

УДК 004.94

Субботин С.А.

Запорожский национальный технический университет. Украина, г. Запорожье

МЕТОД ОТБОРА ОБРАЗЦОВ В ЗАДАЧАХ АВТОМАТИЗАЦИИ НЕРАЗРУШАЮЩЕГО КОНТРОЛЯ

Решена задача автоматизации отбора образцов для построения диагностических моделей в задачах неразрушающего контроля. Предложен метод отбора образцов, который обеспечивает сохранение в сформированной выборке важнейших топологических свойств исходной выборки, не требуя при этом загрузки в память ЭВМ исходной выборки, а также многочисленных проходов по исходной выборке, что позволяет сократить объём выборки и уменьшить требования к ресурсам ЭВМ, а также упростить построение диагностических моделей по прецедентам.

Ключевые слова: отбор образцов, выборка, редукция данных, неразрушающий контроль.

Введение

При решении задач автоматизации неразрушающего контроля, в частности, задач построения диагностических моделей на основе прецедентов [1–4] на практике нередко приходится оперировать выборками данных большого объема. Это влечёт за собой существенные затраты времени на обработку данных, а также требует наличия значительных объемов оперативной и дисковой памяти ЭВМ. Поэтому актуальной задачей является сокращение размерности выборок данных [1, 3–5].

Традиционным и наиболее широко применяемым подходом при решении данной задачи является использование методов отбора информативных признаков [1, 3–5], которые удаляют из исходного набора наименее информативные признаки, и методов конструирования признаков [5, 6], которые заменяют исходный набор признаков рассчитанным на его основе набором искусственных признаков меньшего размера. Однако, если изначально заданный набор признаков не является избыточным, либо объем выборки (число экземпляров в ней) чрезвычайно велик для представления и обработки в памяти ЭВМ, применение этих методов оказывается чрезвычайно затруднительным, а результаты их работы либо приводят к потере существенной для дальнейшего анализа информации, либо не позволяют сохранить исходную интерпретируемость данных.

Другим, существенно реже используемым на практике, подходом при решении данной задачи является сокращение объема выборки. Как правило, это реализуется посредством извлечения случайных подвыборок из исходной выборки [7–9], что может приводить к формированию нерепрезен-

тативных в топологическом смысле выборок вследствие не включения в них редко встречающихся экземпляров на границах классов, представленных в исходной выборке.

В [10–13] предложены переборные и эволюционные методы формирования выборок, а также модель (комплекс критериев) качества выборки, которые позволяют обеспечить формирование из исходной выборки подвыборок меньшего объема, обладающих в системе используемых критериев наилучшими свойствами. Однако для выборок большого объема применение данных методов и модели оказывается весьма затратным как с вычислительной точки зрения, так и с точки зрения ресурсов оперативной и дисковой памяти.

Целью данной работы является создание метода формирования и редукции выборок, позволяющего обрабатывать исходные выборки большого объема.

Постановка задачи

Пусть мы имеем исходную выборку $X = \langle x, y \rangle$ – набор S прецедентов о зависимости $y(x)$, $x = \{x^s\}$, $y = \{y^s\}$, $s = 1, 2, \dots, S$, характеризующихся набором N входных признаков $\{x_j\}$, $j = 1, 2, \dots, N$, где j – номер признака, и выходным признаком y . Каждый s -й образец (прецедент, экземпляр) представим как $\langle x^s, y^s \rangle$, $x^s = \{x_j^s\}$, где x_j^s – значение j -го входного, а y^s – значение выходного признака для s -го образца выборки, $y^s \in \{1, 2, \dots, K\}$, где K – число классов, $K > 1$.

Тогда задача сокращения объема выборки может быть представлена как задача формирования (выделения) из исходной выборки $X = \langle x, y \rangle$ подвыборки X^* , $X^* \subset X$, меньшего объема $S^* < S$, обла-



дающей наиболее важными свойствами исходной выборки.

