References

1. Born, M., Wolf, E. (1999). Principles of Optics: 7th Edition. Cambridge University Press, 987.
2. He, B., Rui, X., Zhang, H. (2012). Transfer Matrix Method for Natural Vibration Analysis of Tree System. Mathematical Problems in Engineering, 1–19. doi:10.1155/2012/393204
3. Moroz, A. (2005). A recursive transfer-matrix solution for a dipole radiating inside and outside a stratified sphere. Annals of Physics, 315 (2), 352–418. doi:10.1016/j.aop.2004.07.002
4. Wu, Z. S., Wang, Y. P. (1991). Electromagnetic scattering for multilayered sphere: Recursive algorithms. Radio Science, 26 (6), 1393–1401. doi:10.1029/91rs01192
5. Gurwich, I., Kleiman, M., Shiloah, N., Cohen, A. (2000). Scattering of Electromagnetic Radiation by Multilayered Spheroidal Particles: Recursive Procedure. Appl. Opt., 39 (3), 470–477.. doi:10.1364/ao.39.000470
6. Grechko, L. G., Lerman, L. B., Shkoda, N. G. (2004). Scattering of electromagnetic waves on multilayered sphere. Bulletin of Kyiv University. Series: Physics & Mathematics, 3, 376–384.
7. Grechko, L. G., Lerman, L. B., Shkoda, N. G. (2004). Multi-layered ellipsoid in electric field. Bulletin of Kyiv University. Series: Physics & Mathematics, 1, 386–394.
8. Lerman, L. B. (1994). Oscillation of flat layered shells with local elastic supports. Int. Appl. Mech, 30 (2), 129−134.
9. Grechko, L. G., Lerman, L. B., Bilokrinizna, L. M. (2006). Surfaces modes in ellipsoidal multi-layered small particles. Bulletin of Kyiv University. Series: Physics & Mathematics, 4, 416–425.
10. Porodko, L. V., Lerman, L. B. (2013). Electrodynamic energy in layered spherical nanoparticles. Technology audit and production reserves, 1 (14), 41–44.

Рекомендовано до публікації д-р фіз.-мат. наук Розенбаум В. М.
Дата надходження рукопису 11.09.2014

Лерман Леонид Борисович, кандидат технических наук, с. н. с., Институт химии поверхности им. А. А. Чуйко НАН Украины, ул. Генерала Наумова, 17, г. Киев, Украина, 03164
e-mail: llerman@yandex.ru

Породько Лилия Владимировна, аспирант, Институт химии поверхности им. А. А. Чуйко НАН Украины, ул. Генерала Наумова, 17, г. Киев, Украина, 03164
e-mail: lilphys@mail.ru

УДК 519.711
DOI: 10.15587/2313-8416.2014.27392

АДАПТИВНАЯ СИСТЕМА КЛАССИФИКАТОРОВ МГУА

© Н. В. Кондрашова, Ал. В. Павлов, Ан. В. Павлов, В. А. Павлов

Показано, что модели самоорганизации, построенные обобщенным релаксационным итерационным алгоритмом (ОРИА), являются наиболее точными при проверке классификаторов на новых данных. Максимальная точность классификации зависит от целевой выборки, вида модели и внешнего критерия МГУА и состава системы классификаторов. Известный многорядный алгоритм с комбинаторной селекцией обобщенных переменных (МАКСО), имеет более гибкую настройку точности на рабочей выборке по сравнению с ОРИА, но гораздо меньшее быстродействие при решении задачи классификации.
Ключевые слова: обобщенный релаксационный итерационный алгоритм (ОРИА), многорядный алгоритм с комбинаторной селекцией обобщенных переменных (МАКСО).

It is shown that the self-organization models, built by Generalized Relaxation Iterative Algorithm (GRIA), are the most accurate when examining the classifiers on new data. The maximum classification accuracy depends on the target sample, the type of model and external criterion of GMDH and classifiers system. Known Multilayer Algorithm with Combinatorial Selection of Variables (MACSoV) has more flexible accuracy adjustment on different parts of sample compared with GRIA, but much lower speed operation in solving the classification problem.
Keywords: generalized relaxation iterative algorithm (GRIA), multilayer algorithm with combinatorial selection of variables (MACSoV).

