i=1,...,I},$$

дает искомую оценку. Распределение вероятностей $\{P(C_i|S)\}_{i=1,...,I}$, которое является результатом решения задачи, позволяет оценить степень надежности этой оценки.

[6] В настоящее время предлагается достаточно много программных продуктов для решения задач численной оптимизации. В частности, пользователи электронной таблицы Excel могут применять программное обеспечение компании Frontline Systems, Inc.

Как указано в разделе 2.1, разрешающая способность полученной оценки определяется длиной интервала между соответствующими смежными уровнями способностей в логитах, которая, в свою очередь, при условии постоянства таких длин задается числом уровней способностей I.

3. Математическая постановка и решение задачи фильтрации Калмана при адаптивном тестировании с использованием марковских моделей

В случае обсуждаемого варианта адаптивного тестирования наблюдаемый процесс представляет историю пребывания в состояниях марковских моделей. Он выражается вектором $\mathbf{x}(t)=(x_0(t),x_1(t),...,x_n(t))^T$, в котором в каждый момент времени один, и только один, из компонентов $x_i(t)$, $i=0,...,n$, соответствующий состоянию, где находится испытуемый, равен единице, а остальные компоненты равны нулю. В свою очередь, исследуемый информационный процесс $\mathbf{P}(t)=(p_0(t),p_1(t),...,p_n(t))^T$ представляет динамику изменения вероятностей пребывания в состояниях модели.

Уравнения информационного и наблюдаемого процессов, используемые при построении многомерного непрерывного фильтра Калмана для моделей рассматриваемого типа[7] $\sum_{k=0}^{n}\frac{d\hat{p}_k(t)}{d}=0$ имеют следующий вид (Тихонов и др., 2009; Шахтарин, 2010):

$$\frac{d\mathbf{P}}{dt}=\mathbf{FP}\ ,$$

$$\mathbf{x}(t)=\mathbf{P}(t)+\mathbf{v}(t),$$

где на случайные ошибки наблюдений $\mathbf{E}(\mathbf{v}(t))$ накладываются условия $\mathbf{E}(\mathbf{v}(t))=\mathbf{0}$ и $\mathbf{E}(\mathbf{v}(t)\mathbf{v}^T(\tau))=\mathbf{R}\delta(t-\tau)$, матрица формирующего фильтра $\mathbf{F}$ размерности $(n+1)\times(n+1)$ есть

$$\begin{pmatrix} -\lambda_0 & \mu_1 & . & . & . & 0 & 0 \\ \lambda_0 & -(\lambda_1+\mu_1) & \mu_2 & . & . & 0 & 0 \\ . & . & . & . & . & . & . \\ 0 & . & \lambda_{k-1} & -(\lambda_k+\mu_k) & \mu_{k+1} & 0 & 0 \\ . & . & . & . & . & . & . \\ 0 & 0 & . & . & \lambda_{n-2} & -(\lambda_{n-1}+\mu_{n-1}) & \mu_n \\ 0 & 0 & . & . & . & \lambda_{n-1} & -\mu_n \end{pmatrix},$$

а $\mathbf{R}$ – симметричная положительно определенная матрица, которую мы далее будем полагать не зависящей от времени. При проведении практических расчетов эта матрица может заменяться на одну их своих выборочных оценок $\hat{\mathbf{R}}$, полученных для каждого из рассматриваемых уровней способностей на основе результатов наблюдений.

Дифференциальное уравнение фильтра Калмана, определяющее несмещенную оценку исследуемого процесса[8] $\hat{\mathbf{P}}(t))=(\hat{p}_0(t),\hat{p}_1(t),...,\hat{p}_n(t))^T$ с минимальным средним квадратом ошибки $\mathbf{e}(t)=\mathbf{P}(t)-\hat{\mathbf{P}}(t)$, представляется в виде:

$$\frac{d\hat{\mathbf{P}}(t)}{dt}=\mathbf{F}\hat{\mathbf{P}}(t)+\mathbf{K_c}(t)\big(\mathbf{x}(t)-\hat{\mathbf{P}}(t)\big)\ ,$$

где $\mathbf{K}(t)$ – матричный коэффициент усиления фильтра Калмана.

