

УДК 004.8, 004.93, 159.95  
DOI: 10.15827/0236-235X.120.601-608

Дата подачи статьи: 20.10.17  
2017. Т. 30. № 4. С. 601–608

## **МЕТОД ПОВЫШЕНИЯ ИНТЕРПРЕТИРУЕМОСТИ РЕГРЕССИОННЫХ МОДЕЛЕЙ НА ОСНОВЕ ТРЕХСТУПЕНЧАТОЙ МОДЕЛИ РАЗВИТИЯ МЫШЛЕНИЯ**

*И.М. Куликовских, к.т.н., доцент, kulikovskikh.i@gmail.com*

*(Самарский национальный исследовательский университет им. академика С.П. Королева,  
Московское шоссе, 34, г. Самара, 443086, Россия)*

Повышение обобщающей способности регрессионных моделей приводит к более эффективному решению задач распознавания, прогнозирования, выявления различных стратегий поведения технических и социальных систем. Известные методы повышения обобщающей способности обладают вычислительной эффективностью, но понижают интерпретируемость модели и конечных результатов.

Предлагаемая статья является попыткой обратиться к данной проблеме, дополняя задачи регрессии и классификации в контексте задач синтеза фильтров из области цифровой фильтрации и тестирования из области психометрии. Принимая во внимание достоинства решений в альтернативных предметных областях, целью данной работы является создание метода повышения интерпретируемости регрессионных моделей через стимулирование процессов внутренней неопределенности в машинном обучении.

Для достижения поставленной цели была создана трехступенчатая модель развития машинного мышления, отражающая динамику анализируемых предметных областей. Эти области связаны единими процессами внутренней неопределенности, что требует создания соответствующего математического аппарата. Данная работа рассматривает процессы внутренней неопределенности с когнитивной точки зрения через процессы забывания и угадывания. Результатом исследования является реализация ступеней синтеза фильтров, тестирования и интегральной ступени регрессии/классификации в рамках созданной трехступенчатой модели развития мышления. При этом ступень синтеза фильтров моделирует техническую среду, ступень тестирования – социальную среду, а интегральная ступень – социально-техническую среду. Тогда как социальная среда моделирует сознание человека, социально-техническая среда вводит понятие сознания машины.

Представленные реализации позволили формализовать метод повышения интерпретируемости регрессионных моделей через формирование принципов перехода от сознания человека к сознанию машины.

**Ключевые слова:** машинное обучение, цифровая фильтрация, психометрия, внутренняя неопределенность, философская система Гегеля.

Создание эффективных методов и алгоритмов с высокой обобщающей способностью является основной задачей машинного обучения [1–3]. С целью повышения обобщающей способности было предложено множество методов усиления моделей регрессии и классификации [1]. К наиболее эффективным методам усиления моделей обучения относятся различные методы понижения размерности, методы регуляризации и композиционные методы. Первые две группы методов способствуют разрешению компромисса между переобучением и недообучением, тогда как усиление обучающих моделей через построение их композиций обеспечивает меньшую восприимчивость к переобучению за счет адаптивности данного класса алгоритмов. Однако, несмотря на их вычислительную эффективность, композиционные методы, как и другие методы усиления, обладают существенным недостатком, связанным с плохой интерпретируемостью моделей [4], а следовательно, и конечных результатов. Данная работа является попыткой обратиться к этой проблеме, дополняя задачи регрессии и классификации в контексте задач синтеза фильтров из области цифровой фильтрации и тестирования из области психометрии соответственно.

Целями слияния понятий в предметных областях являются углубление понимания проблемы интерпретируемости и формирование соответствующего математического аппарата для нахождения более эффективных решений. В качестве инстру-

мента слияния задач регрессии и синтеза фильтров явились понятия *информативности признаков* и *сложности фильтров*. При этом первое понятие задается совокупностью важных признаков в модели обучения, а второе – порядком модели фильтра. В свою очередь, инструментом слияния задач классификации и тестирования явились понятия эффектов «пола» и «потолка» [5–11].

Наиболее эффективное решение задачи синтеза фильтров связано с оптимизацией полюсов модели фильтра [12, 13], тогда как эффективность результатов тестирования связывают с необходимостью выработки метакогнитивной стратегии у обучаемых (тестируемых) [9, 10]. Но оба решения можно представить как введение в рассмотрение дополнительного параметра исследуемой модели, описывающего ее внутреннюю неопределенность (неоднозначность). Принимая во внимание достоинства данных альтернативных решений, целью данной работы является создание метода повышения интерпретируемости регрессионных моделей через стимулирование процессов внутренней неопределенности в машинном обучении.

### **Трехступенчатая модель развития машинного мышления**

Для достижения поставленной цели была создана трехступенчатая модель развития машинного мышления (см. рис. 1), определяющая новый под-

