

На правах рукописи



Малашин Иван Павлович

**БИНАРНЫЙ ГЕНЕТИЧЕСКИЙ АЛГОРИТМ
С ДИНАМИЧЕСКОЙ
КОРРЕКЦИЕЙ ОБЛАСТИ ПОИСКА
ДЛЯ СЛОЖНЫХ ЗАДАЧ ОПТИМИЗАЦИИ
СО СМЕШАННЫМИ ПЕРЕМЕННЫМИ**

2.3.1 – Системный анализ, управление и обработка
информации, статистика

Автореферат
диссертации на соискание учёной степени
кандидата технических наук

Красноярск — 2026

Работа выполнена в федеральном государственном бюджетном образовательном учреждении высшего образования «Сибирский государственный университет науки и технологий имени академика М.Ф. Решетнева», г. Красноярск.

Научный руководитель: доктор технических наук, доцент
Сопов Евгений Александрович

Официальные оппоненты: **Демидова Лилия Анатольевна**,
доктор технических наук, профессор,
ФГБОУ ВО «МИРЭА – Российский технологический университет», г. Москва,
профессор кафедры корпоративных информационных систем

Скобцов Юрий Александрович,
доктор технических наук, профессор,
ФГАОУ ВО «Санкт-Петербургский государственный университет аэрокосмического приборостроения», г. Санкт-Петербург, профессор кафедры компьютерных технологий и программной инженерии

Ведущая организация: федеральное государственное бюджетное образовательное учреждение высшего образования «Воронежский государственный технический университет»

Защита состоится 9 октября 2026 года в 14:00 часов на заседании диссертационного совета 24.2.403.01, созданного на базе ФГБОУ ВО «Сибирский государственный университет науки и технологий имени академика М.Ф. Решетнева» по адресу: 660037, г. Красноярск, пр. им. газеты «Красноярский рабочий» 31, зал заседаний диссертационного совета, ауд. Л-205

С диссертацией можно ознакомиться в научной библиотеке ФГБОУ ВО «Сибирский государственный университет науки и технологий имени академика М.Ф. Решетнева» и на сайте <https://www.sibsau.ru>

Отзывы на автореферат в двух экземплярах, заверенные печатью, просим отправлять по адресу: 660037, Россия, г. Красноярск, просп. им. газеты «Красноярский рабочий», 31, Сибирский государственный университет науки и технологий имени академика М.Ф. Решетнева» (СибГУ им. М.Ф. Решетнева), Диссертационный совет. E-mail: dissovet@sibsau.ru

Автореферат разослан « ____ » _____ 2026 г.

Ученый секретарь
диссертационного совета

Панфилов Илья Александрович

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования. Задачи оптимизации приобретают все большее значение в различных областях науки и техники в связи с развитием систем интеллектуального анализа данных и усложнением процессов проектирования сложных систем. Методы эволюционных вычислений, в том числе генетические алгоритмы (ГА), широко применяются к решению подобных задач. Преимущества современных модификаций ГА заключаются в возможности решать задачи со смешанными переменными, выраженными в произвольных шкалах, путем представления решения в наиболее простой шкале – бинарной. Подобные алгоритмы позволяют решать сложные задачи глобальной оптимизации большой размерности, не накладывая жестких требований к свойствам целевой функции и функций ограничений. Однако бинарные генетические алгоритмы сталкиваются с проблемой выбора параметров кодирования вещественных переменных, таких как границы области поиска и шаг дискретизации. Слишком малый шаг сетки увеличивает размерность бинарного пространства поиска, что усложняет решение задачи. В то же время слишком крупный шаг дискретизации может привести к тому, что узлы сетки могут располагаться недостаточно близко к оптимумам в исходном вещественном пространстве. Эффективнее было бы использовать более плотную сетку в перспективных регионах вещественного пространства поиска и пропускать менее перспективные регионы. Эту проблему можно преодолеть, используя методы динамической коррекции границ области поиска, основанные на анализе плотности распределения решений популяции в пространстве поиска, в частности, подходы, использующие кластеризацию для выделения перспективных регионов пространства.

Степень разработанности темы исследования. Исследованием и развитием теории и практики применения ГА активно занимаются отечественные и зарубежные ученые. Основоположником идеи генетических алгоритмов является Дж. Холланд, который заложил теоретический фундамент для всех последующих исследований. Дальнейшее развитие ГА связано с работами Д. Голдберга, К. Дэба и К. Де Йонга, которые предложили ряд новых операторов и схем применения ГА для бинарного пространства поиска, а также расширили область применения (многокритериальная оптимизация, задачи большой размерности и другие), включая бинарное кодирование. Основопологающие работы по алгоритмам роя частиц и дифференциальной эволюции принадлежат Дж. Кеннеди и Р. Эберхарту. Данные алгоритмы М. Гонгом и А. Тьюсоном были адаптированы для бинарного поиска. В работах Д. Уитни предложены схемы перекодирования решений в ГА для повышения точности поиска, а в работах Н. Шраудольф и Р. Белью исследован ГА с динамическим кодированием решений. Среди российских исследователей необходимо выделить В.М. Курейчика, В.В. Курейчика и Л.А. Гладкова, работы которых посвящены развитию теории ГА. Работы А.В. Еремеева посвящены оценкам сложности ГА и разработке эффективных операторов поиска для различных представлений решений. В работах А.П. Карпенко представлены систематизация и анализ современного состояния области эволюционных вычислений. Работы научной школы Е.С. Семенкина посвя-

щны созданию самоконфигурируемых и самонастраивающихся эволюционных и генетических алгоритмов. Работы Л.А. Демидовой и Ю.А. Скобцова развивают квантово-инспирированные генетические алгоритмы. В работах Е.А. Сопова исследованы способы бинарного кодирования для многих классов задач оптимизации.

Тем не менее, проблема адаптивной коррекции области поиска в бинарных ГА остается недостаточно изученной, а эффективные подходы для ее решения на текущий момент не предложены, что и подтверждает актуальность настоящего диссертационного исследования.

Объектом диссертационного исследования являются генетические алгоритмы, используемые для решения сложных задач глобальной оптимизации с алгоритмически заданными целевыми функциями и смешанными переменными; **предметом исследования** — метод динамической коррекции пространства поиска, применяемый в бинарных генетических алгоритмах.

Целью диссертационного исследования является повышение точности решения сложных задач оптимизации со смешанными переменными бинарным генетическим алгоритмом за счет динамической коррекции области поиска с сохранением размерности бинарного пространства.

Поставленная цель достигается путем решения следующих **задач**:

1. Провести анализ текущего состояния развития методов эволюционной оптимизации и методов кодирования пространства поиска в бинарных генетических алгоритмах.

2. Обосновать и разработать метод динамической коррекции области поиска в генетическом алгоритме.

3. Обосновать и разработать алгоритм корректировки границ кластеров на основе относительной пригодности индивидов для адаптивного расширения или сужения области поиска в зависимости от качества решений на границах.

4. Обосновать и разработать процедуру динамической модификации схемы бинарного кодирования генотипов для объединения разрозненных областей поиска в едином бинарном представлении.

5. Провести экспериментальные исследования для оценки эффективности предложенного подхода на наборе эталонных тестовых задач.

6. Апробировать разработанный метод на практических задачах структурного синтеза моделей машинного обучения.

Научная новизна работы состоит в следующем:

1. Предложен новый метод динамической коррекции области поиска генетического алгоритма, отличающийся от известных процедурой итеративной кластеризации популяции, позволяющей за счет выявления перспективных и неперспективных кластеров определять необходимость сужения области поиска (пункт 4 паспорта специальности).

