УДК 519.246

СТАТИСТИЧЕСКИЕ КРИТЕРИИ СОГЛАСИЯ НА ОСНОВЕ ОЦЕНКИ СКОЛЬЗЯЩЕГО ЭКЗАМЕНА

В. М. Неделько

В работе исследуются статистические свойства оценки скользящего экзамена как критерия выбора модели решения (метода построения решающей функции). Для задачи дисперсионного анализа доказано, что критерий скользящего экзамена эквивалентен критерию Фишера для проверки гипотезы однородности при некотором уровне значимости.

Установлено также, что применение критерия скользящего экзамена для выбора решающей функции из некоторого однопараметрического класса эквивалентно построению оптимальной решающей функции в рамках байесовской модели с нормальным распределением на значениях параметров.

Ключевые слова: оценка скользящего экзамена, регрессионный анализ, проверка статистических гипотез, критерий Фишера

Keywords: crossvalidation, statistical test, regression

MSC: 62H30 – Classification and discrimination; cluster analysis, 62H15 – Hypothesis testing, 62H25 – Factor analysis and principal components; correspondence analysis.

Введение

Метод скользящего экзамена является фактически основным из методов выбора модели, используемых при решении практических задач анализа данных [3] [5].

Вместе с тем, данный метод остаётся в значительной степени эвристическим [9], несмотря на большой число исследований в этой области [13].

Для обоснования применения скользящего экзамена полезно проследить его связь с методами проверки статистических гипотез [6]. Статистический критерий предназначен для выбора вероятностной модели данных. Скользящий экзамен — это критерий выбора модели решения.

В классическом подходе статистического анализа данных (дискриминантный и регрессионный анализ) модель решения выбирается на основе вероятностной модели данных, например, для случая нормальных классов решающей функцией является линейный или квадратичный дискриминант.

Однако, многие методы остаются эффективными и за рамками предположений, на основе которых они построены. В частности, линейная решающая функция используется и без предположений нормальности. Пример — классический дискриминант Фишера. Другой пример — метод опорных векторов.

В задачах анализа данных нас часто не интересует полная вероятностная модель данных (вид распределения), поскольку целью является построение решающей функции.

Таким образом, вместо гипотезы о форме виде распределений мы можем делать предположения непосредственно о виде оптимальной решающей функции [1]. Вид решающей функции в этом случае выступает в роли гипотезы. Такой подход можно считать в некотором смысле обобщением классической теории проверки гипотез.

В данной работе рассматривается случай, когда выбор вида решения однозначно связан с выбором вероятностной модели данных. В этом случае критерий скользящего экзамена фактически является статистическим критерием проверки гипотез. В работе будут исследованы свойства данного критерия.

1. Основные понятия

1.1. Задача построения решающей функции

Пусть X — пространство значений переменных, используемых для прогноза, а Y — пространство значений прогнозируемых переменных, и пусть C — множество всех вероятностных мер на мер на заданной σ-алгебре подмножеств множества $D = X \times Y$. При каждом $c \in C$ имеем вероятностное пространство $\langle D, \mathfrak{B}, \mathsf{P}_c \rangle$, где $\mathfrak{B}$ — σ-алгебра, P_c — вероятностная мера. Параметр c будем называть стратегией природы.

Решающей функцией называется соответствие $\lambda \colon X \to Y$.

Качество принятого решения оценивается заданной функцией потерь $\mathcal{L} \colon Y^2 \to [0, \infty)$. Под риском будем понимать средние потери:

$$R(c, \lambda) = \mathsf{E}\mathcal{L}(y, \lambda(x)) = \int_D \mathcal{L}(y, \lambda(x)) \; \mathsf{P}_c(dx, dy), \quad x \in X, y \in Y. \tag{1.1}$$

Задача машинного обучения (интеллектуального анализа данных) заключается в построении решающей функции, которая бы минимизировала риск.