---

[7] Особенностями этих моделей являются: отсутствие информационного шума, равенство размерностей информационного процесса и процесса наблюдений и единичная матрица наблюдений.

[8] Выход фильтра Калмана.

В классическом случае этот коэффициент задается уравнением:

$$\mathbf{K}_c(t)=\mathbf{U}(t)\mathbf{R}^{-1},$$

в котором ковариационная матрица ошибок $\mathbf{U}(t)=\mathbf{E}(\mathbf{e}(t)\mathbf{e}^T(t))$ является решением одной из матричных форм уравнения Риккати:

$$\frac{d\mathbf{U}(t)}{dt}=\mathbf{FU}(t)+\mathbf{U}(t)\mathbf{F}^T-\mathbf{U}(t)\mathbf{R}^{-1}\mathbf{U}(t).$$

Однако, поскольку в рассматриваемой задаче компоненты оценки информационного процесса $\hat{\mathbf{P}}(t)$ представляют собой нормированные величины – вероятности пребывания в состояниях сети Маркова с суммой, равной единице, – необходима коррекция коэффициента усиления $\mathbf{K}_c(t)$, обеспечивающая поддержание данного условия.

Если нормализующее условие $\sum_{k=0}^{n}\hat{p}_k(t)=1$ выполняется в начальный момент времени $t=0$, а правая часть уравнения фильтра Калмана такова, что при $t\geq 0$ обеспечивается равенство $\sum_{k=0}^{n}\frac{d\hat{p}_k(t)}{dt}=0$, то указанное нормализующее условие выполняется в любой момент времени $t\geq 0$. Очевидно, что условие $\sum_{k=0}^{n}\frac{d\hat{p}_k(t)}{dt}=0$ равносильно равенству нулю суммы компонентов вектора, заданного матричным выражением $\mathbf{F}\hat{\mathbf{P}}(t)+\mathbf{K}_c(t)(\mathbf{x}(t)-\hat{\mathbf{P}}(t))$. Поскольку нулевая сумма компонентов вектора $\mathbf{F}\hat{\mathbf{P}}(t)$ обеспечивается приведенной выше структурой матрицы $\mathbf{F}$, то для равенства нулю суммы компонентов всего указанного матричного выражения необходимо и достаточно нулевой суммы компонентов вектора $\mathbf{K}_c(t)(\mathbf{x}(t)-\hat{\mathbf{P}}(t))$.

Сумма компонентов вектора $\mathbf{x}(t)-\hat{\mathbf{P}}(t)$ равна нулю по условиям рассматриваемой задачи, так как эти величины интерпретируются как вероятности. Учитывая данный факт, несложно доказать, что достаточным условием нулевой суммы компонентов вектора $\mathbf{K}_c(t)(\mathbf{x}(t)-\hat{\mathbf{P}}(t))$ является равенство сумм элементов матрицы $\mathbf{K}_c(t)$ во всех ее столбцах. Таким образом, если матричный коэффициент усиления $\mathbf{K}_c(t)$ в уравнении фильтра Калмана заменить на близкий к нему нормализованный коэффициент $\mathbf{K}_n(t)$ с равными во всех столбцах суммами элементов, то условие $\sum_{k=0}^{n}\frac{d\hat{p}_k(t)}{dt}=0$ будет выполнено. Матрицу $\mathbf{K}_n(t)$ можно получить, домножив справа матрицу $\mathbf{K}_c(t)$ на диагональную матрицу $\mathbf{D}$, элементы которой вычисляются по формуле:

$$d_{jj}=\frac{\sum_{l,m=0}^{n}k_{lm}}{(n+1)k_{*j}},$$

где d_{jj} – j-й диагональный элемент матрицы $\mathbf{D}$; k_{lm}, $l,m=0,...,n$ – элементы матрицы $\mathbf{K}_c(t)$; k_{*j} – сумма элементов в j-м столбце матрицы $\mathbf{K}_c(t)$. Такая замена корректна, если $\mathbf{K}_n(t)=\mathbf{U}(t)\mathbf{R}^{-1}\mathbf{D}$ лежит в допустимых границах вариаций коэффициента $\mathbf{K}_c(t)$, обусловленных ошибками выборочных оценок матрицы $\mathbf{R}$, что проверяется с помощью подходящих критериев согласия.