2. Впервые предложен алгоритм корректировки границ кластеров при динамической коррекции области поиска генетического алгоритма на основе относительной пригодности индивидов, обеспечивающий адаптивное расширение

или сужение области поиска по каждой координате в зависимости от качества решений на границах (пункт 3 паспорта специальности).

3. Разработан новый способ бинарного кодирования вещественных переменных с помощью процедуры динамической модификации схемы кодирования, который, в отличие от известных способов бинарного кодирования вещественных переменных, объединяет идентифицированные перспективные области поиска в едином бинарном представлении, что обеспечивает повышение точности без увеличения размерности бинарного пространства поиска (пункт 5 паспорта специальности).

Теоретическая значимость работы заключается в получении новых знаний о способах решения сложных задач глобальной оптимизации со смешанными переменными. Разработанный метод динамической адаптации области поиска в рамках генетического алгоритма на основе кластерного анализа решений дополняет множество метаэвристических подходов оптимизации. Предложенный подход позволяет формализовать процесс локализации перспективных областей поиска, так как вместо случайного поиска для идентификации областей применяются методы кластерного анализа. Анализ количества, состава и характеристик кластеров позволяет лучше понимать структуру пространства поиска. Полученные результаты могут быть использованы при разработке новых алгоритмов оптимизации.

Практическая значимость работы обусловлена тем, что разработанный метод динамической коррекции области поиска позволяет создавать адаптивные алгоритмы оптимизации, не требующие от пользователя ручного управления схемой кодирования. Предложенный метод динамической коррекции в составе генетического алгоритма внедрен в рамках экспериментальных систем анализа данных по проектам «Классификация государственных контрактов на основе методов обработки естественного языка», «Интеллектуальный анализ юридических лиц для выявления недостоверных сведений в Едином государственном реестре юридических лиц», «Модели и алгоритмы для выявления налогоплательщиков, имеющих признаки незаконного использования схем аренды контрольно-кассовой техники», «Прогнозирование применения необходимых финансовых мер в рамках финансового мониторинга крупных налогоплательщиков». Применение предложенного подхода для проектирования моделей машинного обучения повысило эффективность решения прикладных задач по сравнению с ранее применяемыми методами проектирования.

Методы исследования. В работе использовались методы системного анализа, теории эволюционных алгоритмов, кластерного анализа, теории вероятностей и математической статистики, теории обработки информации, нейросетевого моделирования, а также подходы к созданию прикладных интеллектуальных информационных систем.

Основные положения, выносимые на защиту:

1. Предложенный алгоритм корректировки границ кластеров на основе относительной пригодности индивидов позволяет автоматически изменять структуру области поиска, что повышает точность бинарных генетических алгоритмов по сравнению с их известными версиями без корректировки границ при одинаковых вычислитель-

ных бюджетах в среднем на множестве эталонных тестовых задач, что показано на примере бинарных ГА, алгоритма роя частиц и дифференциальной эволюции.

2. Процедура динамической модификации схемы бинарного кодирования фенотипов (решений) обеспечивает объединение разрозненных после кластеризации областей вещественного пространства поиска в едином бинарном представлении и увеличивает точность бинарного кодирования в перспективных областях при сохранении размерности бинарного пространства поиска.

3. Разработанный метод динамической коррекции области поиска бинарного генетического алгоритма обеспечивает повышение точности решения задач глобальной оптимизации «черного ящика» со смешанными переменными по сравнению со стандартным бинарным генетическим алгоритмом без изменения размерности бинарного пространства поиска и при одинаковом вычислительном бюджете. Повышение точности достигается как для генетического алгоритма с фиксированными настройками, так и для его самоконфигурируемой версии.

Степень достоверности полученных результатов и выводов обеспечивается корректным применением методов сбора, обработки и статистического анализа данных экспериментального исследования предложенных подходов на репрезентативном множестве эталонных задач оптимизации и результатами решения практических задач.

Апробация работы. Процесс разработки и результаты, представленные в диссертации, докладывались и обсуждались на научных конференциях различного уровня, среди которых XXXI Международная открытая научная конференция «Modern Informatization Problems in the Technological and Telecommunication Systems Analysis and Synthesis (MIP-2026'AS)» (г. Йелм, штат Вашингтон, США, январь 2026 г.); XXX Байкальская Всероссийская с международным участием конференция «Информационные и математические технологии в науке и управлении» (г. Иркутск, 2025); XII Международная научная Четаевская конференция «Аналитическая механика, устойчивость и управление» (г. Казань, 2025 г.); Третья молодежная конференция с международным участием «Комплексная автоматизация проектирования и производства 2025» (КАПП(М)-2025) (г. Москва, 2025 г.); 2024 IEEE International Conference on Data Mining Workshops (ICDMW) (г. Абу-Даби, ОАЭ, 2024 г.); XXVIII Международная научно-практическая конференция «Решетневские чтения» (г. Красноярск, 2024 г.); The Thirteenth International Workshop on Mathematical Models and their Applications, Krasnoyarsk (г. Красноярск, 2024 г.).

Диссертационная работа обсуждалась на научных семинарах кафедры «Системного анализа и исследования операций» СибГУ им. М.Ф. Решетнева, кафедры автоматизированных и вычислительных систем Воронежского государственного технического университета, НОЦ «ФНС России» и НОЦ «Технологии искусственного интеллекта» при МГТУ им. Н.Э. Баумана»

Публикации. По теме данного исследования опубликовано 14 печатных работ, среди которых 5 статей в научных журналах, входящих в перечень ВАК, 4 статьи в журналах, входящих в Q1 Web of Science и Scopus, 1 публикация в материалах кон-

ференции, индексируемой Web of Science и Scopus, 4 публикации в других изданиях. Получены 2 свидетельства о государственной регистрации программ для ЭВМ.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **введении** обосновывается актуальность проводимого исследования, определяются его цель и основные задачи; раскрывается научная новизна, практическая значимость работы и положения, выносимые на защиту; приводится информация об апробации полученных результатов на научных конференциях и в рамках научных проектов; указываются сведения о внедрении результатов, а также дается характеристика опубликованных исследований.

В **первой главе** представлены теоретические основы и современные подходы к решению задач глобальной оптимизации со смешанными переменными, обоснованы сложности их реализации в условиях высокой размерности и многоэкстремальности. Показана актуальность применения эволюционных алгоритмов, в частности ГА, для структурно-параметрического синтеза моделей машинного обучения. Выполнен сравнительный анализ методов эволюционной оптимизации, выявлены их ключевые преимущества: устойчивость к шуму, простота параллельной реализации и применимость к задачам «черного ящика». Кроме того, различные метаэвристики, применяемые совместно с методами бинаризации, способны обрабатывать разнотипные переменные. Показано, что для задач глобальной оптимизации «черного ящика» с бинарными переменными ГА обычно превосходит другие подходы.

В работе рассматривается задача глобальной оптимизации

$$f(x) \rightarrow \min_{x \in S} (\max), \quad (1)$$

где $x = (x^r, x^z, x^o, x^b, x^c)$ — вектор переменных задачи, $x^r \in \mathbb{R}^{n_r}$ — вещественные переменные, $x^z \in \mathbb{Z}^{n_z}$ — целочисленные переменные, $x^o \in O^{n_o}$ — порядковые переменные, $x^b \in B_2^{n_b}$ — бинарные переменные, $x^c \in C^{n_c}$ — категориальные переменные, $S = \mathbb{R}^{n_r} \times \mathbb{Z}^{n_z} \times O^{n_o} \times B_2^{n_b} \times C^{n_c}$ — пространство поиска.