1. Введение

Самоорганизация регрессионных моделей или метод группового учета аргументов (МГУА) в бóльшей степени относится к детерминированным методам. Входные и выходные данные, как правило, представляют собой зашумленные, непрерывные переменные, представленные дискретно. Наличие помехи приводит к построению адекватной шуму – неистинной модели [1]. С помощью МГУА, на основании значений признаков (категориальных и метрических переменных) объект может быть причислен к одному из нескольких заданных заранее классов, которые атрибутируются, как категориальные переменные. Тип переменной определяется в соответствии с классификацией измерительных шкал, предложенной в [2]. Покажем на примере, что при решении задачи классификации с категориальными и метрическими переменными целесообразно применять МГУА, обычно используемый при работе с непрерывными переменными. Учитывая, что построение моделей линейных по параметрам и входным переменным не всегда дает большую точность (в работе [3] показано, что точность классификации на всей выборке была равна около 51 %), будем строить модель в расширенном преобразованном пространстве и там находить релевантные признаки. Когда общее количество наблюдений относительно числа классов невелико, используется подход «искусственного» увеличения объема обучающей выборки для построения одного на каждый класс бинарного классификатора. За счет расширения пространства признаков, подхода «один против всех», построения адаптивной системы бинарных классификаторов и применения метода самоорганизации моделей повышается точность классификации на «свежих» данных по сравнению с другими методами [4].

2. Математическая постановка задачи классификации

Рассмотрим детерминистскую постановку задачи классификации, в которой считается что имеющееся множество признаков (переменных) достаточно для разделения классов, т. е. $R_i \cap R_j = \varnothing$ – выполняется гипотеза о компактности классов, где $R_i, R_j, i \neq j$ – области размещения большинства объектов в пространстве признаков. В качестве исходных задано множество наблюдений $\boldsymbol{d} \in W$, которые являются вектор-строками матрицы данных $\mathbf{X} \subset \Re^M$, $\dim \mathbf{X} = n \times M$, вектор-столбцы которой – $\mathbf{x}_i$, $i = 1,...,M$; M – количество входных переменных; n – количество наблюдений.

Заданы также векторы фиктивной переменной – выхода $\breve{\mathbf{h}}_k$, $\dim \breve{\mathbf{h}}_k = n \times 1$, элементы $\breve{h}_{k,i}$, $i = \overline{1,n}$ которого задают для различных наблюдений принадлежность к классам $k = \overline{1,K}$ значениями h^+, h_-, означающими соответственно: объект принадлежит или не принадлежит классу. Классификация производится по принципу отделения каждого класса от всех остальных (англ. “One vs All”). Поскольку дальнейшее описание справедливо для каждого из получаемых классификаторов, индекс класса k опускается.

Предполагается, что $h_i \in \Re$ является суммой значений известной функции $\breve{h}(\mathbf{d}_i)$ $i = \overline{1,n}$ и неизвестной случайной величины ξ_i с нулевым математическим ожиданием и ограниченной дисперсией $\sigma^2 \le \infty$,

$h_i = \breve{h}(\mathbf{d}_i) + \xi_i$.

Шум ξ и функция $\breve{h}(\mathbf{d})$ не зависят друг от друга, $\mathbf{d} \in \mathbf{X}_W$; где вектор $\xi = (\xi_1,...,\xi_n)^T$, $\dim \xi = n \times 1$.

В общем случае матрица признаков может содержать не только наблюдаемые признаки, но и их нелинейные преобразования. Предполагается, что функции преобразований $f_j(\mathbf{d}) \in \Phi$, переводят нас из пространства признаков – как количественных, так и категориальных – размерности M в пространство признаков размерности m, $f_j(\mathbf{d}) : \Re^M \to \Re^m$, где $\Phi = \{f_j(\mathbf{d})\}_{j=\overline{1,m}}$ – заданное конечное множество функций преобразований, $\mathbf{d} \in \mathbf{X}_W$, $m > M$.

