ИССЛЕДОВАНИЯ, ПРОВЕДЕННЫЕ ПРИ ПОДДЕРЖКЕ РОССИЙСКОГО НАУЧНОГО ФОНДА

Проект РНФ 17-71-10176 **«Параллельные вычисления и методы целочисленного программирования для задач кластеризации и интеллектуального анализа больших массивов данных»**

№ гос. регистрации: АААА-А17-117100670002-3

Руководитель проекта – к.ф.-м.н. А.В. Ушаков

В рамках проекта были разработаны и программно реализованы параллельные и распределенные гибридные масштабируемые алгоритмы поиска субоптимальных решений в задаче кластеризации, известной как задача о k-медоидах (k-медиане). Разработанные параллельные алгоритмы протестированы на искусственно сгенерированных примерах, тестовых задачах, взятых из литературы, а также в приложении к кластеризации больших коллекций изображений лиц людей (VGGFace, IJB).

Проект РНФ № 18-71-10001 **«Методология и инструментальная платформа разработки систем извлечения данных из произвольных электронных таблиц»**

№ гос. регистрации: АААА-А18-118082090003-8

Руководитель проекта – к.т.н. А О. Шигаров

Разработан новый предметно-ориентированный язык правил анализа и интерпретации таблиц CRL. По сравнению с оригинальным синтаксисом CRL-диалекта, разработанным для системы исполнения правил Drools Expert, были исключены конструкции управления машинным выводом. Разработана ANTLR грамматика и интерпретатор языка правил анализа и интерпретации таблиц – CRL. Спроектирована архитектура и компонентная модель инструментальной платформы извлечения и трансформации данных произвольных электронных таблиц.

Проведен аналитический обзор современного состояния исследований в области генерации онтологий и связанных данных в форматах RDF/OWL на основе анализа и трансформации структурированных табличных данных. Проведен аналитический обзор существующих подходов и программных средств семантической интерпретации (аннотирования) структурированных табличных данных на основе глобальных онтологий (облака связанных данных – Linking Open Data, LOD). Разработана концепция программного средства аннотирования электронных таблиц и генерации на их основе онтологий в формате RDF/OWL. Разработан прототип модуля (парсера) анализа XSLX-документов электронных таблиц. Разработан прототип модуля взаимодействия с глобальными таксономиями (онтологиями) с целью получения необходимых концептов по запросу. Произведено формирование тестовых наборов данных на основе документов экспертизы промышленной безопасности.

Проект РНФ №16-11-00053 **«Разработка методов исследования и построения иерархических систем децентрализованного интеллектуального управления группами автономных поисково-обследовательских роботов в условиях неопределенности»**

№ гос. регистрации: АААА-А16-116031110118-2

Руководитель проекта – ак. И.В. Бычков

Разработанные на прошлых этапах проекта методы и подходы в области автоматизации и повышения эффективности логических выводов в исчислении ПОФ реализованы в программной системе для автоматического доказательства теорем (прувере).

Основные характеристики и возможности системы. 1) Для выделенного подкласса ПОФ реализован поиск кратчайших вариантов доказательства формулы благодаря системе полного перебора пространства поиска в ширину. 2) Доказательство формулы в интерактивном режиме, т.е. с возможностью выбора человеком необходимых ответов или выбора вопроса, для которого необходимо вычислять возможные варианты ответов. «Вопрос» и «Ответ» – термины из стандартного описания ПОФ. 3) Решение крупных задач благодаря подсистемам индексирования термов, а также экономии памяти, таких как подсистема полного доступа к одинаковым структурам данных и «ленивое» расщепление по-формулы. Подсистемы индексирования позволяют быстро извлекать необходимые термы из крупной базы термов. В случае исчисления ПОФ базой термов является постоянно пополняющаяся база формулы, и при каждой процедуре поиска ответов на вопрос необходимо производить поиск термов в базе, отвечающих заданным критериям: выбрать термы, являющиеся основным примером заданного терма; выбрать термы, являющиеся унифицируемыми с заданным термом; найти точную копию заданного терма. Подсистема полного доступа к одинаковым структурам данных позволяет экономить память на столько, на сколько это возможно. Например, если один и тот же подтерм встречается в двух разных термах, то в в этих термах используется ссылка на одну и ту же область памяти. «Ленивое» расщепление предназначено для эффективной обработки следующего свойства позитивно-образованной формулы: если вопрос содержит дизъюнктивное ветвление и система принимает решение произвести ответ на этот вопрос, то база формулы расщепляется (дублируется) на столько баз, сколько подформул в дизъюнктивном ветвлении. Для некоторых задач рост размера формулы может стать значительным, что даже с использованием полного доступа к одинаковым структурам данных, представляющих части по-формулы, не помогает избежать значительного роста требуемой оперативной памяти. «Ленивое» расщепление позволяет полностью отложить процесс расщепления формулы и расщеплять ее по мере необходимости, опровергая полученные базы одну за одной. 4) Реализована система инкрементального вывода, позволяющего мгновенно производить возврат как назад, так и вперед в любую точку логического вывода. Инкрементальность подразумевает изолированное добавление данных с каждым шагом логического вывода, так что можно четко понять, какие новые данные были добавлены на очередном шаге логического вывода.