Предполагается, что аналитическое выражение целевой функции отсутствует, а её вычисление возможно только с помощью численных методов или натурального эксперимента. Для подобных задач цель формулируется как построение алгоритма, способного после выполнения конечного числа вычислений целевой функции вернуть значение $f(x)$, отличающееся от оптимального не больше чем на некоторую заданную величину. Для решения задачи (1) обычно применяются различные эвристические алгоритмы, включая методы эволюционных вычислений. При этом переменные всех типов выражаются в единой шкале — бинарной. Используется стандартное бинарное кодирование порядковых переменных, которое аналогичным образом применяется и к целочисленным переменным. Для неупорядоченных переменных наименований эффективных методов кодирования не предложено. Вещественные переменные переводятся в дискретные с некоторой точностью (обычно равномерной сеткой), далее применяется тот же способ кодирования, что и для порядковых переменных. Таким образом, одним из «узких мест» применения метаэвристик для задач со смешанными переменными является способ кодирования вещественных переменных.

Анализ как классического варианта, так и существующих модификаций ГА показал существенный недостаток: фиксированные границы области поиска ограничивают эффективность метода при решении сложных оптимизационных задач. Основная трудность проявляется в следующем. При использовании традиционного способа бинаризации даже с высокой точностью точки дискретной сетки могут попасть не во все оптимумы исходного пространства или вовсе не попасть ни в один из оптимумов, и поэтому статический способ не гарантирует нахождение глобального или достижение локального оптимума.

Для преодоления указанного недостатка в настоящем исследовании обосновывается и предлагается идея динамической коррекции области поиска в ГА, основанная на кластеризации генотипов для выявления групп решений с высокой приспособленностью с целью сосредоточить поиск вблизи потенциальных глобальных оптимумов, а также идея адаптивного сужения/расширения границ области поиска в зависимости от плотности популяции и скорости сходимости. Показана необходимость разработки методов динамической коррекции области поиска для современных приложений машинного обучения (ML), включая оптимизацию архитектур нейросетей и выбор гиперпараметров в автоматизированном машинном обучении (AutoML), а также при проектировании сложных инженерных систем.

Вторая глава посвящена разработке и исследованию нового метода динамической коррекции области поиска в бинарном генетическом алгоритме, для чего на каждом поколении его работы после оценки значений целевой функции индивидов выполняется кластеризация решений, на основе которой реализуются процедуры сегментации, расширения и сужения области поиска (рисунок 1).

На первом этапе осуществляется кластеризация решений (индивидов) в области поиска. Кластеры, имеющие малые размеры и содержащие индивидов, средняя пригодность которых ниже, чем средняя пригодность по популяции, исключаются из дальнейшего рассмотрения алгоритма. В результате успешной проверки критерия возможности коррекции области поиска, отражающего соотношение областей незанятых кластерами к занятым, уточняются области поиска на основе сформированных зон интереса, и осуществляется последующее удаление из популяции индивидов, не попадающих в новые границы. Для обеспечения корректной работы генетических операторов производится объединение разрозненных областей поиска в едином бинарном представлении. Это достигается применением процедуры динамической модификации схемы бинарного кодирования генотипов, на основе которой перекодируются индивиды текущего поколения.

Введем следующие обозначения. Пусть \vec{X} — это вектор решения задачи оптимизации, содержащий только вещественные переменные из исходной постановки со смешанными переменными:

$$\vec{X} = \{x_i, i = 1, \dots, n\}, \quad (2)$$

где n - размерность пространства поиска, x_i — переменная.

Условия для переменных — область определения x_i :

$$x_i \in D_{x_i} = [x_i^l, x_i^h], \quad (3)$$

где x_i^l , x_i^h — нижнее и верхнее значение по переменной x_i , соответственно.

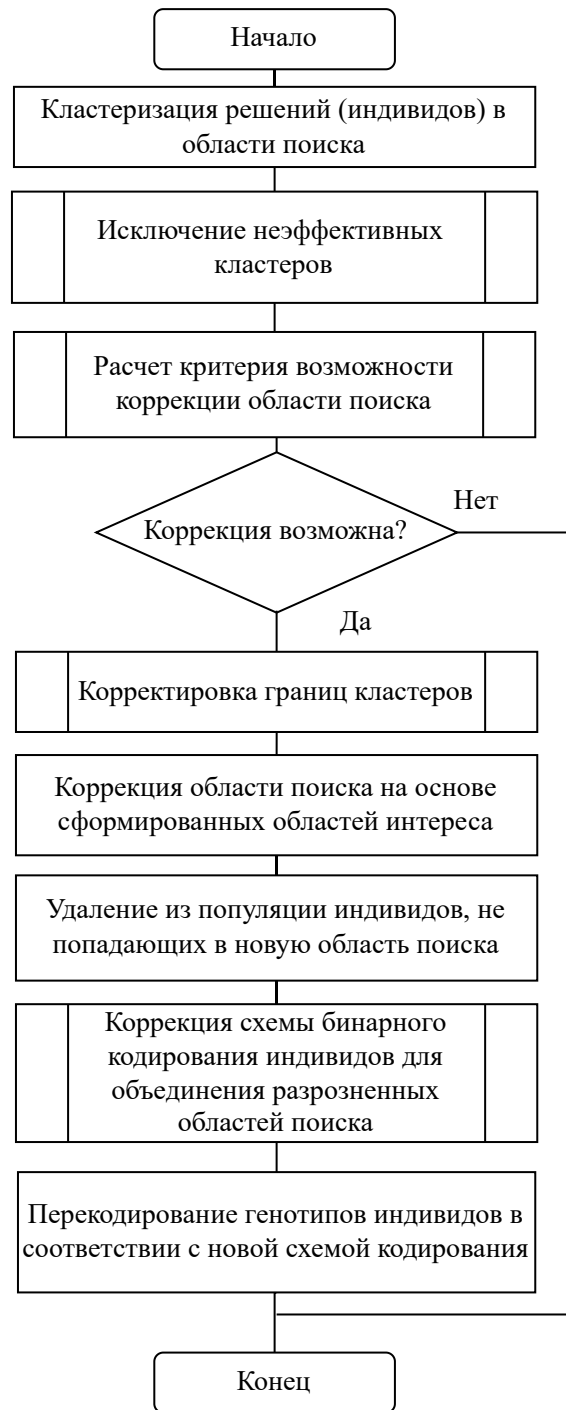


Рисунок 1 — Алгоритм динамической коррекции области поиска на основе кластеризации решений.

В классическом (стандартном) генетическом алгоритме D_{x_i} не изменяется для всех x_i . В предлагаемом подходе D_{x_i} динамически изменяется для всех x_i .

Обозначим как G — общее число поколений ГА, g — номер текущего поколения ГА, M — общее количество индивидов в популяции, m — номер рассматриваемого индивида, K — общее число кластеров, k — номер текущего кластера, $D_{\vec{x}}^{g,k} = (D_{x_1}^{g,k}, \dots, D_{x_n}^{g,k})$ — область k -го кластера на поколении g в n -мерном

области поиска, MS — минимальное количество точек в кластере (требуется, если используется алгоритм кластеризации плотности — Density-Based Spatial Clustering of Applications with Noise, DBScan), Ind_m^g — m -ый индивид в поколении g ($m=1, \dots, M$) из числа всех индивидов популяции, M_k^g — количество индивидов в кластере k на поколении g , $Ind_m^{g,k}$ — индивид m в поколении g из кластера k ($m=1, \dots, M_k^g$), $F_{Ind_m^g}$ — пригодность индивида Ind_m^g , $IndSet^g = \bigcup_{k=1}^K IndSet_k^g$ — множество всех индивидов на поколении g , $IndSet_k^g$ — множество всех индивидов в кластере k на поколении g , $N_{IndSet_k^g}$ — число индивидов в кластере k на поколении g , $\bar{F}_{IndSet_k^g}$ — средняя пригодность индивидов в кластере k на поколении g .

