

Моделирование работы суперкомпьютерного кластера для оценки эффективности интеллектуальных алгоритмов в задачах диспетчеризации

В.Н. Сеннов, А.А. Лукашин, Е.П. Петухов

Санкт-Петербургский политехнический университет Петра Великого

Повышение эффективности использования суперкомпьютерных центров (СКЦ) является крайне важной задачей, особенно в условиях растущего спроса на высокопроизводительные вычисления и дефицита суперкомпьютерных ресурсов.

За последние годы большое количество работ было посвящено моделям оценки времени выполнения задач в СКЦ, в том числе с применением методов машинного обучения [1, 2]. Проведенные исследования показали, что пользователи суперкомпьютерных ресурсов, как правило, не используют параметр времени исполнения задачи или устанавливают его с большим запасом [3]. Данный параметр используется планировщиком ресурсов суперкомпьютерного кластера для постановки задач в очередь и формирование расписания.

Таким образом, разработка моделей для улучшения пользовательской оценки времени выполнения является важной и актуальной задачей. Однако метрики качества (обычно, средняя абсолютная или относительная ошибка), используемые при исследовании получившейся модели оценки не отвечают на вопрос, насколько эффективнее станет работать алгоритм диспетчеризации при использовании исследуемой модели.

Эффективность работы алгоритмов диспетчеризации оценивается при помощи таких метрик, как среднее количество простаивающих узлов или среднее время ожидания задачи в очереди. Исследование эффективности работы алгоритмов оценки времени исполнения задач и поиска наилучших настроек планировщика проводить на работающем кластере крайне сложно, поэтому необходимо иметь возможность провести моделирование работы суперкомпьютерного кластера с выбранным алгоритмом оценки времени и собрать метрики.

В данной работе предложен метод симуляции работы виртуального кластера на основе виртуализации служебных и вычислительных узлов кластера, а также показаны результаты симуляции на данных СКЦ «Политехнический».

Исследование выполнено при финансовой поддержке Минобрнауки России в рамках научно-исследовательского проекта «Разработка и исследование моделей машинного обучения для решения фундаментальных задач искусственного интеллекта в топливно-энергетическом комплексе» (FSEG-2024-0027). Результаты исследования получены с использованием вычислительных ресурсов суперкомпьютерного центра «Политехнический» (<https://scc.spbstu.ru>).

Литература

1. Баранов А.В., Николаев Д.С. Применение машинного обучения для прогнозирования времени выполнения суперкомпьютерных заданий // Программные продукты и системы. 2020. Т. 33, № 2. С. 218–220. DOI: 10.15827/0236-235X.130.218-228.
2. Muliukha V.A., Vostrov A.V., Motorin D.E. et al. Supercomputer resources management using machine learning methods under constraints // Computing, Telecommunications and Control. 2024. Vol. 17, no. 3. P. 32–41. DOI: 10.18721/JCSTCS.17303.
3. Malov S.V., Lukashin A.A. Count time series analysis of jobs scheduling in the hybrid supercomputer center // Computing, Telecommunications and Control. 2024. Vol. 17, no. 3. P. 42–53. DOI: 10.18721/JCSTCS.17304.