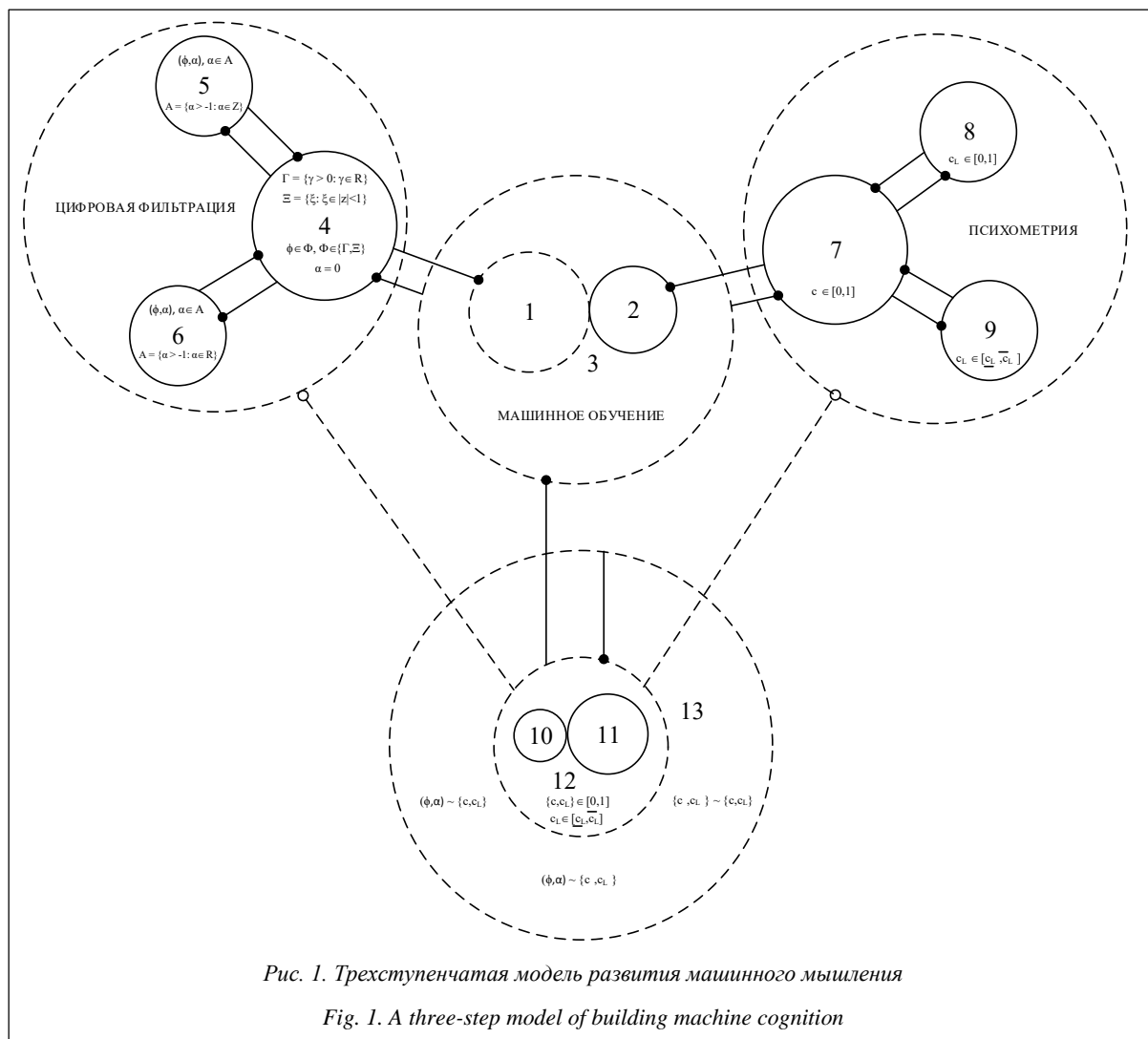


Рис. 1. Трехступенчатая модель развития машинного мышления

Fig. 1. A three-step model of building machine cognition

ход к построению моделей обучения, делая акцент на когнитивной науке и формировании машинного сознания [14]. Наиболее полно идея трехступенчатой модели развития, или триады, была раскрыта в философии Гегеля [15], согласно которой развитие мышления происходит в три этапа: 1) тезис, задающий исходный момент; 2) антитезис – отрицание тезиса; 3) отрицание антитезиса, задающее исходный момент последующего развития мышления. Последний этап объединяет в себе характеристики двух предыдущих ступеней, отражая их на более высоком уровне. В терминах описанной триады тезисом в данной работе являются методы цифровой фильтрации, которые отвечают за построение строго заданных моделей поведения различных динамических систем. Антитезисом, в свою очередь, являются методы психометрии, которые, главным образом, нацелены на описание поведения социальных систем и принципов их взаимодействия при обучении (тестировании). Отрицание антитезиса предполагает их слияние, что позволяет сформировать метод повышения эффективности машинного обучения. При этом, тогда как антитезис отвечает

за моделирование сознания человека, отрицание антитезиса формирует сознание машины.

Данная трехступенчатая модель развития машинного мышления, отражающая динамику анализируемых предметных областей, связана единими процессами внутренней неопределенности, что требует создания соответствующего математического аппарата (рис. 1). С этой целью в работе рассматриваются процессы внутренней неопределенности с когнитивной точки зрения через процессы забывания и угадывания. В цифровой фильтрации забывание описывается через параметр затухания (забывания) [16–18] при оценивании коэффициентов модели фильтра методом взвешенных наименьших квадратов. В когнитивной психологии и психометрии стимулирование процессов забывания и угадывания [19–23] и психометрии [24–26] позволяет исключить эффекты «пола» и «потолка» при оценивании уровня мышления обучаемых. Модели оценивания уровня мышления обучаемых имеют вид логистической регрессии, коэффициенты которой также оцениваются методом наименьших квадратов с итерационным перевзвешиванием. Нако-

нец, в машинном обучении присутствие эффектов «пола» и «потолка» может указывать на наличие идеальной разделимости классов, а следовательно, слабой обучаемости модели [27, 28]. Таким образом, предлагаемая интерпретация в терминах забывания и угадывания позволяет свести разработку методов стимулирования процессов внутренней неопределенности к единому математическому аппарату.

Приведем интерпретацию цифровых обозначений, используемых на рисунке 1.

1. Анализ методов усиления моделей регрессии (модель линейной регрессии).

2. Анализ методов усиления моделей классификации (модель логистической регрессии).

3. Методы стимулирования внутренней неопределенности в машинном обучении.

4. Методы стимулирования внутренней неопределенности в цифровой фильтрации через параметр  $\phi$ . При этом предполагается, что  $\alpha = 0$ .

5. Методы стимулирования внутренней неопределенности в цифровой фильтрации через пару параметров  $(\phi, \alpha)$ , где  $\alpha \in Z$ .

