УДК 519.95

ОБ ЭФФЕКТИВНОСТИ МЕТРИЧЕСКИХ АЛГОРИТМОВ РАСПОЗНАВАНИЯ НА БАЗЕ k БЛИЖАЙЩИХ СОСЕДЕЙ

Игнатьев Н.А., *Турсунмуротов Д.Х.

*mr.davrbek@mail.ru

Национальный университет Узбекистана имени Мирзо Улугбека,
100174, Узбекистан, г. Ташкент, ул. Университетская, д. 4.

Рассматривается решение задачи распознавания с двумя классами с применением метрических алгоритмов на базе k ближайших соседей. Вариантами решения предлагаются: классический алгоритм KNN на основе непараметрической оценки плотности распределения вероятностей; метод Надарая - Ватсона NW вычисления регрессионной зависимости по определяемым значениям целевого признака; вычисления оценок в локальной окрестности объектов и разбиения их значений на непересекающиеся интервалы с доминированием в них представителей одного из двух классов. Критериями оптимальности служили правило мажоритарности для KNN, точность разделение значений целевого признака на числовой оси для NW и устойчивости оценок объектов в границах интервалов. При поиске оптимальных значений k для перечисленных вариантов применялся метод скользящего контроля. Результаты полученных решений демонстрируется на выборке GERMAN из репозитория по описаниям 1000 объектов в разнотипном признаковом пространстве. Доказана эффективность выбора значений k по показателю устойчивости оценок объектов на числовой оси. Обоснование эффективности основывается на теории нечеткой логики.

Ключевые слова: Метод Надарая – Ватсона, решающее правило, целевая функция, устойчивость признака.

Цитирование: *Игнатьев Н.А, Турсунмуротов Д.Х.* Об эффективности метрических алгоритмов распознавания на базе k ближайших соседей // Проблемы вычислительной и прикладной математики. – 2024. – № 4(58). – С. 117-125.

1 Введение

Методы классификации являются важнейшей частью интеллектуального анализа данных (ИАД). Выбор методов зависит от предметной области, степени соответствия гипотезе о компактности, результатов проведения вычислительных экспериментов. Важность учёта соответствия положениям гипотезы о компактности видна из публикаций по разработке и обоснованию метрических алгоритмов классификации. Одним из индикаторов соответствия служат вычисление значений мер компактности классов и выборки в целом. Повысить обобщающую способность метрических алгоритмов предлагалось за счёт удаления шумовых объектов, отбора информативных признаков, вычисления параметров локальных метрик, применения регуляризаторов на основе анализа отступов между классами. Формирование баз прецедентов с применением регуляризаторов для поиска экстремума меры компактности позволило значительно увеличить эффективность реализации алгоритмов метода ближайший сосед (NN) [1] Относительно метода распознавания по k ближайшим соседям (KNN) исследований по формированию базы прецедентов не проводилось. Для реализации

KNN не рассматривались понятия шумовой, эталонный и граничный объекты, параметры локальных метрик, отступы между классами. С перечисленными выше понятиями для метода NN связано формирование баз прецедентов с использованием регуляризаторов [1].

В работе рассматривается:

1. Адаптация алгоритма Надарая – Ватсона (NW) [2] для решения проблемы повышения обобщающей способности распознаванию алгоритмов метода KNN. Необходимо было исследовать проблему преобразования значений целевого признака из интервальных шкал в номинальную;

2. Вычисления оценок объектов в окрестности k ближайших соседей и разбиение объектов на непересекающиеся группы в виде интервалы на числовой оси. Значения функция принадлежности оценок объектов к интервалам служат основой для решающего правила.

Процесс принятия решения о допустимом объекте в методе Надарая – Ватсона зависит от вкладов k ближайщих соседей. Вклады вычисляются по значениям заданной меры близости в ядерных функциях. Для перехода от задачи регрессии к задаче классификации необходимо иметь:

- значения целевого признака в интервальной шкале измерений;

- оптимальную по точности распознавания границу между двумя классами;

- решающие правила для определения принадлежности объектов к классу. Выбор и обоснование методов вычисления значений целевого признака является отдельной задачей. Граница между классами определяется без каких – либо предположений о природе среды данных. Как правило, параметры гипотетический функции плотности распределения неизвестны. Функция плотности изначально полагается многоэкстремальной. Для решения задач регрессии существует различные методы, в том числе и нелинейные. К нелинейным относятся метод Надарая – Ватсона, нейронные сети KAN, реализующие восстановление зависимостей на основе вывода их из теоремы Колмогорова и Арнольда [3]. В работе предлагается специальный критерий, оптимальное значение которого будет использоваться для выбора границы между классами.

