

FEATURE SELECTION FOR PREDICTING LIVE MIGRATION CHARACTERISTICS OF VIRTUAL MACHINES

DOI: 10.36724/2072-8735-2021-15-7-62-70

Denis E. Kirov,
MTUCI, Moscow, Russia, denikirov@yandex.ru

Natalia V. Toutova,
MTUCI, Moscow, Russia, e-natasha@mail.ru

Anatoly S. Vorozhtsov,
MTUCI, Moscow, Russia, as.vorjcov@mail.ru

Iliya A. Andreev,
MTUCI, Moscow, Russia, lc@mtuci.ru

Manuscript received 22 April 2021;
Accepted 27 May 2021

Keywords: virtualization, correlation analysis, live migration, data centers, total migration time, downtime, support vector regression, linear regression

Virtual machine migration is widely used in cloud data centers to scale and maintain the stability of cloud services. However, the performance metrics of virtual machine (VM) applications during migration that are set in the Service Level Agreements may deteriorate. Before starting a migration, it is necessary to evaluate the migration characteristics that affect the quality of service. These characteristics are the total migration time and virtual machine downtime, which are random variables that depend on a variety of factors. The prediction is based on the VM monitoring data. In this paper, we select the most suitable factors for forecasting five types of migrations: precopy migration, postcopy migration, and modification of precopy migration such as CPU throttling, data compression, and delta compression of modified memory pages. To do this, we analyzed a dataset that includes data on five types of migrations, approximately 8000 records of each type. Using correlation analysis, the factors that mostly affect the total migration time and the VM downtime are chosen. These characteristics are predicted using machine learning methods such as linear regression and the support vector machine. It is shown that the number of factors can be reduced almost twice with the same quality of the forecast. In general, linear regression provides relatively high accuracy in predicting the total migration time and the duration of virtual machine downtime. At the same time, the observed nonlinearity in the correlations shows that it is advisable to use the support vector machine to improve the quality of the forecast.

Information about authors:

Denis E. Kirov, postgraduate, MTUCI, Moscow, Russia

Natalia V. Toutova, assistant professor, MTUCI, Moscow, Russia

Anatoly S. Vorozhtsov, assistant professor, MTUCI, Moscow, Russia

Iliya A. Andreev, head of the department, assistant professor, MTUCI, Moscow, Russia

Для цитирования:

Киров Д.Е., Тутова Н.В., Ворожцов А.С., Андреев И.А. Выбор факторов для прогнозирования характеристик живой миграции виртуальных машин // Т-Сomm: Телекоммуникации и транспорт. 2021. Том 15. №7. С. 62-70.

For citation:

Kirov D.E., Toutova N.V., Vorozhtsov A.S., Andreev I.A. (2021) Feature selection for predicting live migration characteristics of virtual machines. *T-Comm*, vol. 15, no.7, pp. 62-70. (in Russian)

1. Introduction

Cloud computing has been developing rapidly in recent years. The availability of the Internet, as well as smartphones, computers and tablets, has led to the popularity of IT services, such as social networks, file sharing services, Internet of Things services, file storage, e-commerce, etc., which has affected the growth of Internet traffic in the world. According to Cisco forecasts, traffic growth will continue and grow from 9.1 zettabytes (ZB) in 2017 to 20.6 ZB in 2021 [1]. Processing and storing "big data" requires the construction of new data centers (DC), as well as the search for the most effective resource management algorithms.

The basis of cloud technologies is virtualization technologies. Between the server hardware and the operating system with applications, there is an additional software layer – a hypervisor, the purpose of which is to create isolated virtual machines on a physical server. Virtualization provide greater flexibility in managing computing resources. If necessary, administrators can add and reduce resources, i.e. apply horizontal or vertical scaling, and move virtual machines from one physical server to another, including to another DC. This process is called VM migration. If the migration occurs without stopping the service, then this migration is called live or dynamic. The use of live migration is preferred in cloud data centers, which are subject to strict requirements for quality of service indicators set in service level agreements (SLAs).

Migration is used in data centers to consolidate virtual machines on fewer physical servers, balance the load, and reduce server overheating [2, 3].

Although migration can be performed without stopping the VM, performance degradation of the VMs and short-term downtime are unavoidable. In [4], it was shown that the downtime can vary significantly from 60 ms to 3 seconds, depending on the size of the virtual machine's RAM, how the application uses the memory, and available network bandwidth. Therefore, before migration, it is necessary to predict migration characteristics such as the total migration time VM downtime, which directly affect the quality of service indicators set in SLA.

This paper analyzes the existing work on predicting the characteristics of live migration, in particular, total migration time and VM downtime. To do this, we analyzed a data set that includes data on five types of migrations, approximately 8000 records of each type. Using correlation analysis, the factors that most affect total migration time VM downtime are selected. It is shown that the number of factors for predicting these characteristics can be reduced almost twice with the same quality of the forecast, which will significantly reduce the complexity of collecting initial data.

This paper is organized as follows. Section 2 discusses the types of VM live migration, such as pre-copy and post-copy migrations and their modifications. Section 3 describes the characteristics of live migration that affect the quality of the cloud service. Section 4 analyzes the live VM migration dataset. The correlation analysis of factors is carried out and the most significant factors influencing the total migration time and VM downtime are selected. Section 5 presents the results of experiments on forecasting by linear regression and the support vector machine (SVM) with a smaller number of factors.

2. Types of live migrations of virtual machines

Currently, two main types of live VM migrations have been developed, which are used in various hypervisors:

- Pre-copy migration [5]: This approach is used in most hypervisors, such as VMware, Xen, and KVM. The migration includes several stages, as shown in Fig. 1. During this migration, the RAM of the VM is copied to the destination VM, while the source VM does not stop: changes to memory pages continue to occur on it. The new VM will not start working until all memory pages are copied. If the speed of modification of the memory pages on the source VM is higher than the available network bandwidth, then the process of copying the modified memory pages will never converge. Therefore, a threshold for the maximum number of pre-copying iterations is used. If it is reached, the VM stops, the remaining memory pages are copied to the destination VM, and then the machine resumes working on the new host. The migration algorithm shows that the number of iterative copy steps is random and depends mostly on the network bandwidth and the speed of memory page modification.