6. Методы стимулирования внутренней неопределенности в цифровой фильтрации через пару параметров  $(\phi, \alpha)$ , где  $\alpha \in R$ .

7. Методы стимулирования внутренней неопределенности в психометрии через вероятностный параметр  $c$ .

8. Методы стимулирования внутренней неопределенности в психометрии через нечеткий параметр  $c$  при индивидуальном обучении.

9. Методы стимулирования внутренней неопределенности в психометрии через нечеткий параметр  $c$  при групповом обучении.

10. Реализация ступени синтеза фильтров, моделирующей техническую среду.

11. Реализация ступени тестирования, моделирующей социальную среду.

12. Реализация интегральной ступени регрессии/классификации, моделирующей социально-техническую среду.

13. Метод повышения эффективности регрессионных моделей через формирование принципов перехода от сознания человека к сознанию машины.

#### Метод повышения интерпретируемости регрессионных моделей

Рассмотрим реализацию ступеней графической модели (рис. 1, цифровые обозначения 10–13) более подробно в терминах пространственной схемы взаимодействия объектов [29]. При этом на данном этапе был использован лишь верхний уровень – пространство функциональных преобразований  $\Pi$ . Рассмотрение соответствующих подпространств функциональных расширений  $\Pi\Pi_{\{\Lambda\}}^{\{\Pi\}}$  и более глубоких уровней вложенности связано с разложе-

нием процессов внутренней неопределенности в зависимости от сложности задания или требуемого уровня когнитивных способностей. Подпространства характеристических расширений пространства функциональных преобразований  $\Pi\Pi_{\{\Lambda\}}^{\{\Pi\}}$  формируются на основе факторов стимулирования процессов внутренней неопределенности для каждого уровня мышления.

**Реализация ступени синтеза фильтров** состоит из четырех пронумерованных функциональных блоков, соединенных между собой ключом, который показывает переход от одного пространства к другому (точкой обозначается исходное пространство или подпространство) [29] (рис. 2).

Функциональные блоки 1–4 представляют собой совокупность алгоритмов для решения поставленных задач для входной пары  $\{D, f\}$ , где  $D \in \{\Lambda, G\}$ ,  $f \in \{F, H\}$  – композиции представлений цифрового фильтра в  $s$ -области и  $z$ -области соответственно. Обозначение  $\psi$  определяет некоторый функционал преобразования с параметрами  $\alpha \in A$ , где  $A = \{\alpha > -1: \alpha \in \mathbf{R}\}$ , и  $\phi \in \Phi$ , где  $\Phi \in \{\Gamma, \Xi\}$ ,  $\Gamma = \{\gamma > 0: \gamma \in \mathbf{R}\}$ ,  $\Xi = \{\xi: \xi \in |z| < 1\}$ . Опишем задачи для каждого блока.

1. Блок оценивания параметров фильтра  $\psi: D \times f \rightarrow f$  и обобщенного фильтра  $\psi^\alpha: (D, \alpha) \times f \rightarrow f$ :

– задача построения модели фильтра в  $s$ -области;

– задача построения модели фильтра в  $z$ -области;

– задача оценивания коэффициентов фильтра в  $s$ -области;

– задача оценивания коэффициентов фильтра в  $z$ -области.

2. Блок понижения эффективной размерности модели фильтра:

– задача оптимизации полюсов фильтра  $\psi^\phi: (D, \phi) \times f \rightarrow f$ ;

– задача понижения размерности фильтра  $\psi^\phi: (D, \phi) \times f \rightarrow D$ .

3. Блок понижения эффективной размерности модели обобщенного фильтра:

– задача оптимизации полюсов обобщенного фильтра  $\psi^{\{\phi, \alpha\}}: (D, \phi, \alpha) \times f \rightarrow f$ ;

– задача понижения размерности обобщенного фильтра  $\psi^{\{\phi, \alpha\}}: (D, \phi, \alpha) \times f \rightarrow D$ .

4. Блок стимулирования внутренней неопределенности  $\psi^\phi: (D, \phi) \times f \rightarrow \phi$  и  $\psi^{\{\phi, \alpha\}}: (D, \phi, \alpha) \times f \rightarrow (\phi, \alpha)$ , который формирует пару выходных ключей  $\{4, -\}$ .

Описанные функциональные блоки соединены ключами  $\{1, 2\}$ ,  $\{1, 3\}$ ,  $\{2, 3\}$ ,  $\{2, 4\}$ ,  $\{3, 4\}$ , по которым передаются входные и выходные параметры, требуемые для решения поставленных задач. Содержание выходных ключей более подробно рассматривается далее при описании метода повышения эффективности регрессионных моделей.