Поскольку для задач интеллектуального анализа данных, связанных с автоматизацией поддержки принятия решений, наиболее важным является сохранение топологии классов, то формируемая подвыборка должна обеспечивать сохранение образцов исходной выборки, находящихся на границах классов.

Метод отбора образцов

Для обнаружения образцов, находящихся на границах классов, в общем случае необходимо решить задачу кластер-анализа, что требует определения расстояний между всеми образцами выборки в пространстве признаков. Это, в свою очередь, требует либо загрузки всей выборки в память ЭВМ (что не всегда возможно из-за ограниченного объема оперативной памяти), либо многократных проходов по исходной выборке (что вызывает значительные затраты машинного времени), а также приводит к необходимости хранить и обрабатывать матрицу расстояний между экземплярами большой размерности.

Для устранения отмеченных недостатков предлагается заменить обработку векторных описаний образцов на обработку их описаний в виде числовых скаляров, которые характеризуют положение экземпляров в пространстве признаков. При этом, заменив образцы, характеризующиеся N признаками, на представления в виде скаляров мы отобразим N -мерное пространство признаков в одномерное пространство.

Исходная выборка, будучи отображенной в одномерное пространство, позволит выделить на одномерной оси интервалы её значений, соответ-

ствующие кластерам разных классов в исходном N -мерном пространстве. Определив границы интервалов на одномерной оси, можно найти ближайшие к ним экземпляры, которые и составят формируемую подвыборку.

Приведенные выше идеи лежат в основе предлагаемого метода.

Этап инициализации. Задать исходную выборку данных $X = \langle x, y \rangle$.

Этап анализа характеристик классов. Разбить выборку X на K подвыборок $X(k)$, отдельных для образцов каждого класса: $X(k) = \{X(k) \cup \langle x^s, y^s \rangle \mid y^s = k\}$, $s = 1, 2, \dots, S$; $k = 1, 2, \dots, K$.

Для каждой подвыборки $X(k)$ определить по каждому признаку его минимальное $\min\{x_j^s \mid x^s \in X(k)\}$, максимальное $\max\{x_j^s \mid x^s \in X(k)\}$ и среднее значения для образцов соответствующего класса:

$$C_j^k = \frac{1}{S^k} \sum_{s=1}^S \{x_j^s \mid y^s = k\}. \quad (1)$$

Для каждой подвыборки $X(k)$, $k = 1, 2, \dots, K$, определить:

– частные поосевые нормированные расстояния от образцов до центров классов:

$$R_{(k)}(s)_j = \frac{C_j^k - x_j^s}{\max_s \{x_j^s \mid x^s \in X(k)\} - \min_s \{x_j^s \mid x^s \in X(k)\}}, \quad (2)$$

$$s = 1, 2, \dots, S, j = 1, 2, \dots, N;$$

– нормированные расстояния образцов до центров классов:

$$R_{(k)}(s) = \frac{1}{N} \sum_{j=1}^N |R_{(k)}(s)_j|, s = 1, 2, \dots, S; \quad (3)$$

– нормированные расстояния между экземплярами:

$$R_{(k)}(s, p) = \begin{cases} R_{(k)}(s, s) = 0, s = 1, 2, \dots, S; \\ R_{(k)}(p, s) = \frac{1}{2N} \sum_{j=1}^N |R_{(k)}(s)_j - R_{(k)}(p)_j|, s = 1, 2, \dots, S, p = s + 1, s + 2, \dots, S. \end{cases} \quad (4)$$

Этап устранения дублирующихся образцов. Целью этапа является выделение подмножеств эквивалентных и существенно похожих образцов и замена каждого такого подмножества на один его образец-представитель.

Чёткий дубляж: из каждой группы одинаковых образцов каждой подвыборки следует оставить только один образец: $X(k) = X(k) \setminus \{x^p \mid x^s \in X(k), x^p \in X(k), R_{(k)}(s) = R_{(k)}(p), R_{(k)}(s, p) = 0\}$, $S^k > 1$, $k = 1, 2, \dots, K$, $s = 1, 2, \dots, S$, $p = s + 1, \dots, S$.