Рассмотрим модели $h(\mathbf{d})$, линейно зависящие от переменных $f_j(\mathbf{d})$, $j = \overline{1,m}$, где $d_j \in \Re$ – элементы вектор-строки $\mathbf{d} \in \mathbf{X}_W$. Обозначим $f_j(\mathbf{d}) \overset{\Delta}{=} f(d)_j$, $j = \overline{1,m}$. Функция $h(\mathbf{d})$ является нелинейной по исходным переменным и линейной по параметрам $\theta_{j,k}$:

$$h(\mathbf{d}, \Theta_q) = \theta_{0,q} + \sum_{j=1}^{m} \theta_{j,q} \cdot f(d)_j, \quad \Theta_q \in \Re^{s_q}, \quad q = \overline{1,Q}. \qquad (1)$$

Векторы неизвестных параметров Θ_q, $q = \overline{1,Q}$ имеют размерность $\dim \Theta_q = (m+1) \times 1$. Сложность структуры модели s_q определяется количеством компонент вектора Θ_q. Цель МГУА – найти оптимальную модель (скалярное произведение) $h^* = \left\langle f^*(\mathbf{d}), \widehat{\Theta}^* \right\rangle$ – результат решения следующих двух задач:

1) Поиска оценок параметров $\widehat{\Theta}_q$, $q = \overline{1,Q}$ $\forall$ $\mathbf{d}_q \in D$ при решении Q задач непрерывной оптимизации:

$$\widehat{\Theta}_q = \arg\min_{\Theta_q \in \Re^{s_q}} QR(h(\mathbf{d}_q), \breve{h}(\mathbf{d}_q, \Theta_q)), \forall\, \mathbf{d}_q \in \mathbf{X}_A \subset \mathbf{X}_W, q = \overline{1,Q}, \qquad (2)$$

где A – обучающая выборка – множество наблюдений, которое используется для оценивания параметров Θ_q, $q = \overline{1,Q}$.

2) Поиска оптимальной модели при решении Q задач дискретной оптимизации:

$$h^*(\mathbf{d}, \widehat{\Theta}^*) = \arg\min_{q=\overline{1,Q}} CR(h(\mathbf{d}_q), \widehat{h}(\mathbf{d}_q, \widehat{\Theta}_q)), \mathbf{d}_q \in \mathbf{X}_B \subset \mathbf{X}_W, \qquad (3)$$

где B – проверочная выборка – множество наблюдений, используемое для оценивания качества модели $\widehat{h}(\mathbf{d}_q, \widehat{\Theta}_q)$

$$\widehat{h}(\mathbf{d}_q, \widehat{\Theta}_q) = \widehat{\theta}_{q,0} + \sum_{j=1}^{m} \widehat{\theta}_{q,j} f(d)_j, \quad q = \overline{1,Q}.$$

Здесь $f = (d)_j$, $j = \overline{1,m}$ - некоторые элементы вектора столбца $f_j(\mathbf{d})$, где $\mathbf{d} \in \mathbf{X}_B$. Исходная выборка $W = A \cup B \cup C$, $A \cap B = \varnothing$, $B \cap C = \varnothing$, $A \cap C = \varnothing$.

Критерий QR – критерий качества оценок параметров $\widehat{\Theta}_k$, CR – критерий качества модели. В качестве критерия QR в МГУА используется остаточная сумма квадратов:

$$RSS_A = \left\| \mathbf{h}_A - \breve{\mathbf{X}}_A \widehat{\Theta}_{A,q} \right\|^2, \ q = \overline{1,Q}.$$

В качестве внешнего критерия в МГУА часто используется *критерий регулярности*:

$$AR_{A/B} = \left\| \mathbf{h}_B - \breve{\mathbf{X}}_B \widehat{\Theta}_{A,q} \right\|^2, \; q = \overline{1,Q}.$$

Полностью заданную функцию $\widehat{h}(\mathbf{d}_q, \widehat{\Theta}_q)$ назовем решением или моделью, а *функцию принадлежности* $h^*(\mathbf{d}, \widehat{\Theta}^*)$ – оптимальным решением задачи (оптимальной моделью).

Задачи (2) и (3) называются проблемой построения структуры и параметров функции принадлежности классификатора МГУА.

При построении системы дифференциальной диагностики по принципу «один против всех» создается система K классификаторов. Матрица исходных данных имеет M исходных переменных и n наблюдений.

3. Обзор работ по применению МГУА для классификации

В работах [5–7] развивается идея синтеза дискриминантной функции для разделения классов на основе данных обучающей и проверочной выборок по МГУА [8]. Идея применения дополнительной выборки для оптимизации пространства исходных признаков не нова, она была опубликована в работе [9], в русском переводе [4]. Известен итерационный алгоритм МГУА для поиска дискриминантной функции по МГУА в задаче классификации [10].

