

На правах рукописи

**ХРИТОНЕНКО ДМИТРИЙ ИВАНОВИЧ**

**АДАПТИВНЫЕ КОЛЛЕКТИВНЫЕ НЕЙРО-ЭВОЛЮЦИОННЫЕ  
АЛГОРИТМЫ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ**

05.13.01 – Системный анализ, управление и обработка информации  
(космические и информационные технологии)

**АВТОРЕФЕРАТ**

диссертации на соискание ученой степени  
кандидата технических наук

Красноярск 2017

Работа выполнена в ФГБОУ ВО «Сибирский государственный университет науки и технологий имени академика М.Ф. Решетнева», г. Красноярск

Научный руководитель: **Семенкин Евгений Станиславович**  
доктор технических наук, профессор

Официальные оппоненты: **Кравец Олег Яковлевич**  
доктор технических наук, профессор  
ФГБОУ ВО «Воронежский государственный технический университет»  
профессор кафедры автоматизированных и вычислительных систем

**Демидова Лилия Анатольевна**  
доктор технических наук, профессор  
ФГБОУ ВО «Рязанский государственный радиотехнический университет»  
профессор кафедры вычислительной и прикладной математики

Ведущая организация: **Федеральный исследовательский центр «Информатика и управление» Российской академии наук, г. Москва**

Защита состоится «22» декабря 2017 г. в 15.00 на заседании диссертационного совета Д 212.249.05, созданного на базе ФГБОУ ВО «Сибирский государственный университет науки и технологий имени академика М.Ф. Решетнева» по адресу 660037 г. Красноярск, проспект имени газеты «Красноярский рабочий», 31.

С диссертацией можно ознакомиться в библиотеке Сибирского государственного университета науки и технологий имени академика М.Ф. Решетнева.

Автореферат разослан «\_\_\_» \_\_\_\_\_ 2017 г.

Ученый секретарь  
диссертационного совета

Илья Александрович Панфилов

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

**Актуальность.** Век информационных технологий и большие объемы накопленной информации, требующие обработки, являются одной из причин возникновения такого направления как интеллектуальный анализ данных. Основная цель данного направления – поиск ранее неизвестных, нетривиальных, практически полезных и легко интерпретируемых закономерностей. Сегодня методы интеллектуального анализа данных развиваются достаточно динамично и используются во многих областях. Прогнозирование характеристик выпускаемого изделия, распознавание изображений, медицинская диагностика, банковский скоринг – лишь некоторые из задач, решаемых методами интеллектуального анализа данных.

Искусственные нейронные сети (ИНС) – один из распространенных сегодня методов интеллектуального анализа данных. Достоинствами метода является широкая область применения и точность получаемых моделей. К недостаткам ИНС стоит отнести высокую вычислительную сложность их формирования и обучения, а также их модель «черного ящика», которая делает интерпретацию результатов затруднительной. Формирование и обучение ИНС может быть сведено к оптимизационной задаче с переменными целочисленного, бинарного, вещественного и категориального типов. Целевая функция в такой задаче задается алгоритмически и, как правило, имеет большую размерность и пространство поиска, что говорит о высокой вычислительной сложности оптимизационной процедуры. Эволюционные алгоритмы (ЭА) – один из эффективных методов решения таких задач. Использование ЭА делает актуальным вопрос о выборе их настроек и параметров. Решение данного вопроса вызывает трудности даже у опытных исследователей и значительно отражается на качестве получаемой модели. Разработка методов, позволяющих в автоматическом режиме выбирать конфигурацию и настраивать параметры алгоритма ставит перед собой несколько целей: повысить эффективность получаемого решения, снизить вычислительные затраты, минимизировать требования к квалификации пользователя и тем самым расширить область применимости. Сегодня уже разработано множество процедур настройки параметров эволюционных алгоритмов, однако их совместное применение при решении сложных задач оптимизации является не всегда успешным и недостаточно исследованным. Кроме того, из-за стремительного развития направления *Big Data* становится необходимым использование алгоритмов снижения размерности и объема данных для применения технологий анализа данных, основанных на ИНС. Поиск, разработка, реализация и практическое применение таких методов является предметом научного исследования.

Учитывая сказанное, можно утверждать, что разработка и исследование методов автоматизированного формирования искусственных нейронных сетей и их коллективов при помощи адаптивных эволюционных алгоритмов с применением методов отбора информативных признаков и обучающих примеров является **актуальной научно-технической задачей.**

**Целью** диссертационной работы является повышение качества нейросетевых моделей интеллектуального анализа данных и снижение вычислительных ресурсов, требуемых для их формирования, за счет использования адаптивных эволюционных алгоритмов оптимизации.

Для достижения поставленной цели необходимо решить следующие задачи:

1. Выполнить обзор и исследовать эффективность методов самонастройки и самоконфигурирования эволюционных алгоритмов оптимизации на репрезентативном множестве тестовых задач.

2. Разработать адаптивный эволюционный алгоритм оптимизации, объединяющий в себе достоинства известных методов самонастройки и самоконфигурирования.

3. Выполнить обзор существующих методов формирования искусственных нейронных сетей с целью выявления наиболее эффективного из них.

4. Разработать эффективные адаптивные алгоритмы формирования искусственных нейронных сетей для решения задач интеллектуального анализа данных.

5. Выполнить обзор существующих методов и разработать эффективный алгоритм автоматического формирования коллективов нейронных сетей.

