



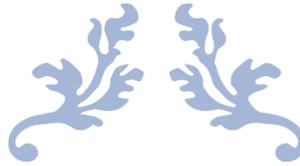
**RAQAMLI TEXNOLOGIYALARNING  
YANGI O‘ZBEKISTON  
RIVOJIGA TA’SIRI**

Xalqaro ilmiy-amaliy  
konferensiyasi to'plami

21 IYUN

2023





---

**RAQAMLI TEXNOLOGIYALARNING YANGI O'ZBEKISTON  
RIVOJIGA TA'SIRI**

**ВЛИЯНИЕ ЦИФРОВЫХ ТЕХНОЛОГИЙ НА РАЗВИТИЕ  
НОВОГО УЗБЕКИСТАНА**

**IMPACT OF DIGITAL TECHNOLOGIES ON THE DEVELOPMENT  
OF NEW UZBEKISTAN**

---

Xalqaro ilmiy-amaliy konferensiyasi maqolalar to'plami



JUNE 21, 2023  
KOKAND UNIVERSITY

"O'zbekiston Respublikasi Oliy ta'lim tizimini 2030 yilgacha rivojlantirish konsepsiyasini tasdiqlash to'g'risida" O'zbekiston Respublika Prezidentining 5847-sonli Farmonida ko'zda tutilgan vazifalardan biri – ilmiy izlanish yutuklarini amaliyotga joriy etish yo'li bilan fan sohalarini rivojlantirish, ya'ni xalqaro ilmiy hamjamiyatda e'tirof etilishiga xizmat qilishdir. Shu va boshqa tegishli farmonlarda va qarorlarda belgilangan vazifalarini amalga oshirish maqsadida 2023 yil 21-iyun kuni Qo'qon universiteti "Raqamli texnologiyalar va matematika" kafedrasida "Raqamli texnologiyalarning Yangi O'zbekiston rivojiga ta'siri" mavzusidagi xalqaro miqyosida o'tkaziladigan ilmiy-amaliy konferensiyasi maqolalar to'plamini e'lon qiladi



### **MAS'UL MUHARRIR**

**Zahidov G'ofurjon Erkinovich** – iqtisodiyot fanlari bo'yicha falsafa doktori, dotsent

### **TAHRIRIYAT HAY'ATI**

**G'ulomov Saidahrur Saidahmedovich** – iqtisodiyot fanlari doktori, akademik;

**Ahmedov Durbek Quدراتillayevich** - iqtisodiyot fanlari doktori, professor;

**Mahmudov Nosir Mahmudovich** – iqtisodiyot fanlari doktori, professor;

**Butaboyev Muhammadjon** - iqtisodiyot fanlari doktori, professor;

**Islamov Anvar Ashirkulovich** - iqtisodiyot fanlari bo'yicha falsafa doktori, dotsent;

**Ruziev Shohrusbek Ravshan o'g'li** - iqtisodiyot fanlari bo'yicha falsafa doktori, dotsent

**Mulaydinov Farxod Murotovich** – Qo'qon universiteti, Raqamli texnologiyalar va matematika kafedrasida mudiri

**Texnik muharrir – Solidjonov Dilyorjon Zoirjon o'g'li**



Ta'lim sifati yangi O'zbekiston taraqqiyotini yanada yuksaltirishning muhim omili / Raqamli texnologiyalarning Yangi O'zbekiston rivojiga ta'siri xalqaro ilmiy-amaliy konferensiyasi to'plami. Kokand university, 2023 yil 21 iyun, - «Innovatsion rivojlanish nashriyot-matbaa uyi» 2023.

© Matn. Mualliflar, 2023.

© Kokand university, 2023.

© «Innovatsion rivojlanish nashriyot-matbaa uyi», original maket, 2023.

