УТВЕРЖДАЮ
Проректор по научной работе
и инновациям
Томского государственного университета
систем управления и радиоэлектроники
канд. техн. наук, доцент

_____ Лощилов Антон Геннадьевич
«29» октября 2019 г.

ОТЗЫВ

ведущей организации на диссертационную работу Поляковой Анастасии Сергеевны «Коллективные методы интеллектуального анализа данных на основе нечеткой логики», представленную на соискание ученой степени кандидата технических наук по специальности 05.13.01 – «Системный анализ, управление и обработка информации (космические и информационные технологии)».

Актуальность темы диссертации. Диссертационная работа Поляковой А.С. посвящена методам коллективного вывода при решении задач классификации и регрессии. В настоящее время известно большое количество методов, решающих подобные задачи: бустинг, бегинг, стэкинг, голосование, взвешенное голосование и др. Современные методы машинного обучения достигли значительного прогресса не только на алгоритмическом, но и на технологическом уровне. Существует большое количество библиотек, позволяющих использовать машинное обучение в конкретных проектах. Логичным продолжением исследований в области машинного обучения является изучение возможностей агрегирования, коллективизации отдельных методов с целью повышения надежности и точности конечного вывода. Однако, большинство известных подходов к принятию решений имеет ограниченные возможности программирования итогового вывода. В работе Поляковой А.С. предложен подход к коллективному принятию решения на основе технологий нечеткой логики, базирующихся на множестве нечетких правил и множестве лингвистических переменных. Аналогично тому, как делает свой вывод эксперт с учетом специфики задачи, данных, условий принятия решения, система коллективного принятия решения на основе нечеткой логике может быть также настроена на решаемую задачу. Кроме того, при настройке процедуры нечеткого вывода исследователь может учитывать не только ту информацию, которую можно извлечь из самой выборки, но и сопутствующие сведения: теоретические

представления о предметной области, пожелания заказчика, режимы работы системы и т.д. Все это позволяет существенно повысить эффективность систем коллективного вывода в задачах интеллектуального анализа данных.

Для достижения поставленных задач, в рецензируемой диссертации, исследуются смежные научные задачи, связанные с автоматизацией формирования систем на основе нечеткой логики и автоматизации выбора эффективных экземпляров в данных. Решение этих задач позволяет существенно снизить трудоемкость формирования систем коллективного вывода на основе нечеткой логики и требования к квалификации пользователя.

Итак, в диссертации Поляковой А.С. решается актуальная научно-техническая задача повышения эффективности систем коллективного принятия решения при интеллектуальном анализе данных за счет введения программируемых процедур логического вывода и разработки процедур автоматизированной настройки основных модулей системы.

Целью диссертации Поляковой А.С. является повышение эффективности интеллектуальных технологий анализа данных путем автоматизированного формирования коллективов алгоритмов с помощью специальных систем на нечеткой логике.

Для достижения данной цели в диссертации решены следующие задачи:

1. Выполнен обзор методов анализа данных и их коллективного взаимодействия.
2. Разработан алгоритм коллективного вывода на основе нечеткой логики для решения задач классификации и регрессии.
3. Разработана процедура выбора алгоритмов классификации или регрессии для включения в состав коллектива.
4. Разработана процедура автоматизированного выбора показательных примеров в опорное множество для формирования коллективного вывода.
5. Реализованы программные средства для процедуры коллективного принятия решения на основе нечеткой логики.
6. Исследована работоспособность предложенного алгоритма на тестовых и практических задачах.

Структура и объем работы. Диссертация состоит из введения, четырех глав, заключения, списка использованных источников, включающего 138 наименований. Объем работы – 150 страниц машинописного текста.

Основные научные результаты, полученные в диссертационной работе, состоят в том, что автором:

1. Разработан и реализован коллективный алгоритм анализа данных на основе нечеткой логики для решения задач классификации и регрессии;
2. Разработана и реализована процедура выбора алгоритмов для включения в состав коллектива.
3. Разработана и реализована процедура автоматизированного выбора показательных примеров в опорное множество для формирования коллективного вывода.
4. Разработана программная система, реализующая предложенный алгоритм для формирования коллектива и подходы к проектированию различных его этапов.
5. Исследована работоспособность предложенного алгоритма на тестовых и практических задачах.