6. Реализовать разработанные подходы в виде программных систем и протестировать их эффективность на репрезентативном множестве тестовых и реальных задач.

**Методы исследования.** В процессе выполнения диссертационной работы использовались методы статистической обработки информации, теории вероятностей, эволюционных вычислений, оптимизации, нейросетевого моделирования, системного анализа, коллективного принятия решений, выявления закономерностей в массивах данных.

**Научная новизна** работы включает следующие пункты:

1. Разработан новый метод адаптивного управления уровнем мутации в эволюционных алгоритмах, отличающийся от известных способом расчета вероятности мутации гена.

2. Разработан новый метод адаптивного управления размером популяции в эволюционных алгоритмах, отличающийся от известных применением динамического показателя успешности подпопуляции.

3. Разработаны и реализованы эволюционные алгоритмы моделирования и оптимизации, отличающиеся от известных использованием эффективной комбинации модификаций самоконфигурирования, адаптации и селекции обучающих примеров.

4. Разработаны новые адаптивные эволюционные алгоритмы формирования искусственных нейронных сетей, отличающиеся от известных методом настройки конфигурации и параметров, а также применением процедур отбора информативных признаков и обучающих примеров.

5. Разработан новый метод построения коллективов искусственных нейронных сетей, отличающийся от известных применением адаптивного алгоритма генетического программирования с механизмом контроля разнообразия внутри коллектива и возможностью использования альтернативных методов анализа данных.

**Теоретическая значимость** результатов диссертационной работы состоит в разработке новых эволюционных алгоритмов моделирования и оптимизации, а также формирования коллективов искусственных нейронных сетей, позволяющих получать эффективные в смысле определенного исследователем критерия модели при помощи алгоритмов адаптации, что представляет собой вклад в теорию и практику интеллектуального анализа данных нейро-эволюционными алгоритмами.

**Практическая ценность.** Разработанные алгоритмы реализованы в виде программных систем, зарегистрированных в Роспатенте, и позволяют эффективно решать задачи классификации, регрессии и оптимизации. Программные системы в автоматическом режиме формируют эффективные структуры искусственных нейронных сетей и их коллективов. Они успешно применены при решении задач банковского скоринга, распознавания эмоций, медицинской диагностики, распознавания изображений и других задач анализа данных.

При помощи реализованных в виде программных систем алгоритмов были решены задачи реальные практические задачи распознавания эмоций по речи и прогнозирования уровня заболеваемости населения. Сравнение с аналогами показало высокую эффективность предлагаемых подходов.

**Реализация результатов работы.** Разработанные в ходе выполнения диссертационного исследования алгоритмы успешно применены при выполнении работ в рамках проектного задания «Разработка теоретических основ автоматизации комплексного моделирования сложных систем методами вычислительного интеллекта» (2.1680.2017/ПЧ), гранта РФФИ № 14-06-00256 «Информационные технологии оценки и прогнозирования экологических рисков», а также в российско-германских проектах (совместно с университетом г. Ульм) «Распределенные интеллектуальные информационные системы обработки и анализа мультилингвистической информации в диалоговых информационно-коммуникационных системах» (ФЦП ИР, ГК №11.519.11.4002) и «Математическое и алгоритмическое обеспечение автоматизированного проектирования аппаратно-программных комплексов интеллектуальной обработки мультилингвистической информации в распределенных высокопроизводительных системах космического назначения» (ФЦП НПК, ГК № 16.740.11.0742). Кроме того, отдельные решения использовались при выполнении проекта № 140/14 «Разработка теоретических основ эволюционного проектирования интеллектуальных информационных технологий анализа данных» тематического плана ЕЗН СибГУ, а также гранта РФФИ № 16-41-243064 «Разработка алгоритмов проектирования кооперативных эволюционно-бионических технологий интеллектуального анализа данных с использованием систем на нечеткой

логике». Диссертационная работа была поддержана Фондом содействия развитию малых форм предприятий в научно-технической сфере в рамках программы «У.М.Н.И.К» по проекту «Разработка нейро-эволюционных алгоритмов коллективного типа для решения задач интеллектуального анализа данных» в 2014-2016 гг., а также Красноярским Краевым фондом науки в рамках проекта «Распределенные самоконфигурируемые эволюционные алгоритмы автоматического формирования искусственных нейронных сетей».

Разработанные в ходе выполнения диссертации 7 программных систем зарегистрированы в Роспатенте. Две из них переданы в инновационные IT-компании г. Красноярска.

#### **Основные защищаемые положения:**

1. Разработанные методы адаптивного изменения размера популяции и величины мутации в эволюционных алгоритмах позволяют более эффективно использовать вычислительные ресурсы при решении задач моделирования и оптимизации этими алгоритмами.

2. Разработанные адаптивные эволюционные алгоритмы моделирования и оптимизации позволяют исследователю отказаться от настройки их основных параметров, что повышает эффективность применения методов, снижает вычислительные затраты и уменьшает влияние квалификации исследователя на ход эволюционного процесса.

3. Разработанные нейро-эволюционные алгоритмы решения задач восстановления регрессии, прогнозирования и классификации не уступают по эффективности известным аналогам.

4. Автоматический отбор информативных признаков внутри эволюционных алгоритмов и селекция обучающих примеров динамическим перераспределением вероятности выбора измерения в обучающую выборку позволяют повысить точность получаемых моделей, снизить их сложность и количество вычислительных ресурсов, затрачиваемых на их формирование.