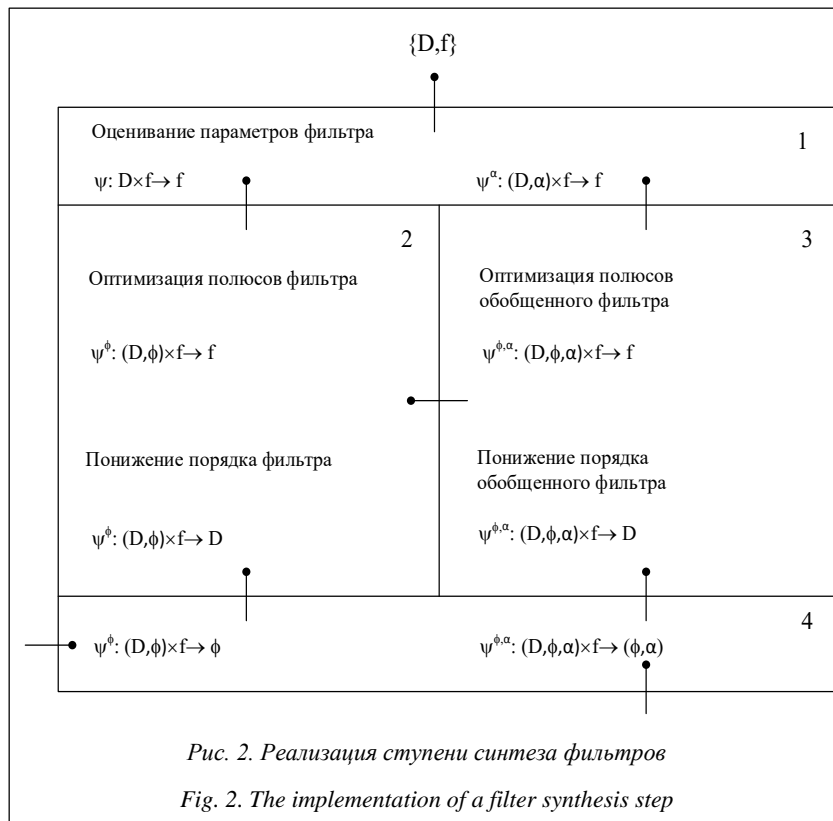


Рис. 2. Реализация ступени синтеза фильтров  
Fig. 2. The implementation of a filter synthesis step

**Реализация ступени тестирования** состоит из пронумерованных функциональных блоков 5–10, соединенных между собой соответствующими ключами: {5, 6}, {5, 7}, {6, 7}, {7, 8}, {7, 9}, {8, 9}, {8, 10}, {9, 10} (рис. 3).

Роль входной пары в данной подсистеме играет  $\{X, u\}$ , где  $X$  – пространство признаков для описания тестируемых (обучаемых);  $u \in U$  – вектор ответов для заданного уровня мышления; модель поведения при тестировании  $h: X \times u \rightarrow u$ .

Приведем описание задач для каждого блока [30].

5. Блок формирования тестовых заданий:

- задача создания инструкций согласно форме тестирования (с множественным выбором) с необходимым количеством правильных и конкурентных неправильных ответов или дистракторов (distractors);
- задача согласования инструкций с таксономией Блума в зависимости от уровня мышления обучаемого и сложности задания.

6. Блок формирования банка ответов:

- задача сбора ответов обучаемых за тестовые задания;

блоков 5 и 6;

- задача проведения статистического анализа: 1) построение описательной статистики для контрольных групп и  $t$ -теста; 2) проверка адекватности

- задача выявления проблемных заданий, изменение формулировки и порядка представления заданий;

- задача обработки событий, возникающих при тестировании, а именно: 1) количества и последовательности нажатий при выборе дистракторов; 2) индикатора активности окна с текущим заданием, возврата к окну с предыдущим заданием, перехода к окну со следующим; 3) отсутствия нажатия в случае пропуска задания;

- задача контроля времени срабатывания событий в выделенный интервал тестирования.

7. Блок создания моделей индивидуального и группового поведения  $h: X \times u \rightarrow u$  [31]:

- задача расширения пространства признаков  $X$  и корректировки вектора ответов  $u$  на основе функциональных



Рис. 3. Реализация ступени тестирования  
Fig. 3. The implementation of an assessment stage

и надежности заданий с  $\alpha$ -Кронбаха; 3) проверка целостности и однородности ответов через выявление аномалий (чистых угадываний и мнений другого участника группового обучения), построение гистограмм и диаграмм размаха;

- задача формулирования и доказательства гипотез (утверждений) для описания поведения обучаемых в условиях внутренней неопределенности;
- задача создания концептуальных и математических моделей поведения на основе гипотез и утверждений.

8. Блок создания модели обучения при вероятностной трактовке параметра внутренней неопределенности  $c^\dagger \in [0, 1]$ :

- задача оценивания уровня когнитивности  $h^{c^\dagger}: X \times (u, c^\dagger) \rightarrow u$ ;
- задача понижения размерности пространства признаков, описывающих поведение при индивидуальном и групповом обучении  $h^{c^\dagger}: X \times (u, c^\dagger) \rightarrow X$ .

9. Блок создания модели обучения при нечеткой трактовке параметра внутренней неопределенности  $c_L^\dagger$  в условиях нечетко заданных когнитивных уровней и/или группового обучения:

- задача оценивания уровня когнитивности  $h^{c_L^\dagger}: X \times (u, c_L^\dagger) \rightarrow u$ ;
- задача понижения размерности пространства признаков, описывающих поведение при индивидуальном и групповом обучении  $h^{c_L^\dagger}: X \times (u, c_L^\dagger) \rightarrow X$ .

10. Блок стимулирования внутренней неопределенности  $h^{c^\dagger}: X \times (u, c^\dagger) \rightarrow c^\dagger$  и  $h^{c_L^\dagger}: X \times (u, c_L^\dagger) \rightarrow c_L^\dagger$ , который формирует пару выходных ключей  $\{10, -\}$ .

**Реализация интегральной ступени регрессии/классификации** дополняет ранее рассмотренные реализации функциональными блоками 11 и 12 (см. рис. 4). Назначение данных блоков – стимулирование методов забывания и угадывания в машинном обучении через адаптацию разработанных моделей стимулирования внутренней неопределенности в цифровой фильтрации и психометрии. Входными пространствами для блоков 11 и 12 являются функциональные блоки 4 и 10, что отмечено соответствующим ключами  $\{4, 11\}$  и  $\{10, 11\}$ . Входная пара в данном случае задана как  $\{X, y\}$ , где  $X$  – матрица «объект-признак»,  $y$  – вектор ответов. Переход между функциональными блоками задан ключом  $\{11, 12\}$ . Ниже рассмотрены задачи, решаемые в рамках указанных блоков.