Мы рассмотрим применение *метода самоорганизации моделей* для построения адаптивной системы классификаторов, полученных релаксационными итерационными алгоритмами МАКСО [11] и ОРИА [12].

4. Решение задачи классификации

Прежде всего, для работы ОРИА необходимо сформировать преобразованное пространство переменных размерности $n \times m$ из исходного пространства переменных $n \times M$, где в модели могут присутствовать все переменные $x_i, i = \overline{1,M}$, их обратные величины $1/x_i$, их перемножения $x_i x_j$ $i, j = \overline{1,M}$ и $x_i(1/x_j)$, $i \neq j$. Пусть M=11, тогда $m = C_{2M-1}^2 - 1 = C_{21}^2 - 1 = 209$, где число переменных, для которых строятся сочетания без повторений, равно (2M-1), т. к. исключаются перемножения $x_i(1/x_i) = 1$). Если присутствуют триплеты $x_i x_j x_\ell$, то оценкой дополнительного числа сочетаний будет $m = C_{21}^3 - 1 = 1329$. Если присутствуют перемножения четырех переменных $x_i x_j x_\ell x_n$, то такой оценкой будет $m = C_{21}^4 - 1 = 5985$. Однако формирование и хранение матриц со всеми четверками сомножителей при работе GRIA невозможно, поскольку оперативная память ограничена. Поэтому построение моделей со сложностью больше трех сомножителей мы не рассматриваем.

МАКСО формирует свои мультипликативные переменные направленным перебором, динамически освобождая оперативную память для формирования следующего одночлена. К его достоинствам относится экономия оперативной памяти, поскольку сохраняются только матрицы, содержащие центрированные исходные и обратные им переменные.

Этими релаксационными алгоритмами итеративно направленным перебором при свободе выбора алгоритма равной 250 строятся модели K бинарных классификаторов «один против всех»

$$\widehat{h}_k(\mathbf{X}_{A,r+1}, \widehat{\Theta}_{r+1}) = \widehat{\varpi}_{r+1} \widehat{h}_k(\mathbf{X}_{A,r}, \widehat{\Theta}_r) + \widehat{\omega}_{r+1} \phi(\mathbf{X}_{A,r+1}),$$

$$k = \overline{1,K}, \quad r = 1, 2, \dots .$$

Кроме задания параметра алгоритма – свобода выбора – в них можно задать наличие или отсутствие обратных функций, максимальное количество перемножений для исходных переменных в функции $\phi(\mathbf{X}_{r+1})$ и правило останова: 1) максимальное число итераций $r<R$, либо 2) правило увеличения значения внешнего критерия $CR_{r+1} > CR_r$, 3) правило достижения малых изменений значений критерия от итерации к итерации.

Начало итерационного процесса имеет вид

$$\widehat{h}_k(\mathbf{X}_{A,1}, \widehat{\Theta}_1) = \widehat{\omega}_1 \phi(\mathbf{X}_{A,1}), \quad k = \overline{1,4}; \quad r = 0.$$

Линейно входящие в модель классификатора параметры представлены вектором, компоненты которого получаются рекуррентно по формуле:

$$\widehat{\Theta}_{r+1} = (\widehat{\varpi}_{r+1} \widehat{\Theta}_r, \widehat{\omega}_{r+1}), \; r = 0, 1, 2, \dots, \; \widehat{\Theta}_0 = \widehat{\omega}_1.$$

Для оценки качества классификации мы используем критерий

$$CR_G = \frac{n_G^*}{n_G} 100\%, \; G = \{A, B, C, U\},$$

где n_G^* – число правильно классифицированных к общему числу n_G объектов в выборке G, где U и C – рабочая и экзаменационная выборки, $U = A \cup B$.