**Публикации.** По теме работы опубликовано более 25 печатных работ, в том числе 3 - в журналах из Перечня ВАК и 2 - в изданиях, индексируемых в международных базах цитирования Web of Science и/или Scopus. Получено 7 свидетельств о регистрации программных систем в Роспатенте.

**Апробация работы.** Результаты диссертационной работы докладывались на 11 Всероссийских и Международных научных конференциях, среди которых: Пятая Международная конференция «Системный анализ и информационные технологии» САИТ-2013 (Красноярск, 2013); 2<sup>nd</sup>, 4<sup>th</sup> and 5<sup>th</sup> International Workshops on Mathematical Models and their Applications (Красноярск, 2013, 2015, 2016), Четырнадцатая Национальная конференция по искусственному интеллекту с международным участием (КИИ, Казань, 2014); Всероссийская научно-практическая конференция «Информационно-телекоммуникационные системы и технологии» (ИТСиТ, Кемерово, 2014, 2015); Международная научно-практическая конференция «Решетневские чтения» (Красноярск, 2014, 2015); International Conference on Environmental Engineering and Computer Application (ICEECA, Hong Kong,

China, 2014), Восьмой Всероссийский форум студентов, аспирантов и молодых ученых (Санкт-Петербург, 2014).

**Структура работы.** Диссертация состоит из введения, четырех глав, заключения, списка литературы и приложений.

## ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **введении** обоснована актуальность диссертационной работы, поставлены цель и задачи исследования, описаны научная новизна, теоретическая и практическая значимости, а также основные положения, выносимые на защиту.

В **первой главе** демонстрируется и обсуждается проблема выбора эффективной конфигурации и параметров эволюционных алгоритмов, для чего описываются стандартные эволюционные алгоритмы: генетический алгоритм однокритериальной безусловной оптимизации и алгоритм генетического программирования для решения задач символьной регрессии. А затем приводятся результаты вычислительных экспериментов по моделированию и оптимизации на репрезентативном множестве задач конкурса Международной конференции по эволюционным вычислениям (Conference on Evolutionary Computation, *CEC'2017*), которые показывают существенную зависимость эффективности эволюционных алгоритмов от комбинации их настроек и параметров, размера популяции и уровня мутации. Такая зависимость объясняется особенностями решаемых задач и свойствами самих генетических операторов, соотношение между которыми остается неопределенным.

Обосновывается, что для достижения большей эффективности эволюционных алгоритмов при проектировании структуры нейронных сетей и настройке их весовых коэффициентов важно учитывать особенности возникающих задач оптимизации, разрабатывать и реализовывать специальные процедуры, позволяющие в автоматическом режиме подстраивать структуру и параметры алгоритма под задачу в ходе ее решения.

В конце главы приводится *NFL*-теорема и обсуждаются следствия из нее.

Во **второй главе** рассматриваются существующие модификации эволюционных алгоритмов, позволяющие решать проблемы, описанные в главе 1, а затем предлагаются новые алгоритмы адаптации величины мутации и размера популяции. Отличительными особенностями предлагаемых подходов адаптации являются:

1. Использование медианы нормированной пригодности для адаптации величины мутации. Подход инвариантен к способу представления решений внутри алгоритма. Выбор медианы обусловлен наибольшим приростом эффективности.

2. Определение текущего размера популяции на основании истории успешности индивидов, порождаемых подпопуляциями алгоритма.

В главе описываются также полезные модификации эволюционных алгоритмов, используемые в ходе работы. Среди них алгоритм самоконфигурирования, а также алгоритм селекции обучающих примеров.

Опишем предлагаемые в главе 2 подходы. Вероятность мутации индивидов рассчитывается следующим образом:  $P_j = P_j^0 + \alpha \cdot \frac{Med(fitness_i)}{n}$ ,

где  $P_j$  – вероятность мутации для  $j$ -го индивида;  $P_j^0$  – вероятность мутации, которая была ранее определена индивиду при помощи алгоритма самоконфигурирования;  $\alpha$  – параметр, задаваемый пользователем;  $n$  – число генов в индивиде;  $fitness_i$  – нормированный массив пригодностей на  $i$ -м поколении;  $Med(f)$  – функция определения значения медианы массива  $f$ .

Работу процедуры определения эффективного размера популяции можно описать при помощи следующих этапов.

1. В начале работы эволюционного алгоритма устанавливаются параметры:  $MaxPopSize$  – максимально возможный размер популяции,  $k$  – число подпопуляций в алгоритме,  $tOpt$  – период работы алгоритма без изменения размера популяции,  $m$  и  $d$  – параметры, отвечающие за изменение размера популяции (коэффициенты сдвига и масштаба в равномерном распределении). Вводятся вектора  $S$  и  $P$ . Компоненты вектора  $S$  приравниваются к нулю, а  $P$  задаются случайно и равномерно на интервале  $[0.05, 1]$ . Параметр времени  $t=0$ .

