

## ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ НА ЗАЩИТЕ ОТ НАВОДНЕНИЙ ДАНЫМИ

**Автор:** Мельник М.А. (Университет ИТМО, Санкт-Петербург)

**Научный руководитель:** Бухановский А.В. (Университет ИТМО, Санкт-Петербург)

Данные являются неотъемлемой частью нашей современной жизни. Мы сталкиваемся с данными напрямую или косвенно при работе с социальными сетями, интернетом вещей, при использовании медиа ресурсов, работе в области аналитики данных. В виду постоянного роста объёмов генерируемых данных, образно можно заявить о наличии проблемы наводнения данными. Суть проблемы заключается в том: что для хранения данных требуются физические ресурсы; не все данные являются полезными; требуется обработка данных для извлечения знаний или иной полезной информации. Для организации распределенных вычислений существует огромное число платформ и технологий, таких как Apache Spark, Storm, Flink, AWS, которые позволяют эффективно масштабировать вычисления. Однако, рост объёма и сложности обработки данных возрастает стремительнее, чем технологии обработки. Таким образом, развитие технологий обработки данных является крайне актуальным направлением для исследований.

Задача планирования является существенной оптимизационной проблемой при организации эффективной обработки данных. Суть задачи заключается в качественном распределении вычислительной нагрузки по узлам инфраструктуры. На данный момент существует множество эвристических и метаэвристических подходов к решению данной задачи, в том числе их гибридные модификации. Однако, несмотря на высокую популярность методов искусственного интеллекта (ИИ) для решения широкого круга задач, полноценные решения задачи планирования на основе методов ИИ отсутствуют. К трудностям реализации таких решений относятся: дискретность задачи планирования; формат состояний и действий (входных и выходных данных). Таким образом, основной идеей данной работы является проектирование модели на основе методов ИИ для решения задачи планирования композитных приложений. Целью работы является достижение разработанной схемой высокой производительности, эффективности, масштабируемости и адаптивности.

Разработанная схема планирования Neural Network Scheduler (NNS) основана на принципах обучения с подкреплением и нейронных сетях в качестве аппроксиматора функции полезности. Для проведения экспериментальных исследований была разработана имитационная модель вычислительной инфраструктуры, включающая в себя моделирование вычислительной нагрузки, ресурсов, средств мониторинга и анализа системы, а также модулей построения моделей производительности. Спроектированная схема планирования NNS подразумевает под собой нейронную сеть, направленную на размещение  $T$  задач из очереди среди  $R$  доступных вычислительных узлов. Для составления входного состояния агента обучения был введен ряд признаков, основанный на характеристиках вычислительной среды. Признаки разделены на несколько групп: общие; признаки задач; признаки ресурсов; признаки производительности задач на ресурсах. Перехода от абсолютных величин к относительным производится на основе введения теоретического худшего времени выполнения, которое используется для нормировки входного состояния агента планирования.

Были проведены серии экспериментальных исследований по планирования как однотипных, так и разнородных композитных приложений. Результат экспериментов показал возможность обучения разработанного агента планирования NNS и совмещения масштабируемости, адаптируемости, производительности и эффективности проводимых вычислений.