Локальная окрестность из k ближайших соседей используется для вычисления оценок по количеству представителей классов. Исследуется предположения о наличии объектов с близкими значениями оценок, из которых можно формировать непересекающиеся группы. Значения функции принадлежности к группам предлагается применять в решающих правилах для определения принадлежности объектов к классам. Показана зависимость выбора оптимального значения k от функций принадлежности к классам.

2 Постановка задачи и метод решения

Представители двух непересекающихся классов K_1 и K_2, заданные в виде выборки объектов $E_0 = \{S_1, ..., S_m\}$, описываются набором разнотипных признаков $X(n) = (x_1, ..., x_n)$. Считается, что на E_0 определены операторы A_1 и A_2 для преобразования $S_i \underset{A_d}{\rightarrow} y_i$, $S_i \in E_0$, $d = 1{,}2$. Набор $Y = (y_1, \ldots, y_m)$ по A_1 рассматривается как множество значений целевого признака для решения задачи регрессии. Оператор A_2 преобразует описание объектов E_0 в значения функции принадлежности [4]. Требуется:

– выбрать метод для восстановления (аппроксимации) значения целевого признака произвольного допустимого объекта S по A_1 и метод вычисления функции принадлежности по A_2;

– по множеству $Y = (y_1, ..., y_m)$ определить значение порога Γ в качестве границы между классами K_1 и K_2;

– выбрать решающее правило для определения принадлежности к классам K_1 или K_2 произвольного допустимого объекта S.

3 Непараметрическая регрессия по методу Надарая – Ватсона

Непараметрическая регрессия используется как инструмент предварительной обработки данных для сглаживания шумов в данных. Для вычисления $f(S) = \alpha$ по Y и k ближайшим соседям применяется метод наименьших квадратов:

$$Q(\alpha, X) = \sum_{i=1}^{m} w_i(S)(\alpha - y_i)^2 \to \min_{\alpha \in Y}, \tag{1}$$

где $w_i(S)$ - вес объекта $S_i \in E_0$ относительно объекта S.

В (1) обходятся без явного задания параметрической модели зависимости от $f(x)$. Неявные модельные предположения закладываются в функцию расстояния $\rho(S, S_i)$. При обучении модели учитываются:

– число ближайших соседей k и ширина окна h;

– выбор метрики;

– веса объектов и веса признаков в метрике.

Веса w_i необходимо задать так, чтобы они убывали по мере увеличения расстояния $\rho(S, S_i)$. Для этого предлагается использовать невозрастающую, гладкую, ограниченную функцию $F : [0, \infty) \to [0, \infty)$, называемую ядром. С помощью ядерной функции вычисление w_i будет представлено в следующем виде: $w_i(S) = F\left(\frac{\rho(S, S_i)}{h}\right)$, где h — ширина окна. Приравняв нулю производную от $Q(\alpha, X)$ как - $\frac{\partial Q}{\partial \alpha} = 0$, и выразив через неё α, получим формулу Надарая-Ватсона.

Ядро $F(*)$ существенно влияет на гладкость функции $f(S)$ и слабо влияет на качество аппроксимации. В отличии от F(*) ширина окна h существенно влияет на качество аппроксимации. Переменная ширина окна по k ближайшим соседям определяется как $h(S) = \rho\left(S, S^{(k+1)}\right)$, где $S(k)$ – k-й сосед объекта S по метрике $\rho(x, y)$

Метод выбора целевого признака имеет принципиальное значение. Чаще всего значение выбирают из {-1,1}. Наиболее приемлемым с точки зрения интеллектуального анализа вариантом является использование вычисления латентных показателей.Источником значений целевого признака являются обобщённые оценки объектов. Обобщённые оценки представляют собой количественный признак, синтезируемый в общем-то из разнотипных сырых признаков на обучающей выборке. Для получения набора зависимых признаков предлагается использовать нелинейные вычисления.