55	ПРОБЛЕМЫ СОВЕРШЕНСТВОВАНИЯ СИСТЕМЫ ОПОВЕЩЕНИЯ НАСЕЛЕНИЯ В ЭКСТРЕМАЛЬНЫХ СИТУАЦИЯХ И ПУТИ ИХ РЕШЕНИЯ - Ходжакулов Мухторжон Назаркулович,	241-246
56	ФОРМАЛИЗАЦИЯ ПРОБЛЕМЫ СНИЖЕНИЯ РАЗМЕРНОСТИ ПРИЗНАКОВОГО ПРОСТРАНСТВА В ЗАДАЧАХ КЛАССИФИКАЦИИ - Х.Ш. Рашидов	247-254
57	YER OSTI KON ISHLARIDA FOYDALANILADIGAN TEXNIK ISHLANMALARNING INNOVATSIYON TEXNOLOGIYALARINI ISHLAB CHIQISH - U.M. Mamirov, U.T. Toshtemirov	255-256
58	RAQAMLI TEXNOLOGIYALARDAN FOYDALANISH ASOSIDA SMART SHAHTALARNI TASHKIL QILISH - O.SH. Yormatov, U.T. Toshtemirov	257-258
<b>5-SHO'BA. QISHLOQ XO'JALIGIDA ZAMONAVIY TEXNOLOGIYALARNI QO'LLASH ORQALI YER UNUMDORLIGI, HOSILDORLIK VA NASLLI CHORVA MOLLARINI YETISHTIRISH</b>		
59	ПАХТА-ТО'QIMACHILIK KLASTERLARINI TASHKIL ETISH ASOSLARI - Urishev Baxtiyor Abdusamatovich, Hojiyev Ixtiyor Bahodirjon o'g'li	260-262
60	BUGUNGI KUNDAGI OROL DENGIZI FOJIASI VA UNING OQIBATLARI - Abdullajonov Davronjon Shokirjon o'g'li, Esonaliyeva Madinabonu Shuxratjon qizi	263-264
61	SUG'ORILADIGAN TIPIK BO'Z TUPROQLARDA BIOMIKROELEMENTLAR MIQDORI VA RAQAMLI XARITALASH - Abduxakimova Xusnidaxon Abdullayevna, Abduvaxobova Ximoyatxon Ilxomjon qizi, Ma'rufjonov Javohir G'ayratjon o'g'li, Solijonova Dilafro'zxon Shuxratjon qizi	265-267
62	SHOHIMARDONSOY KONUS YOYILMA TUPROQLARIDA BIOMIKROELEMENTLAR BIOGEOKIMYOSI - Abduxakimova Xusnidaxon Abdullayevna, Abdulazizov Asilbek Abdulahad o'g'li, Toxirova Madinaxon Rasuljon qizi	268-270
63	ZAMONAVIY DEHQONCHILIKDA TOMCHILATIB SUG'ORISH - Yusupova Mohidil Abdumutalibovna, Abdullaaxatov Asadbek Ilhomjon o'g'li	271-274
64	QUMLI TUPROQLARNING EKOLOGIK-MELIORATIV HOLATI VA XOSSALARINING O'ZGARISHLARI - Yusupova Mohidil Abdumutalibovna, Ibroximov Shahroz Alisher o'g'li	275-277
65	MEVA VA SABZAVOTLI PASTILLALARNI QURITISH JARAYONLARINI TAKOMILLASHTIRISH SAMARADORLIGI - Xolikov M.M., Rahbar: Nasirova Sh.N.	278-280
66	QISHLOQ XO'JALIK MAHSULOTLARINI RIVOJLANISH OMILLARI - Xoliqov M.M.	281-282
67	ГЕОИНФОРМАЦИОННЫЕ МЕТОДЫ ОЦЕНКИ КАЧЕСТВА ПОДЗЕМНЫХ ВОД - Мадримов А.Ш.	283-291
68	SHOLINING UZROS-7-13 NAVI HOSILDORLIGIGA URUG' EKISH MUDDATLARINI TA'SIRI - M.A.Mirzayeva, M.I.Teshaboyeva	292-295
69	СУФОРИЛАДИГАН ТИПИК БЎЗ ТУПРОҚЛАР ШАРОИТИДА БАҲОРГИ МУДДАТДА ЭКИЛГАН СОЯНИНГ 1000 ДОНА ДОН МАССАСИГА ТАСИРИ - Холиқов Аброр Тожимуродович, Қурбонқулова Умида Хусанқизи	296-298
70	QISHLOQ XO'JALIGINI RAQAMLASHTIRISHNING STRATEGIYASI VA RIVOJLANISH YO'NALISHLARI - Kenjayev Toshbolta Aminovich	299-303
<b>6-SHO'BA. ZAMONAVIY AXBOROT-KOMMUNIKATSIYA TEXNOLOGIYALARI SOHALARINING MUAMMOLARI VA YECHIMLARI</b>		
71	ZAMONAVIY AXBOROT-KOMMUNIKATSIYA TEXNOLOGIYALARI YORDAMIDA TURIZM SOHASIDAGI MUAMMOLARGA YECHIMLAR - Ro'ziyev Shohrusbek Ravshanjon o'g'li	305-308