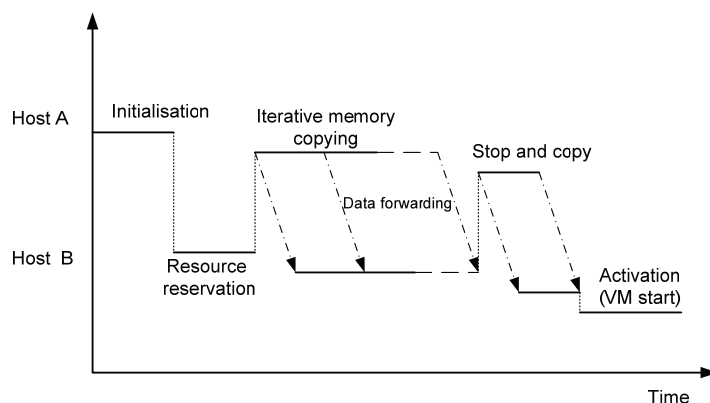


Figure 1. Stages of precopy migration

- Post-copy migration differs from pre-copy migration in that the VM on the destination host is almost immediately activated: first, the minimum necessary data is transmitted to start the virtual machine (processor registers, I / O device status, etc.), after which the virtual machine is immediately restarted on the destination server, while copying all memory pages occurs in the background. In [6], we consider four ways to extract memory pages from the source server. The easiest way is to copy on demand. Accessing pages that have not yet been copied triggers a memory error on the VM, processing which retrieves the desired page on demand. These page errors can lead to high VM CPU usage during the resume phase. Post-copy migration is attractive because it has a short and constant downtime and transfers each memory page only once. However, performance degradation during the resume phase limits its use in an environment with strict SLA agreements. The migration steps are shown in Fig. 2.

Since these two types of migration have their own advantages and disadvantages, many papers have recently appeared that consider hybrid migrations [7-10].

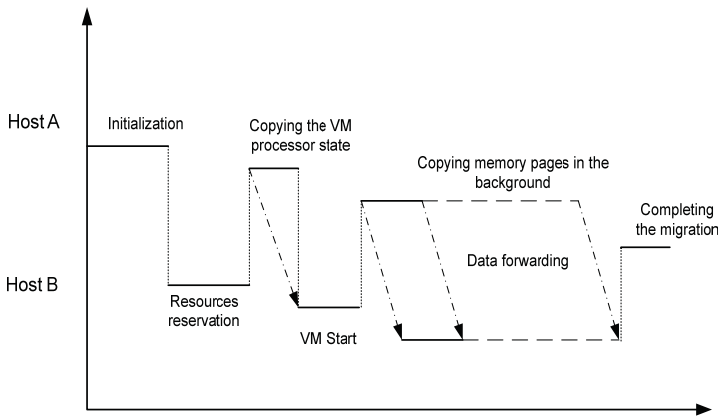


Figure 2. Stages of postcopy migration

To improve various characteristics of migration, such as total migration time, VM downtime, the required network bandwidth, in a number of works, it is proposed to use additional tools:

- Processor throttling (THR) [11] is a method of deliberate reducing the allocated CPU time to a VM to slow down the speed of memory page modification and thus ensure the convergence of the iterative page copying process.
- Delta compression of memory pages (DLTC) [12] – delta compression is applied to partially modified memory pages. This method may require a significant amount of additional memory to store intermediate results of calculations.
- Data Compression Optimization (DTC) [13] compresses pages before transmission to reduce the amount of data being transmitted. This method requires a significant amount of computing resources and thus may not be a viable option when the host CPU utilization is high.

3. Characteristics of virtual machine migration

The efficiency of live migration is estimated by several characteristics [6]. The main characteristics considered in this paper are:

- Total migration time: this is the total time of all migration stages. During the migration, the system performance decreases, which affects the quality of services provided and the process of monitoring resources.
- Downtime: this is the time when the VM is completely stopped. It is one of the main indicators in SLA agreements. At the very least, this is the time to transfer the processor state for a post-copy migration. For a pre-copy migration, this period also includes the transfer of all RAM and iterative copying of the modified memory pages.

4. Live VM migration dataset

4.1. Number of migrations

This paper analyzes a set of virtual machine migration data collected and provided by a research team at the National University of Seoul [14], who used it to predict migration characteristics using machine learning methods such as linear regression and the support vector machine (SVM). This dataset contains 40000 records of various types of migrations – PRE-copy migrations, POST-copy migrations, and modifications to pre-copy

migrations, such as processor throttling (THR), delta compression (DLTC), and data compression (DTC).

The number of migration records of different types in the dataset is shown in Figure 3.

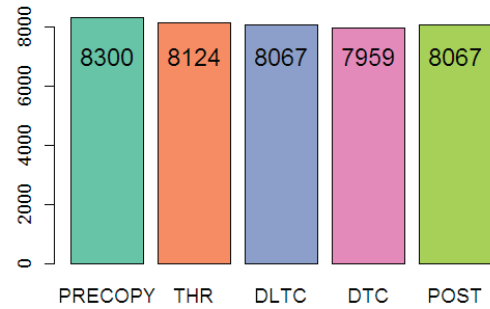


Figure 3. Number of migrations of different species in the dataset

4.2. Applications

The dataset was obtained as a result of benchmarking. The following test workloads were used:

- SPECweb to simulate working with a web server that hosts banking and e-commerce services;
- OLTPBench is an online transaction processing database application.
- Memcached – is a key-value cache store in memory,
- Dacapo is a collection of Java applications,
- PARSEC contains a set of multithreaded application workloads.
- Bzip is used as a compute-and data-intensive application;
- mplayer-simulates the operation of a multimedia player.
- synthetic-synthetic loads;
- idle–no load.