Исследована обобщенная архитектура децентрализованного управления дискретно-событийными системами (ДСС), используемыми на верхнем уровне иерархической системы управления группировками автономных подводных роботов (АПР). На сегодняшний день такая архитектура совмещает в себе принципы «конъюнкция разрешенных локальными супервизорами событий и разрешение общих событий по умолчанию» и «дизъюнкция

разрешенных локальными супервизорами событий и запрет общих событий по умолчанию». При этом выбор принципа обработки того или иного события задается априори. Оба принципа требуют проверки условий существования децентрализованного супервизора как принципиальной возможности построения управления. В качестве этих условий выступают управляемость и так называемая ко-наблюдаемость формального языка, описывающего ограничения на функционирование системы. При невыполнении условий существования возможно построение наибольших подъязыков спецификаций, являющихся управляемыми и ко-наблюдаемыми. Если эти языки не отвечают минимально допустимым требованиям к системе, требуется коррекция множеств управляемых и наблюдаемых событий.

Для построения множеств событий, нарушающих условия существования децентрализованного супервизора, используется метод проверки управляемости языка спецификации на основе разработанной на предыдущих этапах проекта формализации ДСС с помощью ПОФ и организации их автоматического вывода. В процессе вывода формулы, включающей в качестве подформул описания системы и спецификации, автоматически формируется множество событий, нарушающих свойство управляемости. Если это множество непусто, в процессе вывода предлагается наибольший возможный управляемый подъязык. В случае его неудовлетворительности необходимо изменить описание системы, возможно, путем увеличения числа управляемых событий. При проверке ко-наблюдаемости формализованное представление управляемой системы составляется из подформул, соответствующих в том числе локальным супервизорам, каждый из которых наблюдает свое подмножество всех событий системы. В случае возникновения конфликта, т.е. ситуации, когда разрешение события может повлечь как разрешенную, так и запрещенную спецификацией последовательность событий, выбирается то из правил, конъюнкция или дизъюнкция, объединения решений локальных супервизоров, которое разрешает конфликт. Таким образом, предложено новое правило построения глобального решения на основе локальных, в котором принцип объединения выбирается на каждом шаге проверки ко-наблюдаемости языка спецификации. В случае сохранения конфликта предлагается исключить из спецификации событие, вызывающее конфликт, генерируя таким образом ко-наблюдаемый подъязык. Для этого может быть использована специальная стратегия для обеспечения направленного и поэтапного поиска логического вывода моделирующей систему ПОФ. Для каждого перехода автомата, генерирующего язык спецификации, стратегия запускает подвывод, чтобы определить нарушение условия ко-наблюдаемости событием, соответствующим этому переходу. Если нарушения нет, то вывод продолжается для проверки остальных переходов до тех пор, пока нарушение не будет найдено, что будет означать отклонение проверяемой спецификации. Далее вывод будет перезапускаться для спецификации, порождающей автомат, который не содержит проверенного перехода.

Разработанный подход обеспечит эффективное построение децентрализованных супервизоров, гарантирующих неблокирование системы и генерирующих наибольший управляемый и ко-наблюдаемый подъязык спецификации в случае отсутствия свойств управляемости и ко-наблюдаемости. Разработанные методы внедрены в программный комплекс моделирования миссий АПР как подсистема верхнего уровня управления.

Разработана двухуровневая система управления группой АПР (рис. 37) для реализации миссии по обследованию придонной области путем обеспечения движения группы по сгенерированным специальным образом траекториям, покрывающим исследуемую область,

с сохранением заданной геометрической конфигурации (формации) группы во время рабочих ходов и избеганием столкновений с препятствиями. Нижний уровень системы управления АПР включает алгоритмы, реализующие требуемое поведение группы в различных режимах, а верхний уровень, в основе которого лежит дискретно-событийная система, ответственен за переключение между режимами в случае возникновения значимых событий, обусловленных изменением состояния внешней среды и роботов. Разработаны алгоритмы управления АПР нижнего уровня, реализующие характерные для такой миссии режимы функционирования группы. Синтез параметров в этих алгоритмах выполнен с использованием программного пакета для анализа и синтеза цифровых систем управления, основанного на методе векторных функций Ляпунова. Регуляторы верхнего уровня для роботов группы, выполняющих функции лидера и ведомого, построены с использованием теории супервизорного управления дискретно-событийными системами. Проведено численное исследование и моделирование построенной двухуровневой системы управления.