**ФОРМАЛИЗАЦИЯ ПРОБЛЕМЫ СНИЖЕНИЯ РАЗМЕРНОСТИ ПРИЗНАКОВОГО  
ПРОСТРАНСТВА В ЗАДАЧАХ КЛАССИФИКАЦИИ****Х.Ш. Рашидов**

НИИ Цифровых технологий и искусственного интеллекта базовый докторант

**Аннотация:** В статье рассматриваются вопросы формализации проблемы снижения размерности признакового пространства в задачах классификации данных, а также приведены результаты анализа наиболее известных эвристических методов выбора признаков. Определены дальнейшие направления исследований по созданию новых методов снижения размерности пространства признаков в задачах классификации.

**Ключевые слова:** Классификация данных, признаковое пространство, выбор признаков, критерий информативности, оптимальный набор признаков.

При решении большинства практических задач классификации возникает необходимость отбора признаков из-за наличия в данных бесполезных и дублирующих друг друга признаков, приводящих к необоснованному усложнению моделей алгоритмов классификации. Стремление специалистов предметных областей учесть как можно больше характеристик и свойств изучаемых объектов приводит к появлению избыточных признаков, отрицательно влияющих на быстродействие алгоритмов. Поэтому проблема отбора признаков (features selection) и в настоящее время остается актуальной, о чем свидетельствует большое количество публикаций по этой проблеме, появляющихся в последние годы (например, [1,2,3]).

Снижение размерности исходного пространства признаков в задачах анализа данных представляет по существу переход от исходной системы признаков  $x = (x^1, x^2, \dots, x^N)$  к новой системе  $z = (z^1, z^2, \dots, z^\ell)$ , включающей меньшее число признаков ( $\ell < N$ ), чем исходная система. Обычно новые признаки формируются в виде функций от исходных признаков, т.е.  $z = F(x)$ , путем решения оптимизационной задачи. Последняя заключается в нахождении такой системы признаков  $z$ , при которой

$$I(\tilde{z}) = \max_{F \in \Omega} \{I(z)\}. \quad (1)$$

Здесь  $I(z)$  – заданная мера информативности  $\ell$ -мерной системы признаков  $z$ , а  $F$  – класс допустимых преобразований исходных признаков  $x^1, x^2, \dots, x^N$ , который в общем случае может быть представлен одним из следующих типов преобразований: линейный, нелинейный, дискретный, непрерывный, логический. Следует отметить, что конкретный



в отобранном наборе признаков или величиной ошибочно расклассифицированных объектов, если выбор признаков осуществляется для конкретного правила классификации. При рассмотрении методов выбора признаков далее будем предполагать, что качество (информативность) признаков и их наборов оценивается по одному из этих критериев.

Задача определения оптимального набора признаков имеет относительно простое решение в условиях независимости исходных признаков  $x^1, x^2, \dots, x^N$ . В этом случае последние ранжируются по значениям критерия информативности каждого из этих признаков, и для заданного  $\ell$  определяется наилучший набор  $\tilde{z} = (\tilde{x}^1, \tilde{x}^2, \dots, \tilde{x}^\ell)$ , включающий первые  $\ell$  наиболее информативных признаков ранжированного ряда. Подробное описание этого метода приводится в [6].

Ситуация существенно усложняется при коррелированности исходных признаков  $x^1, x^2, \dots, x^N$ : для определения наилучшего набора из  $\ell$  признаков необходимо сравнить все возможные  $C_N^\ell$  наборов из  $\ell$  признаков. Если  $\ell$  не фиксировано, то количество вариантов достигает величины  $P = \sum_{\ell=1}^N C_N^\ell = 2^N - 1$ .

Отсюда следует, что метод полного перебора хотя и гарантирует определение оптимального набора признаков  $\tilde{z} = (\tilde{x}^1, \tilde{x}^2, \dots, \tilde{x}^\ell)$ , однако его можно использовать лишь при небольшом числе исходных признаков.