The number of migrations for different types of loads is shown in Figure 4.

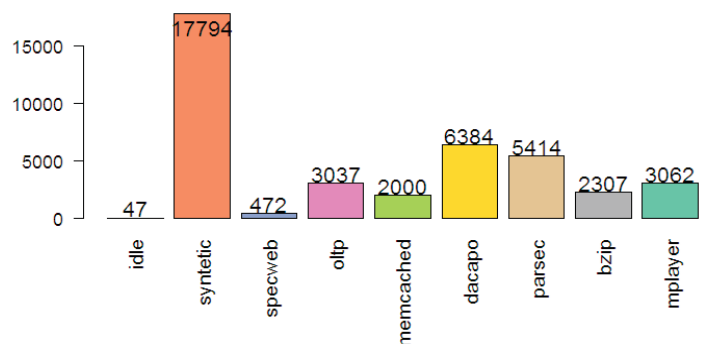


Figure 4. Number of migrations for different load types

4.3. Input features

The list of input features is given in Table 1. The first two columns describe the feature, and the third column shows where the factor value can be obtained from.

The features for the forecasting model were selected in such a way that they cover all the aspects that can affect the estimated characteristics. Since dynamic migration requires the transfer of virtual machine memory to the destination host over the network, the size of the allocated memory (VM_Size), and the available

network bandwidth for transmission (PTR) are important features for all models.

Table 1

Migration parameters

№	Metric	Calculation
1.	The amount of RAM in the VM (VM_size). The number of memory pages allocated in M.	Virtual Machine Monitor (VMM)
2.	Average number of modified memory pages per second (VM_pdr).	VMM
3.	The size of the working memory set is the number of memory pages changed during the VM initialization period (VM_wss).	VMM
4.	The entropy of the working memory set (VM_wse).	VMM
5.	The entropy of the non-working memory set (VM_nwse).	VMM
6.	Number of modified words on modified pages (VM_mwpp)	VMM
7.	Number of processor instructions per second (VM_pmu_instr)	Source host
8.	Network bandwidth reserved for live migration (VM_ptr)	Source host
9.	Virtual Machine Processor Load (VM_cpu_util)	Source host
10.	Network utilization on a VM (VM_net_util)	Source host
11.	CPU utilization on the source host (src_cpu_avail)	Source host
12.	Memory utilization on the source host (src_mem_avail)	Source host
13.	CPU utilization on the destination host (dst_cpu_avail)	Destination Host
14.	Memory utilization on the destination host (dst_mem_avail)	Destination Host

Pre-copy-based algorithms that iteratively copy dirtied memory pages to the destination host require knowledge of the memory page dirty rate (VM_pdr) and the working set size (VM_wss) of memory pages. There is a relationship between the page dirty rate (VM_pdr) and working set size (VM_wss): the first one represents the number of pages dirtied in one profiling period, while the second one represents all pages dirtied over the entire profiling period. For methods involving data compression, the entropies of the working set (VM_wse) and non-working set (VM_nwse), as well as the CPU and memory utilization of the source (src_cpu_avail src_mem_avail) and destination host (dst_cpu_avail, dst_mem_avail) are included. To evaluate the effectiveness of delta compression, the number of modified words on the page (VM_mwpp) is measured.

4.4. Composed features

In addition to input features given in Table 1 composed features were used, which were calculated on the basis of input features [14]. The list of composed features is given in Table 2.

Table 2

Composed features

№	Metric	Calculation
1.	Weighted relative page transfer rate (RPTR) — weighted relative page transfer rate to page dirty rate	IF $\frac{VM_wss}{VM_ptr} * ((\frac{VM_pdr}{VM_ptr}) < VM_wss)$ THEN $RPTR = \frac{VM_wss}{((\frac{VM_pdr}{VM_ptr}) * (VM_pdr / VM_ptr))}$ ELSE $RPTR = VM_wss$
2.	Non-working set size (VM_nwss) — number of not modified pages during profiling period	$VM_nwss = VM_size - VM_wss$
3.	Expected benefit of delta compression technique (DLTC_benefit)	$DLTC_benefit = VM_wss * VM_mwpp / (4096 / 2)$
4.	Expected benefit of CPU throttling technique (THR.BF)	$THR_benefit = VM_pdr * \min((VM_cpu_util / 400.0), 1.0)$
5.	Expected size of WSS after compression (VM_e_wss)	$VM_e_wss = VM_wss * VM_wse$
6.	Expected size of NWSS after compression (VM_e_nwss)	$VM_e_nwss = VM_nwss * VM_nwse$

5. Correlation analysis of features

Collecting such a large amount of raw data requires significant resources, so it is important to reduce the number of features in the models while maintaining the accuracy of the forecast. To find a solution to this problem, consider the relationship between input features and metrics. The correlation matrices of features for pre-copy migration are shown in Figure 4, 5. For the THR, DLTC, and DTC migration types, the matrices are similar in many ways to the pre-copy migration, so they are not given.

Visually, it can be determined that the considered features are not distributed normally. At the same time, there is a significant correlation of the output characteristics (qemu_tt and qemu_dt) with the features VM_wss, VM_pdr, VM_size, VM_nwss, VM_nwse, VM_ptr. In addition, such features as VM_wss and VM_pdr, as well as VM_size, VM_nwse and VM_wss are pairwise correlated, and the some features, such as VM_pmu_instr, src_cpu_avail src_mem_avail, dst_mem_avail do not affect the output features.

The correlation matrices of features for the precopy migration are shown in Figure 6, 7.