Область поиска по переменной x_i на поколении g представляется объединением:

$$D_{x_i}^g = \bigcup_{k=1}^K D_{x_i}^{g,k}, \quad (4)$$

где $D_{x_i}^{g,k}$ — область поиска по переменной x_i на поколении g для кластера k .

В общем случае коррекция границ области поиска может быть представлена в следующем виде:

$$D_{x_i}^{g,k} = [x_i^{g,kl} + \Delta_{x_i^{g,k,l}}, x_i^{g,kh} + \Delta_{x_i^{g,k,h}}], \quad (5)$$

где $x_i^{g,kl}$, $x_i^{g,kh}$ — нижняя и верхняя граница по текущей переменной x_i для кластера k в поколении g , $\Delta_{x_i^{g,k,l}} \in \mathbb{R}$, $\Delta_{x_i^{g,k,h}} \in \mathbb{R}$ — это величина коррекции нижней и верхней границ соответственно.

Кластеризация решений (индивидов) в области поиска выполняется одним из методов кластерного анализа. Количество кластеров либо задается априорно (например, для K-Means, K-Medoids, агломеративной кластеризации, спектральной кластеризации и т.д.), либо определяется автоматически (например, для DBScan задается минимальное число точек в кластере и радиус окрестности вокруг точки, для MeanShift — ширина полуядра и т.д.).

После проведения кластеризации и оценки границ областей пространства поиска, соответствующих выделенным кластерам, производится исключение неэффективных кластеров («схлопывание» малых кластеров). При этом эффективность оценивается по двум критериям: количеству индивидов, попавших в кластер, и отношению пригодности индивидов кластера к общей пригодности популяции. Если кластер оказался неэффективным, то соответствующая область $D_{x_i}^{g,k}$ полностью исключается для всех $i=1, \dots, n$.

Пошаговое представление вышеуказанной процедуры:

Шаг 1. Для всех $k=1, \dots, K$ рассчитать число индивидов $N_{IndSet_k^g}$.

Шаг 2. Рассчитать среднее число индивидов по всем кластерам

$$\bar{N}_{IndSet^g} = \frac{1}{K} \sum_{k=1}^K N_{IndSet_k^g}.$$

Шаг 3. Рассчитать среднюю пригодность по всей популяции

$$\bar{F}_{\text{Ind}^g} = \frac{1}{M} \sum_{m=1}^M F_{\text{Ind}_m^g}.$$

Шаг 4. Перейти к последовательной проверке кластеров, положив $k=1$.

Шаг 5. Если

$$N_{\text{IndSet}_k^g} < \bar{N}_{\text{IndSet}^g},$$

то перейти к шагу 6, иначе — к шагу 8.

Шаг 6. Рассчитать среднюю пригодность индивидов в k -м кластере

$$\bar{F}_{\text{IndSet}_k^g} = \frac{1}{M_k^g} \sum_{m=1}^{M_k^g} F_{\text{Ind}_m^{g,k}}.$$

Шаг 7. Если

$$\bar{F}_{\text{IndSet}_k^g} < \bar{F}_{\text{Ind}^g},$$

то положить $D_{x_i}^{g,k} = \emptyset$ для всех $i=1, \dots, n$.

Шаг 8. Если $k < K$, то положить $k=k+1$ и перейти к шагу 5. Иначе расчёт завершён:

$$D_{x_i}^g = \bigcup_{k=1}^K D_{x_i}^{g,k}.$$

Для оценки целесообразности применения процедуры коррекции области поиска ГА разработан критерий, основывающийся на отношении мощностей множеств, представляющих собой задействованную и незадействованную кластерами области пространства поиска.

Область $D_{\vec{x}}^{g-1, \bar{k}}$, не задействованная кластерами, может быть найдена в соответствии с формулой (6):

$$D_{\vec{x}}^{g-1, \bar{k}} = \left(D_{x_1}^{g-1} \setminus \bigcup_{k=1}^K D_{x_1}^{g,k}, \dots, D_{x_n}^{g-1} \setminus \bigcup_{k=1}^K D_{x_n}^{g,k} \right). \quad (6)$$

Если размер области поиска на поколении $g-1$, не задействованной кластерами с решениями на поколении g , превышает суммарный размер областей, занятых кластерами на поколении g , выполняется процедура коррекции области поиска (7):

$$\text{card}\left(D_{\vec{x}}^{g-1, \bar{k}}\right) > \sum_{k=1}^K \text{card}\left(D_{\vec{x}}^{g,k}\right). \quad (7)$$

где $\text{card}(\cdot)$ обозначает мощность дискретного множества, то есть число допустимых точек в соответствующей области пространства поиска. В случае бинарного кодирования переменной x_i , представленной b_i битами, выполняется

$$\text{card}(D_{x_i}) = 2^{b_i}.$$

Для многомерного пространства поиска $D_{\vec{x}} = \prod_{i=1}^n D_{x_i}$ мощность области определяется как

$$\text{card}(D_{\vec{x}}) = \prod_{i=1}^n \text{card}(D_{x_i}).$$

Алгоритм корректировки границ кластеров. В результате выполнения предыдущих шагов алгоритма сформировано множество эффективных кластеров с решениями. При этом в оценках границ областей пространства поиска не определены значения коррекции. С целью расширения области поиска в сторону эффективных решений, а также редукции областей в малозначимых регионах предлагается процедура корректировки границ кластеров.

Введем следующие обозначения: $\text{Ph}_m^{g,k} = \text{Ph}^g(\text{Ind}_m^{g,k}) = (x_1^{g,k,m}, \dots, x_n^{g,k,m})$ — фенотип m -го индивида в кластере k на поколении g , $\text{Ph}^g(\cdot)$ — схема перевода генотипа в фенотип на поколении g , $\text{Ind}_m^{g,k} = \text{Bin}^g(\text{Ph}_m^{g,k})$ — генотип m -го индивида в кластере k на поколении g , $\text{Bin}^g(\cdot)$ — схема перевода фенотипа в генотип на поколении g .

Рассмотрим произвольное поколение g работы ГА, кластер k и переменную пространства поиска x_i . Шаги нижеизложенной процедуры необходимо повторить для всех переменных по всем кластерам, для которых выполняется $D_x^{g,k} \neq \emptyset$ ($k = \overline{1, K}$, $i = \overline{1, n}$).

Шаг 1. Рассчитать относительную пригодность $RE_m^{g,k}$ всех индивидов во всех кластерах:

$$RE_m^{g,k} = RE(\text{Ind}_m^{g,k}) = \frac{F_{\text{Ind}_m^{g,k}}}{\sum_{p=1}^M \sum_{q=1}^K F_{\text{Ind}_q^{g,p}}}. \quad (8)$$

Шаг 2. В каждом кластере найти по каждой переменной пространства поиска x_i ($i = \overline{1, n}$) индивида $\text{Ind}L_{x_i}^{g,k}$ с наименьшим значением x_i и индивида $\text{Ind}H_{x_i}^{g,k}$ с наибольшим значением x_i :

$$\text{Ind}L_{x_i}^{g,k} = \text{Bin}^g \left(\min_{m=\overline{1, M_k^g}} \text{Ph}_m^{g,k} \right), \quad \text{Ind}H_{x_i}^{g,k} = \text{Bin}^g \left(\max_{m=\overline{1, M_k^g}} \text{Ph}_m^{g,k} \right). \quad (9)$$

Шаг 3. Рассчитать значения коррекции верхней $\Delta_{x_i}^{g,k,h}$ и нижней $\Delta_{x_i}^{g,k,l}$ границ в соответствии с шагами 3.1-3.5.

