

Простые динамические модели предсказания времени работы вариационных автокодировщиков

А. А. Березкин

Санкт-Петербургский государственный
университет телекоммуникаций
им. проф. М. А. Бонч-Бруевича

berezkin.aa@sut.ru

А. А. Ченский

Санкт-Петербургский государственный
университет телекоммуникаций
им. проф. М. А. Бонч-Бруевича

chenskii.aa@sut.ru

Аннотация. Одним из решений проблемы передачи видеопотока по сетям с низкой пропускной способностью являются нейросетевые кодеки. Нейросетевые кодеки используют для кодирования и декодирования кадров видеопотока ряд методов, к которым относятся нейронные сети вида вариационный автокодировщик. Большое время работы вариационных автокодировщиков приводит к пропуску части кадров видеопотока. Для разработки методов стабилизации интенсивности видеопотока требуется предсказывать время работы вариационного автокодировщика. В настоящем исследовании рассмотрены простые динамические модели предсказания как времени кодирования, так и времени декодирования вариационным автокодировщиком модели KL-f4.

Ключевые слова: вариационный автокодировщик; предсказание; временной ряд; время выполнения; кодирование; декодирование; нейросетевой кодек; предсказательная модель

I. ВВЕДЕНИЕ

При управлении беспилотными воздушными судами (БВС) от первого лица (FPV-управление – first person view) возникает проблема передачи видеопотока между БВС и станцией внешнего пилота (СВП). В отдалённых районах Российской Федерации сети с высокой пропускной способностью, такие как мобильные сети 3G и 4G, могут быть не доступны. Соответственно, для FPV-управления БВС потребуется использовать альтернативные сети. Полоса пропускания самоорганизующихся ZigBee сетей, спутниковых сетей и Infrared-сетей (сетей на инфракрасном диапазоне) в отдельных случаях составляет сотни Кбит/с или единицы Мбит/с [1]. В то же время для обеспечения передачи видео с высоким качеством стандартными средствами требуется полоса пропускания в несколько Мбит/с [2].

Одним из решений проблемы передачи видеопотока по сетям с низкой полосой пропускания является использование нейросетевых кодеков [3–4]. Они содержат нейросетевые кодеры, используемые для кодирования кадров, и нейросетевые декодеры, используемые для декодирования кадров. Нейросетевые кодеки используют ряд техник: интерполяцию кадров видеопотока; кодирование и декодирование кадров видеопотока вариационными автокодировщиками; квантование и деквантование тензоров ЛПП.

Научная статья подготовлена в рамках прикладных научных исследований СПбГУТ, регистрационный номер 1024032900105-1-2.2.4; 2.2.5; 2.2.6 в ЕГИСУ НИОКТР.

получаемых в результате кодирования кадров видеопотока; дальнейшее сжатие тензоров ЛПП. Тем не менее, вариационные автокодировщики вызывают проблему пропусков кадров.

Пусть интенсивность видеопотока на СВП равна FPS кадров в секунду (к/с). Тогда восстановленные кадры видеопотока должны отображаться внешнему пилоту (ВП) на экране СВП с периодом

$$T = \frac{1000}{FPS} (\text{мс}). \quad (1)$$

Таким образом, при $FPS = 10$ к/с необходимо поступление кадров каждые 100 мс, при $FPS = 25$ к/с – каждые 40 мс, при $FPS = 30$ к/с – 33,33 мс, а при $FPS = 60$ к/с уже 16,67 мс.

В свою очередь среднее время кодирования t_{CD} вариационными автокодировщиками семейств KL и VQ даже при размещении их на видеокarte A100 [5], обладающей очень высокой производительностью, лежит в интервале [16,96; 131,29] мс, а среднее время декодирования t_{DCD} – в интервале [26,93; 55,63] мс.

За счёт времени выполнения прочих операций суммарное время работы нейросетевого кодера больше, чем время кодирования t_{CD} , а суммарное время работы нейросетевого декодера больше, чем время декодирования t_{DCD} . В случае, когда

$$t_{CD} > T, \quad (2)$$

нейросетевой кодер не сможет отправлять кадры видеопотока на декодер с заданной интенсивностью. В другом случае, когда

$$t_{DCD} > T, \quad (3)$$

нейросетевой декодер не сможет обработать все кадры видеопотока заданной интенсивности. В результате интенсивность отображаемого ВП на экране СВП будет ниже требуемой.

Возникает проблема стабилизации и увеличения интенсивности видеопотока. Есть несколько способов её решения.