В частности, для этого можно:

- сгенерировать множество выборочных оценок ковариационной матрицы $\mathbf{R}$, соответствующих доверительным интервалам для заданного объема выборки N;

- вычислить, используя эти оценки, выборку матриц $\{\mathbf{K}_{\mathbf{ni}}(t)\}_{i=1,...,M}$;

- вычислить выборочное распределение евклидовой нормы разностей $\{\|\mathbf{K}_{\mathbf{ni}}(t)-\mathbf{K}_{\mathbf{c}}(t)\|_E\}_{i=1,...M}$ классического и нормированного коэффициентов усиления;

- учитывая, что полученное выборочное распределение при достаточно большом числе элементов в матричных коэффициентах усиления приблизительно соответствует нормальному, построить для него выборочные оценки математического ожидания и дисперсии и оценить вероятность p превышения евклидовой нормы разности $\|\mathbf{K}_{\mathbf{n}}(t)-\mathbf{K}_{\mathbf{c}}(t)\|_E$.

Если $p \geq 0{,}05$, то использование нормализованного коэффициента $\mathbf{K}_{\mathbf{n}}(t)$ является допустимым. Рассмотренный метод может быть совмещен с процедурой кластеризации, использующей самоорганизующиеся карты Кохонена (Куравский и др., 2011; Kuravsky et al., 2011).

В соответствии с представленной выше процедурой адаптивного тестирования, фильтрация Калмана выполняется автономно для каждого из уровней способностей, учитываемых при постановке решаемой задачи.

В заключение следует отметить, что существует ряд интересных аналогий между фильтром Калмана и скрытыми марковскими моделями (Куравский и др., 2010; Kuravsky et al., 2010), частично рассмотренных в обзорах зарубежных авторов (см., напр.: Roweis, Ghahramani, 1999).

4. Программная реализация

Рассмотренная процедура фильтрации реализована в среде графического программирования LabVIEW (см. рис. 2). При этом интегрирование матричного уравнения Риккати и уравнения фильтра Калмана выполнено численными методами[9], а для оценки начального состояния ковариационной матрицы ошибок $\mathbf{U}(0)$, о которой наблюдения дают, как правило, мало полезной информации, использованы следующие предположения:

- $\mathbf{E}(\mathbf{e}(0))=\mathbf{0}$;

- компоненты вектора ошибок фильтрации $\mathbf{e}(0)$ статистически независимы;

- дисперсии компонентов вектора ошибок фильтрации $\mathbf{e}(0)$ пропорциональны соответствующим дисперсиям компонентов случайного шума наблюдения $\mathbf{v}(t)$.

5. Основные результаты и выводы

1. Разработан и программно реализован вероятностный метод фильтрации искажающих результаты артефактов при адаптивном тестировании, построенном на использовании обучаемых структур в форме марковских моделей с непрерывным временем.

2. Устранение артефактов, обусловленных различными формами некорректного целенаправленного вмешательства в процедуру испытаний, выполняется на основе сравнения наблюдаемых и прогнозируемых результатов ответов на вопросы для разных уровней способностей испытуемых с помощью фильтра Калмана, адаптированного для задачи адаптивного тестирования.

[9] Следует отметить, что процедура фильтрации, в которой $\mathbf{U}(t)$ определяется путем интегрирования уравнения Риккати, является более корректной, чем используемые в значительном числе приложений аналогичные процедуры, где $\mathbf{U}(t)$ находится как решение уравнения $\mathbf{FU}(t)+\mathbf{U}(t)\mathbf{F}^T-\mathbf{U}(t)\mathbf{R}^{-1}\mathbf{U}(t)=0$ для стационарного случая.