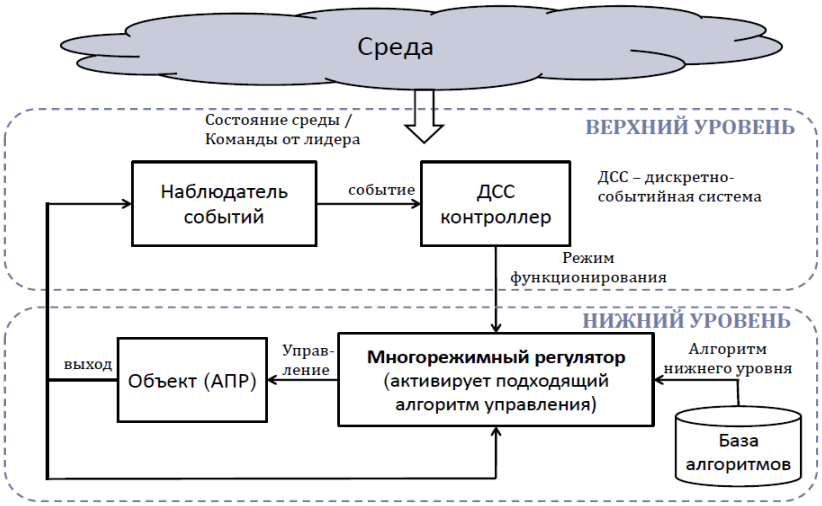

Рис. 37. Структура двухуровневой системы управления АПР

Произведено дальнейшее расширение и развитие модели задачи группового патрулирования акватории АПР, исследованной на первых двух этапах проекта. Задача группового патрулирования в новой усложненной регулярной постановке содержит в себе особенности сразу нескольких известных NP-трудных транспортных задач: систематического покрытия и патрулирования (*persistent coverage and tasks patrolling problem*), мультикоммивояжера и нескольких вариаций задачи маршрутизации транспорта (ЗМТ). Как и в случае систематического патрулирования задачей группы АПР является долгосрочный и непрерывный (отсутствует горизонт планирования в его классическом понимании) обход заданного множества целей с частотой, равной заданной или превышающей ее. В то же время между постановками существует ряд значительных различий. Во-первых, исследуемая в рамках проекта задача регулярного патрулирования посвящена не поиску минимального состава группы АПР, гарантирующего своевременный обход целей без опозданий, а генерации групповой траектории максимальной эффективности для текущего фиксированного состава группы роботов. Во-вторых, при решении задачи учитываются временные затраты на обследование всех целей, при этом требуемое время может различаться в зависимости от приоритета подобласти, в которой находится обследуемая цель. Ограничение по оборудованию при проведении обследований, введенное

в постановку задачи патрулирования на второй год работы над проектом, было расширено для учета требования на наличие одновременно нескольких видов исследовательской аппаратуры. В связи с тем, что в группе может не быть аппаратов (или быть в недостаточном количестве), полностью удовлетворяющих таким требованиям для отдельных подобластей, в модель подводной миссии также добавлена возможность совместного согласованного обследования целей одновременно несколькими роботами группы, в совокупности обладающими требуемым набором бортового оборудования.

Кроме того, модель задачи группового патрулирования была переработана для работы в трехмерном пространстве: добавлены трехмерные координаты всех объектов миссии, адаптирована процедура расстановки контрольных точек (целей) в пространстве. Для построения траекторий движения роботов на трехмерной карте с учетом рельефа в условиях значительно возросшей размерности в связи с добавлением третьего измерения предлагается использовать подход на основе декомпозиции карты акватории на несколько регионов с использованием эвристики шлюзов (*gateway heuristic*), когда на основе карты высот определяются наиболее рациональные места переходов между выделенными регионами, что значительно уменьшает вычислительную нагрузку при поиске путей. При генерации траекторий учитываются динамические ограничения на движение АПР: угловая скорость, углы дифферента и др.

Разработанная на первых этапах проекта модификация эволюционных алгоритмов для решения задачи патрулирования акватории группой АПР была расширена и дополнена новыми эвристиками и вычислительными процедурами, позволяющими не только учитывать новые добавленные требования и ограничения, но и повысить скорость и эффективность подхода в целом. Во-первых, разработанные ранее параллельные конструктивные эвристики были модифицированы для генерации допустимых рациональных маршрутов в условиях жестких ограничений на оборудование. Во-вторых, все используемые генетические операторы были адаптированы с целью сохранения допустимости групповых маршрутов при осуществлении скрещивания и мутации.