As it can be seen in Figure 7, post-copy migration has a short and constant downtime (with the exception of some outliers), since the downtime is only associated with the transmission of the current processor state over the network. At the same time, the highest correlation of the output characteristic qemu_tt (total migration time) is observed with the parameters VM_size, VM_wss, VM_nwse, VM_ptr and src_mem_avail. Such features as VM_ptr, VM_wse, VM_mvpp, VM_pmu_instr, VM_cpu_util, dst_cpu_avail, dst_mem_avail do not affect the output characteristic.

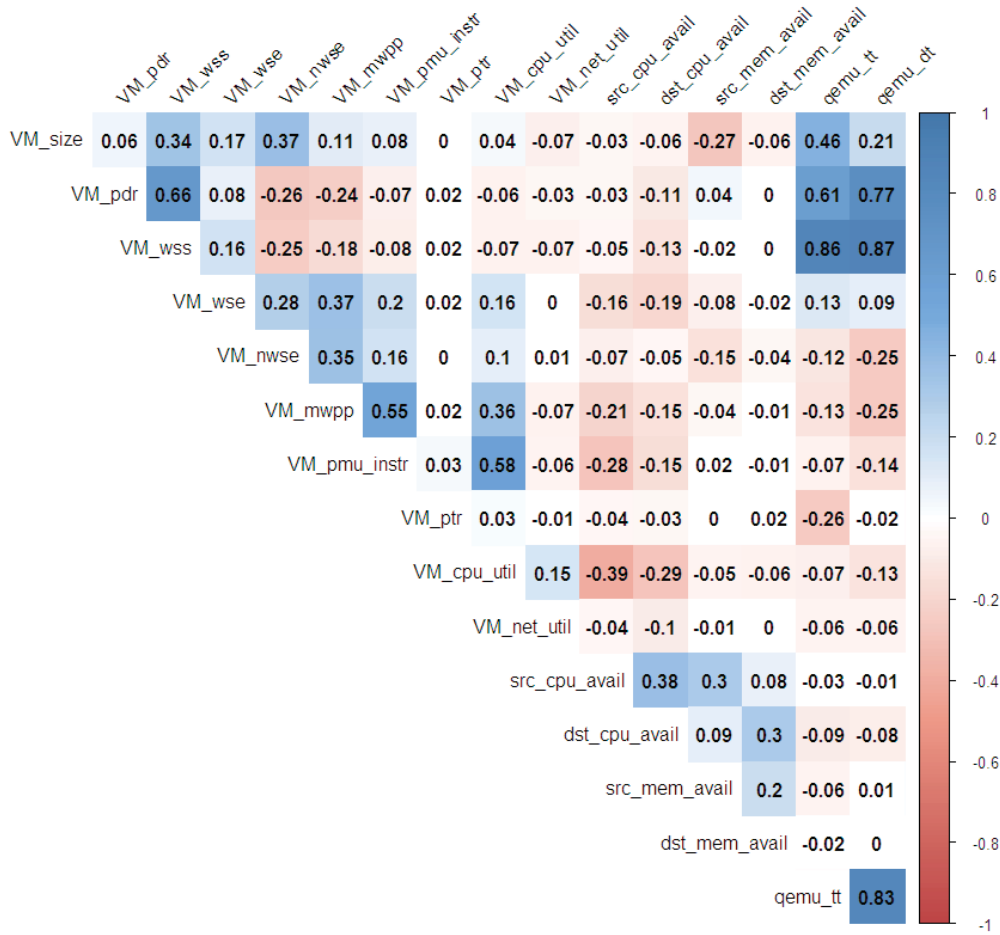


Figure 4. Correlation matrix of features for the precopy migration

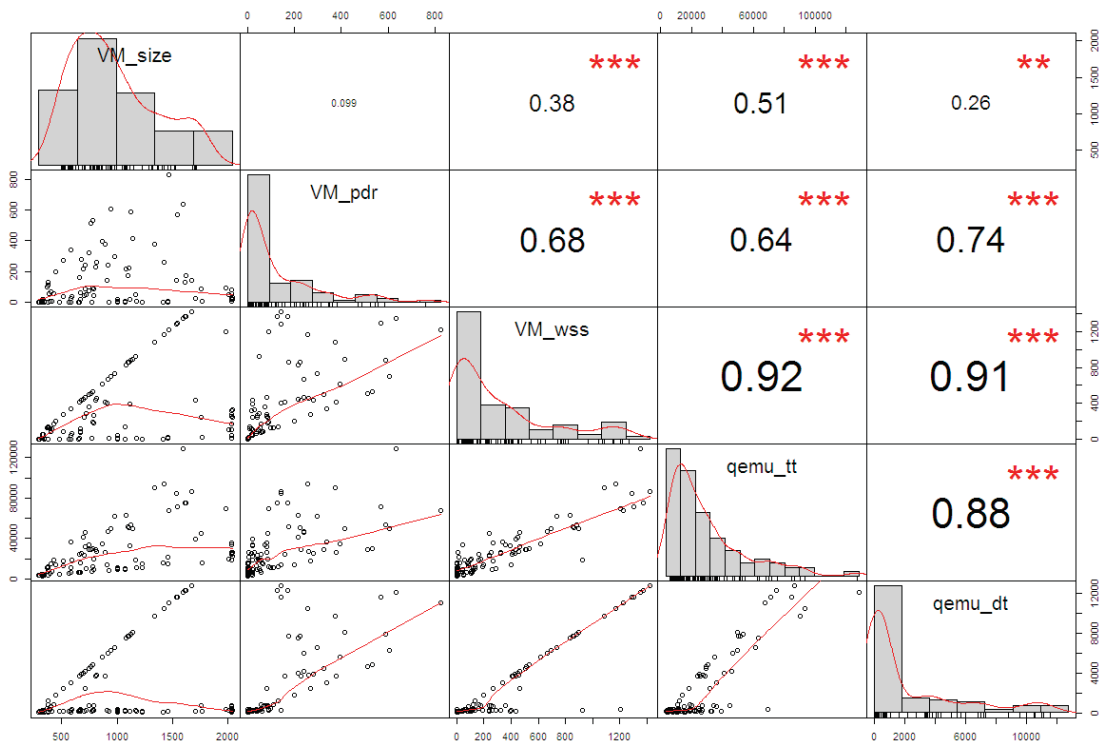


Figure 5. Pairwise correlations of features for the precopy migration for the first 100 records of the dataset