- Использование более легковесной модели автокодировщика (не обязательно вариационного). В этом случае возможно добиться уменьшения t_{CD} и t_{DCD} . Тем не менее, в этом случае ухудшится качество

восстанавливаемых кадров. К этому способу относится дистилляция [6] моделей.

- Оптимизация используемой модели вариационного автокодировщика. При этом качество восстанавливаемых кадров не изменится существенно, но и уменьшение t_{CD} и t_{DCD} будет сравнительно небольшим. К способам оптимизации моделей относятся прунинг [7] и квантование [6] моделей.
- Восполнение утраченных кадров механизмами компенсации падения интенсивности видеопотока. На нейросетевом декодере составляется буфер кадров для отображения, куда добавляются декодированные кадры. Пропущенные (ввиду невозможности декодирования в срок) кадры видеопотока интерполируются либо предсказываются. В этом случае за счёт увеличения задержки между временем поступления на нейросетевой кодер исходного кадра и временем отображения кадра на СВП стабилизируется интенсивность видеопотока.

При использовании компенсаторных механизмов нужно составлять буфер кадров и оценить дополнительную задержку перед отображением кадров из буфера. Для оценки дополнительной задержки перед отображением необходимо знать: время работы нейросетевого кодера, время работы нейросетевого декодера, процент пропущенных кадров, время интерполяции и предсказания кадров. Время работы нейросетевых кодера и декодера в значительной степени определяется временем кодирования t_{CD} и декодирования t_{DCD} используемой модели вариационного автокодировщика.

Тем не менее, t_{CD} и t_{DCD} не являются постоянными величинами. Они меняются в зависимости от оборудования, на котором размещена используемая модель вариационного автокодировщика, и загрузки оборудования в момент выполнения модели. Таким образом, появляется задача предсказания времени кодирования t_{CD} и декодирования t_{DCD} для некоторой модели вариационного автокодировщика на некотором оборудовании.

Целью настоящего исследования является нахождение лучших динамических моделей предсказания значений t_{CD} и t_{DCD} .

II. ПРОСТЫЕ ДИНАМИЧЕСКИЕ МОДЕЛИ ПРЕДСКАЗАНИЯ ВРЕМЕННЫХ РЯДОВ

A. Модель простого среднего

В настоящем исследовании рассматриваются простейшие предсказательные модели временных рядов применительно к задаче предсказания значений t_{CD} и t_{DCD} . Рассматриваются модели: простого среднего, простого скользящего среднего и экспоненциального сглаживания. Оригинально данные модели используются для сглаживания временных рядов, но в данной работе они применяются для предсказания последующих значений временных рядов t_{CD} и t_{DCD} .

Модель простого среднего заключается в предположении, что следующие значения временного

ряда $y(t)$ будут приблизительно равны его среднему значению:

$$\hat{y}_{t+1} = \frac{\sum_{i=0}^{n-1} y_i}{n} \quad [8], \quad (4)$$

где $\hat{y}_{t+1}$ – предсказываемое следующее значение временного ряда, n – количество членов ряда, y_i – значение члена временного ряда с индексом i .

При поступлении каждого нового кадра на нейросетевой декодер требуется рассчитывать новое среднее значение ряда

$$\bar{y} = \frac{\sum_{i=0}^{n-1} y_i}{n}. \quad (5)$$

Это неэффективно с вычислительной точки зрения и требует хранения в памяти буфера значений t_{CD} и t_{DCD} за всё время работы нейросетевого кодера. Так после часа работы нейросетевого декодера и стабильном получении видеопотока интенсивности 30 FPS потребуется хранить до 108 000 значений t_{DCD} . Это займёт около 211 Кб (при использовании структуры данных float16 для хранения одного значения) или 422 (при float32) Кб оперативной памяти.

Во избежание излишних трат памяти и вычислительных ресурсов рекомендуется пересчитывать среднее значение ряда $\bar{y}$ при поступлении каждого нового кадра. Пусть нейросетевым декодером был получен $m + 1$ по счёту кадр, и время его декодирования составило $t_{DCD}(m+1)$ мс. Перед этим кадром было получено всего m кадров со средним временем декодирования $\overline{t_{DCD}}(m)$. Тогда новое среднее значение времени декодирования вычисляется по рекуррентному соотношению:

$$\overline{t_{DCD}}(m+1) = \frac{\overline{t_{DCD}}(m) \times m + t_{DCD}(m+1)}{(m+1)} (\text{мс}). \quad (6)$$

При расчёте среднего значения t_{DCD} (и по аналогии t_{CD}) по указанному соотношению достаточно хранить актуальное среднее значение t_{DCD} и счётчик кадров, поступивших на нейросетевой декодер.

