

В заключение необходимо отметить, что научное исследование эффективности реализации какого-либо социального процесса всегда достаточно длительно и плохо поддается статистической обработке, но если опираться на основы теории передачи данных по сети как на упрощенную модель успешности реализации социального процесса, то можно гораздо эффективнее разрабатывать методики достижения поставленной цели как в вопросах коммуникационного взаимодействия, так и в вопросах восприятия любого другого значимого для человека сигнала.

Yu. B. Kozlova

A MODEL FOR INFORMATION COMMUNICATION

In this work the author dwells upon informatics domain knowledge projection on social and pedagogical processes. Besides, the author performs comparative analysis of information signal transmission theory by means of networks and pedagogical influence of information signal during lectures, and identifies potential domain in the global utilization of the information signal formation theory in social systems.

Keywords: data signal, pedagogical process, social systems.

© Козлова Ю. Б., 2011

УДК 519.24

А. В. Лапко, В. А. Лапко, И. И. Струков, А. А. Гусаров

НЕПАРАМЕТРИЧЕСКИЙ КЛАССИФИКАТОР И КРИТЕРИЙ КОЛМОГОРОВА В ЗАДАЧЕ СРАВНЕНИЯ ЭМПИРИЧЕСКОЙ И ТЕОРЕТИЧЕСКОЙ ФУНКЦИЙ РАСПРЕДЕЛЕНИЯ ОДНОМЕРНОЙ СЛУЧАЙНОЙ ВЕЛИЧИНЫ

Рассматривается применение непараметрического алгоритма распознавания образов в задаче сравнения эмпирической и теоретической функций распределения одномерной случайной величины. По результатам вычислительных экспериментов проводится анализ предлагаемой методики и критерия Колмогорова.

Ключевые слова: непараметрическая статистика, распознавание образов, проверка статистических гипотез, распределение случайных величин, критерий Колмогорова.

Непараметрические алгоритмы распознавания образов, соответствующие критерию максимального правдоподобия, успешно использовались при сравнении эмпирических функций распределения случайных величин [1]. Разработанная на их основе методика позволяет обойти трудно формализуемую проблему разбиения области возможных значений случайной величины на интервалы, которая свойственна критерию согласия Пирсона. Предлагаемая методика при проверке гипотезы о тождественности эмпирических законов распределения одномерных случайных величин имеет результаты, сопоставимые с результатами, полученными с помощью критерия Смирнова [2].

Цель данной работы состоит в развитии методики проверки гипотез о распределениях, основанных на непараметрических алгоритмах распознавания образов, и ее обобщении для решения задачи сравнения эмпирического и теоретического распределения случайных величин.

Библиографические ссылки

1. Олифер В. Г., Олифер Н. А. Компьютерные сети. Принципы, технологии, протоколы. СПб. : Питер, 2001.
2. Козлова Ю. Б., Адольф В. А. О методике совершенствования педагогического мастерства // Высш. образование сегодня. 2007. № 9. С. 35–37.
3. Козлова, Ю. Б. Организация совершенствования информационно-педагогического общения // Вестник СибГАУ. 2008. Вып. 4 (21). С. 210–214.

Критерий Колмогорова. Пусть $F_2(x)$ – известная функция распределения одномерной случайной величины x , предполагаемая непрерывной. Имеется реализация $V_1 = (x^i, i = \overline{1, n_1})$ случайной величины, которая определяет эмпирическое распределение $\bar{F}_1(x)$. Необходимо проверить либо опровергнуть гипотезу

$$H_0 : F_2(x) \equiv F_1(x)$$

о тождественности законов распределения.

Для проверки статистической гипотезы H_0 на основе критерия Колмогорова оценим по выборке V_1 функцию распределения $F_1(x)$ случайной величины x :

$$\bar{F}_1(x) = \frac{1}{n_1} \sum_{i=1}^{n_1} 1(x - x^i),$$

где

$$1(x - x^i) = \begin{cases} 0, & \text{если } x - x^i < 0; \\ 1, & \text{если } x - x^i \geq 0. \end{cases}$$

Анализируя значения эмпирической $\bar{F}_1(x)$ и теоретической $F_2(x)$ функций распределения, определим максимальное расхождение между ними:

$$\bar{D}_{12} = \max_x |\bar{F}_1(x) - F_2(x)|.$$

В соответствии с критерием Колмогорова [3] сравним полученное максимальное расхождение $\bar{D}_{12}$ с пороговым:

$$D_\alpha = \sqrt{-\ln\left(\frac{\alpha}{2}\right) / (2n_1)}, \quad (1)$$

где α – принятый уровень доверия (риск отвергнуть гипотезу H_0 , например, $\alpha = 0,05$).