Заметим, что значение риска зависит от c — распределения, которое неизвестно. Поэтому в приведённой формулировке задача некорректна. Однако полностью строгой общей постановки задачи построения решающей функции в настоящее время не существует [9]. Здесь уместно привести аналогию с ситуацией в математической статистике, где не существует понятия оптимального во всех случаях статистического критерия и, соответственно, нельзя сформулировать задачу поиска такого критерия. На практике либо решаются более частные строго поставленные задачи, либо разрабатываются эвристические методы, например, методы, минимизирующие выборочную оценку риска [14].

Пусть $V_N = \{(x^i, y^i) \in D \,|\, i = 1, \ldots, N\}$, $V_N \in D^N$ — случайная независимая выборка из распределения P_c. В дальнейшем объём выборки N будет, как правило, фиксированным, поэтому этот параметр в обозначении выборки обычно будем опускать.

Метод (алгоритм) построения решающих функций есть отображение $Q \colon \mathcal{V} \to \Lambda$, где Λ — заданный класс решающих функций, $\mathcal{V} = \bigcup\limits_{N=1}^{\infty} D^N$ — множество всевозможных выборок. а $\lambda_{Q,V}$ — функция, построенная по выборке V методом Q.

Критерием качества метода построения решающих функций может служить средний риск

$$\mathcal{F}_N(c, Q) = \mathsf{E}_N R(c, \lambda_{Q,V}).$$

Усреднение производится по выборкам объёма N.

Для оценивания риска обычно используются эмпирические функционалы качества, т. е. точечные оценки риска, такие как эмпирический риск, оценка скользящего экзамена, оценка bootstrap и т.п.

Эмпирический риск определяется как средние потери на выборке:

$$\widetilde{R}(V, \lambda) = \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}(y^i, \lambda(x^i)).$$

Выражение для эмпирического риска является частным случаем формулы 1.1, если в неё подставить эмпирическое распределение.

1.2. Оценка скользящего экзамена

Функционал скользящего экзамена определяется как:

$$\breve{R}(V,Q) = \frac{1}{N}\sum_{i=1}^{N} \mathcal{L}(y^i, \lambda_{Q,V_i'}(x^i)),$$

где $V_i' = V\backslash\{(x^i, y^i)\}$ — выборка, получаемая из V удалением i-го наблюдения,

Оценка скользящего экзамена является несмещённой оценкой риска для решения, построенного по выборке объёма $N-1$, а именно

$$\mathsf{E}\breve{R}(V_N, Q) = \mathcal{F}_{N-1}(c, Q).$$

Несмещённость скользящего экзамена является одной из основных причин его широкого использования в качестве оценки риска.

1.3. Задача восстановления зависимостей

Задача восстановления зависимостей [7] [10] [11] возникает, когда в качестве значений целевой переменной Y выступает множество действительных чисел.

В задаче оценивания регрессии используется квадратичная функция потерь $\mathcal{L}(y, y') = (y - y')^2$. В этом случае риск есть средний квадрат отклонения.

Обозначим $\overline{y} = \frac{1}{N}\sum_{i=0}^{N} y_i$ – среднее значение y по всей выборке.

Пусть метод Q_0 в качестве решения возвращает $\overline{y}$. Тогда оценка качества этого метода посредством скользящего экзамена даст величину

$$\breve{R}(V, Q_0) = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \frac{N\overline{y} - y_i}{N-1}\right)^2 = \frac{N}{(N-1)^2}\sum_{i=1}^{N}(y_i - \overline{y})^2.$$

Для демонстрации техники работы с оценкой скользящего экзамена проведём некоторые выкладки. Вычислим

$$\mathsf{E}R(c, \lambda_{Q_0,V}) = \mathsf{E}(y - \overline{y})^2 = \mathsf{D}y + \mathsf{D}\overline{y} = \mathsf{D}y \cdot \frac{N+1}{N}.$$

Используя несмещённость оценки скользящего экзамена, получаем

$$\mathsf{E}\breve{R}(V, Q_0) = \frac{N}{N-1}\widetilde{\mathsf{D}}y, \quad \widetilde{\mathsf{D}}y = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \overline{y})^2.$$

Теперь с учётом $\mathsf{E}R(c, \lambda_{Q_0,V_{N-1}}) = \mathsf{D}y \cdot \frac{N}{N-1}$ получаем общеизвестный факт

$$\mathsf{D}y = \mathsf{E}\widetilde{\mathsf{D}}y.$$

Мы доказали несмещённость стандартной оценки дисперсии $\widetilde{\mathsf{D}}y$ как следствие несмещённости оценки скользящего экзамена. Это доказательство не проще общепринятого, однако полезно как пример использования свойств скользящего экзамена.