Границу между двумя классами определим через разбиение на непересекающиеся интервалы $[\pi_1; \pi_2], (\pi_2; \pi_3]$ значений целевого признака из Y по критерию

$$\left(\frac{\sum_{d=1}^{2}\sum_{i=1}^{2}(u_i^d - 1)u_i^d}{\sum_{i=1}^{2} K_i(|K_i| - 1)}\right)\left(\frac{\sum_{d=1}^{2}\sum_{i=1}^{2} u_i^d\left(|K_{3-i}| - u_{3-i}^d\right)}{2|K_1||K_2|}\right) \to \max_{\pi_1 < \pi_2 < \pi_3}, \tag{2}$$

как $\Gamma = \frac{\pi_2+b}{2}$, где b - ближайшее к π_2 значение из $(\pi_2; \pi_3]$.

Выбор между линейным и нелинейным принципом вычисления обобщенных оценок в машинном обучении зависит от конкретной решаемой задачи. Процедура выбора реализуется на основе специальных критериев. Значения латентных признаков рассматриваются как обобщённые оценки объектов классов [4–10]. Основные причины выбора обобщённых оценок:

- исходные признаки измеряются в интервальных и номинальный шкалах;

- методы вычисления инвариантны к масштабам измерений количественных признаков.

- значения оценок можно использовать при дискриминантном анализа классов по (2).

Рассмотрим мультипликативный принцип вычисления обобщённых оценок объектов, в котором используются значения весов признаков и их попарных комбинаций. Веса количественных и номинальных признаков определяются разными методами. Для данных (латентных и исходных) в интервальных шкалах измерений используется разбиения на интервалы из [3].

Пусть для признака $x_c \in X(n)$ построена упорядоченная по не убыванию значений последовательность

$$r_1, ..., r_j, ..., r_m. \tag{3}$$

Значение границ между классами по (3) определяется по критерию (2). Для номинального признака $x_i \in X(n)$ значение границ между классами вычисляется по функции принадлежности к классам. Значение функции принадлежности по градации μ, $\mu \in \{1, ..., \eta(i)\}$ определяется как

$$f_c(\mu) = \frac{d_{1c}(\mu)/|K_1|}{d_{1c}(\mu)/|K_1| + d_{2c}(\mu)/|K_2|}, \tag{4}$$

а граница между классами

$$G_c = (q_1 + q_2)/2, \tag{5}$$

где

$$\begin{aligned} q_1 &= \min\{f_c(\mu) | 1 - f_c(\mu) < 0.5, \mu = 1, ..., p_c\}, \\ q_2 &= \max\{f_c(\mu) | 0.5 - f_c(\mu) > 0,\ \mu = 1, ..., p_c\}. \end{aligned} \tag{6}$$

Реализация процесса попарного объединения признаков для формирования множества из непересекающихся групп $\Omega = \{G_1, ..., G_t\}, t \geqslant 1$ основана на использовании экстремальных значений критерия (2) и соответствующих им границ двух непересекающихся интервалов.

Обозначим через $Y(n-q)$ набор из латентных и исходных признаков на q-ом шаге иерархической группировки, $Y(n) = X(n)$ при q=0. Проверке условия включения признака $y_p \in Y(n-q)$ в группу $G \in \Omega$ предшествует нормирование его значений с использованием границ интервала $[\pi_{1p}; \pi_{2p}](\pi_{2p}; \pi_{3p}]$ по формуле

$$a_p = (y_p - \pi_{2p})/(\pi_{3p} - \pi_{1p}). \tag{7}$$

Для количественного признака значения границ интервалов $[\pi_{1p}; \pi_{2p}](\pi_{2p}; \pi_{3p}]$ определяются по критерию (2). После замены градаций номинального признака на значения функции принадлежности (4) выбор границ π_{1p},π_{3p} аналогичен (5), а в качестве значения π_{2p} используется (6).

Синтез значений латентного признака $y \in Y(n-q-1)$ с нормированием по (7) для пары $(y_i, y_j) \subset Y(n-q)$ производится по формуле

$$y = \eta_{ij}(t_i w_i a_i + t_j w_j a_j) + (1-\eta_{ij}) t_{ij} w_{ij} (a_i a_j - \pi_{2ij})/(\pi_{3ij} - \pi_{1ij}), \quad t_i, t_j, t_{ij} \in \{-1, 0, 1\}, \eta_{ij} \in [0; 1], \tag{8}$$

где w_i,w_j,w_{ij} – веса признаков и значения границ $[\pi_{1ij}; \pi_{2ij}](\pi_{2ij}; \pi_{3ij}]$, определяемые по (8).