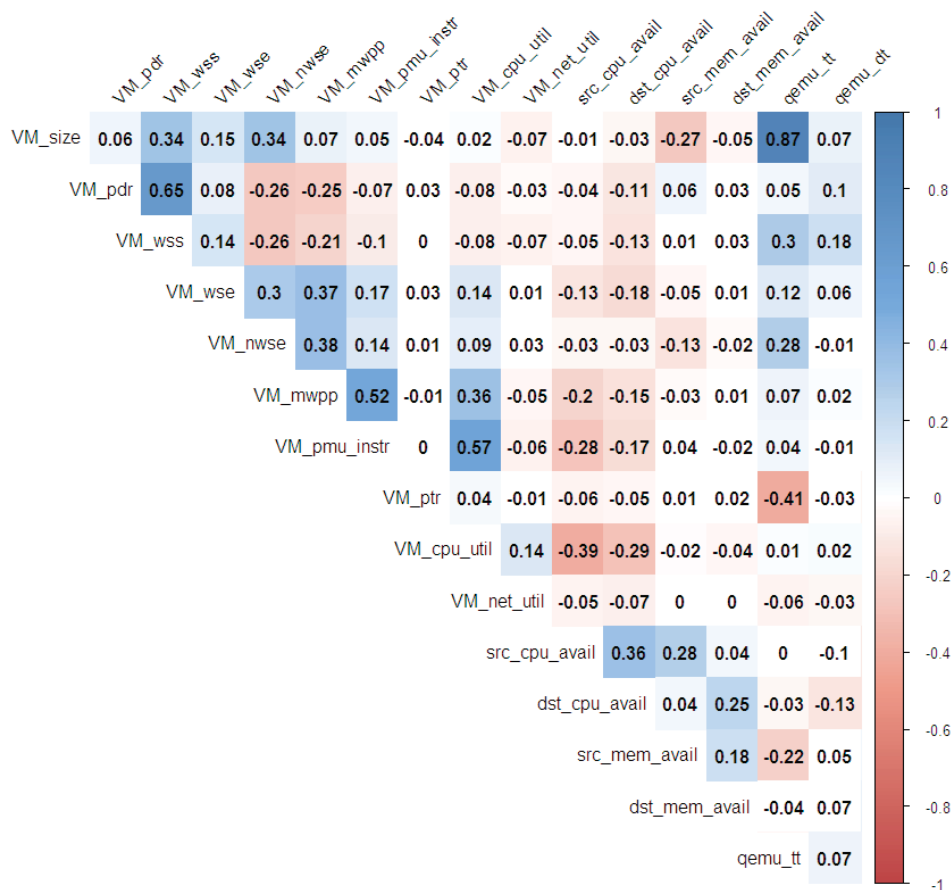


Figure 6. Correlation matrix of features for post-copy migration

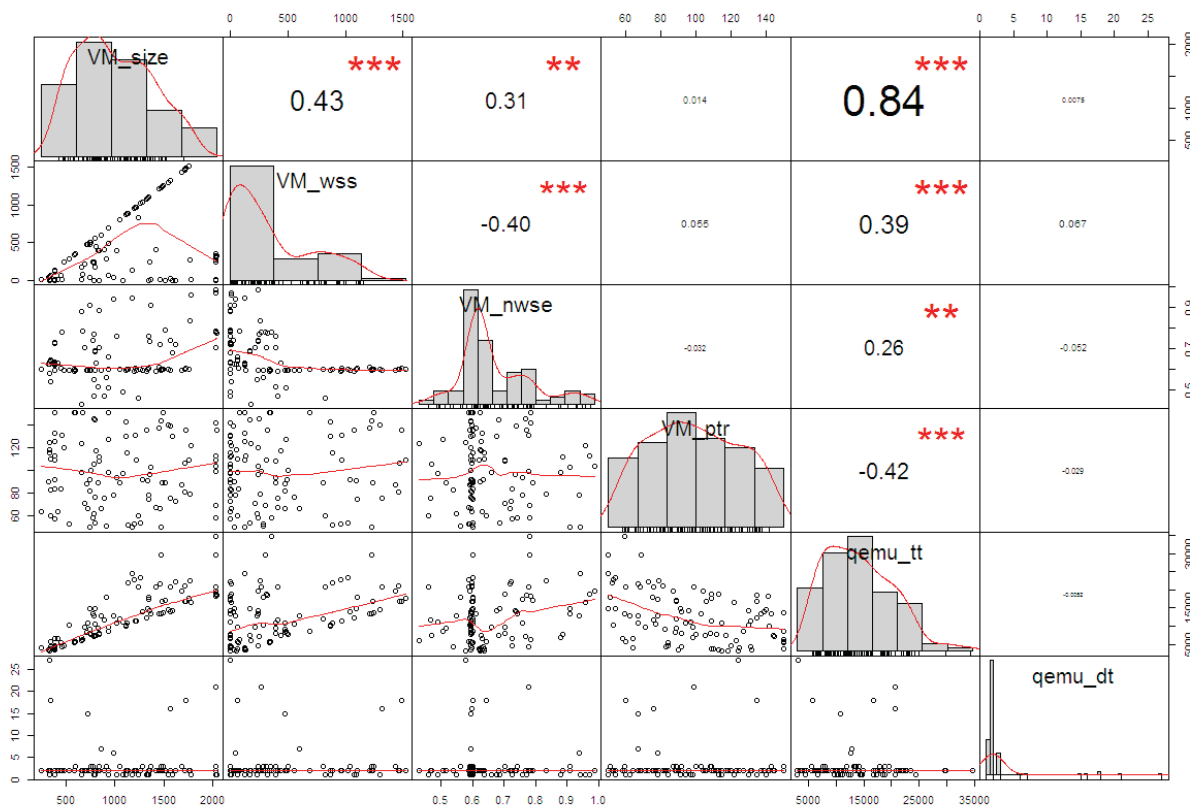


Figure 7. Pairwise correlations of features for post-copy migration (POST) for the first 100 records of the dataset