Научная новизна результатов диссертационной работы состоит в разработанных соискателем новых алгоритмах формирования систем коллективного принятия решения на основе нечеткой логики: система на основе нечеткой логики для формирования коллективов моделей и алгоритмов анализа данных для решения задач классификации и регрессии, эволюционная процедура выбора агентов для формирования эффективных коллективов, эволюционная процедура автоматизированного формирования базы правил с применением двух уровней эволюции и новым способом представления решений в бинарном пространстве поиска, схема формирования коллективного вывода на основе нечеткой логики с применением иерархической процедуры коллективного вывода, комплексная процедура автоматизированного формирования системы коллективного вывода с возможностью эффективного перераспределения вычислительных ресурсов.

Значимость основных научных результатов для науки и производства. Предложенные алгоритмы формирования коллективного принятия решения, разработанные на основе нового подхода, включающего применение процедуры нечеткого логического вывода, процедур автоматизированного формирования базы нечетких правил, лингвистических переменных, процедур автоматизированного выбора эффективных агентов и эффективных экземпляров данных позволяют существенно дополнить арсенал исследователя в области интеллектуального анализа данных. Применение предложенных алгоритмов в ряде научных и исследовательских проектов может являться обобщением работы отдельных агентов при решении сложных задач.

Так, решение сложных задач анализа данных с помощью методов машинного обучения может привести к ситуации, когда каждая из

применяемых моделей достигает определенного уровня точности и дальнейшее увеличение вычислительных ресурсов либо не улучшают результат, либо ухудшают его за счет эффекта «переобучения» модели. Улучшения полученного результата можно добиться агрегированием результатов различных моделей за счет процедур коллективного вывода. Часто, полученные агрегации за счет явного или неявного усреднения приводят к тому, что итоговый вывод ухудшает значения лучших агентов. В рецензируемой диссертационной работе показано, что применение технологий нечеткого вывода и предложенных процедур автоматизированной настройки баз правил и лингвистических переменных позволяет подстраивать коллективный вывод под специфику решаемой задачи и получать эффективный вывод коллектива. Так в ряде сложных практических задач, таких как распознавание лиц по изображениям, прогнозирование эмоций человека по аудиоданным, прогнозирование криолитового соотношения в металлургической отрасли, было получено более эффективное решение по сравнению с лучшими агентами коллектива.

Кроме того, автором предлагаются новые процедуры автоматизированного выбора эффективных агентов в коллектив и выбора эффективных экземпляров данных на основе генетических алгоритмов. Применение этих процедур позволяют существенно снизить объемы необходимых вычислительных ресурсов и при этом, в ряде случаев, повысить точность коллективного решения.

Теоретическая значимость работы заключается в разработке комплексного подхода к решению задачи интеллектуального анализа данных, обобщающего отдельные результаты различных моделей и алгоритмов машинного обучения. Кроме того, предложенный подход легко масштабируется и на те задачи, где уже получены результаты с помощью отраслевых моделей, построенных экспертами предметной области с применением соответствующих теоретических сведений и собственных представлений экспертов о решаемой задачи, но требуется улучшить полученный результат за счет повышения его точности или снижения остаточной дисперсии модели.

Интеграция современных подходов в области эволюционных вычислений, нечетких логических систем, методов машинного обучения и анализа данных позволила получить новые алгоритмы и модели, повысить эффективность отдельных методов машинного обучения и снизить требования к квалификации конечного пользователя в области анализа данных и построения регрессионных и классификационных моделей при решении сложных практических задач.

Практическая значимость работы заключается в следующем:

- разработанные алгоритмы и модели были реализованы в виде программной системы в среде Python v3.7 с применением свободно распространяемых библиотек;

- эффективность разработанных библиотек подтверждена в процессе решения ряда задач (распознавание лиц по изображению, прогнозирование эмоционального состояния человека по аудиоданным, прогнозирование криолитового соотношения в металлургической отрасли);

- полученные инструменты используются в учебном процессе для проведения практических занятий и лабораторных работ по курсам «Интеллектуальный анализ данных» и «Методы анализа экспериментальных данных» на кафедре САИО СибГУ им. М.Ф. Решетнева.

Полнота опубликования результатов, соответствие автореферата содержанию диссертации. Результаты диссертации достаточно полно опубликованы в открытой печати. Основные положения исследования отражены в 19 публикациях автора, из них три статьи в журналах перечня ВАК РФ и три в изданиях, индексируемых в международных базах цитирования Web of Science и/или Scopus.

Основные теоретические и практические результаты обсуждались на всероссийских и международных конференциях.

Автореферат в основном соответствует содержанию диссертации.