Оказалось, что наиболее точный набор классификаторов на обучающей выборке содержит классификаторы, полученные веерным алгоритмом [13]. Идея веерного алгоритма была реализована на базе МАКСО [11] для построения мультипликативных членов. В качестве критерия для поиска оптимальной модели (3) в МАКСО используется комбинированный критерий точности выбора лучших моделей

$$CR = \alpha RSS_A + (1-\alpha) AR_{B/A}, \; \alpha \in [0;1],$$

где α – вес ошибок обучающей выборки; (1-α) – вес ошибок проверочной выборки. Это позволяет получать адаптивную систему классификаторов в зависимости от целевой выборки, на данных которой обеспечивается максимальная точность. Так как функция принадлежности принимает действительные значения, $h_k^*(\mathbf{d}, \widehat{\Theta}^*) \in \Re$, то можно применить простое правило: принадлежность объекта наблюдения к классу выбирать по максимальному значению, что позволяет избегать конфликтных ситуаций в случае ошибок k-х классификаторов первого и второго рода

$$k^*(\mathbf{d}) = \arg \max_{k=\overline{1,K}} h_k^*(\mathbf{d}, \widehat{\Theta}^*).$$

Изначально фиктивная переменная

принадлежности $h_k(\cdot)$ имеет одинаковое абсолютное значение принадлежности объекта к классу в каждом из k-х классификаторов и одинаковое абсолютное значение отнесения его к другим классам. Поэтому получаемые по моделям значения функций принадлежности по исходным заданиям не имеют приоритета друг перед другом, а получаются в результате моделирования взаимосвязи признаки-класс.

В данной работе строятся системы $S_G(a)$ классификаторов вида μ-ОРИА и (K-μ)-МАКСО, где μ- число классификаторов ОРИА, a – алгоритм

$S_G^*(a) = \arg \max\limits_{a \in \Omega_a, \mu = \overline{0,K}} CR_G(\mu, \Omega_a)$, Ω_a ={ОРИА, МАКСО}, G={A, B, C, U,W}.

5. Результаты моделирования систем классификаторов

На рис. 1 видно, что все четыре классификатора ОРИА$_3$ системы имеют наибольшую точность на проверочной (CR_B=100 %) и на экзамене (CR_C=80 %), а на обучающей выборке точность CR_A=81,66 %. Индекс 3 у ОРИА$_3$ означает максимальное число сомножителей в одночленах. Чтобы повысить точность системы CR_A с 81,7 до 85 %, не потеряв максимума точности на экзамене CR_C, можно в составе классификаторов ОРИА один из четырех классификаторов заменить на МАКСО-классификатор (на рисунке это обозначено, как 3-ОРИА$_3$ и 1-МАКСО$_4$). При этом точность на проверочной выборке ОРИА$_3$ в комбинации с различным числом МАКСО-классификаторов, остается максимальной CR_B=100 %. Хотя точность CR_A растет, точности CR_C и CR_B падают, и для системы четырех классификаторов МАКСО$_4$ (обозначено 4-МАКСО$_4$) они минимальны CR_C=60 %, CR_B = 90 %. На рисунке видно, что точность классификации ОРИА падает, если использовать более простые нелинейные модели, например, только с ковариациями переменных. Точность четырех ОРИА-классификаторов, когда присутствует умножение максимально двух переменных (обозначено 4-ОРИА$_2$) остается высокой только на проверочной выборке (CR_B=100 %), на экзамене она уменьшается до CR_C= 60 %. Особенно значительно падает точность обучения ОРИА$_2$ по сравнению с ОРИА$_3$-классификаторами (до CR_A=68,3 %).

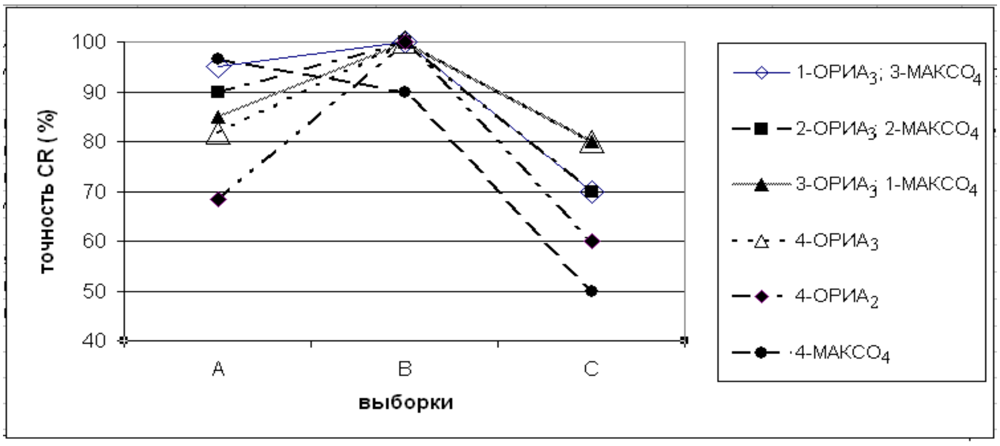

Рис. 1. Точность классификации моделями МАКСО и ОРИА системой из четырех классификаторов