В связи с этим обстоятельством в прикладных исследованиях обычно используются методы частичного перебора, позволяющие получить субоптимальный набор признаков при сравнительно малых вычислительных затратах. Согласно [6] среди них наиболее распространенными считаются методы, использующие процедуры последовательного отбрасывания признаков (алгоритм «последовательной селекции назад»), последовательного присоединения признаков (алгоритм «последовательной селекции вперед»), а также комбинации этих процедур.

Результаты сравнительного анализа этих методов по вычислительным затратам приведены в [6], где показано, что в условиях представительной выборки процедура последовательного присоединения обеспечивает результаты, более близкие к оптимальному, чем процедура последовательного отбрасывания. Каждому из этих методов присущи свои недостатки. Например, при реализации процедуры последовательного присоединения нельзя отбросить признак, включенный в оптимальный набор на предыдущих шагах, а при использовании процедуры последовательного отбрасывания не учитывается статистическое влияние исключенных ранее признаков.

Другим широко применяемым методом выбора информативных признаков является случайный поиск с адаптацией (метод СПА) [7,8], использующий псевдослучайный датчик для генерирования булевских векторов  $\lambda$  таких, где  $\sum_{j=1}^N \lambda^j = \ell$ . При этом вероятность выпадения каждого из  $N$  признаков в начале одинакова и равна  $\frac{1}{N}$ . Иными словами, на основе вектора вероятностей  $p = (p^1, p^2, \dots, p^N)$ , где  $p^j$ -вероятность выпадения признака  $j$ , датчик

генерирует некоторый псевдослучайный вектор  $\lambda = (\lambda^1, \lambda^2, \dots, \lambda^N)$ ,  $\lambda^j \in \{0,1\}, j = \overline{1, N}$ ;  $\sum_{j=1}^N \lambda^j = \ell$ , причем в начале  $p^j = \frac{1}{N}, j = \overline{1, N}$ . Метод СПА осуществляет пошаговое изменение вектора вероятностей  $p$  до тех пор, пока в нем не останется  $\ell$  ненулевых компонент.

На каждом шаге посредством псевдослучайного датчика с учетом текущего вектора вероятностей  $p = (p^1, p^2, \dots, p^N)$  генерируются  $k$  векторов  $\lambda^1, \lambda^2, \dots, \lambda^k$  (по рекомендации автора  $k = 10 \div 15$ ). Среди них выбирается пара векторов  $\lambda_{min}, \lambda_{max}$  на которых функционал  $I(\lambda)$  принимает соответственно минимальное и максимальное значение.

Далее производится изменение вектора вероятностей. Для каждой  $j$ , равной единице, соответствующая компонента вектора вероятностей уменьшается на некоторую величину  $h \ll \frac{1}{N}$ , называемую штрафом, если последняя не становится отрицательной. В противном случае уменьшение выполняется до нуля.

В то же время этому методу присущи свои недостатки. Например, неизвестно, как его использовать при конкурирующих группах признаков. Кроме того, в условиях малой обучающей выборки решения, полученные с помощью этого метода, могут быть статистически неустойчивыми.

Рассмотренные методы выбора признаков отличаются друг от друга способом организации перебора различных сочетаний исходных признаков. В то же время они обладают общей особенностью: каждый из этих методов может использовать для оценки качества признаков и их наборов различные критерии информативности. Однако такая универсальность порождает проблему выбора эффективного критерия, так как результат выбора признаков существенно зависит от используемого критерия информативности.

В настоящее время разработано несколько методов выбора признаков, которые ориентированы на использование конкретного критерия информативности. Один из таких методов предложен и исследован в [9,10]. Суть его заключается в использовании меры важности исходного признака, представляющей собой обработанную надлежащим образом степень уменьшения так называемых «голосов» при удалении этого признака. Авторами метода эта мера названа информационным весом признака. Для определения информационных весов используются специальные вычислительные алгоритмы, которые позволяют учитывать взаимосвязи признаков.

Интересный подход к выбору информативных признаков рассмотрен в [11]. В рамках этого подхода предложен ряд методов определения информативных наборов признаков с учетом ограничений, связанных с затратами на создание технических средств измерения этих признаков. Мера информативности последних задается критериями, основанными на использовании евклидова расстояния.