Возможность использования результатов работы. Результаты диссертационной работы могут быть использованы во многих организациях, разрабатывающих и использующих интеллектуальные технологии анализа данных. Результаты диссертации представляют также интерес для образовательных организаций: СФУ, СибГУ им. М.Ф. Решетнева, СПГЭТУ «ЛЭТИ» им. В.И. Ульянова (Ленина), МГТУ им. Баумана, МФТИ, НГТУ, ТПУ, ТУСУР и др.

Замечания и недостатки диссертационной работы. Отметим следующие недостатки диссертационной работы:

1. В диссертационной работе не приводится строгого обоснования применения именно эволюционных алгоритмов для построения оптимизационных процедур, являющихся основой автоматизации выбора основных параметров системы коллективного принятия решения.
2. В работе не приводится постановка математических задач и обоснование выбора тестовых наборов данных классификации и

регрессии, на которых автор проводит сравнение эффективности метода коллективного вывода.

3. Представленные результаты по точности коллективных алгоритмов на некоторых примерах задач распознавания и регрессии в сравнении с отдельными алгоритмами без учета среднеквадратического отклонения по точности или доверительных интервалов (Таблица 1.2, Таблица 1.3, Таблицы 2.5-2.10) не позволяют однозначно судить об эффективности предлагаемого коллективного метода принятия решений на основе нечетких множеств. Желательно было бы накопление статистики по нескольким вариантам созданных обучающих, тестовых, валидационных выборок.
4. Открытым остается вопрос, способен ли предлагаемый автором диссертации метод, объединяя набор только слабых агентов принятия решений, повысить точность принятия решения либо данный подход требует наличия сильных методов, комбинация которых и повышает точность.
5. Приведенные данные свидетельствуют о том, что повышение количества агентов для принятия решения снижает качество принимаемого решения с помощью нечетких множеств. Вероятно, более эффективным методом была бы настройка весов каждого из агентов при взвешенном принятии решений и назначения некоторых весов агентов равным или близким к нулю или их модификация. Попытка выбрать агенты и потом применять стандартный алгоритм взвешивания с вычислением веса на основе качества решения агентами на близком к объекту из опорной выборки является, видимо, частным случаем, когда какие-то полученные веса обнуляются и затем нормируются.

Указанные замечания не снижают научной значимости и общей положительной оценки диссертационной работы.

Заключение. Диссертационная работа Поляковой Анастасии Сергеевны является завершенной квалификационной научно-исследовательской работой, выполненной на актуальную тему, содержащей новые научные результаты.

В целом диссертация написана грамотным научным языком и четко структурирована. Каждая глава содержит содержательные выводы, что облегчает понимание материала. Цель и задачи диссертации полностью соответствуют полученным результатам. Все утверждения диссертации обоснованы и подтверждены вычислительным экспериментами.

Диссертационная работа «Коллективные методы интеллектуального анализа данных на основе нечеткой логики» отвечает требованиям пунктов 9-14 «Положения о порядке присуждения ученых степеней» ВАК в части требований, предъявляемым к кандидатским диссертациям, и соответствует паспорту научной специальности специальности 05.13.01 – «Системный анализ, управление и обработка информации (космические и информационные технологии)», а ее автор - Полякова Анастасия Сергеевна - заслуживает присуждения ученой степени кандидата технических наук по указанной специальности.

Диссертационная работа, автореферат и отзыв на диссертацию обсуждены и одобрены на заседании кафедры «Автоматизированные системы управления» ФГБОУ ВО «Томский государственный университет систем управления и радиоэлектроники» (протокол № 12 от «29» октября 2019 г.).

Заведующий кафедрой
«Автоматизированные системы управления»
ФГБОУ ВО «Томский государственный университет
систем управления и радиоэлектроники»
доктор технических наук, профессор,
заслуженный деятель науки РФ

Кориков Анатолий Михайлович

Профессор кафедры
«Автоматизированные системы управления»
ФГБОУ ВО «Томский государственный университет
систем управления и радиоэлектроники»
доктор технических наук, профессор

Катаев Михаил Юрьевич

Подписи профессоров Корикова Анатолия Михайловича и Катаева Михаила Юрьевич заверяю,

Ученый секретарь ТУСУР Е.В. Прокопчук

ФГБОУ ВО «Томский государственный университет
систем управления и радиоэлектроники»
634050, г. Томск, пр. Ленина, 40
Тел. (3822) 51-05-30, office@tusur.ru