2. Проверяем критерий останова. Если он достигнут, алгоритм завершает работу;

3. Положим  $t = t + 1$ . Если  $t$  не кратно  $tOpt$ , переходим к шагу 5;

4. Находим  $best = \arg \max_i (S_i)$ . Случайным образом выбираем  $j$ -ю

компоненту вектора  $P$ . Вычисляем  $P_j = Rav(m, d) + P_{best}$ . Здесь  $Rav(m, d)$  – функция равномерного распределения с коэффициентом сдвига  $m$  и масштаба  $d$ .  $P_i \in [0.05, 1] \forall i = \overline{1, k}$ . Компоненты вектора  $P$ , отличные от  $j$ , не претерпевают никаких изменений. Далее  $S_i = 0 \forall i = \overline{1, k}$  и переходим к шагу 5.

5. Определяем количество потомков, порождаемых в поколении  $t$ :  $subPopSize = MaxPopSize \cdot \max_i (P_i)$ . Для каждого потомка случайно выбирается

компонента вектора  $P_j$ . Из всех индивидов случайным образом формируется подпопуляция размера  $MaxPopSize \cdot P_j$ . Далее к новой подпопуляции последовательно применяются стандартные операторы селекции, скрещивания и мутации с учетом алгоритма самоконфигурации. Если полученный потомок по пригодности превосходит родителей, увеличивается показатель «успешности» подпопуляции:  $S_j = S_j + 1$ . Как только было порождено  $subPopSize$  потомков, переходим к шагу 2.

Разработанные модификации позволили генетическому алгоритму занять 5-е место среди 8-ми представленных алгоритмов оптимизации

(победителей конкурса *SEC'2017*), что можно считать успехом, т.к. на сложных функциях вещественных переменных, представленных на конкурсе, бинарный генетический алгоритм обычно не составляет серьезной конкуренции специализированным алгоритмам, ориентированным на вещественные переменные.

В алгоритме генетического программирования предлагаемые модификации позволили существенно снизить ошибку моделирования. Их совместное применение с методом селекции обучающих примеров, оператором равномерного скрещивания и самоконфигурированием позволило наряду с уменьшением ошибки аппроксимации снизить сложность получаемых моделей и время работы алгоритма. Показана эффективность в сравнении с известными аналогами, реализованными в среде анализа данных *RapidMiner*. Алгоритм ГП проиграл по точности двум методам (*Neural Net* и *SVM*). Однако преимуществом алгоритма является приемлемый уровень точности с возможностью получения легко интерпретируемых моделей в символьной форме.

Таким образом, во второй главе были разработаны эффективные модификации стандартных и самоконфигурируемых эволюционных алгоритмов, позволяющие в автоматическом режиме определять уровень мутации и размер популяции и, как следствие, увеличивать точность моделирования и оптимизации.

В **третьей** главе разрабатывается и исследуется адаптивный эволюционный алгоритм для автоматического проектирования искусственных нейронных сетей. Для этого на основе анализа существующих алгоритмов обучения и проектирования искусственных нейронных сетей предлагается применить прямые методы кодирования ИНС, подходящие для использования в эволюционных алгоритмах.

Пример кодирования в алгоритме ГП представлен на рисунке 1.

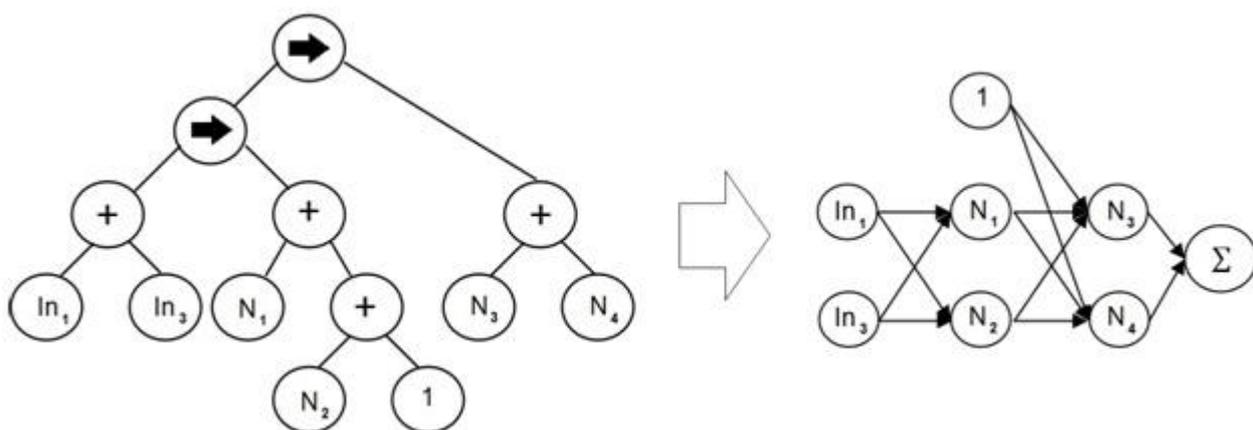


Рисунок 1 - Кодирование ИНС алгоритмом ГП

В алгоритме функциональное множество включает два вида операторов: объединения в слой (на рисунке 1 представлен знаком плюс) и упорядочивания слоев (представлен стрелкой). Левая ветвь оператора упорядочивания имеет порядковый номер слоя  $i-1$ , в то время как правая  $i$ .

Выходы всех терминальных элементов слоя  $i-1$  являются входами в нейроны слоя  $i$ . Терминальное множество ГП содержит нейроны с различными типами активационных функций, константы, атрибуты задачи. Здесь  $N_l$  – нейрон с  $l$ -ой активационной функцией,  $In_i$  – атрибут задачи под номером  $i$ . Для корректного функционирования алгоритма вводится функция адаптации, которая проверяет генерируемые программной системой решения на выполнение условия замкнутости. В случае его нарушения она производит замену функциональных операторов.