2. Гипотеза однородности

Предположим, что единственная переменная X принимает целые значения $\{1, ..., M\}$.

Обозначим $N_x = \sum_{i=1}^{N} I(x_i = x)$ – количество выборочных точек с $x_i = x$ и $\overline{y}_x = \frac{1}{N_x}\sum_{i=1}^{N} y_i \cdot I(x_i = x)$ – среднее по выборке значение y в точке x. Здесь $I(\cdot)$ — индикаторная функция (равна 1, когда условие истинно, и 0, когда ложно).

Обозначим $\overline{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$ – среднее значение y по всей выборке.

Статистика Фишера для проверки гипотезы однородности в этих обозначениях запишется как

$$F(V) = \frac{(N-M)\sum\limits_{x=1}^{M} N_x(\overline{y}_x - \overline{y})^2}{(M-1)\sum\limits_{i=1}^{N}(y_i - \overline{y}_{x_i})^2}.$$

Пусть метод Q_1 в качестве решения возвращает $\lambda(x) = \overline{y}_x$. Тогда оценка качества этого метода посредством скользящего экзамена даст величину

$$\breve{R}(V,Q_1) = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \frac{N_{x_i}\overline{y}_{x_i} - y_i}{N_{x_i}-1}\right)^2 = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{N_{x_i}}{N_{x_i}-1}\right)^2 (y_i - \overline{y}_{x_i})^2.$$

Введём обозначения:

$$z_i = y_i - \overline{y}_{x_i}, \qquad \overline{z}_x = \overline{y}_x - \overline{y}, \qquad d_x^2 = \sum_{i=1}^{N} z_i^2 \cdot I(x_i = x).$$

В этих обозначениях

$$\breve{R}(V,Q_0) = \frac{N}{(N-1)^2}\sum_{i=1}^{N}(z_i + \overline{z}_{x_i})^2 = \frac{N}{(N-1)^2}\sum_{i=1}^{N}(z_i^2 + \overline{z}_{x_i}^2) = \frac{N}{(N-1)^2}\sum_{x=1}^{M}(d_x^2 + N_x\overline{z}_x^2).$$

Аналогично

$$\breve{R}(V,Q_1) = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{N_{x_i}}{N_{x_i}-1}\right)^2 z_i^2 = \frac{1}{N}\sum_{x=1}^{M}\left(\frac{N_x}{N_x-1}\right)^2 d_x^2.$$

Также преобразуем выражение для статистики Фишера

$$F(V)\cdot(M-1)\sum_{x=1}^{M} d_x^2 = (N-M)\sum_{x=1}^{M} N_x\overline{z}_x^2.$$

Рассмотрим статистику

$$\breve{S}(V) = N\breve{R}(V,Q_0) - N\breve{R}(V,Q_1) = \left(\frac{N}{N-1}\right)^2\sum_{x=1}^{M}(d_x^2 + N_x\overline{z}_x^2) - \sum_{x=1}^{M}\left(\frac{N_x}{N_x-1}\right)^2 d_x^2.$$

После подстановки статистики Фишера

$$\breve{S}(V) = \left(\frac{N}{N-1}\right)^2\left(1 + F(V)\cdot\frac{M-1}{N-M}\right)\sum_{x=1}^{M} d_x^2 \; - \sum_{x=1}^{M}\left(\frac{N_x}{N_x-1}\right)^2 d_x^2. \tag{2.2}$$

Скользящий экзамен является общепринятым способом выбора метода Q при решении практических задач. Из нескольких методов выбирается тот, который даёт наименьшую ошибку на скользящем контроле.