Таблица 1

Результаты быстродействия и точности

Название	Время счета	Точность CR_A, %	Точность CR_B, %	Точность CR_C, %	Процессор
4-ОРИА$_3$	18 мин.	81,7	100	80	Pentium III
4-ОРИА$_4$	≈22 дня	96,3	90	50	Core i3 M

Табл. 1 представляет результаты сравнительного быстродействия и точности релаксационных итерационных алгоритмов, строящих нелинейные по исходным переменным модели. В ней видно преимущество ОРИА: он считает и точнее, и намного быстрее.

6. Выводы

Продемонстрирована целесообразность применения метода группового учета аргументов при решении задачи классификации по категориальным и метрическим исходным данным. Существует оптимальные величины степени нелинейности моделей и свободы выбора алгоритма, при которой точность на обучении и экзамене системы классификаторов максимальна.

Результаты модифицированного МАКСО и ОРИА, приспособленных для решения задачи дифференциальной диагностики, показывают преимущество последнего в точности на новых данных (экзамене) в 1,6 раза и особенно в быстродействии в 1760 раз. Для увеличения точности обучения можно составить комбинированную систему бинарных классификаторов, к которой некоторые из классификаторов, полученные ОРИА, заменить на МАКСО-классификаторы. Однако при замене больше, чем одного классификатора ОРИА на классификатор МАКСО наряду с увеличением точности обучения падает точность на экзамене.

Литература

1. Ивахненко, А. Г. Помехоустойчивость моделирования [Текст] / А. Г. Ивахненко, В. С. Степашко. – К.: Наукова думка, 1985. – 214 с.
2. Stevens, S. S. On the Theory of Scales of Measurement [Text] / S. S. Stevens // Science. – 1946. – Vol. 2684, Issue 103. – P. 677–680.
3. Кондрашова, Н. В. Решение задачи медицинской диагностики линейным дискриминантным анализом и МГУА [Текст] / Н. В. Кондрашова, В. А. Павлов, А. В. Павлов // УСиМ. – 2013. – № 2. – С. 79–88.
4. Ким, Дж.-О. Факторный, дискриминантный и кластерный анализ: Пер. с англ. [Текст] / Дж.-О. Ким, Ч. У. Мьюллер, У. Р. Клекка и др.; под ред. И. С. Енюкова. – М.: Финансы и статистика, 1989. – 215 с.
5. Сарычев, А. П. Дискриминантный анализ с обучающими и проверочными подвыборками наблюдений [Текст] / А. П. Сарычев // Автоматика. – 1991. – № 2. – С. 55–59.
6. Сарычев, А. П. Решение задачи дискриминантного анализа в условиях структурной неопределенности на основании МГУА [Текст] / А. П. Сарычев // Проблемы управления и информатики. – 2008. – № 3. – С. 100–112.
7. Sarychev, A. P. GMDH-Based Criterion for Optimal Set Features Determination in Nonlinear Discriminant Analysis [Text] : proc. of 3d intern. conf. ICIM’2010 / A. P. Sarychev, L. V. Sarycheva // Inductive Modeling. Yevpatoria, (Ukraine), 2010. – P. 40-43.
8. Ивахненко, А. Г. Системы эвристической самоорганизации в технической кибернетике [Текст] / А. Г. Ивахненко. – К.: Техніка, 1971. – 364 с.
9. Klecka, W. R. Discriminant Analysis [Text] / W. R. Klecka – // Quantitative Applications in the Social Sciences. – Thousand Oaks, CA: Sage Publications, 1980. http://dx.doi.org/10.4135/9781412983938
10. Сарычев, А. П. Итерационный алгоритм МГУА для синтеза разделяющей функции в задаче дискриминантного анализа [Текст] / А. П. Сарычев // Автоматика. – 1988. – № 2. – С. 20–24.
11. Павлов, О. В. Алгоритми самоорганізації в задачах підвищення інформативності геометричних моделей процесів, заданих точковим каркасом [Текст] : дис. ...канд. техн. наук / О. В. Павлов. — К., 2006. — 197 с.
12. Павлов, А. В. Обобщённый релаксационный итерационный алгоритм МГУА [Текст] : зб. наук. праць / А. В. Павлов // Індуктивне моделювання складних систем. – Київ: МННЦІТС НАНУ, 2011. – С. 95–108.
13. Кондрашова, Н. В. Многорядный алгоритм веерных решений [Текст] / Н. В. Кондрашова, В. А. Павлов, А. В. Павлов // Вісник Національного технічного університету України „КПІ”. Інформатика, управління та обчислювальна техніка. – 2006. – № 45. – С. 218–227.