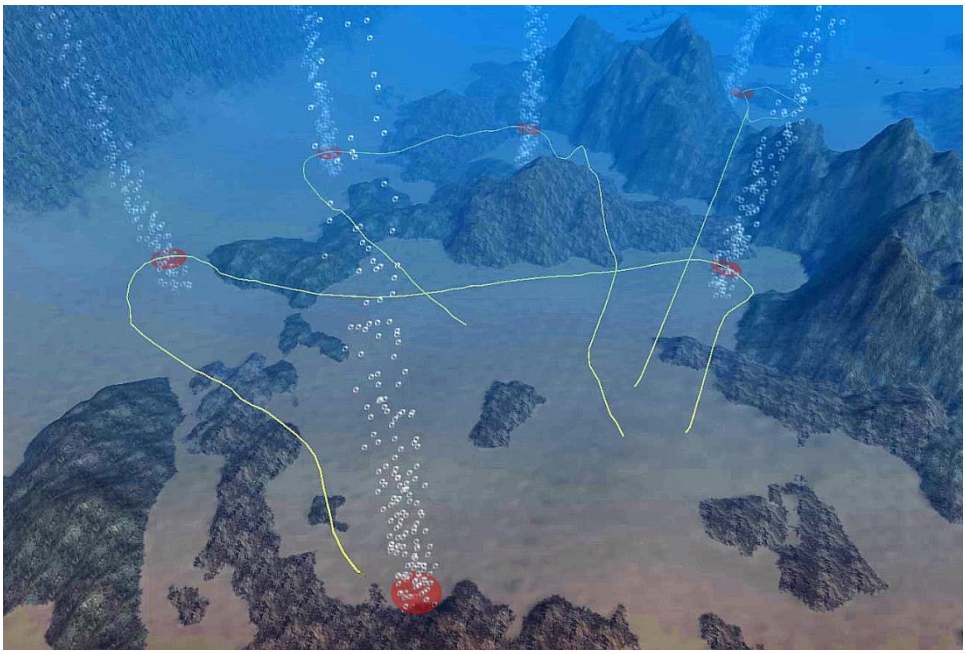

Рис. 38. Моделирование группового обследования набора контрольных точек в трехмерном пространстве

Проект РНФ №16-11-10046 **«Применение параллельных и распределенных алгоритмов решения проблемы булевой выполнимости (SAT) к криптоанализу, поиску комбинаторных структур и исследованию дискретных моделей коллективного поведения»**

Руководитель проекта – к.т.н. А.А. Семенов

На основании развитой ранее техники автоматического построения guess-and-determine атак описаны новые алгебраические атаки на ряд блочных алгоритмов шифрования (DES, GOST 28147-89, PRESENT, Simon). В основе новых атак лежит идея использовать уравнения, описывающие как прямые (зашифрование), так и обратные (расшифрование) преобразования в рассматриваемом шифре. Часть переменных, используемых в таких кодировках, связываются соотношениями, смысл которых аналогичен соотношениям согласования в атаках из класса meet in the middle («встреча посередине»). Получаемая система уравнений над полем GF(2) сводится к проблеме выполнимости булевой формулы в КНФ, к которой применяются современные алгоритмы решения SAT. Более точно, используется автоматическая техника построения guess-and-determine атаки с оценкой трудоемкости, которая была бы лучше времени полного перебора пространства ключей. С этой целью задача поиска множества угадываемых бит (guessed bits set) ставится в форме задачи минимизации псевдобулевой функции, которая является формализацией понятия SAT-иммунность, введенного ранее Н. Куртуа. Применение данной техники к новым пропозициональным кодировкам (которые можно условно назвать meet in the middle - кодировками) позволило построить атаки, существенно более эффективные в сравнении с известными аналогами. Так, полученная атака для 6-раундовой версии шифра PRESENT-80 оказалась примерно в 1000 раз эффективнее лучшей известной алгебраической атаки на данный шифр, опубликованной в 2017 г.