Кроме того, для формирования критерия эффективности системы распознавания авторами [11] вводятся следующие меры близости между объектами:

$$S_p(\lambda) = \sqrt{\frac{2}{m_p} \cdot \frac{1}{m_p-1} \sum_{j=1}^{m_p} \sum_{t=1}^{m_p} d_E^2(x_{pj}, x_{pt})_\lambda}, \quad (4)$$

характеризующая среднеквадратичный разброс объектов внутри данного класса  $X^p, p = \overline{1, k}$ ;

$$R_{p,q}(\lambda) = \sqrt{\frac{1}{m_p m_q} \sum_{j=1}^{m_p} \sum_{t=1}^{m_q} d_E^2(x_{pj}, x_{qt})_\lambda}, \quad (5)$$

характеризующая среднеквадратичный разброс данной пары классов  $X^p$  и  $X^q, p, q = \overline{1, k}; p \neq q$ .

В качестве алгоритма распознавания выбирается классификатор евклидового расстояния, определяющий величину

$$L(x, X^p)_\lambda = \sqrt{\frac{1}{m_p} \sum_{j=1}^{m_p} d_E^2(x, x_{pj})_\lambda}, \quad (6)$$

которая является среднеквадратичным расстоянием между распознаваемым объектом  $x$  и объектами класса  $X^p, p = \overline{1, k}$  по набору признаков, определяемому вектором  $\lambda$ . Решающее правило этого классификатора состоит в следующем:  $x \in X^p$ , если  $L(x, X^p)_\lambda = \min_{q=\overline{1, k}} L(x, X^q)_\lambda$ .

Рассматривая в качестве критерия эффективности системы распознавания функционал, зависящий в общем случае от функций (5), (6) и выбранного алгоритма распознавания (7), задачу определения информативного набора признаков можно свести к оптимизационной задаче

$$I_1(\lambda) = I[S_p(\lambda); R_{p,q}(\lambda); L(x, X^p)] \rightarrow \underset{\lambda}{extr} \quad (7)$$

при

$$C(\lambda) = \sum_{i=1}^N C_i \lambda_i \leq C_0, \quad (8)$$

где  $C_i$ -затраты на создание технического средства, предназначенного для определения признака  $x^i$ ;  $C_0$ -ресурсы, выделенные на создание технических средств наблюдений.

В прикладных исследованиях обычно можно ограничиться решением оптимизационных задач для частных видов функционала  $I_1$ , например,

$$I_2(\lambda) = \max_{t=\overline{1, k}} [S_t(\lambda): R_{p,q}(\lambda) \geq R_{p,q}; p, q = \overline{1, k}] \rightarrow \underset{\lambda}{min}, \quad (9)$$

$$I_3(\lambda) = \min_{p,q=1,k} [R_{p,q}(\lambda): S_t(\lambda) \geq S_t^c; t = \overline{1,k}] \rightarrow \max_{\lambda} \quad (10)$$

$$I_4(\lambda) = \min_{p,q=1,k} \left[ \frac{R_{p,q}^2(\lambda)}{S_p(\lambda)S_q(\lambda)} \right] \rightarrow \max_{\lambda} \quad (11)$$

при ограничении (9).

Решение каждой из задач (10) - (12) направлено на определение информативных наборов признаков, улучшающих разделимость заданных классов, что, в конечном итоге, повышает эффективность системы распознавания при соблюдении ограничения на общую сумму стоимостей технических средств наблюдений. Причем допустимой областью решений этих задач является подмножество вершин единичного гиперкуба, число которых равно  $C_N^{\ell}$ .

Один из предлагаемых в [11] методов решения задач (10)-(12) основан на построении непрерывного аналога для задачи дискретной оптимизации. В этом случае вместо дискретной задачи

$$\begin{cases} I(\lambda) \rightarrow \max_{\lambda} \text{extr}, \\ \sum_{i=1}^N \lambda_i = \ell; \\ \lambda_i \in \{0,1\} \end{cases} \quad (12)$$

решается задача непрерывной оптимизации

$$\begin{cases} I(\lambda) \rightarrow \max_{\lambda} \text{extr}, \\ \sum_{i=1}^N \lambda_i = \ell; \\ \lambda_i \in [0,1], \end{cases} \quad (13)$$

в которой параметры  $\lambda_i$  могут принимать произвольные значения из отрезка  $[0,1]$ .