3. Выбор фильтра Калмана для устранения артефактов является оптимальным среди близких по содержанию подходов, поскольку он наилучшим образом согласуется с принятой концепцией адаптивного тестирования и контекстом ее использования.

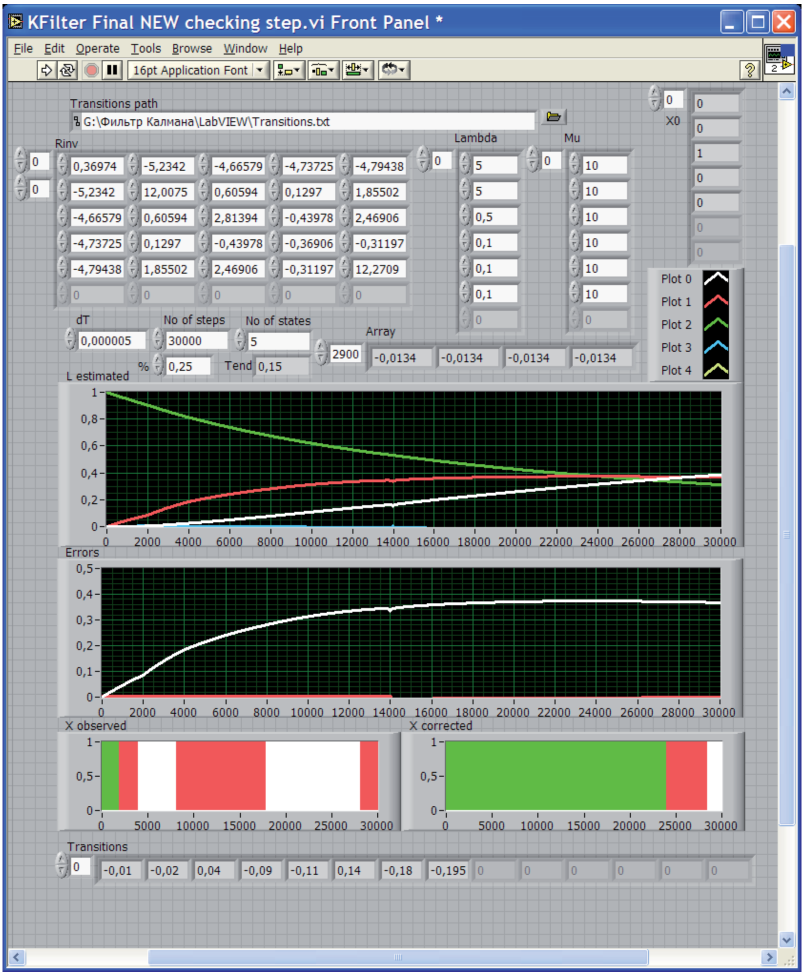

Рис. 2. Результаты фильтрации Калмана для марковской модели с пятью состояниями

Литература

Аванесов В. С. Педагогическое измерение латентных качеств // Педагогическая диагностика. 2003. № 4. С. 69–78.

Карданова Е. Ю. Моделирование и параметризация тестов: основы теории и приложения. М.: ФГУ «Федеральный центр тестирования», 2008.

Крамер Г. Математические методы статистики. М.: Мир, 1976.

Куравский Л. С., Баранов С. Н. Синтез сетей Маркова для прогнозирования усталостного разрушения // Нейрокомпьютеры: разработка и применение. 2002. № 11. С. 29–40.

Куравский Л. С., Баранов С. Н. Применение нейронных сетей для диагностики и прогнозирования усталостного разрушения тонкостенных конструкций // Нейрокомпьютеры: разработка и применение. 2001. № 12. С. 47–63.