При реализации правил иерархической агломеративной группировки используется принцип динамического программирования. Группа G и латентный признак y, полученный на её основе по (8) считаются сформированными, если не существует пары (y, y_i), $y_i \in Y(n-q)$ при объединении которой в $y^* \in Y(n-q-1)$, веса признаков по (8) $w(y^*) \geqslant w(y)$.

Предлагается выбор критерия для оптимизации $g(S_i, E_0 \backslash S_i)$ – оценки объекта S_i методом скользящего экзамена. Определим предикат

$$\beta\,(S_i) = [S_i \in K_1 \; and \; g\,(S_i; E_0 \backslash \{S_i\}) > \Gamma \; or \; S_i \in K_2 \; and \; g\,(S_i; E_0 \backslash \{S_i\}) < \Gamma].$$

Критерий запишется таким образом

$$LOO\,(h, E_0) = \sum_{i=1}^{m} \beta\,(S_i) \to \max_h . \tag{9}$$

4 Выбор оптимального значения k ближайших соседей по значениям устойчивости признака

Устойчивость признака как исходного так и латентного вычисляется по значениям функции принадлежности (4). Значениями латентного признака $c(k)$ далее будем считать множества оценок $\{z(S_i, k)\}$ объектов E_0 в локальной окрестности из k ближайших соседей. К упорядоченному множеству (последовательности) оценок $c(k)$ вида (3) применим разбиение на непересекающиеся интервалы по критерию

$$\left| \frac{d_{tc}(u, v)}{|K_t|} - \frac{d_{3-t,c}(u, v)}{K_{3-t}} \right| \to \max, \tag{10}$$

где $d_{tc}(u, v)$, $d_{3-t}, c(u, v)$ – количество представителей классов K_t, K_{3-t} в интервале $[r_u; r_v]^i$, $i \in \{1, ..., p_c\}$, p_c – число интервалов для покрытия (3).

Обозначим через b_r значение функции принадлежности $f(\mu)$ к классу K_1 по (4), если значения признака $c(k)$ объекта S_r согласно (10) принадлежит интервалу $[r_u; r_v]^\mu$. Устойчивость признака $c(k)$ вычисляется как

$$\varphi\,(c\,(k)) = \frac{1}{m} \sum_{r=1}^{m} \begin{cases} b_r, b_r > 0.5, \\ 1 - b_r, b_r < 0.5. \end{cases} \tag{11}$$

Множество допустимых значений (11) принадлежат (0.5;1]. Оптимальное значение k выбирается согласно критерия

$$k_{opt} = \arg\max_k \varphi\,(c(k)). \tag{12}$$

В качестве примера рассмотрим два способа вычисления оценок в окрестности объекта $S \in K_i$, i=1,2 из k ближайших соседей

$$\frac{d_i(k)}{|K_i|} - \frac{d_{3-i}(k)}{|K_{3-i}|}, \tag{13}$$

и

$$\frac{d_i(k)}{|K_i|} - \frac{d_{3-i}(k)}{|K_{3-i}|}, \tag{14}$$

при условии $d_i(k) \geqslant 0$, $d_i(k) + d_3 - i(k) = k$. Оценка (13) является аддитивной, (14) – мультипликативной.

5 Вычислительный эксперимент

Для демонстрации результатов эксперимента использованы данные German [11]. Описание каждого из 1000 объектов выборки при соотношении классов $|K_1|$: $|K_2|$=700:300 производилось 7-ю количественными и 13-ю номинальными признаками. Значения количественных признаков в данных дробно-линейным преобразованием отображались в диапазоне [0; 1].

Целевой признак выбирается из числа латентных, вычисляемых по (8). Выбор основывается на максимальном значении критерия (2). Число сырых признаков и в какой последовательности синтезировали латентный признак по (2) также представляет интерес для исследования. Результаты выбора латентных признаков показаны в табл.1.

Значение точности (см. табл.1), полученное по первому латентному признаку является ориентиром для сравнения её с метрическими алгоритмами. В качестве меры расстояния использовалась метрика Журавлева. Получить дополнительную информацию о разделимости признаков по классам через устойчивость (11) можно из табл. 2. В скобках указаны порядковый по возрастанию номер в последовательности.

Цель анализа результатов из табл. 2 демонстрация плохой (по близости устойчивости к 0.5) разделимости объектов классов в GERMAN. Признаки num_dependents и residence_since изначально попадают под определение неинформативных.