Генетический алгоритм кодирует ИНС при помощи матрицы смежности  $M_{n \times n}$  нейронов. В работе используется треугольная матрица, т.к. рекуррентные нейронные сети не рассматриваются. При необходимости добавляются две матрицы, отвечающие за отбор информативных признаков  $I_{k \times n}$  и кодирования типа активационных функций  $A_{3 \times n}$ .  $k$  - количество атрибутов задачи,  $n$  - максимальное число нейронов сети. Тип каждой активационной функции нейрона кодируется при помощи 3 бит информации. Далее из трех представленных матриц формируется бинарная строка, позволяющая генетическому алгоритму производить процедуру оптимизации структуры.

В главе описываются меры качества, используемые в задачах классификации, восстановления регрессии и прогнозирования, а именно точность, полнота,  $F$ -мера, среднеквадратичная и среднеабсолютная ошибки аппроксимации. Приводится описание тестовых и реальных задач с репозиториях машинного обучения *KEEL* и *UCI*, на которых проверялась эффективность разрабатываемого подхода (таблица 1).

Таблица 1. Тестовые задачи классификации

Задача	Число измерений	Количество атрибутов	Число классов
Классификация			
<i>Magic</i>	19020	10	2
<i>Page-Blocks</i>	5472	10	5
<i>Australian credit</i>	690	14	2
<i>German credit</i>	1000	20	2
<i>Texture</i>	5500	40	11
<i>Twonorm</i>	7400	20	2
<i>Segment</i>	2310	19	7
<i>Ring</i>	7400	20	2
<i>Penbased</i>	10992	16	10
<i>Satimage</i>	6435	36	6
Регрессия			
<i>Forest Fires</i>	517	12	-
<i>Weather</i>	1609	9	-
<i>Concrete</i>	1030	8	-
<i>Pole Telecom.</i>	14998	26	-
<i>House</i>	22784	16	-

Отметим, что для формирования ИНС используются оба из разработанных эволюционных алгоритмов, т.е. ГП и ГА. Их эффективность по критерию Вилкоксона была признана равной, поэтому в таблице 2 представлен результат усреднения по 100 запускам каждого из алгоритмов. Поведение алгоритмов ГА и ГП при проектировании структуры ИНС сильно отличается. На ранних этапах работы алгоритм ГА значительно проигрывает алгоритму ГП, т.к. кодирование ИНС в ГА создает матрицы с сильной разреженностью, что требует выделения значительных вычислительных ресурсов. Несмотря на это, использование ГА является перспективным, т.к. при формировании коллективов требуется поддержание разнообразия ИНС. Дополнительным преимуществом ГА является возможность модификации для проектирования рекуррентных нейронных сетей. В таблицах ниже представлена ошибка классификации, выраженная в процентах при сравнении разработанных подходов с известными аналогами.

Таблица 2. Сравнение эффективности на задачах банковского скоринга (ошибка в %)

Название алгоритма	<i>Australian credit</i>	<i>German credit</i>	Название алгоритма	<i>Australian credit</i>	<i>German credit</i>
<i>SC_GP</i>	9.78	20.50	<i>Bayesian</i>	15.3	32.10
<i>MGP</i>	10.15	21.25	<i>Boosting</i>	24.0	30.00
<i>2SGP</i>	<b>9.73</b>	<b>19.85</b>	<i>Bagging</i>	15.3	31.60
<i>GP</i>	11.11	21.66	<i>RSM</i>	14.8	32.30
<i>Fuzzy classifier</i>	10.90	20.60	<i>CCEL</i>	13.4	25.40
<i>C4.5</i>	10.14	22.27	<i>CART</i>	12.56	24.35
<i>LR</i>	13.04	21.63	<i>MLP</i>	10.14	23.82
<i>k-NN</i>	28.50	28.49	<i>ANNEA</i>	9.97	21.70

Разработанный алгоритм (*ANNEA*) превосходит по точности стандартные методы классификации, однако уступает специализированным (*2SGP*, *SC\_GP*).

Реализация метода селекции обучающих примеров (*ANNEA+IS*), позволила существенно сократить вычислительную сложность алгоритма по сравнению с базовой версией (*ANNEA*) и снизить ошибку классификации для большинства из задач (таблица 3).

Таблица 3. Эффективность алгоритма с селекцией обучающих примеров

Задача	<i>ANNEA</i>			<i>ANNEA+IS</i>		
	Ошибка на обуч., %	Ошибка на тест., %	Время, мин.	Ошибка на обуч., %	Ошибка на тест., %	Время, мин.
<i>Magic</i>	14.83	15.08	734	13.75	<b>14.72</b>	229
<i>Page-Blocks</i>	3.72	<b>4.23</b>	180	3.96	<b>4.23</b>	51
<i>Australian credit</i>	8.71	10.99	67	10.07	<b>9.97</b>	19
<i>German credit</i>	19.56	23.76	129	18.91	<b>21.70</b>	47

<i>Texture</i>	6.93	8.05	805	5.90	<b>7.27</b>	146
<i>Twonorm</i>	5.21	5.97	481	3.67	<b>4.24</b>	131
<i>Segment</i>	5.89	6.14	312	4.69	<b>5.10</b>	81
<i>Ring</i>	6.37	6.54	454	4.78	<b>4.99</b>	97
<i>Penbased</i>	5.36	7.91	470	3.30	<b>3.55</b>	137
<i>Satimage</i>	12.90	14.09	691	12.45	<b>13.82</b>	115

В конце главы делается вывод, что адаптивный эволюционный алгоритм автоматического генерирования искусственных нейронных сетей при сравнении со своей базовой версией и аналогами, известными из научной литературы, позволяет получать эффективные модели, используемые для решения задач классификации и регрессии. Реализованные и предлагаемые модификации позволяют существенно сэкономить вычислительные ресурсы, повысить точность моделирования, отказаться от ручной настройки основных параметров.