6. Results of experiments

It is known that the cross-correlation of input features has a bad effect on the quality of the forecast. Indeed, if we build a model only on these factors, without introducing composed factors, the quality of the obtained forecast by Linear Regression (LR) and the support vector regression (SVR) is low (Table 3.)

Table 3

The results of the forecast for 14 input features for precopy migration

Type of migration	Machine learning method	Predicted factor	MAE	M	100%-WAPE
PRE		TT	7260.371	0.486	30.63 %
		DT	120.4	1.998	37.06 %
	SVR	TT	3260.371	0.228	64.35 %
		DT	158.0	1.403	52.29 %

To compare the quality of forecasts obtained by various machine learning methods, the following indicators were used: Mean Absolute Error (MAE), Mean Relative Error (MRE). For clarity, another indicator of forecast accuracy was introduced in this paper, which was calculated as the inverse of the mean absolute percentage error (MAPE):

$$100\% - \frac{1}{N} \sum_{i=1}^N \frac{|A_i - F_i|}{A_i} \cdot 100\% ,$$

where A_i is the actual value of the output characteristic from the test dataset;

F_i is the forecast value;

N is the number of values in the test set.

By introducing composed features, the authors of [13] were able to significantly improve the accuracy of the forecast (Table 4).

Table 4

The results of the forecast for 20 input features

Type of migration	Machine learning method	Predicted factor	MAE	MRE	100%-WAPE
PRE		TT	3131.785	0.164	81.99 %
		DT	151.0	0.268	85.37 %
	SVR	TT	1213.210	0.064	92.27 %
		DT	128.0	0.227	89.33 %
THR		TT	3047.512	0.164	81.30 %
		DT	239.4	0.497	72.99 %
	SVR	TT	1154.471	0.062	92.22 %
		DT	137.6	0.285	86.56 %
DLTC		TT	2414.955	0.169	78.03 %
		DT	120.4	1.070	7.17 %
	SVR	TT	764.970	0.054	91.72 %
		DT	39.7	0.353	64.90 %
DTC		TT	7770.499	0.222	73.37 %
		DT	555.7	0.834	66.38 %
	SVR	TT	3903.717	0.112	86.42 %
		DT	333.5	0.501	84.04 %
POST	LR	TT	1207.952	0.121	85.46 %
		DT	1.1	0.520	28.06 %
	SVR	TT	288.219	0.029	96.11 %
		DT	0.6	0.289	44.70 %

A correlation matrices with composed features for precopy, delta compression, and post-copy migration types are shown in Figure 8.

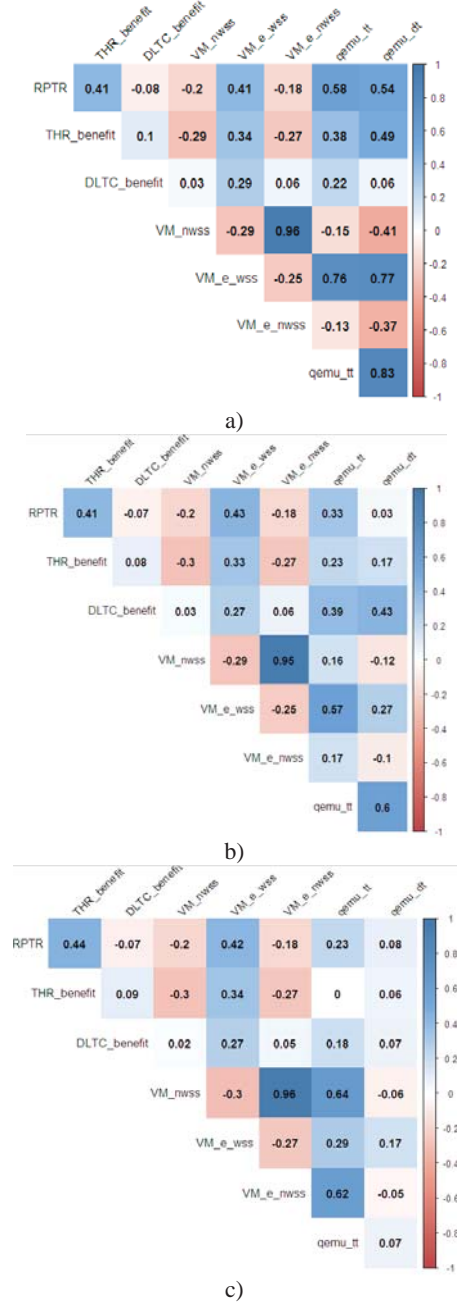


Figure 8. Correlation matrices of composed features with (a) pre-copy, (b) delta compression, and (c) post-copy migrations

The results shown in Table 5 were obtained by removing features uncorrelated with the output, such as VM_pmu_instr, VM_net_utils, src_cpu_avail, dst_cpu_avail, src_mem_avail, dst_mem_avail.

As can be seen from Tables 4 and 5, if we remove factors that are uncorrelated with the output features, the accuracy of the forecast practically does not change, except for the type of migration with data compression (DTC). Also, the SVR machine learning method shows more accurate results compared to linear regression.