11. Блок усиления алгоритмов машинного обучения с учетом внутренней неопределенности:

- задача классификации  $g_c: (X, c) \times (y, c) \rightarrow y$ , где  $c \in [0, 1]$  задает вероятностный характер внутренней неопределенности при переходе от сознания человека  $c^\dagger \in [0, 1]$  к сознанию машины;
- задача классификации  $g_{cL}: (X, c_L) \times (y, c_L) \rightarrow y$ , где  $c_L \in [0, 1]$  задает нечеткий характер внутренней неопределенности при переходе от сознания человека  $c_L^\dagger \in [0, 1]$  к сознанию машины;

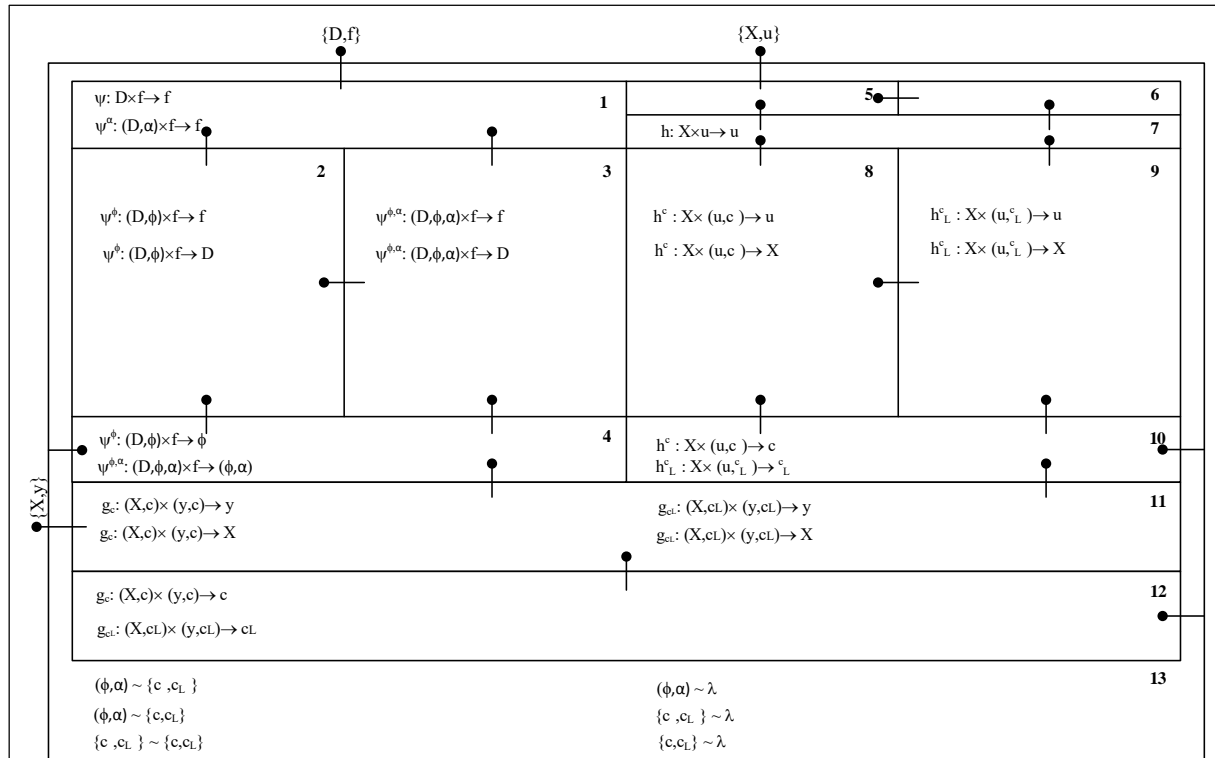


Рис. 4. Реализация трехступенчатой модели

Fig. 4. The implementation of a tree-step model

– задача понижения размерности пространства признаков  $g_c: (X, c) \times (y, c) \rightarrow X$ , где  $c \in [0, 1]$  задает вероятностный характер внутренней неопределенности;

– задача понижения размерности пространства признаков  $g_{cL}: (X, c_L) \times (y, c_L) \rightarrow X$ , где  $c_L \in [0, 1]$  задает нечеткий характер внутренней неопределенности.

12. Блок стимулирования внутренней неопределенности:

– задача стимулирования процессов забывания и угадывания при вероятностной трактовке  $g_c: (X, c) \times (y, c) \rightarrow c$ ;

– задача стимулирования процессов забывания и угадывания при нечеткой трактовке  $g_{cL}: (X, c_L) \times (y, c_L) \rightarrow c_L$ .

Блок стимулирования внутренней неопределенности при нечеткой трактовке учитывает не только уровень тестируемого (обучаемого), но и уровень сложности задания по таксономии Блума [32–34], меру влияния группового обучения, а также различные факторы стимулирования для достижения наилучшего результата. Как видно на рисунке 4, функциональный блок также формирует выходной ключ  $\{12, -\}$ .

**Метод повышения эффективности регрессионных моделей** реализован в виде функционального блока 13, собирающего информацию с выходных ключей:  $\{4, 13\}$ ,  $\{10, 13\}$ ,  $\{12, 13\}$  (рис. 4).

Основное назначение функционального блока – формирование вектора внутренней неопределенности для описания перехода от сознания человека к сознанию машины. С этой целью надсистема анализирует информацию с выходных ключей подсистем в виде моделей:

- 1)  $\psi^\phi: (D, \phi) \times f \rightarrow \phi$ ;
- 2)  $\psi^{(\phi, \alpha)}: (D, \phi, \alpha) \times f \rightarrow (\phi, \alpha)$ ;
- 3)  $h^{c^\dagger}: X \times (u, c^\dagger) \rightarrow c^\dagger$ ;
- 4)  $h^{c_L^\dagger}: X \times (u, c_L^\dagger) \rightarrow c_L^\dagger$ ;
- 5)  $g_c: (X, c) \times (y, c) \rightarrow c$ ;
- 6)  $g_{cL}: (X, c_L) \times (y, c_L) \rightarrow c_L$ ,

описывающих методы стимулирования вектора внутренней неопределенности  $\{\phi, (\phi, \alpha), c^\dagger, c_L^\dagger, c, c_L\}$ , сформированного в задачах цифровой фильтрации, тестирования и машинного обучения.