Если выполняется условие $\bar{D}_{12} < D_\alpha$, то гипотеза H_0 справедлива, иначе анализируемые законы распределения различаются.

Непараметрический алгоритм распознавания образов в задаче проверки гипотезы H_0 . Известно, что если при решении двальтернативной задачи распознавания образов вероятность ошибки классификации ρ равна 0,5, то законы распределения случайных величин в области определения классов совпадают. Поэтому появляется возможность перехода от задачи сравнения законов распределения случайных величин к проверке гипотезы H_0 о равенстве вероятности ошибки распознавания образов значению 0,5.

На основании априорной информации осуществим синтез непараметрического распознавания образов, соответствующего критерию максимального правдоподобия [4]:

$$\bar{m}(x): \begin{cases} x \in \Omega_1, & \text{если } \bar{f}_{12}(x, c) \leq 0; \\ x \in \Omega_2, & \text{если } \bar{f}_{12}(x, c) > 0, \end{cases} \quad (2)$$

где $\bar{f}_{12}(x, c) = p_2(x) - \bar{p}_1(x, c)$ – непараметрическая оценка уравнения разделяющей поверхности между классами Ω_1, Ω_2 ; $p_2(x)$, $\bar{p}_1(x, c)$ – плотность вероятности распределения x в классе Ω_2 и оценка плотности вероятности $x \in \Omega_1$. При формировании $\bar{f}_{12}(x, c)$ используем непараметрическую оценку плотности вероятности одномерной случайной величины типа Розенблатта–Парзена [5]

$$\bar{p}_1(x, c) = \frac{1}{n_1 c} \sum_{i=1}^{n_1} \Phi\left(\frac{x - x^i}{c}\right), \quad (3)$$

восстанавливаемой по статистическим данным $V_1 = (x^i, i = \overline{1, n_1})$. Ядерные функции $\Phi(u)$ в статистике (3) удовлетворяют условиям

$$\Phi(u) = \Phi(-u), \quad 0 \leq \Phi(u) < \infty,$$

$$\int \Phi(u) du = 1, \quad \int u^2 \Phi(u) du = 1,$$

$$\int u^m \Phi(u) du < \infty, \quad 0 \leq m < \infty,$$

а значения их коэффициентов размытости c убывают с ростом количества n_1 элементов выборки V_1 . Бесконечные пределы интегрирования в приведенных условиях опускаются.

Выбор оптимального значения $\bar{c}$ коэффициента размытости непараметрического решающего правила $\bar{m}(x)$ осуществляется из условия минимума оценки вероятности ошибки распознавания образов

$$\bar{\rho}(c) = \frac{1}{n_1} \sum_{i=1}^{n_1} 1(\sigma(t), \bar{\sigma}(t)),$$

где индикаторная функция

$$1(\sigma(t), \bar{\sigma}(t)) = \begin{cases} 0, & \text{если } \sigma(t) = \bar{\sigma}(t); \\ 1, & \text{если } \sigma(t) \neq \bar{\sigma}(t), \end{cases}$$

где $\sigma(t)$ – «указание» о принадлежности наблюдения x^t из выборки V_1 классу Ω_1 . При вычислении $\bar{\rho}(c)$ «решение» $\bar{\sigma}(t)$ алгоритма (2) об отнесении наблюдения x^t к тому или иному классу определяется в соответствии со знаком статистики

$$\bar{f}_{12}(x^t) = p_2(x^t) - \frac{1}{n_1 c} \sum_{\substack{i=1 \\ i \neq t}}^{n_1} \Phi\left(\frac{x^t - x^i}{c}\right),$$

т. е. ситуация x^t , которая подается на контроль, исключается из процесса обучения непараметрического алгоритма (2).

В соответствии с критерием Колмогорова проверим гипотезу $\bar{H}_0$: $\rho(c) = 0,5$. Для этого сравним его пороговое значение (1) с отклонением $\bar{D}_{12} = |0,5 - \bar{\rho}(c)|$ при вероятности α отвергнуть правильную гипотезу $\bar{H}_0$.

Если выполняется соотношение $\bar{D}_{12} < D_\alpha$, то гипотеза $\bar{H}_0$ справедлива, иначе она отвергается.

Анализ результатов вычислительных экспериментов. Сравнение эффективности предложенной методики проверки гипотезы о распределениях случайных величин и критерия Колмогорова проводилось по данным вычислительных экспериментов. Последовательность случайных наблюдений $V_1 = (x^i, i = \overline{1, n_1})$ формировалась на основе датчиков случайных величин с равномерным $x^i = \varepsilon^i$ и нормальным $x^i = 0,5 + 0,15 \left(\sum_{j=1}^{12} \varepsilon^j - 6 \right)$, $i = \overline{1, n}$ законами распре-

ления. Случайные величины ε с равномерным законом распределения определены на интервале $[0, 1]$. При их формировании использовался стандартный датчик псевдослучайных величин среды визуального программирования Delphi.