Нечёткий дубляж: из каждой группы неодинаковых подобных образцов каждой подвыборки следует оставить только один образец: $X(k) = X(k) \setminus \{x^p$

$\mid p \neq s, x^s \in X(k), x^p \in X(k), \mid R_{(k)}(s) - R_{(k)}(p) \mid \leq \varepsilon_1(k), R_{(k)}(s, p) \leq \varepsilon_2(k, s, p)\}$, $S^k > 1$, $k = 1, 2, \dots, K$, $s = 1, 2, \dots, S$, $p = s + 1, \dots, S$, где $\varepsilon_1(k) = (\ln S^k)^{-1}$, $\varepsilon_2(k, s, p) = \exp(-\mid R_{(k)}(s) - R_{(k)}(p) \mid \ln S^k)$.

Этап выделения граничных образцов. Целью данного этапа является выявление образцов, находящихся вблизи границ классов, что позволит устранить остальные образцы, находящиеся внутри области класса.

Вначале необходимо определить индексы для всех образцов выборки относительно центров всех подвыборок:

$$I^s(k) = \text{round}(R_{(k)}(s) \ln S^k) + \frac{1}{\pi} \arccos \left(\frac{\sum_{j=1}^N C_j^k x_j^p}{\sqrt{\sum_{j=1}^N (C_j^k)^2} \sqrt{\sum_{j=1}^N (x_j^p)^2}} \right), \quad (5)$$

где round – функция округления до ближайшего целого числа.

Это позволит отобразить исходную выборку на одномерные оси $I(k)$ (заметим, что при этом произойдет потеря части информации вследствие неявного квантования пространства признаков при преобразовании).

Просматривая каждую одномерную ось $I(k)$ можно выделить скопления (области пространства) близко расположенных образцов одного класса, выделив интервалы для каждого из них $I(k) = \{i_l(k)\}$, где $i_l(k)$ – l -й интервал k -й оси, либо для простоты разбить эту ось на несколько равных

интервалов и определить доминирующий класс в каждом из них.

До тех пор, пока $\exists k, k = 1, 2, \dots, K : X(k) \neq \emptyset$, выполнять в цикле:

– если для области пространства, где расположены образцы k -го класса не существует попавших в нее образцов других классов ($\neg \exists s, s = 1, 2, \dots, S : y^s = k, F(k) \leq P(k), \forall p = 1, 2, \dots, S, s \neq p, y^p = k$), то данный класс расположен компактно и отделен от других классов. Следовательно, из образцов k -го класса в новую выборку следует включить лишь те образцы, которые находятся вблизи его внешней границы:

$$X^* = X^* \cup \{x^p \mid p = 1, 2, \dots, S : x^p \in X(k), I^p(k) \geq \frac{\alpha}{S^k} \sum_{s=1}^S \{I^s(k) \mid y^s = k\}, \quad (6)$$

где α – задаваемый пользователем коэффициент, регулирующий долю помещаемых в новую выборку экземпляров k -го класса (например, можно рекомендовать задавать $\alpha = 1$).

После чего необходимо исключить образцы k -го класса из дальнейшего рассмотрения: $X(k) = X(k) \setminus \{x^p \mid y^p = k, p = 1, 2, \dots, S$.

– если для области пространства, где расположены образцы k -го класса, существуют попавшие в нее