Шаг 3.1. Рассчитать внутри-кластерную эффективность (Intra-Cluster Efficiency) — долю относительной эффективности m -го индивида в кластере k на поколении g :

$$\text{ICE}(\text{Ind}_m^{g,k}) = \frac{\text{RE}(\text{Ind}_m^{g,k})}{\sum_{q=1}^M \text{RE}(\text{Ind}_q^{g,k})}. \quad (10)$$

Шаг 3.2. Сформировать множество AI_k^g индивидов, для которых $\text{ICE}(\text{Ind}_m^{g,k}) \geq \frac{1}{M_k^g}$.

Шаг 3.3. Если $\text{ICE}(\text{Ind}L_{x_i}^{g,k}) \geq \frac{1}{M_k^g}$, то выполнить процедуру расширения нижней границы и рассчитать ее коррекцию:

$$\Delta_{x_i}^{g,k,l} = - \left| x_i^{g,k,h} - x_i^{g,k,l} \right| \text{ICE}(\text{Ind}L_{x_i}^{g,k}). \quad (11)$$

Перейти к шагу 3.5.

Иначе — перейти к шагу 3.4.

Шаг 3.4. Выполнить процедуру сужения нижней границы.

Заменить текущий $\text{Ind}L_{x_i}^{g,k}$ на индивид из множества AI_k^g в соответствии с формулой:

$$\text{Ind}L_{x_i}^{g,k} = \text{Bin}^g \left(\min_{\text{Ind}_m^{g,k} \in AI_k^g} \text{Ph}^g(\text{Ind}_m^{g,k}) \right). \quad (12)$$

Перейти к шагу 3.3.

Шаг 3.5. Если $\text{ICE}(\text{Ind}H_{x_i}^{g,k}) \geq \frac{1}{M_k^g}$, то выполнить процедуру расширения верхней границы и рассчитать ее коррекцию:

$$\Delta_{x_i^{g,k,h}} = \left| x_i^{g,k,h} - x_i^{g,k,l} \right| \text{ICE}(\text{Ind}H_{x_i}^{g,k}). \quad (13)$$

Закончить расчет.

Иначе — перейти к шагу 3.6.

Шаг 3.6. Выполнить процедуру сужения верхней границы.

Заменить текущий $\text{Ind}H_{x_i}^{g,k}$ на индивид из множества AI_k^g в соответствии с формулой:

$$\text{Ind}H_{x_i}^{g,k} = \text{Bin}^g \left(\max_{\text{Ind}_m^{g,k} \in AI_k^g} \text{Ph}^g(\text{Ind}_m^{g,k}) \right). \quad (14)$$

Перейти к шагу 3.5.

Описанная выше процедура позволяет не только производить сужение границ областей поиска (пример показан на рисунке 2), но и расширять поисковое пространство в направлении расположения эффективных решений.

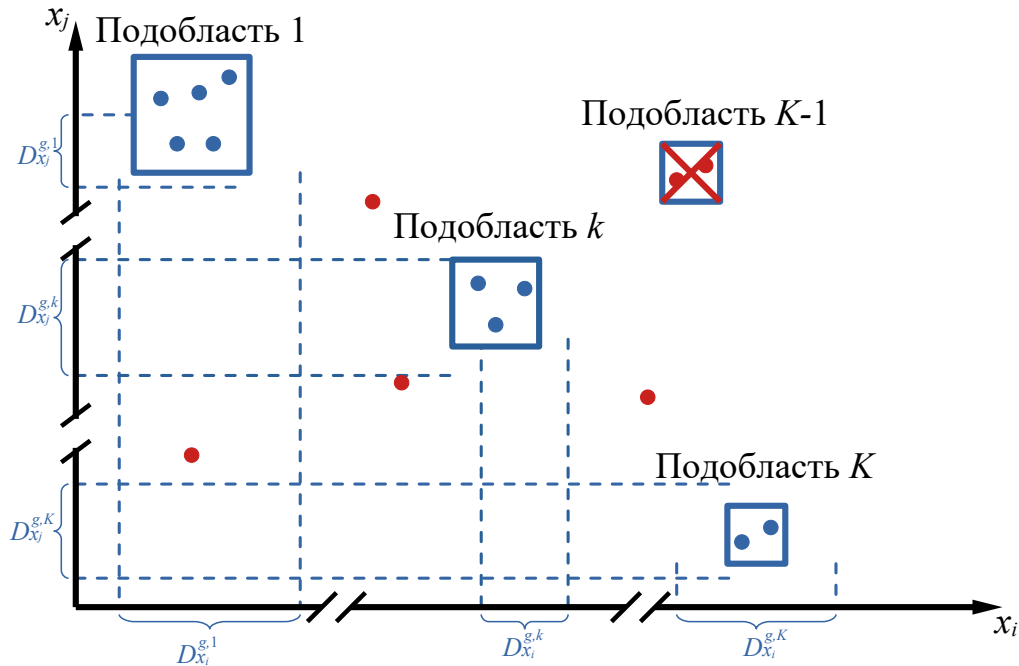


Рисунок 2 — Коррекция границ области поиска.

Процедура динамической модификации схемы бинарного кодирования генотипов. Для корректной работы генетических операторов в условиях реализации оптимизационной процедуры в разрозненных областях поиска без внесения дополнительных ограничений в постановку оптимизационной задачи предлагается динамически модифицировать схему бинарного кодирования фенотипов индивидов с целью объединения разрозненных областей поиска в едином бинарном представлении. Кроме того, предлагаемая процедура позволяет в условиях сохранения постоянной длины генотипа индивида уменьшать шаг дискретизации получаемых решений в пространстве вещественных переменных.

Обозначим $\text{Bin}^1(\cdot) = (\text{Bin}_{x_1}^1(\cdot), \dots, \text{Bin}_{x_n}^1(\cdot))$ — классическая схема бинарного кодирования, которая постоянна в течение всех поколений работы стандартного ГА и является оператором над множеством фенотипов решений, $N\text{Bit}_{x_i}$ — количество бит, необходимое для кодирования точек перебора области поиска по каждой переменной x_i , $\text{step} = \frac{x^n - x^1}{2^n - 1}$ — скорректированный шаг с учетом максимального количества кодируемых точек для заданного $N\text{Bit}_{x_i}$, Ph_{m,x_i}^g — фенотип индивида по координате x_i , Ind_{m,x_i}^g — компонент индивида по координате x_i :

$$\text{Ind}_{m,x_i}^g = \left(\frac{\text{Ph}_{m,x_i}^g - x_i^l}{\text{step} C_{x_i}^1} \right)_{10 \rightarrow 2}, \quad (15)$$

где $(\cdot)_{10 \rightarrow 2}$ — операция перевода вещественного числа в двоичное представление с $N\text{Bit}_{x_i}$ разрядами, при этом первые незначимые разряды имеют значение 0.

Для поколений, в которых не происходит корректировки границ области поиска, схема кодирования не претерпевает изменений, т.е. $\text{Bin}^g(\cdot) = \text{Bin}^{g-1}(\cdot)$.