Заметим, однако, что методы Q_0 и Q_1 фактически основаны на гипотезах соответственно равенства и неравенства средних. Иными словами, выбор метода Q_1 по сути означает, что гипотеза однородности выборки отвергается.

Таким образом, метод скользящего экзамена даёт статистический критерий, который отвергает гипотезу равенства средних при $\breve{S}(V) > 0$.

Теорема 1. *Пусть* $N_x = \frac{N}{M}$. *Тогда условие* $\check{S}(V) > 0$ *равносильно*

$$F(V) > 1 + \frac{N-1}{N-M}.$$

Доказательство непосредственно следует из 2.2.

Мы выявили достаточно любопытный факт, а именно, что метод скользящего экзамена эквивалентен некоторому статистическому критерию, уровень значимости для которого определяется параметрами данных и не может устанавливаться произвольно.

В случае, когда все N_x равны, что метод скользящего экзамена эквивалентен критерию Фишера, с пороговым значением статистики $1 + \frac{N-1}{N-M} \approx 2$, что при $M = 2$ соответствует уровню значимости около 0.157.

Заметим, что с ростом M уровень значимости уменьшается (стремится к 0), так как число степеней свободы растёт, а пороговое значение остаётся примерно одинаковым.

3. Байесовская модель

Для содержательной интерпретации дальнейших выкладок полезно рассмотреть пример с реальными данными.

Рассмотрим задачу Zillow Prize: Zillow's Home Value Prediction (Zestimate) (https://www.kaggle.com/c/zillow-prize-1).

Данные представляют собой значения 58 переменных, измеренные для 86795 объектов.

Целевая переменная Y' представляет собой некоторую характеристику, вычисляемую на основе цены объекта недвижимости.

В качестве переменной X будет выступать код региона, который принимает 317 различных значений.

К переменной Y' для удобства применим квантильное преобразование, т.е. введём переменную $Y = \widetilde{\mathsf{F}}(Y')$, $\widetilde{\mathsf{F}}(\cdot)$ – эмпирическая функция распределения.

Как и раньше, $\overline{y}_x = \frac{1}{N_x} \sum_{i=1}^{N} y_i \cdot I(x_i = x)$ – среднее по выборке значение y в точке x.

Введём величины $\overline{u}_x = \sqrt{12 \cdot N_x} \cdot (\overline{y}_x - 0.5)$.

Если справедлива гипотеза однородности (т.е. если Y не зависит от X), то величины $\overline{u}_x$ имеют распределение, близкое к нормальному с параметрами 0, 1.

На рисунке приведена (круглые серые маркеры) эмпирическая функция распределения $\widetilde{\mathsf{F}}(\overline{u}_x)$. Сплошная чёрная кривая изображает функцию нормального распределения с параметрами -0.06, 2.07, оцененными по обучающей выборке.