образцы других классов $\exists s, p : s = 1, 2, \dots, S, p = 1, 2, \dots, S, s \neq p, x^s \notin X^*, x^p \notin X^*, y^s \neq k, y^p = k, F(k) \leq P(k)$, но число образцов других классов $S^s(k)$, попавших в область k -го класса невелико: $S^s(k) \leq \beta S(k)$, где $S(k)$ – число образцов k -го класса, β – заданный коэффициент ($0 < \beta < 1$), то из образцов k -го класса в новую выборку следует включить лишь те образцы, которые находятся вблизи его внешней границы, а также образцы, ближайшие к образцам других классов:

$$X^* = X^* \cup \{x^p \mid p = 1, 2, \dots, S : x^p \in X(k), I^p(k) \geq \frac{\alpha}{S^k} \sum_{s=1}^S \{I^s(k) \mid y^s = k\}, \quad (7)$$

$$X^* = X^* \cup \left\{ x^q \mid \begin{array}{l} q = 1, 2, \dots, S, x^q \notin X^*, y^q = k, \exists p, p = 1, 2, \dots, S : \\ y^p \neq k, |I^q(k) - I^p(k)| \leq |I^s(k) - I^p(k)|, \forall s = 1, 2, \dots, S, y^s = k, s \neq q \end{array} \right\}. \quad (8)$$

Все экземпляры других классов, попавшие в область k -го класса, также следует включить в новую выборку: $X^* = X^* \cup \{x^s \mid s = 1, 2, \dots, S, p = 1, 2, \dots, S, s \neq p, x^s \notin X^*, y^s \neq k, y^p = k, F(k) \leq P(k)\}$;

– если для области пространства, где расположены экземпляры k -го класса, существуют попавшие в нее экземпляры других классов $\exists s, p : s = 1, 2, \dots, S, p = 1, 2, \dots, S, s \neq p, x^s \notin X^*, x^p \notin X^*, (y^s \neq k, y^p = k,$

$F(k) \leq P(k)$), но число экземпляров других классов, попавших в область k -го класса велико ($S^s(k) > \beta S(k)$), то на оси $I(k)$ следует выделить отдельные скопления экземпляров каждого класса и включить в новую выборку лишь те экземпляры, которые находятся вблизи его внешней границы, а также граничные экземпляры каждого интервала и экземпляры, ближайшие к ним:

$$X^* = X^* \cup \left\{ x^p \mid p = 1, 2, \dots, S : x^p \in X(k), I^p(k) \geq \frac{\alpha}{S^k} \sum_{s=1}^S \{I^s(k) \mid y^s = k\} \right\}, \quad (9)$$

$$X^* = X^* \cup \left\{ x^q \cup x^p \mid \begin{array}{l} q, p = 1, 2, \dots, S, x^q \notin X^*, x^p \notin X^*, y^q = k, y^p \neq k, \\ |\tau(x^q, k) - \tau(x^p, k)| = 1, \\ |I^q(k) - I^p(k)| \leq \gamma(k, \tau(x^q, k), \tau(x^p, k)) |I^s(k) - I^g(k)|, \\ y^s \in i_{\tau(x^q, k)}(k), y^g \in i_{\tau(x^p, k)}(k), s, g = 1, 2, \dots, S \end{array} \right\} \quad (10)$$

где $\tau(x^s, k) = \begin{cases} l, x^s \in i_l(k), l = 1, 2, \dots, L(k); \\ 0, \text{ в противном случае;} \end{cases}$

$L(k)$ – число интервалов значений, на которые разбита ось $I(k)$, γ – заданный коэффициент, регули-

$$\omega(k, l, m) = \begin{cases} \frac{\max_{\substack{s=1, 2, \dots, S; \\ g=1, 2, \dots, S}} \{ \|I^s(k) - I^g(k)\| y^s \in i_l(k), y^g \in i_m(k) \}}{\min_{\substack{s=1, 2, \dots, S; \\ g=1, 2, \dots, S}} \{ \|I^s(k) - I^g(k)\| y^s \in i_l(k), y^g \in i_m(k) \}}, & \min_{\substack{s=1, 2, \dots, S; \\ g=1, 2, \dots, S}} \{ \|I^s(k) - I^g(k)\| y^s \in i_l(k), y^g \in i_m(k) \} > 0; \\ 0, & \min_{\substack{s=1, 2, \dots, S; \\ g=1, 2, \dots, S}} \{ \|I^s(k) - I^g(k)\| y^s \in i_l(k), y^g \in i_m(k) \} = 0. \end{cases}$$