Для поколений, в которых происходит корректировка границ области поиска, $\text{Bin}^g(\cdot)$ определяется как нижеследующая процедура:

Шаг 1. Упорядочивание кластеров по возрастанию левых границ $x_i^{g,k,l} + \Delta_{x_i^{g,k,l}}$.

Шаг 2. Для каждого кластера k найти показатель удаленности границы минимального значения по x_i относительно начала $D_{x_i}^{g,l}$ пространства поиска $D_{x_i}^g$, которая определяется как сдвиг $S_{x_i}^{g,kl}$:

$$\begin{aligned} S_{x_i}^{g,kl} = & \left(x_i^{g,1,h} + \Delta_{x_i^{g,1,h}} \right) - \left(x_i^{g,1,l} + \Delta_{x_i^{g,1,i}} \right) + \\ & + \sum_{\substack{e=2,k-1 \\ x_i^{g,e,h} > x_i^{g,d,h}, \forall d=\overline{1,e-1} \\ x_i^{g,e,l} > x_i^{g,d,h}, \forall d=\overline{1,e-1}}} \left(\underbrace{\left(x_i^{g,e,h} + \Delta_{x_i^{g,e,h}} \right) - \left(x_i^{g,e,l} + \Delta_{x_i^{g,e,i}} \right)}_{\text{неперекрывающиеся отрезки}} \right) + \\ & + \sum_{\substack{e=2,k-1 \\ x_i^{g,e,h} > x_i^{g,d,h}, \forall d=\overline{1,e-1} \\ x_i^{g,e,l} \leq x_i^{g,d,h}, \forall d=\overline{1,e-1}}} \left(\underbrace{\left(x_i^{g,e,h} + \Delta_{x_i^{g,e,h}} \right) - \max_{d=\overline{1,e-1}} \left(x_i^{g,d,h} + \Delta_{x_i^{g,d,i}} \right)}_{\text{отрезки с перекрытиями}} \right). \quad (16) \end{aligned}$$

При этом:

- если $x_i^{g,e,l} > x_i^{g,e-1,h}$, то в расчет берется вся длина отрезка;
- если $x_i^{g,e,h} > x_i^{g,d,h}, \forall d = \overline{2..e-1}$, то длина отрезка не учитывается;
- если $x_i^{g,e,l} \leq x_i^{g,e-1,h}, \forall d = \overline{2..e-1}$, то в расчет берется неперекрывающаяся часть длины отрезка.

Шаг 3. Скорректировать шаг по каждой переменной x_i :

$$\text{step}C_{x_i}^1 = \frac{\left(x_i^{g,k,h} + \Delta_{x_i^{g,k,h}}\right) - \left(x_i^{g,k,l} + \Delta_{x_i^{g,k,l}}\right)}{2^{N\text{Bit}_{x_i}}}. \quad (17)$$

Шаг 4. Для всех решений, оставшихся после процедуры удаления индивидов и не попадающих в новую область поиска, выполнить процедуру:

$$\text{Ind}_{m,x_i}^{g+1,k} = \begin{cases} \frac{\text{Ph}_{m,x_i}^{g,k} - \left(x_i^{g,1,l} + \Delta_{x_i^{g,1,l}}\right)}{\text{step}C_{x_i}^1}, & k = 1 \\ S_{x_i^{g,k,l}} + \frac{\text{Ph}_{m,x_i}^{g,k} - \left(x_i^{g,k,l} + \Delta_{x_i^{g,1,l}}\right)}{\text{step}C_{x_i}^1}, & k = \overline{2..K}. \end{cases} \quad (18)$$

Результаты исследования эффективности предложенного подхода на функциях CEC2017. Для анализа эффективности метода динамической коррекции проведены сравнения ряда классических и современных метаэвристических алгоритмов, позволяющих работать с бинаризованными задачами. Каждый алгоритм реализован и исследован в своем исходном варианте и в версии с динамической коррекцией области поиска (ДКОП). В качестве оценки снизу проведено исследование классического ГА. Современные подходы: самонастраивающийся генетический алгоритм с адаптацией на основе истории успеха SelfCSHAGA (Шерстнев, Семенкин, 2025), бинарный алгоритм роя частиц с полной информацией Binary Fully Informed PSO (Kessentini, 2024) и бинарный самонастраивающийся алгоритм дифференциальной эволюции на основе доли успешных решений L-SRTDE (Становов, 2024). Алгоритмы роя частиц и дифференциальной эволюции напрямую не работают с бинарными переменными, поэтому используется стандартный подход: каждый бит представляется вещественной переменной в интервале $[0, 1]$, выбор бита осуществляются относительно порога 0.5. Эксперименты проводились на наборе эталонных тестовых задач IEEE CEC-2017, включающем 30 функций различной сложности. Для каждого алгоритма выполнялось по 50 независимых запусков на каждой функции.

В таблицах 1, 2 и 3 представлены агрегированные показатели для бинарных ГА, метода роя частиц и дифференциальной эволюции по четырём группам тестовых функций набора CEC 2017: унимодальным, многоэкстремальным, гибридным и композиционным. Для каждой группы указано число задач, в которых модификация с ДКОП продемонстрировала статистически значимое по критерию Манна-Уитни-Уилкоксона ($p < 0.05$) улучшение точности по сравнению с базовой версией алгоритма (победы), отсутствие значимых различий (ничьи) и большую эффективность базового алгоритма (поражения).

Как показали численные эксперименты, на наборе задач IEEE CEC 2017 алгоритмы с ДКОП превосходят свои базовые алгоритмы без коррекции

Таблица 1 — Анализ эффективности бинарных ГА с ДКОП в сравнении с базовым алгоритмом без ДКОП.

Базовый алгоритм, размерность	Показатель	Унимод. (f1–f2)	Многоэктр. (f3–f9)	Гибридные (f10–f19)	Композиц. (f20–f30)	Итого
GA $D=10$	Победы	2	7	8	10	27
	Ничьи	0	0	2	0	2
	Поражения	0	0	0	1	1
SelfCSHAGA $D=10$	Победы	1	7	6	10	24
	Ничьи	1	0	3	0	4
	Поражения	0	0	1	1	2
SelfCSHAGA $D=30$	Победы	1	7	9	9	26
	Ничьи	1	0	0	1	2
	Поражения	0	0	1	1	2
SelfCSHAGA $D=50$	Победы	1	7	7	9	24
	Ничьи	1	0	2	0	3
	Поражения	0	0	1	2	3

Таблица 2 — Анализ эффективности бинарных PSO с ДКОП в сравнении с базовым алгоритмом без ДКОП.

Базовый алгоритм, размерность	Показатель	Унимод. (f1–f2)	Многоэктр. (f3–f9)	Гибридные (f10–f19)	Композиц. (f20–f30)	Итого
BPSO, $D=10$	Победы	1	6	5	3	15
	Ничьи	1	1	5	8	15
	Поражения	0	0	0	0	0
BPSO, $D=30$	Победы	2	5	6	10	23
	Ничьи	0	1	4	1	6
	Поражения	0	1	0	0	1
BPSO, $D=50$	Победы	2	7	6	9	24
	Ничьи	0	0	3	2	5
	Поражения	0	0	1	0	1

Таблица 3 — Анализ эффективности бинарных L-SRTDE с ДКОП в сравнении с базовым алгоритмом без ДКОП.