Перечислим задачи, решаемые в функциональном блоке 13.

13. Блок формирования вектора внутренней неопределенности:

– задача нормирования вектора неопределенности  $\{\phi, (\phi, \alpha), c^\dagger, c_L^\dagger, c, c_L\}$ ;

– задача создания метрик подобия:

1)  $(\phi, \alpha) \sim c^\dagger, (\phi, \alpha) \sim c_L^\dagger, \{c^\dagger, c_L^\dagger\} \sim \phi, \{c^\dagger, c_L^\dagger\} \sim \alpha$ ;

2)  $(\phi, \alpha) \sim \{c, c_L\}, \{c, c_L\} \sim \{c^\dagger, c_L^\dagger\}$ ;

3)  $(\phi, \alpha) \sim \lambda, \{c^\dagger, c_L^\dagger\} \sim \lambda, \{c, c_L\} \sim \lambda$ ;

– задача упаковки исходных данных  $\{D, f\}$ ,  $\{X, u\}$ ,  $\{X, y\}$  с целью группировки заданий для распре-

деленных вычислений в блоках с выходными ключами, то есть 4, 10, 12.

При этом метрики подобия группы 1 отражают степень схожести описания процессов внутренней неопределенности между параметрами технической среды (область цифровой фильтрации) и социальной (область психометрии). Метрики группы 2 направлены на выявление схожести между интегральным параметром, описывающим социально-техническую среду (область машинного обучения), и каждой средой в отдельности. Наконец, метрики группы 3 рассматривают степень схожести введенных параметров с регуляризационными методами в рамках одной среды (технической, социальной или социально-технической).

### Выводы

В работе предложен метод повышения интерпретируемости регрессионных моделей, который дополняет задачи регрессии и классификации в контексте задач синтеза фильтров из области цифровой фильтрации и тестирования из области психометрии. Данные области связаны едиными процессами внутренней неопределенности, позволяющими сформировать трехступенчатую модель развития мышления и соответствующий математический аппарат. С целью формализации предлагаемого метода в работе также приведены реализации ступени синтеза фильтров, ступени тестирования и интегральной ступени регрессии/классификации в рамках пространственной схемы взаимодействия объектов.

### Благодарности

Автор выражает благодарность д.т.н., профессору С.А. Прохорову за ценные замечания и рекомендации, способствующие повышению качества представления результатов исследований.

Работа выполнена при поддержке Министерства образования и науки РФ, грант № 074-U01.

### Литература

1. Воронцов К.В. Математические методы обучения по прецедентам (теория обучения машин). URL: <http://www.machinelearning.ru/wiki/images/6/6d/Voron-ML-1.pdf> (дата обращения: 30.05.2017).
2. Vapnik V.N. The nature of statistical learning theory. Springer India, 1998, 314 p.
3. Vapnik V.N. Statistical learning theory. Wiley-Interscience, 1998, 768 p.
4. Lipton Z.C. The mythos of model interpretability. URL: [arxiv.org/abs/1606.03490](http://arxiv.org/abs/1606.03490) (дата обращения: 30.05.2017).
5. Agresti A. Foundations of linear and generalized linear models. Wiley Series in Probability and Statistics, 2015, 472 p.
6. Hastie T., Tibshirani R., Friedman J. The elements of statistical learning: Data mining, inference, and prediction. Springer Series in Statistics, 2013, 745 p.
7. Donnelly S., Verkuilen J. Empirical logit analysis is not logistic regression. Jour. of Memory and Language, 2017, no. 94, pp. 28–42.
8. Everitt B.S. The Cambridge dictionary of statistics. Cambridge, Cambridge Univ. Press, 2010, 480 p.