Куравский Л. С., Баранов С. Н., Корниенко П. А. Обучаемые многофакторные сети Маркова и их при-

менение для исследования психологических характеристик // Нейрокомпьютеры: разработка и применение. 2005. № 12. С. 65–76.
Куравский Л. С., Баранов С. Н., Малых С. Б. Нейронные сети в задачах прогнозирования, диагностики и анализа данных: Учеб. пособие. М.: РУСАВИА, 2003.
Куравский Л. С., Баранов С. Н., Юрьев Г. А. Синтез и идентификация скрытых марковских моделей для диагностики усталостного разрушения // Нейрокомпьютеры: разработка и применение. 2010. № 12. С. 20–36.
Куравский Л. С., Ушаков Д. В., Мармалюк П. А., Панфилова А. С. Исследование факторных влияний на развитие психологических характеристик с применением нового подхода к оценке адекватности моделей наблюдениям // Информационные технологии. 2011. № 11 (в печати).
Куравский Л. С., Юрьев Г. А. Адаптивное тестирование как марковский процесс: модели и их идентификация // Нейрокомпьютеры: разработка и применение. 2011 а. № 2. С. 21–29.
Куравский Л. С., Юрьев Г. А. Использование марковских моделей при обработке результатов тестирования // Вопросы психологии. 2011 б. № 2. С. 98–107.
Овчаров Л. А. Прикладные задачи теории массового обслуживания. М.: Машиностроение. 1969.
Саати Т. Л. Элементы теории массового обслуживания и ее приложения. М.: ЛИБРОКОМ, 2010.
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.
Тюменева Ю. А. Психологическое измерение. М.: Аспект-Пресс, 2007.
Шахтарин Б. И. Случайные процессы в радиотехнике. Т. 1. Линейные преобразования. М.: Горячая линия–Телеком, 2010.
Baker F. B. The Basics of Item Response Theory. ERIC Clearinghouse on Assessment and Evaluation, University of Maryland, College Park, MD, 2001.
Gregory R. J. Psychological testing: History, principles, and applications (5th edition). N.Y.: Pearson, 2007.
Gulliksen H. Theory of Mental Tests. John Wiley & Sons Inc, 1950.
Kuravsky L. S., Malykh S. B. Application of Markov models for analysis of development of psychological characteristics // Australian Journal of Educational & Developmental Psychology. 2004. V. 2. P. 29–40.
Kuravsky L. S., Baranov S. N. Condition monitoring of the structures suffered acoustic fatigue failure and forecasting their service life // Proc. Condition Monitoring 2003, Oxford, United Kingdom. P. 256–279, July 2003.
Kuravsky L. S., Baranov S. N. Neural networks in fatigue damage recognition: diagnostics and statistical analysis // Proc. 11th International Congress on Sound and Vibration, St.-Petersburg, Russia. P. 2929–2944, July 2004.
Kuravsky L. S., Baranov S. N. The concept of multifactor Markov networks and its application to forecasting and diagnostics of technical systems // Proc. Condition Monitoring 2005, Cambridge, United Kingdom. P. 111–117, July 2005.
Kuravsky L. S., Baranov S. N., Yuryev G. A. Synthesis and identification of hidden Markov models based on a novel statistical technique in condition monitoring // Proc. 7th International Conference on Condition Monitoring & Machinery Failure Prevention Technologies, Stratford-upon-Avon, England, June 2010.
Kuravsky L. S., Marmalyuk P .A., Panfilova A. S. Estimation of goodness-of-fit measures for identification of unrestricted factor models employing arbitrarily distributed observed data // Proc. 8th International Conference on Condition Monitoring & Machinery Failure Prevention Technologies, Cardiff, UK, June 2011.
Rasch G. Probabilistic models for some intelligence and attainment tests // Copenhagen, Danish Institute for Educational Research / Expanded edition with foreword and afterword by B. D. Wright. Chicago: The University of Chicago Press. 1980.
Roweis S., Ghahramani Z. A unifying review of linear Gaussian models // Neural Computation. V. 11. № 2. 1999. P. 305–345.
Wright B. D., Masters G. N. Rating scale analysis. Rasch measurements. Chicago: MESA Press, 1982.
Wright B. D., Stone M. N. Best Test Design. Chicago: MESA Press, 1979.