В четвертой главе сначала рассматриваются известные методы формирования коллективов. Затем предлагается подход автоматического формирования коллективов нейросетевых технологий анализа данных, на основе разработанного адаптивного алгоритма генетического программирования. Для его функционирования предлагается использовать регрессионную модель, в которой терминальное множество состоит из настраиваемых констант и сгенерированных нейросетевых (либо внешних) моделей, а функциональное – различные математические операции. С целью поддержания разнообразия коллектива и удаления одинаковых ИНС предлагается следующий алгоритм отбора.

1. Сгенерировать  $K$  нейросетей любым из представленных методов.
2. Каждой из  $K$  нейросетей ставится в соответствие вектор  $\bar{X}^k = \{x_1^k, x_2^k, \dots, x_m^k\}$ ,  $k = 1, \dots, K$ , показывающий ошибку на каждом из измерений выборки данных;  $m$  – число измерений в выборке. Так, если имеет место задача классификации, то единица на  $i$ -й позиции данного вектора может говорить об ошибочно принятом решении текущей ИНС. В случае решения задачи регрессии в таких позициях может стоять, например, разница между истинным и полученным при помощи ИНС значением.
3. Выбирается ИНС с минимальной суммой по вектору  $\bar{X}$  (в случае регрессии – сумма по модулю);
4. Далее, руководствуясь заранее определенной метрикой  $d(\bar{X}^i, \bar{X}^j)$  и максиминным критерием  $\max_i (\min_j (d(\bar{X}^i, \bar{X}^j)))$ , выбираются остальные ИНС.

Здесь  $i \in ENS$ ,  $j \in PL$ .  $ENS$  – множество уже отобранных в коллектив ИНС,  $PL$  – множество претендентов в коллектив. Примером выбираемых метрик могут служить метрика Хемминга для задачи классификации, а также среднеквадратичный критерий для решения задач восстановления регрессии;

5. Пункт 4 выполняется до тех пор, пока не будет отобрано необходимое количество ИНС, либо метрика  $d(\bar{X}^i, \bar{X}^j)$  окажется ниже некоторого порога для всех  $i$  и  $j$ .

Далее в главе 4 приводится описание реализованной программной системы (*ENS+ANN*). На описанных в главе 3 задачах проводится тестирование и статистический анализ полученных результатов. Сравнение с аналогами и базовой версией (*ANNEA+IS*) показывает конкурентоспособность разработанного подхода (таблица 4, 5).

Таблица 4. Сравнение эффективности на задачах классификации (ошибка в %)

Задача	<i>ANNEA+IS</i>	<i>ENS+ANN</i>	<i>HEFCA_IS</i>	<i>GP-Coach</i>	<i>IVFS-Amp</i>	<i>FARC-HD</i>
<i>Magic</i>	14.72	<b>14.3</b>	15.08	20.18	20.82	15.49
<i>Page-Blocks</i>	4.23	4.07	<b>3.25</b>	8.77	5.84	4.99
<i>Texture</i>	7.27	7.29	<b>4.45</b>	-	-	7.11
<i>Twonorm</i>	4.24	<b>4.21</b>	4.81	15.17	-	4.72
<i>Segment</i>	<b>5.10</b>	5.13	5.19	24.04	-	-
<i>Ring</i>	4.99	<b>4.73</b>	<b>5.08</b>	-	16.89	5.92
<i>Penbased</i>	3.55	<b>3.19</b>	3.81	17.80	21.73	3.96
<i>Satimage</i>	13.82	13.46	12.93	27.50	-	<b>12.68</b>

В задачах регрессии используется критерий  $MSE = \frac{1}{n} \cdot \sum_{i=1}^n (y_i - \bar{y}_i)^2$ .

Таблица 5. Сравнение эффективности на задачах регрессии (величина *MSE*)

Задача	<i>ANNEA+IS</i>	<i>ENS+ANN</i>	<i>TSK-IRL</i>	<i>Linear-LMS</i>	<i>ANFIS-SUB</i>	<i>LEL-TSK</i>	<i>METSK-HD</i>
<i>Forest Fires</i>	317	272	-	2013	<b>204.8</b>	1407	5587
<i>Weather</i>	1.06	0.83	-	1.2	<b>0.8</b>	1.6	1.2
<i>Concrete</i>	26.9	21.7	<b>19.1</b>	54.7	188.2	31.4	23.8
<i>Pole Telecom.</i>	67.4	<b>58.1</b>	-	465	131.7	-	61
<i>House</i>	7.9	8.2	-	10.4	<b>7.6</b>	-	8.6

Результаты численных экспериментов на задаче классификации эмоций по звуковому сигналу представлены в таблице 6.