Базовый алгоритм, размерность	Показатель	Унимод. (f1–f2)	Многоэктр. (f3–f9)	Гибридные (f10–f19)	Композиц. (f20–f30)	Итого
L-SRTDE, $D=10$	Победы	1	4	5	5	15
	Ничьи	1	3	5	6	15
	Поражения	0	0	0	0	0
L-SRTDE, $D=30$	Победы	2	4	5	5	16
	Ничьи	0	3	5	6	14
	Поражения	0	0	0	0	0
L-SRTDE, $D=50$	Победы	2	4	8	7	21
	Ничьи	0	2	1	3	6
	Поражения	0	0	0	1	1

практически на всех задачах, что делает выбор алгоритма с ДКОП более предпочтительным при решении задач с алгоритмически заданными функциями.

Для сравнения эффективности разных метаэвристик с ДКОП результаты их решения задач были проранжированы по средним значениям целевой функции, полученным по серии независимых запусков (меньший ранг соответствует лучшему результату). Далее ранги усреднялись по всем функциям набора для каждой размерности задачи. Результаты сравнения приведены в таблице 4.

Таблица 4 — Ранжирование результатов разных метаэвристик с ДКОП на задачах CEC 17.

Размерность D	BPSO+ДКОП	L-SRTDE+ДКОП	SelfCSHAGA+ДКОП
10	2.567	1.800	1.633
30	2.567	2.067	1.367
50	2.533	2.300	1.167

Как видно из таблицы, наилучшие значения среднего ранга на всех рассмотренных размерностях демонстрирует подход SelfCSHAGA+ДКОП. Несмотря на то, что алгоритм базовый L-SRTDE является ведущим для решения вещественных задач, его бинарная версия существенно уступает бинарному ГА. Наиболее сильное отставание бинарных версий метода роя частиц и дифференциальной эволюции наблюдается с ростом размерности, что объясняется способом бинаризации, применяемым в этих алгоритмах.

Учитывая полученные результаты, дальнейшие эксперименты проводятся с использованием самонастраиваемого ГА (SelfCSHAGA).

Решение тестовых прикладных задач. Эффективность подхода также оценена при решении прикладных задач оптимизации со смешанными переменными, используемых на соревнованиях по оптимизации IEEE CEC: задача оптимизации пружины растяжения-сжатия (TCSД), проектирование редуктора скорости (SRD) и проектирование резервуара высокого давления (PVDP).

На рисунке 3 представлено сравнение распределений лучших значений целевых функций, полученных SelfCSHAGA и SelfCSHAGA+ДКОП по сериям независимых запусков

Для задачи TCSД медианное значение целевой функции при использовании SelfCSHAGA+ДКОП оказалось ниже, чем у SelfCSHAGA, приблизительно на 5–7%, при одновременном снижении межквартильного размаха. В задаче PVDP применение SelfCSHAGA+ДКОП также приводит к снижению медианы целевой функции примерно на 8–12%, при этом сокращается число выбросов, присутствующих при использовании SelfCSHAGA.

В **третьей главе** представлены результаты апробации предложенного метода динамической коррекции области поиска при решении ряда практических задач оптимизации.

Оптимизация гиперпараметров моделей машинного обучения в системе классификации текстовых материалов госконтрактов. Рассмат-

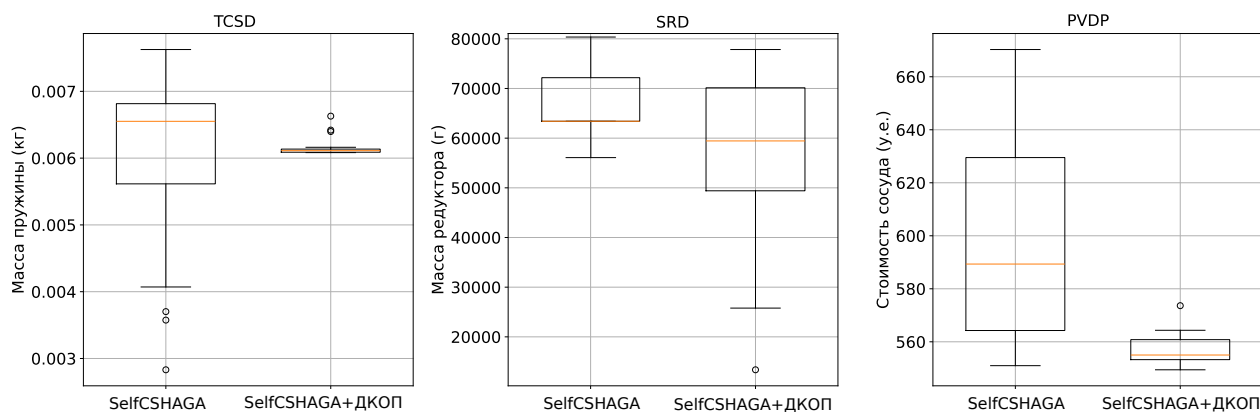


Рисунок 3 — Сравнение распределений лучших значений целевой функции для алгоритмов SelfCSHAGA и SelfCSHAGA+ДКОП при решении задач TCS, SRD и PVDP.

ривалась задача оптимизации гиперпараметров моделей в интеллектуальной системе классификации текстовых описаний государственных контрактов по кодам ОКВЭД. Набор данных включал около 20 млн записей. Целевая функция формулировалась как максимизация F1-меры для нейронной сети и точности классификации для случайного леса на валидационной выборке. Вектор переменных включал параметры различной природы: дискретные (выбор алгоритма классификации), целочисленные (глубина деревьев, число нейронов, число слоёв), вещественные (скорость обучения, коэффициенты регуляризации), а также бинарные параметры (использование отдельных методов предобработки). Таким образом, задача представлена переменными разных типов, а целевая функция – моделью «черного ящика».

Использование предложенного в диссертационном исследовании метода позволило локализовать перспективные области пространства гиперпараметров и обеспечить повторяемость результатов при многократном запуске алгоритма. В результате для модели случайного леса точность классификации увеличилась с 0.71 до 0.79, а для многослойного перцептрона достигнута F1-мера 0.816, что превышает результаты, полученные при использовании стандартного генетического алгоритма.

Задача оптимизации гиперпараметров модели бинарной классификации несбалансированных данных юридических лиц. Решалась задача оптимизации гиперпараметров модели XGBoost для бинарной классификации юридических лиц в ходе налоговых проверок по косвенным признакам. Набор данных включал 97095 организаций, при этом доля положительного класса (юр. лица, выявленные для камеральной проверки) составляла около 22%, что формировало выраженный дисбаланс. Целевая функция формулировалась как максимизация F1-score положительного класса. Вектор оптимизируемых переменных включал вещественные параметры (learning rate, subsample, colsample_bytree), целочисленные параметры (max_depth, n_estimators), а также параметры регуляризации (коэффициенты L1 и L2).

Применение предложенного в работе метода динамической коррекции области поиска позволило уменьшить число поколений, необходимых для достижения стационарного значения целевой функции, и повысить значение F1-score с 0.744 до 0.756. Точность классификации увеличилась с 0.79 до 0.82.

ЗАКЛЮЧЕНИЕ

В рамках диссертационного исследования получены новые научные результаты, включающие разработку метода и алгоритма, направленных на повышение точности решения сложных задач оптимизации со смешанными переменными бинарным генетическим алгоритмом за счет динамической коррекции пространства поиска. Цель диссертационного исследования достигнута путем решения поставленных задач, а именно:

1. Проведен анализ современного состояния методов эволюционной оптимизации, который показал, что, несмотря на широкое применение бинарных ЭА в задачах глобальной оптимизации со смешанными переменными, их эффективность ограничивается фиксацией области поиска. Это приводит к увеличению числа вычислений целевой функции вследствие нерационального распределения решений и к снижению точности получаемых решений.

2. Разработан новый метод динамической коррекции области поиска, который позволяет перераспределять вычислительные ресурсы алгоритма в пользу перспективных областей пространства решений.