9. Groth-Marnat G., Wright A.J. Handbook of psychological assessment. Wiley, 2016, 768 p.
10. Lewis-Beck M.S., Bryman A., Liao T.F. The SAGE encyclopedia of social science research methods. SAGE Publ., 2003, 1528 p.
11. Kulikovskikh I.M., Prokhorov S.A. Modifications of log-likelihood to measure floor and ceiling effects // Информационные технологии и нанотехнологии (ИТНТ-2017): сб. тр. III Междунар. конф. 2017. С. 1849–1853 (на англ.).
12. King J.J., O'Canainn T. Optimum pole positions for Laguerre-function models. Electronics Letters, 1969, no. 5, pp. 601–602.
13. Prokhorov S.A., Kulikovskikh I.M. Unique condition for generalized Laguerre functions to solve pole position problem. Signal Processing, 2016, vol. 108, pp. 25–29.
14. Величковский Б.М. Когнитивная наука: Основы психологии познания. М.: Академия, 2006. Т. 1. 448 с.
15. Hegel G.W.F. Hegel's science of logic. Prometheus Books, 1991, 844 p.
16. Andersson P. Adaptive forgetting in recursive identification through multiple models. Intern. Jour. of Control, 1985, no. 42, pp. 171–185.
17. Kulhavy R. Zarrop M.B. On a general concept of forgetting. Intern. Jour. of Control, 1993, no. 58, pp. 905–924.
18. Vahidi A., Stefanopoulou A., Peng H. Recursive least squares with forgetting for online estimation of vehicle mass and road grade: theory and experiments. Vehicle System Dynamics: Intern. Jour. of Vehicle Mechanics and Mobility, 2005, no. 43, pp. 31–55.
19. Alberini C.M., LeDoux J.E. Memory reconsolidation. Current Biology, 2013, no. 23, pp. R746–R750.
20. Alberini C.M. The role of reconsolidation and the dynamic process of long-term memory formation and storage. Front. Behav. Neurosci., 2011, no. 5, p. 12.
21. Elliott G., Isaac C.L., Muhlert N. Measuring forgetting: A critical review of accelerated long-term forgetting studies. Cortex, 2014, no. 54, pp. 16–32.
22. Chan J.C.K. When does retrieval induce forgetting and when does it induce facilitation? Implications for retrieval inhibition, testing effect, and text processing. Jour. of Memory and Language, 2009, no. 61, pp. 153–170.
23. Cowan N. What are the differences between long-term, short-term, and working memory? Prog. Brain Res, 2008, no. 169, pp. 323–338.
24. Lord F.M. Applications of item response theory to practical testing problems. Mahwah, NJ, Lawrence, Erlbaum Associates, Inc., 1980, 274 p.
25. van der Linden W.J. Forgetting, guessing, and mastery: The MacReady and Dayton models revisited and compared with a latent trait approach. Jour. of Educational Statistics, 1978, no. 3, pp. 305–317.
26. Ken Cor M., Sood G. Guessing and forgetting: A latent class model for measuring learning. Political Analysis, 2016, no. 24, pp. 226–242.
27. Schapire R.E., Valiant L. The strength of weak learnability. Machine Learning, 1990, no. 5, pp. 197–227.
28. Shalev-Shwartz S., Singer Y. On the equivalence of weak learnability and linear separability: New relaxations and efficient boosting algorithms. Machine Learning, 2010, no. 80, pp. 141–163.
29. Прохоров С.А., Куликовских И.М. Создание комплекса программ на основе пространственной схемы взаимодействия объектов // Программные продукты и системы. 2012. № 3. С. 5–8.
30. Куликовских И.М., Прохоров С.А., Сучкова С.А., Матвеев Е.В. Комплексная система коллаборативного обучения на основе нечетких моделей для описания поведения систем с частичным знанием // Изв. ШИЦ РАН. 2016. Т. 18. № 4. С. 760–765.
31. Kulikovskikh I.M., Prokhorov S.A., Suchkova S.A. Promoting collaborative learning through regulation of guessing in clickers. Computers in Human Behavior, 2017, vol. 75, pp. 81–91.
32. Prudencio R.B.C., Hernandez-Orallo J., Martinez-Uso A. Analysis of instance hardness in machine learning using item response theory. URL: [http://users.dsic.upv.es/~flip/LMCE2015/Papers/LMCE\\_2015\\_submission\\_1.pdf](http://users.dsic.upv.es/~flip/LMCE2015/Papers/LMCE_2015_submission_1.pdf) (дата обращения: 30.05.2017).
33. Smith R.M., Martinez T., Giraud-Carrier C. An instance level analysis of data complexity. Machine Learning, 2014, no. 95, pp. 225–256.
34. Rijmen F., Tuerlinckx F., De Boeck P., Kuppens P. A non-linear mixed model framework for Item Response Theory. Psychological Methods, 2003, no. 8, pp. 185–205.

Software &amp; Systems

DOI: 10.15827/0236-235X.120.601-608

Received 20.10.17

2017, vol. 30, no. 4, pp. 601–608

#### A METHOD FOR IMPROVING INTERPRETABILITY OF REGRESSION MODELS BASED ON A THREE-STEP BUILDING COGNITION MODEL

**I.M. Kulikovskikh**<sup>1</sup>, Ph.D. (Engineering), Associate Professor, [kulikovskikh.i@gmail.com](mailto:kulikovskikh.i@gmail.com)

<sup>1</sup> Samara National Research University, Moskovskoe Highway 34, Samara, 443086, Russian Federation

**Abstract.** Increasing generalization performance of regression models leads to a more effective solution for the problems of recognition, prediction, and extraction of social and engineering behavior strategies. A number of known methods for improving the generalization properties demonstrate computational effectiveness, however they reduce interpretability of a model and results.

This study is an attempt to approach this problem looking at the methods of regression and classification from *digital filtering* and *psychometrics* points of view. Considering the advantages of the methods for solving the interpretability problem in these areas, this research is aimed at defining a method to improve the interpretability of regression models by promoting learner's internal uncertainty in *machine learning*.

In order to solve the problem, the author has developed a three-step model of building cognition. This model reflects direct relations among *digital filtering*, *psychometrics*, and *machine learning*. These research areas employ the same sources of internal uncertainty that makes creating consistent mathematical models that connects the areas possible. For this purpose, the paper considers internal uncertainty from a cognitive point of view as processes of forgetting and guessing. The findings of this study provide the implementations of the following steps in accordance with the tree-step model: a filter synthesis step, a psychological assessment step, and an integrated regression/classification step. While the first step models an *engineering* environment and the second step presents a *social* environment, the integrated step helps to create a *social-engineering* environment. In addition, in contrast to the social environment that may simulate *human cognition*, the social-engineering environment seems promising in introducing *machine cognition*.

The proposed implementations allow formalizing the method for improving interpretability of regression models changing from one kind of cognition to the other.

**Keywords:** machine learning, digital filtering, psychometrics, internal uncertainty, Hegel's philosophical system.

**Acknowledgements.** The author is grateful to Dr.Sc. Prokhorov for valuable comments and recommendations that contribute to improving the research results presentation quality.

The work has been supported by the Ministry of Education and Science of the Russian Federation, grant no. 074-U01.