PROBABILISTIC METHOD OF FILTERING ARTIFACTS IN ADAPTIVE TESTING

KURAVSKY L. S., *Moscow City University of Psychology and Education, Moscow*
YUR'EV G. A., *Moscow City University of Psychology and Education, Moscow*

In this article we present a method of filtering the results of adaptive testing, built on the use of structures in the form of Markov models with continuous time. The elimination of artifacts caused by various forms of incorrect direct intervention in the procedure of testing is performed on the basis of comparison of the observed and the predicted results of the answers to the questions using the Kalman filter, adapted for the solution of the considered problem.

Keywords: adaptive testing, Markov models, Kalman filter.

Transliteration of the Russian references

Avanesov V. S. Pedagogicheskoe izmerenie latentnyh kachestv // Pedagogicheskaja diagnostika. 2003. № 4. S. 69–78.
Kardanova E. Ju. Modelirovanie i parametrizacija testov: osnovy teorii i prilozhenija. M.: FGU «Federal'nyj centr testirovanija», 2008.
Kramer G. Matematicheskie metody statistiki. M.: Mir, 1976.
Kuravskij L. S., Baranov S. N. Sintez setej Markova dlja prognozirovanija ustalostnogo razrushenija // Nejrokomp'jutery: razrabotka i primenenie. 2002. № 11. S. 29–40.
Kuravskij L. S., Baranov S. N. Primenenie nejronnyh setej dlja diagnostiki i prognozirovanija ustalostnogo razrushenija tonkostennyh konstrukcij // Nejrokomp'jutery: razrabotka i primenenie. 2001. № 12. S. 47–63.
Kuravskij L. S., Baranov S. N., Kornienko P. A. Obuchaemye mnogofaktornye seti Markova i ih primenenie dlja issledovanija psihologicheskih harakteristik // Nejrokomp'jutery: razrabotka i primenenie. 2005. № 12. S. 65–76.
Kuravskij L. S., Baranov S. N., Malyh S. B. Nejronnye seti v zadachah prognozirovanija, diagnostiki i analiza dannyh: Ucheb. posobie. M.: RUSAVIA, 2003.
Kuravskij L. S., Baranov S. N., Jur'ev G. A. Sintez i identifikacija skrytyh markovskih modelej dlja diagnostiki ustalostnogo razrushenija // Nejrokomp'jutery: razrabotka i primenenie. 2010. № 12. S. 20–36.
Kuravskij L. S., Ushakov D. V., Marmaljuk P. A., Panfilova A. S. Issledovanie faktornyh vlijanij na razvitie psihologicheskih harakteristik s primeneniem novogo podhoda k ocenke adekvatnosti modelej nabljudenijam // Informacionnye tehnologii. 2011. № 11 (v pechati).
Kuravskij L. S., Jur'ev G. A. Adaptivnoe testirovanie kak markovskij process: modeli i ih identifikacija // Nejrokomp'jutery: razrabotka i primenenie. 2011 a. № 2. S. 21–29.
Kuravskij L. S., Jur'ev G. A. Ispol'zovanie markovskih modelej pri obrabotke rezul'tatov testirovanija // Voprosy psihologii. 2011 b. № 2. S. 98–107.
Ovcharov L. A. Prikladnye zadachi teorii massovogo obsluzhivanija. M.: Mashinostroenie. 1969.
Saati T. L. Jelementy teorii massovogo obsluzhivanija i ee prilozhenija. M.: LIBROKOM, 2010.
Tihonov V. I., Shahtarin B. I., Sizyh V. V. Sluchajnye processy. Primery i zadachi. T. 5. Ocenka signalov, ih parametrov i spektrov. Osnovy teorii informacii. M.: Gorjachaja linija–Telekom, 2009.
Tjumeneva Ju. A. Psihologicheskoe izmerenie. M.: Aspekt-Press, 2007.
Shahtarin B. I. Sluchajnye processy v radiotehnike. T. 1. Linejnye preobrazovanija. M.: Gorjachaja linija–Telekom, 2010.