3. Разработан алгоритм корректировки границ кластеров, основанный на вычислении относительной пригодности индивидов, который позволяет адаптивно расширять или сужать область поиска по каждой координате в зависимости от качества решений на границах, исключая менее перспективные области поиска.

4. Разработана процедура динамической модификации схемы бинарного кодирования генотипов, которая позволяет за счет объединения разрозненных областей поиска в едином бинарном представлении гарантировать корректное применение стандартных генетических операторов и обеспечить повышение точности получаемых решений без увеличения размерности бинарного вектора.

5. Проведены экспериментальные исследования на наборах тестовых функций, которые включают стандартные задачи, задачи чемпионата мира по эволюционной оптимизации IEEE CEC 2017 и прикладные тестовые задачи с соревнований по оптимизации IEEE CEC. Результаты экспериментов показали, что применение предложенного подхода повышает эффективность базового алгоритма, который использует стандартное статическое бинарное кодирование. Поскольку метод динамической коррекции не меняет структуру алгоритма, повышение эффективности достигается на всех исследованных алгоритмах, включая как стандартный ГА, так и ведущие на текущий момент подходы: бинарный алгоритм роя частиц с полной информацией, бинарный самонастраивающийся алгоритм дифференциальной эволюции на основе доли успешных решений L-SRTDE и самонастраивающийся генетический алгоритм SelfCSHAGA.

6. Проведена апробация на практических задачах структурно-параметрического синтеза моделей машинного обучения для классификации госконтрактов и выявления недостоверных сведений о юридических лицах, в которых получены новые практические результаты, превосходящие показатели стандартных алгоритмов.

Таким образом, в диссертационном исследовании поставлена и решена актуальная научная задача повышения эффективности решения сложных задач глобальной оптимизации алгоритмически заданных целевых функций со смешанными переменными, что имеет существенное значение для теории и практики системного анализа, управления и обработки информации.

СПИСОК РАБОТ АВТОРА ПО ТЕМЕ ДИССЕРТАЦИИ

Публикации в изданиях, рекомендованных ВАК России:

1. Малашин И.П. Методика корректировки границ кластеров при динамическом формировании области поиска в генетическом алгоритме // Системы управления и информационные технологии, №3(101), 2025. С. 96-100

2. Малашин И.П. Процедура динамической модификации схемы бинарного кодирования индивидов в генетическом алгоритме // Моделирование, оптимизация и информационные технологии. – 2025. – Т. 13, № 3. — DOI: 10.26102/2310-6018/2025.50.3.040

3. Малашин И.П. Применение методов кластерного анализа для динамической коррекции области поиска в генетическом алгоритме / И.П. Малашин, В.С. Тынченко // Сибирский аэрокосмический журнал. 2025. Т. 26, № 3. С. 318–333

4. Малашин И.П. Анализ эффективности самоконфигурируемого бинарного генетического алгоритма с модифицированным методом динамической коррекции области поиска / И.П. Малашин, Е.А. Сопов // Моделирование, оптимизация и информационные технологии. – 2026. – Т. 14, № 2. — DOI: 10.26102/2310-6018/2026.53.2.001

5. Малашин И.П. Анализ причин немонотонной сходимости генетического алгоритма с динамической коррекцией области поиска / И.П. Малашин, Е.А. Сопов // Системы управления и информационные технологии, №1(103), 2026. С. 34-37

Публикации в журналах, входящих в Q1/Q2 Web of Science и Scopus:

6. Malashin I.P. Adaptive search space scaling in self-adaptive genetic algorithms for high-dimensional global black-box optimization problems / I.P. Malashin, E.A. Sopov, V.A. Nelyub, A.S. Borodulin, A.P. Gantimurov, V.N. Tynchenko // Discover Artificial Intelligence. – 2026. - Т. 6. - Ст. 491. — DOI: 10.1007/s44163-026-01186-3

7. Malashin I. Generation of Natural-Language Explanations for Static-Analysis Warnings Using Single-and Multi-Objective Optimization // Computers. – 2025. – Т. 14. – № 12. – С. 534. — DOI: 10.3390/computers14120534

8. Malashin I.P. Confidence-Aware Rejection Class Minimization in Indirect Tax Audit Classification / I.P. Malashin, I.S. Masich, V.S. Tynchenko, A.P. Gantimurov, V.A. Nelyub, A.S. Borodulin // IEEE Access. — 2024. — Т. 12. — С. 192554–192571. — DOI: 10.1109/ACCESS.2024.3519333.

9. Malashin I. Application of Natural Language Processing and Genetic Algorithm to Fine-Tune Hyperparameters of Classifiers for Economic Activities Analysis / I. Malashin, I. Masich, V. Tynchenko, V. Nelyub, A. Borodulin, A. Gantimurov // Big Data and Cognitive Computing. – 2024. – Т. 8, № 6. – Ст. 68. – DOI: 10.3390/bdcc8060068.

Публикации, индексируемые в Web of Science и Scopus:

10. Malashin I.P. Anomaly Detection in Financial Data Using GA-Optimized MLP Models / I.P. Malashin, I.S. Masich, V.S. Tynchenko, A.P. Gantimurov, V.A. Nelyub, A.S. Borodulin // 2024 IEEE International Conference on Data Mining Workshops (ICDMW). – IEEE, 2024. – С. 278-285. - DOI: 10.1109/ICDMW65004.2024.00042

Публикации в других изданиях:

11. Малашин И.П. Кластерно-ориентированный подход к оптимизации проектных решений в системах автоматизированного проектирования / И.П. Малашин, В. С. Тынченко // Третья молодежная конференция с международным участием «Комплексная автоматизация проектирования и производства 2025» (КАПП(М)-2025) : сборник статей, Москва, 16 мая 2025. С. 359-361.

12. Малашин И.П. Моделирование динамического изменения пространства поиска в эволюционных алгоритмах с использованием кластерного анализа / И.П. Малашин, В.С. Тынченко // XII Международная научная Четаевская конференция «Аналитическая механика, устойчивость и управление» : сборник статей, Казань, 2025. С. 132-136

13. Малашин И.П. Адаптивное исключение неэффективных кластеров в эволюционной оптимизации / И.П. Малашин, В.С. Тынченко, С.О. Курашкин, Д.А. Мартысюк // XXX Байкальская Всероссийская с международным участием конференция «Информационные и математические технологии в науке и управлении» : сборник статей, Иркутск, 2025. С. 54-55.

14. Малашин И.П. Модификация самонастраивающегося генетического алгоритма на основе динамической коррекции области поиска / И.П. Малашин, Е.А. Сопов // Тридцать первая международная открытая научная конференция «Современные проблемы информатизации» (Modern Informatization Problems), 1 ноября 2025 г. – 31 января 2026 г. — Yelm, WA, USA : Science Book Publishing House, 2026.

Свидетельства о государственной регистрации программы для ЭВМ:

15. Малашин И.П. Модуль кластеризации налогоплательщиков по данным контрольнокассовой техники для выявления аномального поведения / Авторы: С. И. Токарев, Д. И. Васильев, А. С. Бородулин, В. А. Нелюб, И. Н. Елисеева и др. // М.: Роспатент. – 2024. Свидетельство №2024691599 от 23.12.2024.

16. Малашин И.П. Программный модуль динамической коррекции области поиска (ДКОП) в генетическом алгоритме для решения многомерных задач оптимизации / Авторы: И.П. Малашин, В.А. Нелюб, В.С. Тынченко, А.С. Бородулин // М.: Роспатент. – 2025. Свидетельство №2025696199 от 16.12.2025.