### References

1. Vorontsov K.V. *Matematicheskie metody obucheniya po pretседentam (teoriya obucheniya mashin)* [Mathematical Methods for supervised Learning (Machine Learning Theory)]. Available at: <http://www.machinelearning.ru/wiki/images/6/6d/Voron-ML-1.pdf> (accessed May 30, 2017).
2. Vapnik V.N. *The Nature of Statistical Learning Theory*. Springer India Publ., 1998, 314 p.
3. Vapnik V.N. *Statistical Learning Theory*. Wiley-Interscience Publ., 1998, 768 p.
4. Lipton Z.C. *The Mythos of Model Interpretability*. Available at: [arxiv.org/abs/1606.03490](http://arxiv.org/abs/1606.03490) (accessed May 30, 2017).
5. Agresti A. *Foundations of Linear and Generalized Linear Models*. Wiley Series in Probability and Statistics, 2015, 472 p.
6. Hastie T., Tibshirani R., Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. Springer Series in Statistics, 2013, 745 p.
7. Donnelly S., Verkuilen J. Empirical logit analysis is not logistic regression. *Jour. of Memory and Language*. 2017, no. 94, pp. 28–42.
8. Everitt B.S. *The Cambridge Dictionary of Statistics*. Cambridge, Cambridge Univ, Press, 2010, 480 p.
9. Groth-Marnat G., Wright A.J. *Handbook of Psychological Assessment*. Wiley Publ., 2016, 768 p.
10. Lewis-Beck M.S., Bryman A., Liao T.F. *The SAGE Encyclopedia of Social Science Research Methods*. SAGE Publ., 2003, 1528 p.
11. Kulikovskikh I.M., Prokhorov S.A. Modifications of log-likelihood to measure floor and ceiling effects. *Proc. of Information Systems and Nanotechnology (ITNT-2017)*. Samara, 2017, pp. 1849–1853.
12. King J.J., O'Canainn T. Optimum pole positions for Laguerre function models. *Electronics Letters*. 1969, no. 5, pp. 601–602.
13. Prokhorov S.A., Kulikovskikh I.M. Unique condition for generalized Laguerre functions to solve pole position problem. *Signal Processing*. 2016, vol. 108, pp. 25–29.
14. Velichkovsky B.M. *Kognitivnaya nauka: Osnovy psikhologii poznaniya* [Cognitive Science: Foundations of Epistemic Psychology]. Moscow, Akademiya Publ., 2006, vol. 1, 448 p.
15. Hegel G.W.F. *Hegel's Science of Logic*. Prometheus Books Publ., 1991, 844 p.
16. Andersson P. Adaptive forgetting in recursive identification through multiple models. *Int. Jour. of Control*. 1985, no. 42 (5), pp. 171–185.
17. Kulhavy R., Zarrow M.B. On a general concept of forgetting. *Int. Jour. of Control*. 1993, no. 58 (4), pp. 905–924.
18. Vahidi A., Stefanopoulou A., Peng H. Recursive least squares with forgetting for online estimation of vehicle mass and road grade: theory and experiments. *Vehicle System Dynamics: Int. Jour. of Vehicle Mechanics and Mobility*. 2005, no. 43 (1), pp. 31–55.
19. Alberini C.M., LeDoux J.E. Memory reconsolidation. *Current Biology*. 2013, no. 23 (17), pp. R746–R750.
20. Alberini C.M. The role of reconsolidation and the dynamic process of long-term memory formation and storage. *Front Behav Neurosci*. 2011, no. 5, pp. 12.
21. Elliott G., Isaac C.L., Muhlert N. Measuring forgetting: A critical review of accelerated long-term forgetting studies. *Cortex*. 2014, no. 54, pp. 16–32.
22. Chan J.C.K. When does retrieval induce forgetting and when does it induce facilitation? Implications for retrieval inhibition, testing effect, and text processing. *Jour. of Memory and Language*. 2009, no. 61 (2), pp. 153–170.
23. Cowan N. What are the differences between long-term, short-term, and working memory? *Prog. Brain Res*. 2008, no. 169, pp. 323–338.
24. Lord F.M. *Applications of item response theory to practical testing problems*. Mahwah, NJ, Lawrence, Erlbaum Associates Inc. Publ., 1980, 274 p.
25. van der Linden W.J. Forgetting, guessing, and mastery: The MacReady and Dayton models revisited and compared with a latent trait approach. *Jour. of Educational Statistics*. 1978, no. 3 (4), pp. 305–317.
26. Ken Cor M., Sood G. Guessing and forgetting: A latent class model for measuring learning. *Political Analysis*. 2016, no. 24 (2), pp. 226–242.
27. Schapire R.E., Valiant L. The strength of weak learnability. *Machine Learning*. 1990, no. 5 (2), pp. 197–227.
28. Shalev-Shwartz S., Singer Y. On the equivalence of weak learnability and linear separability: New relaxations and efficient boosting algorithms. *Machine Learning*. 2010, no. 80, pp. 141–163.
29. Prokhorov S.A., Kulikovskikh I.M. Creation of software on the basis of spatial object interaction diagram. *Programmnye produkty i sistemy* [Software & Systems]. 2012, no. 3, pp. 5–8 (in Russ.).
30. Kulikovskikh I.M., Prokhorov S.A., Suchkova S.A., Matytsin E.V. Complex system for collaborative learning based on fuzzy models to describe systems behavior with partial knowledge. *Izvestiya SNTs RAN* [Proc. Samara Scientific Center of RAS]. 2016, vol. 18, no. 4 (4), pp. 760–765 (in Russ.).
31. Kulikovskikh I.M., Prokhorov S.A., Suchkova S.A. Promoting collaborative learning through regulation of guessing in clickers. *Computers in Human Behavior*. 2017, no. 75, pp. 81–91.
32. Prudencio R.B.C., Hernandez-Orallo J., Martinez-Uso A. *Analysis of instance hardness in machine learning using item response theory*. Available at: [http://users.dsic.upv.es/~flip/LMCE2015/Papers/LMCE\\_2015\\_submission\\_1.pdf](http://users.dsic.upv.es/~flip/LMCE2015/Papers/LMCE_2015_submission_1.pdf) (accessed May 30, 2017).
33. Smith R.M., Martinez T., Giraud-Carrier C. An instance level analysis of data complexity. *Machine Learning*. 2014, no. 95 (2), pp. 225–256.
34. Rijmen F., Tuerlinckx F., De Boeck P., Kuppens P. A nonlinear mixed model framework for Item Response Theory. *Psychological Methods*. 2003, no. 8 (2), pp. 185–205.