Table 5

Forecast results with composed features

Migration type	Machine learning method	Predicted characteristic	MAE	MRE	100%-WAPE
PRE		TT	3131.785	0.164	81.36 %
		DT	151.0	0.268	79.53%
	SVR	TT	1213.210	0.064	92.70 %
		DT	128.0	0.227	88.68 %
THR		TT	7187.524	0.386	53.81 %
		DT	961.5	1.998	2.64 %
	SVR	TT	1205.449	0.065	92.09 %
		DT	177.1	0.367	77.51 %
DLTC		TT	3260.371	0.228	54.10 %
		DT	158.0	1.403	58.21 %
	SVR	TT	809.267	0.057	89.86 %
		DT	42.4	0.377	49.43 %
DTC		TT	16105.914	0.461	40.53 %
		DT	2151.1	3.231	56.3 %
	SVR	TT	5297.040	0.152	78.79 %
		DT	397.2	0.596	77.57 %
POST	LR	TT	1214.535	0.122	85.00 %
		DT	1.1	0.551	24.83 %
	SVR	TT	227.194	0.023	97.02 %
		DT	0.5	0.259	46.33 %

The most significant features are those that are involved in the formation of composed features that can significantly improve the accuracy of forecasting. This allows us to simplify the process of generating a dataset and simplify the model itself without significantly losing the accuracy of the forecast. Thus, for 5 different types of migration, the main selected features are listed in Table 6.

Table 6

Selected features for forecasting

Type of migration	Predicted factor	Features
PRE	TT	VM_size, VM_pdr, VM_wss, VM_ptr Composed Features RPTR, NWSS
	DT	VM_size, VM_pdr, VM_wss, VM_ptr Composed Features RPTR, NWSS
THR		VM_size, VM_pdr, VM_wss, VM_ptr, VM_mwpp Composed Features RPTR, NWSS, THR.BF
	DT	VM_size, VM_pdr, VM_wss, VM_ptr, VM_cpu_util Composed Features RPTR, NWSS, THR.BF
DLTC		VM_size, VM_pdr, VM_wss, VM_mwpp, VM_ptr Composed Features DLTC.BF
	DT	VM_size, VM_pdr, VM_wss, VM_mwpp, VM_ptr Composed Features DLTC.BF
DTC	TT	VM_size, VM_pdr, VM_wss, VM_wse, VM_nwse, VM_ptr Составные факторы: E.WSS, E.MWSS

	DT	VM_size, VM_pdr, VM_wss, VM_wse, VM_nwse, VM_ptr Composed Features VM_e_wss, E.MWSS
POST	TT	VM_size, VM_pdr, VM_wss, VM_ptr Composed Features RPTR, VM_e_nwss
	DT	-

Conclusions

1. The features that most significantly affect the total migration time and the VM downtime are selected. The correlation analysis made it possible to reduce the number of features by about half compared to the known work, which significantly reduces the complexity of collecting and processing the dataset without losing the quality of the forecast.

2. The introduction of composed features into the model made it possible to reduce the pair correlation of features and significantly improve the accuracy of the forecast.

3. The SVR method generally provides a more accurate forecast compared to linear regression and can be recommended for use in predicting the characteristics of live migration.

References

1. Cisco V. N. I. (2018) *Cisco visual networking index: Forecast and trends*, 2017–2022. White Paper, V. 1.
2. Mishra M. et al. (2012) *Dynamic resource management using virtual machine migrations*. IEEE Communications Magazine, V. 50, №. 9, pp. 34–40.
3. Tutov A.V., Tutova N.V., Vorozhtsov A.S. (2017). Modeling of resource allocation in cloud data centers. *T-Comm*, vol. 11, no.4, pp. 76–80.
4. Clark C. et al. (2005) *Live migration of virtual machines*. Proceedings of the 2nd conference on Symposium on Networked Systems Design & Implementation-Volume 2, pp. 273–286.
5. Hines M. R., Gopalan K. (2009) *Post-copy based live virtual machine migration using adaptive pre-paging and dynamic self-ballooning*. Proceedings of the 2009 ACM SIGPLAN/SIGOPS international conference on Virtual execution environments, pp. 51–60.
6. Aleksankov S. M. (2015) *Models of dynamic migration with an iterative approach and network migration of virtual machines*. Scientific and Technical Bulletin of Information Technologies, Mechanics and Optics. - 2015. - Vol. 15. - No. 6. (in Russian).
7. Lei Z. et al. (2017) *A novel hybrid-copy algorithm for live migration of virtual machine* Future Internet, T. 9, №. 3, p. 37.
8. Tikhomirov P. O. et al. (2014) *Minimizing the downtime of processes when migrating them to cloud hosting*. Bulletin of the Novosibirsk State University. Series: Information Technologies, Vol. 12, No. 4 (in Russian).
9. Sahni, S., & Varma, V. (2012, October). *A hybrid approach to live migration of virtual machines*. In 2012 IEEE International Conference on Cloud Computing in Emerging Markets (CCEM), pp. 1–5.
10. Liu Z. et al. (2010) *Xen live migration with slowdown scheduling algorithm* //2010 International Conference on Parallel and Distributed Computing, Applications and Technologies. IEEE, pp. 215–221.
11. Svård, P., Hudzia, B., Tordsson, J., & Elmroth, E. (2011). *Evaluation of delta compression techniques for efficient live migration of large virtual machines*. ACM Sigplan Notices, 46(7), pp. 111–120.
12. Jin, H., Deng, L., Wu, S., Shi, X., & Pan, X. (2009, August). *Live virtual machine migration with adaptive, memory compression*. In 2009 IEEE International Conference on Cluster Computing and Workshops, pp. 1–10.
13. Jo C., Cho Y., Egger B. (2017) *A machine learning approach to live migration modeling*. Proceedings of the 2017 Symposium on Cloud Computing, pp. 351–364.