После чего из дальнейшего рассмотрения следует исключить экземпляры k -го класса, а также те экземпляры остальных классов, которые были включены в новую выборку: $X(q) = X(q) \setminus \{x^s \mid s = 1, 2, \dots, S, p = 1, 2, \dots, S, s \neq p, x^s \notin X^*, y^s = q, y^s \neq k, y^p = k, F(k) \leq F^p(k)\}$, $q = 1, 2, \dots, K, X(k) = X(k) \setminus \{x^p \mid x^p \in X^*, p = 1, 2, \dots, S\}$.

В результате выполнения предложенного метода будет сформирована выборка $X^* \subseteq X$.

Эксперименты и результаты

Для экспериментальной проверки работоспособности предложенного метода была разработана его программная реализация, с помощью которой проводились эксперименты по автоматизации отбора образцов для построения моделей в различных практических задачах неразрушающего контроля в технике и медицине [4, 14], характеристики которых приведены в таблице 1.

Результаты проведенных экспериментов подтвердили работоспособность и практическую применимость предложенного метода, а также реализующего его программного обеспечения. Как видно из таблицы 1, использование предложенного метода позволяет существенно сократить объём выборки, не требуя при этом многочисленных проходов по исходной выборке, что существенно снижает требования к ресурсам ЭВМ, обеспечивая при этом сохранение в сформированной подвыборке важнейших для последующего анализа топологических свойств исходной выборки.

рующий размер области вблизи межклассовых границ, экземпляры из которой включаются в формируемую выборку, $0 < \gamma(k, l, m) < \omega(k, l, m)$, где

Заключение

В работе решена актуальная задача формирования и редукции выборок для интеллектуального анализа данных.

Научная новизна результатов работы заключается в том, что впервые предложен метод формирования и редукции выборок, который обеспечивает сохранение в сформированной подвыборке важнейших для последующего анализа топологических свойств исходной выборки, не требуя при этом загрузки в память ЭВМ исходной выборки, а также многочисленных проходов по исходной выборке, что позволяет существенно сократить объём выборки, существенно уменьшает требования к ресурсам ЭВМ.

Практическая значимость результатов работы состоит в том, что разработано программное обеспечение, реализующее предложенный метод формирования и редукции выборок, а также проведены эксперименты по их исследованию при решении практических задач, результаты которых позволяют рекомендовать разработанный метод для использования на практике при решении задач интеллектуального анализа данных.

Дальнейшие исследования могут быть сосредоточены на разработке новых способов формирования описаний экземпляров в виде обобщённых показателей, разработке реализаций предложенного метода для параллельных вычислительных систем и распределенной обработки данных.

Таблица 1

Характеристики исходных и сформированных выборок

Задача	K	N	S	n	S^*	S^*/S
Неразрушающий контроль лопаток газотурбинных авиадвигателей [4]	2	10000	32	320000	8	0,25
Отбор лопаток газотурбинных авиадвигателей для упрочнения алмазным выглаживанием [4]	2	13	56	728	11	0,2
Диагностирование патологий плода по кардиотокограмме [14]	3	23	2126	48898	182	0,09

Работа выполнена в рамках госбюджетных научно-исследовательских тем Запорожского национального технического университета «Методы, модели и устройства принятия решений в системах распознавания образов» (№ гос. регистрации 0111U000059) и «Интеллектуальные информационные технологии автоматизации проектирования, моделирования, управления и диагностирования производственных процессов и систем».