Точность распознавания по методам из табл.3 не было выше 75.9, полученного (см. табл.1) по латентному признаку (8).

Таблица 1 Латентные признаки, сформированные по (8).

Номер латентного признака	**Порядок формирования латентного признака**	**Значение критерия (2)**	**Границы интервалов**	**Точность в %**
1	$(((x_1, x_2), x_4), x_{20})$	0.3766	$[-0.2796; 0.0]$ $(0.0; 0.0971]$	75.9
2	$((((x_3, x_5), x_6, x_8), x_{10}), x_{16})$	0.3427	$[-0.0734; 0.0]$ $(0.0; 0.2694]$	72.3
3	$((((x_7, x_{12}), x_{13}), x_{14}), x_9, x_{18})$	0.2930	$[-0.1330; 0.0]$ $(0.0; 0.1601]$	65.9
4	(x_{15}, x_{17})	0.2749	$[-0.0024; 0.0]$ $(0.0; 0.2726]$	63.6
5	(x_{11}, x_{19})	0.2523	$[-0.1678; 0.0]$ $(0.0; 0.085]$	59.5

Таблица 2 Устойчивость признаков в German

№	Название признака	Тип шкалы	Устойчивость (11)
1	checking_status	категориальный	0.6901 (20)
2	duration	числовый	0.5976 (18)
3	credit_history	категориальный	0.5894 (14)
4	purpose	категориальный	0.5913 (16)
5	credit_amount	числовый	0.6017 (19)
6	savings_status	категориальный	0.5978 (17)
7	employment	категориальный	0.5592 (12)
8	installment_commitment	числовый	0.5385 (8)
9	personal_status	категориальный	0.5493 (10)
10	other_parties	категориальный	0.5135 (3)
11	residence_since	числовый	0.5112 (2)
12	property_magnitude	категориальный	0.5583 (11)
13	age	числовый	0.5808 (15)
14	other_payment_plans	категориальный	0.5464 (9)
15	housing	категориальный	0.5648 (13)
16	existing_credits	числовый	0.5246 (7)
17	job	категориальный	0.5164 (4)
18	num_dependents	числовый	0.5011 (1)
19	own_telephone	категориальный	0.5195 (6)
20	foreign_worke	категориальный	0.5187 (5)

Таблица 3 Результаты вычисления оптимального k с помощью скользящего экзамена

Метод	Целевой признак	Значение k	Точность в %
KNN	-	11	74.7
Надаряй - Ватсона	по (8)	10	75.0
	из $\{-1, 1\}$	15	73.1

Таблица 4 Устойчивость по оценкам (14) в окрестности k ближайших соседей

Значения k	Число интервалов	Устойчивость (11)	Точность в %
2	3	0.9772	95.6
3	3	0.7538	86.1
4	5	0.9977	99.5
5	3	0.8135	85.5
6	3	0.8305	89.0

Таблица 5 Значения функции принадлежности к K_1 по (4) при $k = 4$

Номер интервала	Граница интервалов по (10)	Значение по (4)
1	[0.0; 0.0]	0.0395
2	[0.001414; 0.002838]	1
3	[0.003319; 0.003319]	0
4	[0.004271; 0.005714]	1
5	[0.006648; 0.013333]	0

Для распознавания принадлежности к классу K_1 или K_2 произвольного допустимого объекта S используется следующее правило. Находится интервал, к которому принадлежат оценка (14) ближайшего к S объекта S' из E_0 и анализируется значение функции принадлежности $f(S')$ по (4) к классу K_1. Если $f(S') > 0.5$, то $S \in K_1$, в противном случае $S \in K_2$. Этим правилом объясняется наличие 5 – ти неправильно распознанных объектов (см. табл. 5) из класса K_1.

6 Заключение

Произведён анализ эффективности использования для распознавания объектов 3 - мя методами реализации метрических алгоритмов на основе k ближайших соседей. Лучше показатели эффективности по точности распознавания у алгоритма, методика реализации которого строится на положениях теории нечёткой логики.

Литература

[1] *Игнатьев Н.А., Турсунмуротов Д.Х.* Цензурирование обучающих выборок с использованием регуляризации отношений связанности объектов классов // Научно-технический вестник информационных технологий, механики и оптики. – 2024. – Т. 24, №2. – С. 322-329. – doi: http://dx.doi.org/10.17586/2226-1494-2024-24-2-322-329.

