

Тематический выпуск

Вступительное слово к специальному выпуску по материалам XXV Международной конференции DAMDID/RCDL-2023

Специальный выпуск журнала “Автоматика и телемеханика” Российской академии наук посвящен научным результатам, представленным на XXV Международной конференции «Аналитика и управление данными в областях с интенсивным использованием данных» (DAMDID/RCDL-2023), прошедшей 24–27 октября 2023 г. в НИУ ВШЭ (Москва, Россия). Читателю предлагаются полные тексты избранных статей, соответствующих тематике журнала и представленных на DAMDID/RCDL-2023.

В 2015 г. в результате трансформации конференции RCDL (“Электронные библиотеки: перспективные методы и технологии, электронные коллекции”) была образована конференция DAMDID для создания форума, посвященного насущным проблемам анализа и управления данными в ходе исследований в различных областях с интенсивным использованием данных. Данная реорганизация была связана с расширением сферы использования электронных библиотек, в частности с активизацией их применения в научных исследованиях, возникновением потребности в технологиях эффективного хранения, обработки и анализа научных данных, интеграции неоднородных данных из множества источников. В итоге преобразования была обеспечена не только тематическая преемственность конференции по отношению к RCDL, но также сохранено RCDL-сообщество, сформировавшееся в течение 16 лет успешной работы данной конференции.

Конференция DAMDID известна как мультидисциплинарный форум исследователей и практиков из различных областей науки и промышленности, содействующий сотрудничеству и обмену идеями и методами в сфере анализа и управления данными, развиваемыми в конкретных областях X -информатики ($X \in \{\text{астро, био, гео, нейро, мед, физика, химия, материаловедение, социо, бизнес, финансы, ...}\}$).

Тематика конференции включает различные треки группы анализа данных, решения задач и организации экспериментов:

- постановки и решение задач в областях с интенсивным использованием данных (методы, процессы, инструменты);
- организация экспериментов (теории, гипотезы, модели, имитационное моделирование, инфраструктуры, потоки работ);
- методы и процедуры анализа данных (разведочный анализ, статистика и машинное обучение, мета-анализ, эффективность и масштабируемость методов);

- концептуальное моделирование предметных областей и представление знаний (семантика, онтологии, схемы баз данных, концептуализация алгоритмов и потоков работ, семантическая интероперабельность);
- поддержка исследований в инфраструктурах данных (функции и архитектуры виртуальных лабораторий и обсерваторий, совместное использование данных в междисциплинарных исследованиях).

В программу конференции также традиционно включена тематика управления данными:

- методы, инструменты и инфраструктуры сбора, хранения, обработки данных (данные и метаданные, их семантика, качество и происхождение, очистка, обнаружение аномалий);
- интеграция данных (модели данных, схемы, разрешение сущностей и слияние данных, виртуальная и материализованная интеграция, хранилища данных, ETL-процессы, многомерные модели данных, разноструктурированные данные);
- извлечение информации из данных наблюдений (полнота и актуальность информации в астрономии, спектроскопии, материаловедении, медицине и т.д.);
- извлечение информации из текстов;
- исследовательские инфраструктуры и их применение (облака, гриды, распределенные кластеры, суперкомпьютеры; реализация, масштабирование и оценка производительности инфраструктур; новые модели программирования, виртуализация);
- роль семантического веба в областях с интенсивным использованием данных.

Программа включала три ключевых и три приглашенных пленарных доклада, из них четыре — научных и два — промышленных.

Доклады были распределены по следующим секциям и направлениям:

- алгоритмы решения задач;
- анализ данных в медицине и когнитивных науках;
- анализ изображений;
- анализ социальных сетей;
- анализ данных в астрономии;
- базы данных в материаловедении;
- извлечение информации из текстов;
- искусственный интеллект и машинное обучение в материаловедении;
- концептуальные модели, онтологии и семантический веб;
- методы машинного обучения и их приложения;
- моделирование в науках о Земле.

Программный комитет конференции, включающий ученых из Великобритании, Германии, Индии, Италии, Испании, Китая, Латвии, России, Сербии, США, Японии, рассмотрел 79 заявок на конференцию. В результате одностороннего слепого рецензирования, проводившегося в два этапа, включая обсуждение итогов рецензирования членами программного комитета, 53 ра-

боты были приняты как полные статьи, 15 — как короткие, 11 — отклонены или сняты.

В конференции приняли участие более 120 человек. Были проведены три пленарных заседания и заседания 18 секций. Заслушано 6 пленарных докладов и 61 секционный доклад. Участники представляли научные и учебные заведения из 11 регионов России: Казани, Москвы, Муром, Новосибирска, Обнинска, Самары, Санкт-Петербурга, Томска, Тюмени, Челябинска, а также зарубежных стран: Армении, Великобритании, Германии, Индии, Ирана, Канады, Китая, Пакистана, ОАЭ, США и Японии.

Как и в предыдущие годы, доклады, не вошедшие в специальные выпуски журналов, будут опубликованы в серии Communications in Computer and Information Science издательством Springer.

На основе рекомендаций программного комитета по итогам рецензирования в соответствии с тематикой журнала и с учетом предпочтений авторов для публикации в специальном выпуске было отобрано 8 статей:

- *Б.Г. Миркин, А.А. Паринов* “Агломеративный консенсусный кластер-анализ с автоматическим выбором числа кластеров”;
- *М. Сохраби, А. Фатоллахи-Фард, В.А. Громов* “Алгоритм геномной инженерии (GEA): эффективный метаэвристический алгоритм для решения задач комбинаторной оптимизации”;
- *Т.М. Биджиев, Д.Е. Намиот* “Атаки на модели машинного обучения, основанные на фреймворке PyTorch”;
- *М.М. Зуева, С.О. Кузнецов* “Индексы интересности как инструмент отбора формальных понятий для построения нейронной сети на основе решетки формальных понятий”;
- *Е.Д. Вязилов, Д.А. Мельников* “Об использовании данных цифровых двойников в моделях, связанных с учетом воздействия окружающей среды на предприятия”;
- *К.А. Найденова, В.А. Пархоменко* “Правдоподобные рассуждения в алгоритме генерации хороших классификационных тестов”;
- *Д.А. Люткин, Д.В. Поздняков, А.А. Соловьев, Д.В. Жуков, М.Ш.И. Малик, Д.И. Игнатов* “Применение трансформеров для определения профильного врача на основе запросов пользователей”;
- *А.Г. Сорока, Г.В. Микельсон, А.В. Мецераков, С.В. Герасимов* “Smart Routes: система для разработки и сравнения алгоритмов решения задачи оптимизации маршрутов с реалистичными ограничениями”.

В центре внимания авторов отобранных статей лежат такие темы, как методы машинного обучения и анализа данных, правдоподобные рассуждения и нейросети, алгоритмы управления и оптимизации, анализ текстов и моделирование на основе цифровых двойников.

Ж. Башариас, Барселона

Д.И. Игнатов, Москва

С.О. Кузнецов, Москва

С.А. Ступников, Москва