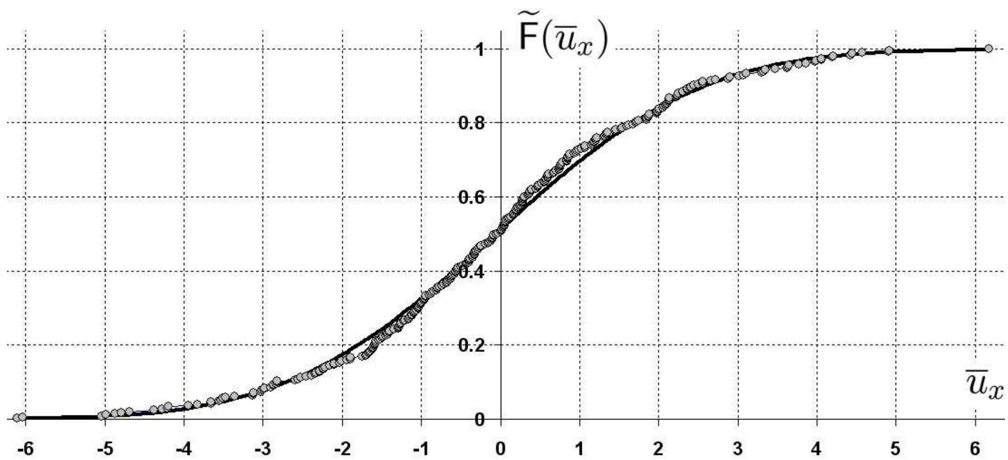

Рис. 1. Эмпирическая функция распределения оценок среднего в задаче Zestimate

Видим, что распределение очень похоже на нормальное, но дисперсия значимо больше 1, поэтому поведение $\overline{u}_x$ не может быть объяснено случайными флуктуациями частот.

Введём обозначения $\xi_x = \mathsf{E}[Y\,|x]$ – условное математическое ожидание целевой переменной в точке x и $\omega_x = \overline{u}_x - \xi_x$.

Предположим, что целевая переменная подчиняется следующей модели $Y = \xi_x + \omega$, $\mathsf{E}\omega = 0$, $\mathsf{D}\omega = \sigma^2$.

Тогда наблюдаемые значения $\overline{y}_x$ являются реализациями случайной величины $\zeta_x = \xi_x + \omega_x$, $\mathsf{E}\omega_x = 0$, $\mathsf{D}\omega_x = \frac{\sigma^2}{N_x}$.

В силу центральной предельной теоремы величина ω_x имеет приближённо нормально распределение.

В исходной постановке задачи значения ξ_x не являются случайными, но $\overline{y}_x$ эмпрически выглядит как величина с нормальным распределением.

Это даёт основания рассмотреть байесовскую постановку, в которой ξ_x будет случайной величиной с нормальным распределением.

В этой модели оптимальное решение есть

$$\lambda(x) = \mathsf{E}[\xi_x\,|\zeta_x = \overline{y}_x] = \mathsf{E}Y + (\overline{y}_x - \mathsf{E}Y)\cdot\alpha_x = \mathsf{E}Y\cdot(1-\alpha_x) + \overline{y}_x\alpha_x,$$

где

$$\alpha_x = 1 - \frac{\mathsf{D}\omega_x}{\mathsf{D}\zeta_x} = 1 - \frac{\sigma^2}{N_x\mathsf{D}\zeta_x}.$$

При построении решающей функции по выборке естественно взять несмещённые оценки неизвестных параметров:

$$\mathsf{E}Y \approx \overline{y},\quad \sigma^2 \approx \frac{1}{N-M}\sum_{i=1}^{N}(y_i - \overline{y}_{x_i})^2,\quad \mathsf{D}\zeta_x \approx \frac{1}{M-1}\sum_{x=1}^{M}(\overline{y}_x - \overline{y})^2.$$

В случае, когда все N_x одинаковы, при подстановке этих оценок в выражение для α_x получаем $\alpha_x = 1 - \frac{1}{F(V)}$. Последнее выражение представляется достаточно интересным. В следующем разделе мы его получим ещё раз, из других соображений.

4. Усреднённое решение

Методы Q_0 и Q_1 строят решения, исходя из того, принимается или отвергается гипотеза однородности. Поскольку мы не можем быть полностью уверены в справедливости гипотез, представляется естественным рассмотреть решения, основанные на усреднении [8] [12] базовых решающих функций.

Пусть метод Q_α в качестве решения возвращает $\lambda(x) = \alpha\overline{y}_x + (1-\alpha)\overline{y}$. Тогда оценка качества этого метода посредством скользящего экзамена даст величину

$$\breve{R}(V, Q_\alpha) = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \frac{N_{x_i}\overline{y}_{x_i} - y_i}{N_{x_i}-1}\cdot\alpha - \frac{N\overline{y} - y_i}{N-1}\cdot(1-\alpha)\right)^2.$$

После преобразований имеем

$$\breve{R}(V, Q_\alpha) = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{N_{x_i}}{N_{x_i}-1}(y_i - \overline{y}_{x_i})\cdot\alpha + \frac{N}{N-1}(y_i - \overline{y})\cdot(1-\alpha)\right)^2.$$