Литература

- [1] Олійник А. О. Інтелектуальний аналіз даних: навчальний посібник [Текст] / А. О. Олійник, С. О. Субботін, О. О. Олійник. — Запоріжжя: ЗНТУ, 2012. — 271 с.
- [2] Рутковская Д. Нейронные сети, генетические алгоритмы и нечёткие системы [Текст] / Д. Рутковская, М. Пилиньский, Л. Рутковский; пер. с польск. И. Д. Рудинского. — М.: Горячая линия — Телеком, 2004. — 452 с.
- [3] Интеллектуальные информационные технологии проектирования автоматизированных систем диагностирования и распознавания образов: монография [Текст] / [С. А. Субботин, Ан. А. Олейник, Е. А. Гофман, С. А. Зайцев, Ал. А. Олейник; под ред. С. А. Субботина]. — Харьков : ООО «Компания Смит», 2012. — 317 с.
- [4] Прогрессивные технологии моделирования, оптимизации и интеллектуальной автоматизации этапов жизненного цикла авиационных двигателей: монография [Текст] / [А. В. Богуслаев, Ал. А. Олейник, Ан. А. Олейник, Д. В. Павленко, С. А. Субботин; под ред. Д. В. Павленко, С. А. Субботина]. — Запорожье: ОАО «Мотор Сич», 2009. — 468 с.
- [5] Субботин С. А. Формирование выборок и анализ качества моделей на основе нейронных и нейро-нечётких сетей в задачах диагностики и распознавания образов [Текст] / С. А. Субботин: монография. — Saarbrücken: LAP Lambert academic publishing, 2012. — 232 с. — (ISBN 978-3-8473-4471-1).
- [6] Jensen R. Computational intelligence and feature selection: rough and fuzzy approaches [Text] / R. Jensen, Q. Shen. — Hoboken: John Wiley & Sons, 2008. — 339 p.
- [7] Chaudhuri A. Survey sampling theory and methods [Text] / A. Chaudhuri, H. Stenger. — New York: Chapman & Hall, 2005. — 416 p.
- [8] Encyclopedia of survey research methods [Text] / ed. P. J. Lavrakas. — Thousand Oaks: Sage Publications, 2008. — Vol. 1–2. — 968 p.
- [9] Кокрен У. Методы выборочного исследования [Текст] / У. Кокрен; пер. с англ. И. М. Сони́на; под ред. А. Г. Волкова, Н. К. Дружинина. — М.: Статистика, 1976. — 440 с.
- [10] Subbotin S. A. The training set quality measures for neural network learning [Text] / S. A. Subbotin // Optical Memory and Neural Networks (Information Optics). — 2010. — Vol. 19. — № 2. — P. 126–139.
- [11] Субботин С. А. Комплекс характеристик и критериев сравнения обучающих выборок для решения задач диагностики и распознавания образов [Текст] / С. А. Субботин // Математичні машини і системи. — 2010. — № 1. — С. 25–39.
- [12] Субботин С. А. Критерии индивидуальной информативности и методы отбора экземпляров для построения диагностических и распознающих моделей [Текст] / С. А. Субботин // Біоніка інтелекту. — 2010. — № 1. — С. 38–42.
- [13] Субботин С. А. Методы формирования выборок для построения диагностических моделей по прецедентам [Текст] / С. А. Субботин // Вісник Національного технічного університету «Харківський політехнічний інститут»: зб. наук. праць. — Харків: НТУ «ХПІ», 2011. — № 17. — С. 149–156.
- [14] Cardiotocography Data Set [Electronic resource]. — Access mode: <http://archive.ics.uci.edu/ml/datasets/Cardiotocography>.



Subbotin S. A

Zaporizhzhya National Technical University. Ukraine, Zaporizhzhya

SAMPLING METHOD IN THE PROBLEMS OF AUTOMATION OF NONDESTRUCTIVE TESTING

The problem of automation of sampling for diagnostic model building in the problems of nondestructive testing was solved. The method of sampling which provides a saving of the most important topological properties of original sample in the formed sample without the need for downloading the original sample to the computer memory, and without numerous passages of the original sample is provosed. It allows to reduce the size of the sample and to reduce the resource requirements to a computer, and to simplify the diagnostic model construction by precedents.

Keywords: sampling, sample, data reduction, nondestructive testing.

