

**О МЕТОДЕ ПРОГНОЗИРОВАНИЯ ИСПОЛЬЗОВАНИЯ  
ВЫЧИСЛИТЕЛЬНЫХ РЕСУРСОВ В ОБЛАКАХ**

*Задорожная Юлия Андреевна, Пашков Василий  
Николаевич*

*Студент, Ассистент кафедры АСВК*

*Факультет ВМК МГУ имени М. В. Ломоносова, Москва, Россия*

*E-mail: s02190029@gse.cs.msu.ru, pashkov@lvk.cs.msu.ru*

**Научный руководитель** — *Пашков Василий Николаевич*

Технология удаленной поставки аппаратного обеспечения позволяет владельцам облачных услуг управлять большим количеством экземпляров виртуальных машин на одном физическом сервере. Из-за этой особенности организации отображения виртуальных ресурсов на физические на сервере возможна ситуация перегрузки, когда реальных физических ресурсов для некоторых виртуальных машин, размещенных на сервере, станет не хватать. В таком случае виртуальная машина начнет деградировать, а сервис, функционирующий в данной виртуальной машине станет недоступным для конечных пользователей.

Подобные ситуации недопустимы для облачных провайдеров, и их количество должно быть сведено к минимуму для обеспечения доступности облачных сервисов для конечных пользователей, поэтому краткосрочный прогноз спроса на ресурсы является необходимой задачей для обеспечения бесперебойной работы сервера.

Пусть сервер обладает следующим набором физических ресурсов, заданным в виде вектора  $H = (V_{CPU}^{host}, V_{RAM}^{host}, V_{SSD}^{host}, \dots)$ , где

- $V_{CPU}$  - объем физических ресурсов CPU сервера;
- $V_{RAM}$  - объем физических ресурсов RAM сервера;
- $V_{SSD}$  - объем жестких SSD дисков;

Пусть на заданном сервере запущено  $N$  виртуальных машин (VM). Каждая виртуальная машина характеризуется аналогичным набором показателей потребления ею физических ресурсов сервера в каждый момент времени:

$$VM_i(t) = (v_i^{CPU}(t), v_i^{RAM}(t), v_i^{SSD}(t), \dots), i = 1..N$$

Таким образом, при нормальном режиме функционирования облака, суммарное количество потребляемых физических ресурсов

сервера всеми запущенными виртуальными машинами на нем не должно превышать его доступных объемов ресурсов. Другими словами, в каждый момент времени должно выполняться условие (например, для CPU):

$$\sum_{i=1}^N v_i^{CPU}(t) < V_{CPU}^{host}$$

Для предотвращения перегрузок и ситуаций нехватки физических ресурсов для виртуальных машин необходимо иметь дополнительное время  $T_{migration}$  для принятия решения о миграции виртуальной машины на сервер в облаке с достаточным количеством физических ресурсов. Это время можно увеличить с помощью прогнозирования потребления ресурсов на некоторый краткосрочный период времени, используя методы машинного обучения [4].

В работе проводится обзор существующих подходов и методов прогнозирования использования основных ресурсов сервера виртуальными машинами и их сравнительный анализ, исследуя такие однородные нейронные сети, как LSTM [1], GRU [3], ARIMA [1], а также гибридные нейронные сети в виде LSTM и CNN [2].

В экспериментальной части исследования предлагаемое решение основывается на сравнение результатов однородных и гибридных нейронных сетей, предсказывая временные ряды физических ресурсов виртуальных машин входного датасета от Microsoft Azure и формируя стандартные метрики показателя качества прогноза (MAE, RMSE) на несколько шагов дискретного времени вперед.

### Литература

1. D. Janardhanan and E. Barrett, "CPU workload forecasting of machines in data centers using LSTM recurrent neural networks and ARIMA models," 2017 12th International Conference for Internet Technology and Secured Transactions (ICITST), Cambridge, UK, 2017, pp. 55-60, doi: 10.23919/ICITST.2017.8356346.
2. Ouham, Soukaina Hadi, Youssef Ulah, Arif. (2021). An efficient forecasting approach for resource utilization in cloud data center using CNN-LSTM model. Neural Computing and Applications. 33. 10.1007/s00521-021-05770-9.
3. Y. Cheng, C. Wang, H. Yu, Y. Hu and X. Zhou, "GRU-ES: Resource Usage Prediction of Cloud Workloads Using a Novel Hybrid Method," 2019 IEEE 21st International Conference on High Performance Computing and Communications;

- IEEE 17th International Conference on Smart City; IEEE 5th International Conference on Data Science and Systems (HPCC/SmartCity/DSS), Zhangjiajie, China, 2019, pp. 1249-1256, doi: 10.1109/HPCC/SmartCity/DSS.2019.00175
4. Hsieh S. Y. et al. Utilization-prediction-aware virtual machine consolidation approach for energy-efficient cloud data centers //Journal of Parallel and Distributed Computing. – 2020. – Т. 139. – С. 99-109.