Выражаем через z_i, $\overline{z}_x$, d_x^2, которые введены в предыдущем разделе.

$$\breve{R}(V, Q_\alpha) = \frac{N}{(N-1)^2}\sum_{i=1}^{N}\left(C_{x_i}z_i\cdot\alpha + (z_i + \overline{z}_{x_i})\cdot(1-\alpha)\right)^2$$

$$= \frac{N}{(N-1)^2} \sum_{x=1}^{M} \left((C_x \alpha + 1 - \alpha)^2 d_x^2 + \overline{z}_x^2 N_x \cdot (1-\alpha)^2 \right),$$

где $C_x = \frac{N_x \cdot (N-1)}{(N_x - 1) \cdot N}$.

Подставим статистику Фишера

$$\breve{R}(V, Q_\alpha) = \frac{N}{(N-1)^2} \sum_{x=1}^{M} \left((C_x \alpha + 1 - \alpha)^2 d_x^2 + F(V) \cdot \frac{M-1}{N-M} \cdot (1-\alpha)^2 d_x^2 \right).$$

Обозначим $\alpha^* = \arg\min_\alpha \breve{R}(V, Q_\alpha)$.

Теорема 2. *При* $N_x = \frac{N}{M}$ *имеет место*

$$\alpha^* = \frac{F(V) - 1}{F(V) + \frac{M-1}{N-M}} \approx 1 - \frac{1}{F(V)}.$$

Д о к а з а т е л ь с т в о. Из $N_x = \frac{N}{M}$ находим $C_x = \frac{N-1}{N-M}$.

Теперь

$$\breve{R}(V, Q_\alpha) = \frac{N}{(N-1)^2} \left(\left(\frac{M-1}{N-M} \alpha + 1 \right)^2 + F(V) \cdot \frac{M-1}{N-M} \cdot (1-\alpha)^2 \right) \sum_{x=1}^{M} d_x^2.$$

Дифференцируя выражение по α и приравнивая производную к 0, получаем искомое.

Заметим, что при $F(V) = 2$, что соответствует пороговому значению, при котором нулевая гипотеза отвергается, имеем $\alpha^* \approx 0.5$. Это вполне логично: когда гипотезы одинаково правдоподобны, решения усредняются с равными весами.

5. Обсуждение результатов

В разделе 2 рассматривалась классическая задача дисперсионного анализа. Пусть, например, задано некоторое множество городов, для каждого из которых имеется выборка объектов недвижимости с известными ценами. Требуется проверить гипотезу, одинакова ли средняя цена недвижимости в этих городах.

Один из стандартных методов решения этой задачи заключается в применении критерия Фишера.

Однако вместо явной проверки гипотез мы можем построить две решающие функции: первую в предположении равенства ожидаемой цены недвижимости по городам, вторую — в предположении, что цены разные. Предположение, при котором решение будет оценено как более точное, мы и примем в качестве гипотезы.

Для оценки точности решения одним из общепринятых подходов является метод скользящего экзамена.

Оказывается, что применение скользящего экзамена для задачи дисперсионного анализа эквивалентно классическому критерию Фишера, причём с «магическим» числом в качестве уровня значимости (которое соответствует значению статистики Фишера, равному 2). Например, для случая двух подвыборок этот уровень значимости примерно равен 0.16. При этом с ростом числа подвыборок (числа городов в примере) этот уровень значимости стремится к 0.

Этот факт выглядит непривычно. В классическом подходе к проверке гипотез уровень значимости выбирается произвольно. Полученный результат можно интерпретировать как то, что некоторые значения уровня значимости могут считаться в определённом смысле «более обоснованными». С другой стороны, слишком малые значения уровня значимости (при больших M) могут свидетельствовать о том, что оценка скользящего экзамена в этих условиях не является адекватной.