[2] *Kloke J., McKean J.* Nonparametric Statistical Methods Using R. – CRC Press, – 2014. – С. 98-106.

[3] *Liu Z. et al.* KAN: Kolmogorov-Arnold Networks // arXiv. – 2024. – arxiv.org/abs/2404.19756.

[4] *Игнатьев Н.А., Акбаров Б.Х.* Оценка близости структур отношений объектов обучающей выборки на многообразиях наборов латентных признаков // Вестник Томского государственного университета.Управление, вычислительная техника и информатика. – 2023. – №65. – С. 69-78. – doi: http://dx.doi.org/10.17223/19988605/65/7.

[5] *Ignatyev N.A., Madrakhimov Sh.F., Saidov D.Y.* Stability of Object Classes and Selection of the Latent Features // International journal of engineering technology and sciences. – 2017. – doi: http://dx.doi.org/10.15282/ijets.7.2017.1.9.1071.

[6] *Саидов Д.Ю.* Нелинейной отображение наборов признаков на числовую ось и дискриминантный анализ данных // Проблемы вычислительной и прикладной математики. – 2017. – №1. – С. 18-23.

[7] *Згуральская Е.Н* Устойчивость разбиения данных на интервалы в задачах распознавания и поиск скрытых закономерностей // Известия Самарского центра Российской академии наук. – 2018. – №4. – С. 826-829.

[8] *Саидов Д.Ю., Игнатьев Н.А.* Вычисление сложности эффективных алгоритмов выбора оптимальных границ интервалов Проблемы информатики и энергетики. – 2014. – №6. – С. 35-40.

[9] *Рахимова М.А., Игнатьев Н.А.* Формирование и анализ наборов информативных признаков объектов по парам классов Искусственный интеллект и принятие . – 2021. – №4. – С. 18-26. – doi: http://dx.doi.org/10.14357/20718594210402.

[10] *Саидов Д.Ю.* Информационные модели на основе нелинейных преобразований признакового пространства в задачах распознавания : дисс.... доктора философии по физ.-мат. наук. – Ташкент, 2017. – 93 с. – doi: http://dx.doi.org/10.14357/20718594210402.

[11] URL: https://archive.ics.uci.edu/dataset/144/statlog+german+credit+data.

Поступила в редакцию 16.07.2024

UDC 519.95

ON THE EFFICIENCY OF METRIC RECOGNITION ALGORITHMS BASED ON k NEAREST NEIGHBOR

Ignatev N.A., ** **Tursunmurotov D.Kh.***

*mr.davrbek@mail.ru

National University of Uzbekistan named after Mirzo Ulugbek,
4, Universitetskaya str., Tashkent, 100174 Uzbekistan.

The article considers a solution to a two-class recognition problem using metric algorithms based on k-nearest neighbors. The proposed solution options are: the classical KNN algorithm based on a nonparametric estimate of the probability density function; the Nadaraya-Watson NW method for calculating the regression dependence based on the determined values of the target feature; calculating estimates in the local neighborhood of objects and dividing their values into non-overlapping intervals with dominance of representatives of one of the two classes. The optimality criteria were the majority rule for KNN, the accuracy of dividing the values of the target feature on the numerical axis for NW and the stability of object estimates within the interval boundaries. When searching for the optimal k values for the listed options, the sliding control method was used. The results of the obtained solutions are demonstrated on the GERMAN sample from the repository based on the descriptions of 1000 objects in a heterogeneous feature space. The efficiency of choosing the k values based on the stability of object estimates on the numerical axis is proven. The rationale for the efficiency is based on the theory of fuzzy logic.

Keywords: Nadaraya-Watson method, decision rule, objective function, feature stability.

Citation: Ignatev N.A., Tursunmurotov D.Kh. 2024. On the efficiency of metric recognition algorithms based on k nearest neighbor. *Problems of Computational and Applied Mathematics.* 4(58): 117-125.

№ 4(58) 2024

ISSN 2181-8460

HISOBLASH VA AMALIY MATEMATIKA MUAMMOLARI

ПРОБЛЕМЫ ВЫЧИСЛИТЕЛЬНОЙ И ПРИКЛАДНОЙ МАТЕМАТИКИ

PROBLEMS OF COMPUTATIONAL AND APPLIED MATHEMATICS

RAQAMLI TEXNOLOGIYALAR VA
SUN'IY INTELLEKTNI RIVOJLANTIRISH
ILMIY-TADQIQOT INSTITUTI

ПРОБЛЕМЫ ВЫЧИСЛИТЕЛЬНОЙ И ПРИКЛАДНОЙ МАТЕМАТИКИ

№ 4(58) 2024

Журнал основан в 2015 году.
Издается 6 раз в год.