ВЫБОР ФАКТОРОВ ДЛЯ ПРОГНОЗИРОВАНИЯ ХАРАКТЕРИСТИК ЖИВОЙ МИГРАЦИИ ВИРТУАЛЬНЫХ МАШИН

Киров Денис Евгеньевич, Московский технический университет связи и информатики, Москва, Россия,
denikirov@yandex.ru

Тугова Наталья Владимировна, Московский технический университет связи и информатики, Москва, Россия,
e-natasha@mail.ru

Ворожцов Анатолий Сергеевич, Московский технический университет связи и информатики, Москва, Россия,
as.vorocov@mail.ru

Андреев Илья Александрович, Московский технический университет связи и информатики, Москва, Россия,
lc@mtuci.ru

Аннотация

Миграция виртуальных машин широко используется в облачных центрах обработки данных для масштабирования и поддержания стабильности облачных сервисов. Однако показатели качества работы приложений виртуальной машины во время миграции, задаваемые в соглашениях об уровне сервиса, могут ухудшиться. Перед тем, как принять решение о запуске миграции, необходимо оценивать характеристики миграции, влияющие на качество работы приложений. Такими характеристиками являются общая длительность миграции и длительность простоя виртуальной машины, которые являются случайными величинами, зависящими от множества факторов. Прогнозирование основывается на данных мониторинга виртуальных машин. В работе выбираются наиболее подходящие факторы для прогноза по пяти видам миграций: миграции с предварительным копированием (precopy migration), миграции с пост копированием (postcopy migration) и модификации миграции с предварительным копированием, с использованием дросселирования процессора, сжатием данных и дельта-сжатием измененных страниц памяти. Для этого проанализирован набор данных, включающий в себя данные по пяти видам миграциям, примерно по 8000 записей каждого вида. С использованием корреляционного анализа выбраны факторы, влияющие на длительность миграции и длительность простоя виртуальной машины. Прогнозируются общая длительность миграции и длительность простоя виртуальной машины во время миграции с использованием таких методов машинного обучения, как линейная регрессия и метод опорных векторов. Показано, что число факторов можно сократить почти в два раза при том же качестве прогноза. В целом линейная регрессия дает относительно высокую точность прогноза общей длительности миграции и длительности простоя виртуальной машины. В то же время наблюдаемая нелинейность в корреляционных связях показывает, что для повышения качества прогноза целесообразно использовать метод опорных векторов.

Ключевые слова: виртуализация, корреляционный анализ, живая миграция, центры обработки данных, длительность миграции, метод опорных векторов, линейная регрессия.

Литература

1. Cisco V. N. I. Cisco visual networking index: Forecast and trends, 2017-2022. White Paper, V. 1. 2018.
2. Mishra M. et al. Dynamic resource management using virtual machine migrations // IEEE Communications Magazine, 2012. V. 50, №. 9, pp. 34-40.
3. Tutov A.V., Tutova N.V., Vorozhtsov A.S. Modeling of resource allocation in cloud data centers. T-Comm, 2017, vol. 11, no.4, pp. 76-80.
4. Clark C. et al. Live migration of virtual machines // Proceedings of the 2nd conference on Symposium on Networked Systems Design & Implementation-2005. Vol. 2, pp. 273-286.
5. Hines M. R., Gopalan K. Post-copy based live virtual machine migration using adaptive pre-paging and dynamic self-ballooning. Proceedings of the 2009 ACM SIGPLAN/SIGOPS international conference on Virtual execution environments, 2009, pp. 51-60.
6. Алексанков С.М. Модели динамической миграции с итеративным подходом и сетевой миграции виртуальных машин // Научно-технический вестник информационных технологий, механики и оптики. 2015. Т. 15. №. 6.
7. Lei Z. et al. A novel hybrid-copy algorithm for live migration of virtual machine Future Internet, 2017. Т. 9, №. 3. С. 37.
8. Тихомиров П.О. и др. Минимизация времени простоя процессов при их миграции в облачном хостинге // Вестник Новосибирского государственного университета. Серия: Информационные технологии. 2014. Т. 12. №. 4.
9. Sahni, S., & Varma, V. (2012, October). A hybrid approach to live migration of virtual machines // In 2012 IEEE International Conference on Cloud Computing in Emerging Markets (CCEM), pp. 1-5.
10. Liu Z. et al. Xen live migration with slowdown scheduling algorithm //2010 International Conference on Parallel and Distributed Computing, Applications and Technologies. IEEE, 2010, pp. 215-221.
11. Sverd, P., Hudzia, B., Tordsson, J., & Elmroth, E. Evaluation of delta compression techniques for efficient live migration of large virtual machines. ACM Sigplan Notices, 46(7), 2011, pp. 111-120.
12. Jin, H., Deng, L., Wu, S., Shi, X., & Pan, X. (2009, August). Live virtual machine migration with adaptive, memory compression. In 2009 IEEE International Conference on Cluster Computing and Workshops, pp. 1-10.
13. Jo C., Cho Y., Egger B. A machine learning approach to live migration modeling. Proceedings of the 2017 Symposium on Cloud Computing, 2017, pp. 351-364.

Информация об авторах:

Киров Денис Евгеньевич, Московский технический университет связи и информатики, аспирант, Москва, Россия

Тугова Наталья Владимировна, Московский технический университет связи и информатики, доцент кафедры "Корпоративные информационные системы", доцент, к.т.н., Москва, Россия

Ворожцов Анатолий Сергеевич, Московский технический университет связи и информатики, доцент кафедры "Корпоративные информационные системы", доцент, к.т.н., Москва, Россия

Андреев Илья Александрович, Московский технический университет связи и информатики, и.о. зав. кафедрой "Корпоративные информационные системы", к.э.н., Москва, Россия