Разработаны автоматические алгоритмы синтеза ослабляющих ограничений (relaxation constrains) для задач обращения неполнораундовых вариантов криптографических хеш-функций, основанных на конструкции Меркля-Дамгарда. На текущем этапе исследовались функции из семейства MD4-k, где через k обозначается число шагов базового алгоритма (алгоритм MD4 в полнораундовой версии имеет k=48 шагов). Новизна предложенного подхода заключается в том, что каждому набору relaxation constraints сопоставляется значение специальной псевдобулевой функции, которое эвристически оценивает эффективность соответствующего набора ограничений. Ограничения с лучшей оценкой эффективности ищутся в процессе работы алгоритма локального поиска, относящегося к классу поиска с запретами (Tabu Search), максимизирующего введенную псевдобулеву функцию. Показано, что каждое синтезируемое ограничение порождает специальную функцию, обозначаемую через g, длина входа которой существенно меньше длины одного блока хешируемого сообщения. Для задачи обращения хеш-функции MD4-39 была построена функция типа g, обозначаемая через g(MD4-39), длина входа которой равна 128 битам. В процессе экспериментов мы установили, что примерно 65% всевозможных 128 битных векторов являются значениями функции g(MD4-39). Тем самым случайное хеш-

значение MD4-39 является значением функции g(MD4-39) с вероятностью, близкой к 65%. Функция g(MD4-39) обращается на обычном ПК менее, чем за минуту с использованием последовательного SAT-решателя. Далее мы исследовали задачу обращения MD4-40. По аналогии с g(MD4-39) мы построили функцию g(MD4-40), которая преобразует 128 битные векторы в 128 битные. Было показано, что не менее 50% 128 битных векторов являются значениями функции g(MD4-40). Тем самым случайное хеш-значение MD4-40 с вероятностью, не менее 1/2, совпадает с выходом функции g(MD4-40). Обращение функции g(MD4-40) уже не под силу последовательным SAT-решателям. Однако нам удалось построить guess-and-determine preimage-атаку на g(MD4-40), используя введенное ранее понятие инверсного множества с лазейкой (Inverse Backdoor Set, IBS). Оценка трудоемкости построенной атаки говорит о возможности ее реализации в проектах добровольных вычислений, подобных проекту SAT@home. Насколько нам известно, это первая preimage-атака с реалистичной оценкой трудоемкости для функции MD4-40.

На основании изучения динамических процессов активационного типа в коллективах, представленных сетями, разработаны новые алгоритмы построения графов, отображающих развитие атак в компьютерных сетях. В предложенной модели развития атак события имеют четко выраженную причинно-следственную взаимосвязь в рамках дискретного динамического процесса, задаваемого графом компьютерной сети. Граф состояний соответствующей дискретной динамической системы (ДДС) является отправной точкой для построения графа, представляющего атаки. По графу состояний ДДС строится формула в КНФ, которая есть, по сути, результат символьного исполнения алгоритма, описывающего развитие атаки. Построенную КНФ можно использовать в качестве основы для порождения графа, представляющего все атаки в сети. Процесс порождения такого графа выглядит как распространение/вывод булевых ограничений из КНФ в результате применения правила единичного дизъюнкта (Unit Propagation rule). Сложность процедуры построения графа атак в рамках введенной модели есть O(n^2), где n – число хостов в сети. Графы атак, которые строятся в рамках новой модели, в отличие от близких аналогов (например, генерируемых системой MulVAL), не содержат циклов и тем самым дают возможность эффективно выделить множество кратчайших атак. Помимо сказанного, были продолжены исследования процессов активационной динамики в случайных и естественных сетях большой размерности (в частности, рассматривались фрагменты сети Twitter). Проведено экспериментальное сравнение ряда алгоритмов, используемых для решения проблемы максимизации влияния в сетях (Influence Maximization Problem). Построена условная иерархия таких алгоритмов. На первом уровне находятся эвристики активации, использующие различные меры центральности (это самые быстрые, но и наименее точные алгоритмы). Затем идут жадные алгоритмы и эволюционные метаэвристики. На последнем уровне иерархии – алгоритмы активации, основанные на SAT-подходе: при большом времени решения задачи расстановки активаторов данные алгоритмы позволяют строить активирующие множества наименьшей мощности.

Сотрудники ИДСТУ СО РАН также проводили исследования по проектам РНФ под руководством ведущих ученых из других научных организаций:

Проект РНФ № 18-12-00128 **«Прецизионное исследование связанных состояний частиц в квантовой теории поля»**
Руководитель проекта – д.ф.-м.н. А.П. Мартыненко,
исполнитель от ИДСТУ СО РАН – А.Е. Раджабов

Проект РНФ № 18-41-06003 **«Карлсруэ-Российская инициатива по работе с астрофизическими данными на протяжении их жизненного цикла»**
Руководитель проекта – к.ф.-м.н. А.П. Крюков,
исполнитель от ИДСТУ СО РАН – ак. И.В. Бычков, к.т.н. А.О. Шигаров