References

1. Ivakhnenko, A. G., Stepashko, V. S. (1985). Pomehoustojchivost' modelirovanija [Simulation noise immunity]. Kyiv, USSR :Naukova dumka, 214.
2. Stevens, S. S. (1946). On the Theory of Scales of Measurement. Science, 2684 (103), 677–680.
3. Kondrashova, N. V., Pavlov, V. A., Pavlov, A. V. (2013). Reshenie zadachi medicinskoj diagnostiki linejnym diskriminantnym analizom i MGUA [Problem solution of medical diagnosis by linear discriminant analysis and GMDH]. USiM, 244 (2), 79–88.
4. Yenyukov, I. S. (Ed.) (1989). Faktornyj, diskriminantnyj i klasternyj analiz [Factorial, Discriminant and Cluster analyses]. Moscow, USSR : Finance and Statistics, 215.
5. Sarychev, A. P. (1991). Diskriminantnyj analiz s obuchajushhimi i proverochnymi podvyborkami nabljudenij [Discriminant analysis with training, and check the subsample of observations]. Avtomatika, 2, 55–59.
6. Sarychev, A. P. (2008). Reshenie zadachi diskriminantnogo analiza v uslovijah strukturnoj neopredelennosti na osnovanii MGUA [Problem solution of discriminant analysis in the conditions of structural uncertainty based on GMDH]. Problemy upravleniya i informatiki, 3, 100–112.
7. Sarychev, A. P., Sarycheva, L. V. (2010). GMDH-Based Criterion for Optimal Set Features Determination in Nonlinear Discriminant Analysis. Proc. of 3d International Conference on Inductive Modeling, ICIM’2010. Yevpatoria (Ukraine), 40–43.
8. Ivakhnenko, A. G. (1971). Sistemy jevristicheskoj samoorganizacii v tehnicheskoj kibernetike [Systems of heuristic self-organization in technical cybernetics]. Kyiv, USSR, Tehnika, 364.
9. Klecka, W. R. (1980). Discriminant analysis. Quantitative Applications in the Social Sciences Series, Thousand Oaks, CA: Sage Publications. http://dx.doi.org/10.4135/9781412983938
10. Sarychev, A. P. (1988). Iteracionnyj algoritm MGUA dlja sinteza razdeljajushhej funkcii v zadache diskriminantnogo analiza [Iterative GMDH algorithm for synthesis separating function in problem of discriminant analysis]. Avtomatika, 2, 20–24.
11. Pavlov, O. V. (2006). Algorytmy samoorganizacii' v zadachah pidvyshhennja informatyvnosti geometrychnyh modelej procesiv, zadanyh tochkovym karkasom [Algorithms for self-organization in problems of increasing informativity geometric models for the processes specified with help of points framework]. Kyiv, 197.
12. Pavlov, A. V. (2011). Obobshhjonnyj relaksacionnyj iteracionnyj algoritm MGUA [Generalized relaxation iterative algorithm of GMDH]. Inductive Modelling of complex Systems. Kyiv: IRTC IT&S, 2, 95–108.
13. Kondrashova, N. V., Pavlov, V. A., Pavlov, A. V. (2006). Mnogorjadnyj algoritm veernyh reshenij [Multilayered algorithm of fan solutions]. Herald National Technical Univ. ’KPI’. Informatics, Management and Computer Science, 45, 218–227.