Вычислительные эксперименты при фиксированных условиях исследования осуществлялись $N = 100$ раз. Если сравниваемые законы распределения отличались, то оценивалась вероятность P_1 отклонения гипотезы H_0 . При априори тождественных законах распределения случайных величин по полученным результатам оценивалась вероятность P_2 выполнения гипотезы H_0 . Риск α отвергнуть гипотезу H_0 принимался равным 0,05.

При синтезе непараметрического классификатора $\bar{m}(x)$ использовались параболические ядерные функции В. А. Епанечникова [6].

Рассмотрим результаты вычислительных экспериментов, когда сравниваемые законы распределения случайных величин разные (рис. 1). При $n_1 > 20$ рассматриваемые критерии безошибочно отклоняют гипотезу H_0 . В интервале малых значений $n_1 < 20$ применение сравниваемых критериев приводит к неудовлетворительным результатам, что, возможно, зависит от качества используемого датчика случайных величин.

Если априори законы распределения случайных величин тождественны, то для достаточно широкого интервала изменения объема $n_1 < 200$ анализируемой выборки оценки вероятности справедливости гипотезы H_0 при использовании критерия Колмогорова и предлагаемой методики сопоставимы (рис. 2). При увеличении $n_1 > 200$ наблюдается снижение эффективности исследуемой методики как при сравнении равномерных законов случайных величин (рис. 2, а),

так и, что особенно характерно, при сравнении нормальных законов случайных величин (рис. 2, б).

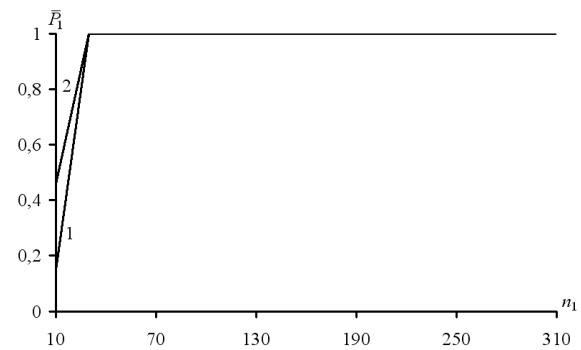
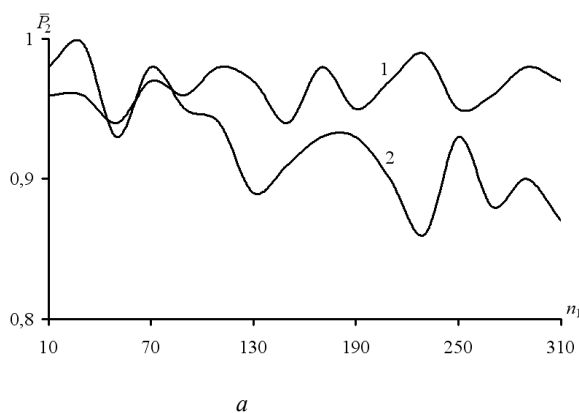


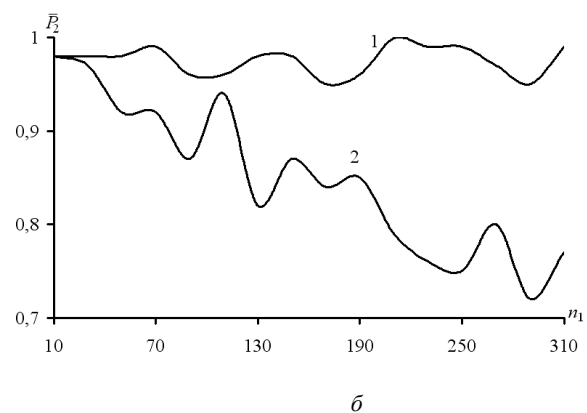
Рис. 1. Зависимости оценок $\bar{P}_1$ вероятностей отклонения гипотезы H_0 от объема n_1 выборки V_1 при сравнении равномерного и нормального законов распределения случайных величин. Кривые 1, 2 получены при использовании соответственно критерия Колмогорова и исследуемой методики

Показана возможность использования непараметрического алгоритма распознавания образов, соответствующего критерию максимального правдоподобия, в задаче сравнения эмпирической и теоретической функций распределения случайных величин. Существуют условия, когда использование предлагаемой методики и критерия Колмогорова приводит к сопоставимым результатам. К ним относятся задачи проверки гипотез при разных законах распределения случайных величин; при одинаковых законах распределения, когда объем сравниваемой последовательности принадлежит относительно малым значениям $n_1 < 200$.