References

- [1] Olijnik A. O. Intelktual'nij analiz danih: navchal'nij posibnik / A. O. Olijnik, S. O. Subbotin, O. O. Olijnik. — Zaporizhzhja: ZNTU, 2012. — 271 p. (In Ukrainian).
- [2] Rutkovskaja D. Nejronnye seti, geneticheskie algoritmy i nechjotkie sistemy / D. Rutkovskaja, M. Pilin'skij, L. Rutkovskij; per. s pol'sk. I. D. Rudinskogo. — M.: Gorjachaja linija — Telekom, 2004. — 452 p. (In Russian).
- [3] Intelktual'nye informacionnye tehnologii proektirovanija avtomatizirovannyh sistem diagnostirovanija i raspoznavanija obrazov: monografija / [S. A. Subbotin, An. A. Olejnik, E. A. Gofman, S. A. Zajcev, Al. A. Olejnik; pod red. S. A. Subbotina]. — Har'kov: OOO "Kompanija Smit", 2012. — 317 p. (In Russian).
- [4] Progressivnye tehnologii modelirovanija, optimizacii i intellektual'noj avtomatizacii etapov zhiznennogo cikla aviacionnyh dvigatelej: monografija / [A. V. Boguslaev, Al. A. Olejnik, An. A. Olejnik, D. V. Pavlenko, S. A. Subbotin; pod red. D. V. Pavlenko, S. A. Subbotina]. — Zaporozh'e: OAO "Motor Sich", 2009. — 468 p. (In Russian).
- [5] Subbotin S. A. Formirovanie vyborok i analiz kachestva modelej na osnove nejronnyh i nejro-nechjotkih setej v zadachah diagnostiki i raspoznavanija obrazov / S. A. Subbotin: monografija. — Saarbrücken: LAP Lambert academic publishing, 2012. — 232 p. — (ISBN 978-3-8473-4471-1). (In Russian).
- [6] Jensen R. Computational intelligence and feature selection: rough and fuzzy approaches / R. Jensen, Q. Shen. — Hoboken: John Wiley & Sons, 2008. — 339 p.
- [7] Chaudhuri A. Survey sampling theory and methods / A. Chaudhuri, H. Stenger. — New York: Chapman & Hall, 2005. — 416 p.
- [8] Encyclopedia of survey research methods / ed. P. J. Lavrakas. — Thousand Oaks: Sage Publications, 2008. — Vol. 1–2. — 968 p.
- [9] Kokren U. Metody vyborochnogo issledovanija / U. Kokren; per. s angl. I. M. Sonina ; pod red. A. G. Volkova, N. K. Druzhinina. — M.: Statistika, 1976. — 440 p. (In Russian).
- [10] Subbotin S. A. The training set quality measures for neural network learning / S. A. Subbotin // Optical Memory and Neural Networks (Information Optics). — 2010. — Vol. 19. — № 2. — P. 126–139.
- [11] Subbotin S. A. Kompleks harakteristik i kriteriev sravnenija obuchajushhih vyborok dlja reshenija zadach diagnostiki i raspoznavanija obrazov / S. A. Subbotin // Matematichni mashini i sistemi. — 2010. — № 1. — P. 25–39. (In Russian).
- [12] Subbotin S. A. Kriterii individual'noj informativnosti i metody otbora jekzemplarov dlja postroenija diagnosticheskikh i raspoznajushhih modelej / S. A. Subbotin // Bionika intelektu. — 2010. — № 1. — P. 38–42. (In Russian).
- [13] Subbotin S. A. Metody formirovanija vyborok dlja postroenija diagnosticheskikh modelej po precedentam / S. A. Subbotin // Visnik Nacional'nogo tehničnogo universitetu "Harkivskij politehničnij institut": zb. nauk. prac'. — Harkiv: NTU "HPI", 2011. — № 17. — P. 149–156. (In Russian).
- [14] Cardiotocography Data Set [Electronic resource]. — Access mode: <http://archive.ics.uci.edu/ml/datasets/Cardiotocography>.