Рекомендовано до публікації д-р техн. наук Степашко В. С.
Дата надходження рукопису 26.09.2014

Кондрашова Нина Владимировна, кандидат технических наук, с. н.с., Отдел информационных технологий и индуктивного моделирования, Международный научно-учебный центр информационных технологий и систем НАН Украины и МОН Украины, пр. Глушкова, 40, Киев, Украина, 03680
E-mail: nkondrashova@ukr.net

Павлов Александр Владимирович, кандидат технических наук, ассистент, кафедра начертательной геометрии, инженерной и компьютерной графики, Национальный технический университет Украины "КПИ", пр. Победы, 37, Киев, Украина, 03056
E-mail: Alexander_mk@ukr.net
Павлов Андрей Владимирович, кандидат технических наук, научный сотрудник, Отдел информационных технологий и индуктивного моделирования, Международный научно-учебный центр информационных технологий и систем НАН Украины и МОН Украины, пр. Глушкова, 40, Киев, Украина, 03680
E-mail: andriypavlove@gmail.com
Павлов Владимир Анатольевич, кандидат технических наук, доцент, кафедра биомедицинской кибернетики, Национальный Технический Университет "КПИ", пр. Победы, 37, Киев, Украина, 03056
E-mail: VPavlo@bk.ru

УДК 681.3.016:519.17:541.66
DOI: 10.15587/2313-8416.2014.27967

ПРОГНОЗИРОВАНИЕ КРИТИЧЕСКИХ СВОЙСТВ ФРЕОНОВ С ИСПОЛЬЗОВАНИЕМ НОВЫХ ИНВАРИАНТОВ ВЗВЕШЕННЫХ ГРАФОВ

© Ю. А. Кругляк, И. В. Передунова

Предлагается новый подход в задачах «структура – свойство» с использованием инвариантов полностью взвешенных графов для количественного описания критических свойств фреонов. Сформулирован общий принцип построения топологических инвариантов полностью взвешенных графов и предложены два новых инварианта, примененных для расчета критических свойств фреонов метанового, этанового и пропанового рядов без привлечения экспериментальных данных и закономерностей.
Ключевые слова: фреон, критические свойства, критическая температура, критическое давление, критический объем, молекулярные графы, инварианты графов, взвешенные графы, индекс паросочетаний.

It is proposed a new approach to the "structure – property" problems with usage the invariants of the fully weighted graphs for quantitative description of the critical properties of freon. A general principle of topological invariants construction of fully weighted graphs is formulated and propose two new invariants applied to calculate the critical properties of freon of methane, ethane and propane series without the involvement of experimental data and patterns.
Keywords: freon, critical properties, critical temperature, critical pressure, critical volume, molecular graphs, graph invariants, fully weighted graphs, matching index.

1. Введение

Проблема установления связи между структурой молекул и свойствами молекулярных веществ сложна и многопланова. Причиной тому является не только разнообразие свойств веществ и трудности их экспериментального исследования как и структуры и свойств составляющих их молекул, но и следующее обстоятельство. Основными методами исследования связи «структура – свойство» являются методы регрессионного и корреляционного анализа и распознавания образов. Эти методы оперируют с численными характеристиками структуры молекул. Естественные численные характеристики молекулярной структуры такие как длины связей, валентные и диэдрические углы и разнообразные квантовохимические расчетные свойства молекул часто с успехом используются в методах распознавания образов, но мало пригодны для регрессионного и корреляционного анализа. Для этих методов предпочтительны интегральные численные характеристики молекулярной структуры. В этом плане большой интерес вызывают топологические инварианты, позволяющие описать структуру молекулы одним числом [1–3].

Под инвариантом молекулярного графа подразумевают такую величину, которая принимает одно и то же численное значение при любой произвольной нумерации вершин графа. Инварианты молекулярных графов получили в литературе название топологических индексов (ТИ). Под молекулярным графом подразумевают такой граф, вершины которого находятся во взаимно-однозначном соответствии с атомами рассматриваемой молекулы, а ребра – с химическими связями. Применение ТИ в качестве численных интегральных характеристик структуры молекул для установления связи «структура – свойство» имеет ряд преимуществ. Во-первых, топологическое описание молекул опирается на хорошо разработанную теорию графов. Во-вторых, ТИ вычисляются только на основе структурной формулы молекул. В-третьих, для таких расчетов не требуется больших вычислительных ресурсов.