Перспективность предлагаемой методики заключается в возможности ее обобщения для решения задач проверки гипотез о распределениях многомерных случайных величин с обходом проблемы разбиения области их значений на интервалы.



а



б

Рис. 2. Зависимости оценок $\bar{P}_2$ вероятностей справедливости гипотезы H_0 от объема n_1 выборки V_1 в условиях априори одинаковых законов распределения: равномерных (а); нормальных (б). Обозначения кривых соответствуют рис. 1

Библиографические ссылки

1. Лапко А. В., Лапко В. А. Непараметрические алгоритмы распознавания образов в задаче проверки статистической гипотезы о тождественности двух законов распределения случайных величин // Автоматизация. 2010. Т. 46. № 6. С. 47–53.
2. Лапко А. В., Лапко В. А. Применение непараметрического алгоритма распознавания образов в задаче проверки гипотезы о распределениях случайных величин // Системы управления и информационные технологии. 2010. Т. 41. № 3. С. 8–11.
3. Пугачев В. С. Теория вероятностей и математическая статистика. М. : Наука, 1979.
4. Непараметрические системы классификации / А. В. Лапко, В. А. Лапко, М. И. Соколов, С. В. Ченцов. Новосибирск : Наука, 2000.
5. Parzen E. On estimation of a probability density function and mode // Ann. Math. Statistic. 1962. Vol. 33. № 3. P. 1065–1076.
6. Епанечников В. А. Непараметрическая оценка многомерной плотности вероятности // Теория вероятности и ее применения. 1969. Т. 14. Вып. 1. С. 156–161.

A. V. Lapko, V. A. Lapko, I. I. Strukov, A. A. Gusarov

NONPARAMETRIC QUALIFIER AND KOLMOGOROV'S CRITERION IN THE TASK OF MATCHING OF EMPIRICAL AND THEORETICAL CUMULATIVE DISTRIBUTION FUNCTIONS OF AN ONE-DIMENSIONAL RANDOM VARIABLE

Application of nonparametric algorithm of pattern recognition in the task of matching of empirical and theoretical cumulative distribution functions of an one-dimensional random variable is considered. By results of computing experiments the analysis of an offered technique and Kolmogorov's criterion is carried out.

Keywords: nonparametric statistics, pattern recognition, check of statistical hypotheses, allocation of random variables, Kolmogorov's criterion.

© Лапко А. В., Лапко В. А., Струков И. И., Гусаров А. А., 2011

УДК 681.3

М. А. Масюк

АНАЛИЗ И ВИЗУАЛИЗАЦИЯ ВЗАИМОСВЯЗЕЙ НОРМАТИВНО-ПРАВОВЫХ ДОКУМЕНТОВ В СПРАВОЧНО-ПРАВОВЫХ СИСТЕМАХ

Рассмотрена сложившаяся в Российской Федерации ситуация, связанная со стремительным ростом количества принимаемых документов законодательского характера, значительная часть которых носит поправочный характер, т. е. содержит в себе ссылки на другие акты. Анализ множества документов с их взаимосвязями является сложным, но при этом рутинным занятием, которое требует наличия высококвалифицированных специалистов. Автором предложен комплексный подход к совершенствованию справочно-правовых систем и электронных баз данных путем интеграции в них системы, реализующей визуальное отображение взаимосвязей документов и их анализ на соответствие нормам законодательства.

Ключевые слова: нормативно-правовой документ, анализ, визуализация.

В последние годы в Российской Федерации и ее субъектах наблюдается стремительный рост законодательской деятельности, который, однако, не свидетельствует о высоком качестве правового регулирования [1]. С развитием законодательной базы существенно возрастает количество производных нормативно-правовых актов: законов, постановлений, указов, значительная часть которых носит поправочный характер, т. е. содержит в себе ссылки на другие документы с описанием вносимых в текст поправок или отменой ранее действующих документов. Такие ссылки одних документов на другие образуют единую связанную структуру – ориентированный граф, который можно рассматривать на множестве документов какой-либо электронной базы данных или справочно-

правовой системы в рамках законодательства Российской Федерации или ее отдельного субъекта. Однако представление общей картины путем анализа текстов является трудоемкой процедурой. Кроме того, существует вероятность возникновения противоречий с формальными правилами и нормами законодательного процесса, закрепленными в виде специальных документов [2; 3].

Автором предлагается комплексный подход к совершенствованию справочно-правовых систем и электронных баз данных путем интеграции в них системы, реализующей визуальное отображение взаимосвязей нормативно-правовых документов и их автоматический анализ на предмет соответствия нормам законодательства. Рассматриваемый подход облегчает про-