Таблица 6. Решение задачи классификации эмоций (доля правильных ответов)

Алгоритм	Точность	Алгоритм	Точность
<i>ENS+ANN</i>	<b>0.725</b>	<i>W-LinearRegression</i>	0.632
<i>Naive_Bayes</i>	0.608	<i>Linear Regression</i>	0.640
<i>Fast_Large_Margin</i>	0.610	<i>W-PLSClassifier</i>	0.650
<i>Linear_Discriminant_Analysis</i>	0.620	<i>Logistic_Regression</i>	0.670
<i>W-MultilayerPerceptron</i>	0.660	<i>W-LMT</i>	0.670
<i>W-FT</i>	0.670	<i>W-LogisticBase</i>	0.670
<i>Neural Net</i>	0.701	<i>W-GaussianProcesses</i>	0.650

Стоит отметить, что именно коллективный подход позволил повысить точность классификации.

Была также решена реальная практическая задача прогнозирования уровня заболеваемости и смертности населения в зависимости от химического состава воздуха в городе. В связи с малым объемом выборки только для 6 из 10 прогнозируемых величин были построены модели с удовлетворительной точностью прогноза. Применение коллективного подхода позволило повысить точность моделирования в среднем на 10% по сравнению с базовым эволюционным алгоритмом проектирования искусственных нейронных сетей. Общий показатель смертности, смертность детей до одного года, число больных злокачественными новообразованиями и смертность от них - характеристики, для которых не удалось построить адекватную модель прогноза. Хотя полученные результаты удовлетворили заказчика, очевидно, что полное решение задачи требует большего объема информации, т.к. предложенный алгоритм не предназначен для функционирования в условиях малого объема данных.

Таким образом, решение различных задач при помощи разработанного адаптивного коллективного нейро-эволюционного алгоритма позволило в большинстве случаев повысить эффективность решения. Алгоритм селекции обучающих примеров снизил вычислительную сложность задачи и частично нивелировал эффект переобучения ИНС. Адаптация позволила отказаться от ручной настройки основных параметров и снизила требования к квалификации пользователя в области эволюционного моделирования, что способствует привлечению новых исследователей в область интеллектуального анализа данных.

**Заключение** диссертации содержит основные результаты работы и выводы.

## **ОСНОВНЫЕ РЕЗУЛЬТАТЫ И ВЫВОДЫ**

1. Обоснована необходимость разработки методов адаптации размера популяции и величины мутации в самоконфигурируемых эволюционных алгоритмах.

2. Разработана новая схема вычисления вероятности мутации в эволюционных алгоритмах, основанная на использовании медианы нормированной пригодности популяции.

3. Разработана новая схема адаптации размера популяции эволюционного алгоритма, основанная на использовании истории успешности подпопуляций с фиксированным размером.

4. Проведено комплексное исследование эффективности разработанных подходов на задачах восстановления регрессии и оптимизации.

5. Разработаны адаптивные эволюционные алгоритмы автоматического формирования искусственных нейронных сетей.

6. Разработанный адаптивный алгоритм генетического программирования применен для автоматического формирования коллективов нейронных сетей, что позволяет повысить эффективность их применения.

7. Обоснована целесообразность использования метода селекции обучающих примеров при формировании искусственных нейронных сетей и их коллективов. Селекция обучающих примеров позволила снизить вычислительную сложность алгоритма без снижения точности моделирования.

8. Разработанные в ходе исследования программные системы, реализующие предложенные подходы, успешно применены для решения реальных практических задач.

Таким образом, в диссертационной работе разработаны, исследованы и апробированы новые адаптивные эволюционные алгоритмы автоматического формирования искусственных нейронных сетей и их ансамблей, используемых для решения задач восстановления регрессии, прогнозирования и классификации, что является существенным вкладом в теорию и практику обработки информации и интеллектуального анализа данных.

## **ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ**

### **Статьи в ведущих рецензируемых научных журналах и изданиях**

1. Хритonenко Д.И., Семенкин Е.С. Адаптивная мутация в самоконфигурируемых эволюционных алгоритмах // Системы управления и информационные технологии, №3(69), 2017. – С. 37-42

2. Khritonenko D., Semenkin E., Sugak E., Potilicina E. Solving the problem of city ecology forecasting with neuro-evolutionary algorithms // Вестник СибГАУ. 2015. Т. 16, № 1. С. 137–143.

3. Хритonenко Д.И., Семенкин Е.С. Distributed self-configuring evolutionary algorithms for artificial neural networks design // Вестник СибГАУ. 2013. № 4 (50), С. 112-116.

### **Публикации в изданиях, индексируемых в международных базах**

4. Khritonenko D., Stanovov V., Semenkin E. Applying an instance selection method to an evolutionary neural classifier design // IOP Conference Series: Materials Science and Engineering. V International Workshop on Mathematical Models and their Applications. – 2017. - Vol. 173. (**Web of Science, Scopus**).

5. Khritonenko D., Semenkin E. Application of artificial neural network ensembles for city ecology forecasting using air chemical composition information // Proceedings of the 2014 International Conference on Environmental Engineering and Computer Application (ICEECA 2014, Hong Kong). (**Scopus**).

### **Публикации в сборниках трудов конференций**

6. Становов В. В., Хритonenко Д. И., Шкраба А. Метод селекции обучающих примеров в нейросетевых классификаторах // Материалы XIX

междунар. науч.-практ. конф., «Решетневские чтения». СибГАУ, 2015. – Красноярск. Т. 2. С. 100-102.