Очевидно, что такая замена имеет смысл в том случае, если функционал  $I(\lambda)$ , определенный для дискретных значений  $\lambda_i$ , имеет свое очевидное и естественное продолжение на непрерывный случай.

Если решением задачи (14) является некоторый вектор  $\lambda^*$ , то решением задачи (13) будет считаться такой вектор  $\lambda^0$ , у которого компоненты, соответствующие  $\ell$  максимальным компонентам  $\lambda^*$ , равны 1, а остальные -0.

Задача (14) является классической задачей нелинейного программирования и ее решение можно получить, привлекая различные численные методы [12]. В [11] для решения задачи (14) предварительно вводятся соответствующие штрафные функции, а затем

используется градиентный метод оптимизации. Преимущество данного подхода к определению информативного набора признаков по сравнению с рассмотренными выше методами частичного перебора заключается в надежности получаемых результатов, которая обеспечивается использованием хорошо развитого математического аппарата численных методов оптимизации.

Завершая рассмотрение наиболее распространенных методов выбора признаков, следует отметить, что, несмотря на достигнутые результаты, решение задачи определения информативных наборов признаков и поныне не потеряло своей актуальности в связи с расширением сферы применения методов распознавания образов для решения новых прикладных задач со специфическими особенностями. Эффективному использованию известных методов выбора информативных наборов признаков при решении таких задач препятствует ряд проблем. Одна из них связана с выбором наиболее подходящей для конкретной задачи комбинации метода и критерия, т.е. фиксированием для этой задачи такой пары метода и критерия, при использовании которых минимизируются вычислительные затраты и улучшается качество распознавания.

#### **ЛИТЕРАТУРА:**

1. Munoz-Romero S., Gorostiaga A., Soguero-Ruiz C., Mora-Jiménez I., Rojo-Alvarez J.L. Informative variable identifier: expanding interpretability in feature selection // Pattern Recognition. – 2020. – Vol. 98. – P. 87-99.
2. Ircio J., Lojo A., Mori U., Lozano J. Mutual information- based feature subset selection in multivariate time series classification // Pattern Recognition. – 2020. – Vol. 108. – URL: <https://10.1016/j.patcog.2020.107525>.
3. Viharos Z.J., Kis K.B., Fodor A., Buki M.I. Adaptive, Hybrid Feature Selection (AHFS) // Pattern Recognition. – 2021. – Vol. 116. – URL: <https://doi.org/10.1016/j.patcog.2021.107932>.
4. Фазылов Ш.Х., Маматов Н.С. Информатив белгилар фазосини куришда Дельталар усули // "Информатика ва энергетика муаммолари" Ўзбекистон журнали. – Тошкент. – 2005. – 6-сон. – 11-16 б.
5. Авалиани Г.В. Эвристические методы в распознавании образов. – Тбилиси: Мецниереба, 1988. – 75 с.
6. К.А.Чепонис, Д.А.Жвиренайте, Б.С.Бусыгин, Л.В.Мирошниченко. Методы, критерии и алгоритмы, используемые при преобразовании, выделении и выборе признаков в анализе данных // Сборник статей. - Вильнюс, 1988. – 150 с.
7. Лбов Г. С. Выбор эффективной системы зависимых признаков. – В сб.: Вычислительные системы. – 1965. – Вып. 19. С. 21-34.
8. Лбов Г. С. Методы обработки разнотипных экспериментальных данных. - Новосибирск: Наука, 1981. – 160 с.

9. Адылова З.Т., Лутфуллаев Р.А. Некоторые свойства меры важности признаков// Вопросы кибернетики. – Ташкент: Институт кибернетики с ВЦ АН УзССР, 1972. – Вып. 49. – С. 12-21.
10. Лутфуллаев Р.А. Об одном способе определения информативных весов признаков// ЖВМ и МФ.– 1974. – Т. 14. – № 5. – С. 701-705.
11. Горелик А.Л., Скрипкин В.А. Методы распознавания. – М.: Высшая школа, 1984. – 208 с.
12. Базарра М., Шетти К. Нелинейное программирование. Теория и алгоритмы. – М.: Мир, 1982. – 583 с.