В следующих разделах рассматривалась задача, которую можно считать одной из постановок задачи регрессионного анализа. В качестве примера можно рассмотреть те же выборки недвижимости по городам, но цель теперь в том, чтобы оценить математическое ожидание цены недвижимости в каждом городе.

Решающая функция, оптимальная (в классе линейных комбинаций двух базовых моделей) с точки зрения критерия скользящего экзамена, оказывается совпадающей (при некоторых условиях) с решением, оптимальным в рамках некоторой байесовской постановки с нормальным распределением параметров модели.

Полученные результаты могут рассматриваться как частичное обоснование метода скользящего экзамена, который в настоящее время имеет статус эвристического.

СПИСОК ЛИТЕРАТУРЫ

1. **Бериков В.Б., Лбов Г.С.** Выбор оптимальной сложности класса логических решающих функций в задачах распознавания образов // Доклады Академии наук. 2007. Т. 417. № 1. С. 26–29.
2. **Бериков В.Б., Пестунов И.А.** Построение кластерного ансамбля для сегментации гиперспектральных изображений // Вычислительные технологии. 2016. Т. 21. № 1. С. 15–24.
3. **Воронцов К.В.** Комбинаторные оценки качества обучения по прецедентам // Докл. РАН, 394:2 (2004), С. 175–178.
4. **Генрихов И.Е., Дюкова Е.В., Журавлёв В.И.** О полных регрессионных решающих деревьях // Машинное обучение и анализ данных. 2016. Т. 2, №1. С. 116–126. doi: 10.21469/22233792.2.1.09.
5. **Загоруйко Н.Г., Лбов Г.С.** Проблема выбора в задачах анализа данных и управления // Сибирский журнал индустриальной математики. 2000. Т. III. № 1. С. 101–109.
6. **Ковалевский А.П., Шаталин Е.В.** Выбор регрессионной модели зависимости массы тела от роста с помощью эмпирического моста // Вестник томского государственного университета, 2015. №5 (37). С. 35–47.
7. **Линке Ю.Ю.** Асимптотические свойства одношаговых взвешенных M-оценок с приложениями к задачам регрессии // Теория вероятн. и ее примен., 62:3 (2017), С. 468–498.
8. **Мазуров Вл.Д., Хачай М.Ю.** Бустинг и полиномиальная аппроксимируемость задачи о минимальном аффинном разделяющем комитете // Тр. Ин-та математики и механики УрО РАН. 2013. Т. 19, № 2. С. 231–236.
9. **Неделько В.М.** Некоторые вопросы оценивания качества методов построения решающих функций // Вестн. Том. гос. ун-та. Управление, вычислительная техника и информатика. 2013. № 3 (24). С. 123–132.
10. **Неделько В.М.** Оценивание значимости переменных в моделях ранговой регрессии // Машинное обучение и анализ данных, 2017. Т. 3. № 2. С. 151–159. doi: 10.21469/22233792.3.2.05
11. **Неделько В.М.** Регрессионные модели в задаче классификации. // Сибирский журнал индустриальной математики. 2014. Т. XVII. №1. С. 86–98.
12. **Spirin N., Vorontsov K.** Learning to Rank with Nonlinear Monotonic Ensemble // Lecture Notes on Computer Science. 10th International Workshop on Multiple Classidier Systems (MCS-10). Naples, Italy, June 15–17, 2011. Pp. 16–25.
13. **Турков П.А., Красоткина О.В., Моттль В.В.** Отбор признаков в задаче классификации при смещении решающего правила // Известия ТулГУ. Естественные науки. 2015. № 4. С. 67–78.
14. **Хачай М.Ю., Поберий М.И.** Схема бустинга в задачах комбинаторной оптимизации, индуцированных коллективными алгоритмами обучения // Автомат. и телемех. 2014. № 4. С. 81–93.

Неделько Виктор Михайлович
канд. физ.-мат. наук, доцент
с.н.с.
Институт математики им. С. Л. Соболева СО РАН
пр. Коптюга, 4, г. Новосибирск, 630090 Россия
e-mail: nedelko@math.nsc.ru

Поступила 22.12.2017