7. Хритonenко Д. И. Самоконфигурируемый эволюционный алгоритм автоматического проектирования рекуррентных нейронных сетей // Материалы Всероссийской научно-практической конференции «Информационно-телекоммуникационные системы и технологии» (ИТСиТ-2015), 2015. – Кемерово. С. 488-489.

8. Хритonenко Д. И., Семенкин Е. С., Сугак Е. В., Потылицина Е. Н. Автоматическое генерирование нейросетевых моделей в задаче прогнозирования уровня заболеваемости населения // Труды четырнадцатой национальной конференции по искусственному интеллекту с международным участием (КИИ-2014), 2014. – Казань. Т. 1, С. 276-285.

9. Хритonenко Д. И. Автоматическое проектирование коллективов искусственных нейронных сетей в задачах анализа данных // Материалы Восьмого Всероссийского форума студентов, аспирантов и молодых ученых «Наука и инновации в технических университетах», 2014. – Санкт-Петербург. С. 60-61.

10. Хритonenко Д. И., Становов В. В. Comparing intelligence information technologies efficiency for solving classification problem // Материалы XIII междунар. науч. конф. бакалавров, магистрантов и аспирантов «Молодежь. Общество. Современная наука, техника и инновации», 2014. – Красноярск, СибГАУ. С. 217–219.

11. Хритonenко Д. И. Программная система автоматического генерирования нейросетевых классификаторов эволюционными алгоритмами // Материалы XVIII Международ. науч. конф. «Решетневские чтения», 2014. – Красноярск, СибГАУ. Ч. 2. С. 290-291

12. Хритonenко Д. И., Семенкин Е.С. Применение распределенных самоконфигурируемых эволюционных алгоритмов в задачах нейросетевого моделирования // Труды V Международной конференции «Системный анализ и интеллектуальные технологии» (САИТ-2013), 2013. – Красноярск. Т. 2, С. 108-114.

13. Хритonenко Д. И. Самоконфигурируемый распределенный генетический алгоритм //: Труды XIII Междунар. науч. Конференции «Интеллект и наука», 2013. – Железногорск. – С. 134-135.

14. Хритonenко Д. И. Применение самонастраивающегося алгоритма генетического программирования в задачах нейросетевого моделирования // Сборник материалов победителей и лауреатов Всероссийского конкурса научно-исследовательских работ студентов и аспирантов в области технических наук, 2012. – Санкт-Петербург. С. 39–40.

### **Зарегистрированные программные системы**

15. Семенкин Е.С., Панфилов И.А., Сопов Е.А., Становов В.В., Хритonenко Д.И., Брестер К. Ю., Семенкина О.Е. Программная система для автоматизированного управления интеллектуальными информационными

сетями. Свидетельство № 2015662581 о гос. регистрации в Реестре программ для ЭВМ от 26.11.2015.

16. Семенкин Е.С., Панфилов И.А., Сопов Е.А., Становов В.В., Хритonenко Д.И., Брестер К. Ю., Семенкина О.Е. Программная система для автоматизированной генерации моделей и алгоритмов решения задач анализа активности пользователя. Свидетельство № 2015662579 о гос. регистрации в Реестре программ для ЭВМ от 26.11.2015.

17. Семенкин Е.С., Панфилов И.А., Сопов Е.А., Становов В.В., Хритonenко Д.И., Брестер К. Ю., Семенкина О.Е. Программная система автоматизированного проектирования интеллектуальных информационных сетей. Свидетельство № 2015662501 о гос. регистрации в Реестре программ для ЭВМ от 25.11.2015.

18. Хритonenко Д. И., Панфилов И.А., Сопов Е.А. Система автоматического генерирования коллективов нейросетевых моделей распределенными самоконфигурируемыми эволюционными алгоритмами. Свидетельство № 2014612727 о гос. регистрации в Реестре программ для ЭВМ от 05.03.2014.

19. Брестер К. Ю., Панфилов И.А., Семенкин Е.С., Семенкина О.Е., Сопов Е.А., Становов В.В., Хритonenко Д.И. Система автоматизированной классификации и категоризации мультилингвистических документов по содержанию. Свидетельство № 2013660992 о гос. регистрации в Реестре программ для ЭВМ от 26.11.2013.

20. Хритonenко Д. И., Семенкин Е. С. Система автоматического генерирования нейросетевых моделей самоконфигурируемыми эволюционными алгоритмами. Свидетельство № 2013619021 о гос. регистрации в Реестре программ для ЭВМ от 24.09.2013.

21. Хритonenко Д. И., Сергиенко Р. Б. Распределенная программная система автоматического формирования нейронных сетей эволюционными алгоритмами. Свидетельство № 2013611034 о гос. регистрации в Реестре программ для ЭВМ от 9.01.2013.

Хритоненко Дмитрий Иванович  
Адаптивные коллективные нейро-эволюционные алгоритмы  
интеллектуального анализа данных

Автореферат

Подписано к печати xx.xx.20xx. Формат 60x84/16

Уч. изд. л. 1.0 Тираж 100 экз. Заказ № \_\_\_\_\_

Отпечатано в отделе копировальной и множительной техники СибГУ им.  
М.Ф. Решетнева.

660037, г. Красноярск, пр. им. газ. «Красноярский рабочий», 31