

Задачи оптимизации использования распределённой облачной вычислительной инфраструктуры

Писковский Виктор Олегович^{1,2*}

vpiskovski@lvk.cs.msu.ru

*Грушо Александр Александрович*²

grusho@yandex.ru

*Забезжайло Михаил Иванович*²

m.zabezhailo@yandex.ru

¹Москва, ВМК МГУ им. М.В. Ломоносова

²Москва, ФИЦ ИУ РАН

В докладе представлены результаты решения задач оптимального использования распределённой облачной вычислительной инфраструктуры. В качестве таких задач рассматривались:

- прогнозирование качества получаемой клиентом сетевой услуги по имеющимся трассам работы оборудования при реализации сервисов,
- оценка времени выполнения вычислительных задач центром обработки данных по журналу регистрации работ за 2023 год системы управления заданиями высокопроизводительного вычислителя SLURM,
- время активации образа виртуальной машины в качестве абонентского облачного терминала, загружаемого с общего разделяемого сетевого ресурса.

В рамках первой задачи [1]-[4] решалась проблема нахождения оптимального метода по имеющимся в распоряжении наборе данных:

$$F_k(\vec{x}) = \operatorname{argmin}_{F_k} \|F_k(\vec{x}) - y_k\|_{l_2}, \quad \vec{x} \in \mathbb{R}^D, \quad \{\vec{x}\}_{n=1}^N, \quad N \sim 20\,000$$

где F_k - искомый метод, $\mathbb{R}^D$ - пространство показателей, y_k - прогнозируемые характеристики, N - количество трасс, используемых для обучения модели

В качестве набора данных были использованы трассы Центра обработки данных Шведского Королевского технологического университета (Стокгольм, Швеция) [5] для сервисов видео по запросу и СУБД общим объёмом десятки тысяч записей (N) по полторы тысячи показателей (D) в каждой.

В результате исследований были отобраны методы:

- на основе случайного леса (Random Forest, RF),
- анализ главных компонент (Principal Component Analysis, PCA),
- усеченного сингулярного разложения (Truncated Singular Value Decomposition, TSVD),
- на основе нейронных сетей: многослойный перцептрон (Multilevel Perceptron, MLP) и свёрточная сеть (Convolution Neuron Network, CNN).

Наилучшие результаты получены для композиции метода анализа главных компонент с методом на основе случайного леса или методами на основе нейронных сетей.

Результат решения первой задачи был применён ко второй проблеме оценки времени выполнения вычислительных задач Центром обработки данных, показав вполне приемлемую точность прогноза.

В рамках третьей задачи [6] исследовались время старта виртуальной машины (ВМ) в режиме расширенного фильтра записи в зависимости от применяемых сетевых характеристик, протоколов доступа и файловых систем. Работа ВМ осуществляется непосредственно из сетевого хранилища, доступного только в режиме чтения. В результате разработаны и экспериментально обоснованы практические рекомендации по оптимальной организации работы с ВМ в рассматриваемых условиях.

В планах работ – расширить комплекс методов техниками матричного разложения для восстановления пропущенных данных и понижения размерности задачи, применить промышленные отечественные платформы машинного обучения ГосНИИАС «Plat2023» / «Платформа-ГНС» [7], использовать методы сжатия информации для уменьшения трафика при обмене данными между узлами инфраструктуры и встраивания методов идентификации анализируемых данных.

- [1] *Лычева Е. О., Писковский В. О.* Прогноз выполнения задач VOD и KV БД на оборудовании сетей доступа // Ломоносовские чтения. Научная конференция. 20 марта – 3 апреля 2024 г. : тезисы докладов. Москва: МАКС Пресс, 2024. — С.119-120.
- [2] *Могиленец В. М., Писковский В. О.* Применение CNN для прогноза выполнения задач VoD и KV БД на оборудовании сетей доступа // Ломоносовские чтения. Научная конференция. 20 марта – 3 апреля 2024 г. : тезисы докладов. Москва: МАКС Пресс, 2024. — С.121-122.
- [3] *Писковский В. О.* Прогнозирование временных характеристик периферийных вычислений при реализации прикладных сетевых сервисов. //Международная конференция «Математика в созвездии наук». Издательство Московского университета, 2024. — 815 с. ил., стр. 639-640
- [4] *Писковский В. О., Грушо А. А.* Прогнозирование временных характеристик прикладных сетевых сервисов.//Труды XIV Всероссийского совещания по проблемам управления (ВСПУ-2024, Совещание), 17-20 июня 2024 года ИПУ РАН (г. Москва, Россия). (Срок выхода публикации – октябрь 2024)
- [5] Data traces from a data center testbed – Kaggle. // [Электронный ресурс] - URL: www.kaggle.com/datasets/jaliltaghia/data-traces-from-a-data-center-testbed (дата обращения: 11.09.2023)
- [6] *Сагалевич В. Д., Писковский В. О.* Исследование времени подготовки АРМ к работе при использовании разделяемого сетевого репозитория образов ВМ // Ломоносовские чтения. Научная конференция. 20 марта – 3 апреля 2024 г. : тезисы докладов. Москва: МАКС Пресс, 2024. — С.122-123.
- [7] Унифицированная программная платформа для разработки конечно ориентированных программных комплексов автоматического распознавания объектов на основе нейросетевых подходов «Платформа-ГНС». // [Электронный ресурс] - URL: <https://www.gosniias.ru/platform.html> (дата обращения: 03.04.2024)