Учредитель:
Научно-исследовательский институт развития цифровых технологий и искусственного интеллекта.

Главный редактор:
Равшанов Н.

Заместители главного редактора:
Азамов А.А., Арипов М.М., Шадиметов Х.М.

Ответственный секретарь:
Ахмедов Д.Д.

Редакционный совет:
Азамова Н.А., Алоев Р.Д., Амиргалиев Е.Н. (Казахстан), Бурнашев В.Ф., Загребина С.А. (Россия), Задорин А.И. (Россия), Игнатьев Н.А., Ильин В.П. (Россия), Исмагилов И.И. (Россия), Кабанихин С.И. (Россия), Карачик В.В. (Россия), Курбонов Н.М., Маматов Н.С., Мирзаев Н.М., Мирзаева Г.Р., Мухамадиев А.Ш., Назирова Э.Ш., Нормуродов Ч.Б., Нуралиев Ф.М., Опанасенко В.Н. (Украина), Расулмухамедов М.М., Расулов А.С., Садуллаева Ш.А., Старовойтов В.В. (Беларусь), Хаётов А.Р., Халджигитов А., Хамдамов Р.Х., Хужаев И.К., Хужаеров Б.Х., Чье Ен Ун (Россия), Шабозов М.Ш. (Таджикистан), Dimov I. (Болгария), Li Y. (США), Mascagni M. (США), Min A. (Германия), Schaumburg H. (Германия), Singh D. (Южная Корея), Singh M. (Южная Корея).

Журнал зарегистрирован в Агентстве информации и массовых коммуникаций при Администрации Президента Республики Узбекистан.
Регистрационное свидетельство №0856 от 5 августа 2015 года.

ISSN 2181-8460, eISSN 2181-046X

При перепечатке материалов ссылка на журнал обязательна.
За точность фактов и достоверность информации ответственность несут авторы.

Адрес редакции:
100125, г. Ташкент, м-в. Буз-2, 17А.
Тел.: +(998) 712-319-253, 712-319-249.
Э-почта: journals@airi.uz.
Веб-сайт: https://journals.airi.uz.

Дизайн и вёрстка:
Шарипов Х.Д.

Отпечатано в типографии НИИ РЦТИИ.
Подписано в печать 30.08.2024 г.
Формат 60x84 1/8. Заказ №4. Тираж 100 экз.

Содержание

Абдурахимов Б.Ф., Джураева Н.Т.
Численное моделирование сингулярно возмущенного уравнения четвертого порядка методом предварительного интегрирования 8

Равшанов Н., Неъматов А., Исмаилов Ш.
Параллельный алгоритм решения задачи фильтрации нефти с учетом изменения проницаемости пласта 18

Эшмаматова Д.Б., Юсупов Ф.А.
Дискретная модель SWMB исследования передачи компьютерных вирусов . 33

Равшанов Н., Набиева И.
Вычислительные эксперименты по оценке и прогнозированию распространения промышленных выбросов в атмосферном бассейне Ташкентской области 48

Узаков З.У., Бегулов О.У.
Применение объектно-ориентированного подхода в задаче трехфазной фильтрации 69

Равшанов Н., Туракулов Ж.
Численное исследование процесса фильтрования малоконцентрированных растворов через пористую среду 84

Фазылов Ш.Х., Мирзаев Н.М., Раджабов С.С., Раббимов И.М.
Интенсиональный метод распознавания объектов, описанных в признаковом пространстве большой размерности 98

Игнатьев Н.А, Турсунмуротов Д.Х.
Об эффективности метрических алгоритмов распознавания на базе k ближайщих соседей 117

Хакимов М.Х., Бекова В.Г.
Морфологический анализ существительных в машинном переводе с английского на узбекский язык 126

Шарипов Д.К., Саидов А.Д.
Использование различных алгоритмов машинного обучения для прогнозирования смертности после острого инфаркта миокарда с использованием медицинских данных GUSTO-1 135

Ёркулов Б.А., Сулюкова Л.Ф.
Модели оценки качества образовательной информационной системы на основе нечетких когнитивных карт 148