B. Модель простого скользящего среднего

Модель простого скользящего среднего предполагает, что следующее значение временного ряда будет приблизительно равно среднему значению q последних членов ряда:

$$\hat{y}_{t+1} = \frac{\sum_{i=0}^{q-1} y_{t-i}}{q} \quad [8]. \quad (7)$$

Следует отличать модель простого скользящего среднего и модель скользящего среднего $MA(q)$ из семейства моделей ARIMA [9]. Данная модель основывается на теореме Вольда: “всякий слабо стационарный временной ряд может быть представлен в виде линейной комбинации белых шумов, с разными весовыми коэффициентами” [10, с. 181]. Значение временного ряда модели $MA(q)$ представляет собой

линейную комбинацию белых шумов с различными весовыми коэффициентами [10, с. 182]. В текущем исследовании данная модель не рассматривается.

Однако использование формулы (6) для нахождения средних значений t_{CD} и t_{DCD} означает, что при кодировании или поступлении каждого кадра потребуется обрабатывать все q значений буфера времени кодирования или декодирования. Это неэффективно с точки зрения трат вычислительных ресурсов. В действительности необходимо обрабатывать только одно, самое последнее и удаляемое значение из буфера.

Пусть нейросетевым декодером был получен $m+1$ кадр видеопотока, а время его декодирования составило $t_{DCD}(m+1)$ мс. До обновления буфера в нём содержались значения из множества $\{t_{DCD}(m), t_{DCD}(m-1), \dots, t_{DCD}(m-q+1)\}$, а среднее значение времени декодирования было равно $\overline{t_{DCD}(m)}$. Тогда новое среднее значение времени декодирования $\overline{t_{DCD}(m+1)}$ находится по формуле

$$\begin{aligned} \overline{t_{DCD}(m+1)} &= \frac{\overline{t_{DCD}(m)} \times q - t_{DCD}(m-q+1) + t_{DCD}(m+1)}{q} = \\ &= \overline{t_{DCD}(m)} + \frac{t_{DCD}(m+1) - t_{DCD}(m-q+1)}{q} \text{ (мс)} \end{aligned} \quad (8)$$

После обновления буфера в нём содержатся значения из множества $\{t_{DCD}(m+1), t_{DCD}(m), \dots, t_{DCD}(m-q+2)\}$.

С. Модель экспоненциального сглаживания

Модель экспоненциального сглаживания предполагает предсказание новых значений с учётом коррекции предыдущего прогноза:

$$\hat{y}_{t+1} = \hat{y}_t + a \times (y_t - \hat{y}_t) = a \times y_t + (1-a) \times \hat{y}_t \quad [8], \quad (9)$$

где $\hat{y}_t$ – предсказываемое текущее значение временного ряда, y_t – реальное текущее значение временного ряда, $a \in [0;1]$ – постоянная сглаживания.

III. МЕТОДИКА ЭКСПЕРИМЕНТОВ

В исследовании используется авторский набор данных, состоящий из 2-х временных рядов: ряд времени кодирования (рис. 1) и ряд времени декодирования (рис. 2).

Данные временные ряды содержат значения t_{CD} и t_{DCD} . Они были измерены во время обработки нейросетевым декодером из исследований [3–4] кадров FPV-управления БВС. Использовалась модель вариационного автокодировщика VQ-f16 из Stable Diffusion [11], размещённая на видеокарте A100 [5].

В рамках экспериментов испытываются прогностические модели:

- Модель простого среднего.
- Модель простого скользящего среднего с $q \in [1;100]$.

- Модель экспоненциального сглаживания с $a \in \{0,01; 0,02; \dots; 1,00\}$.

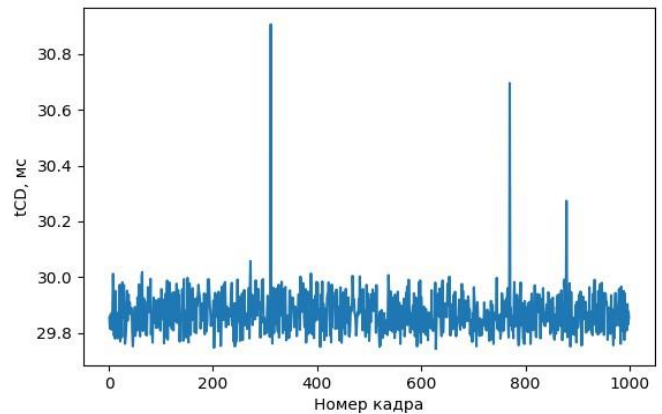


Рис. 1. Тренировочный ряд времени кодирования

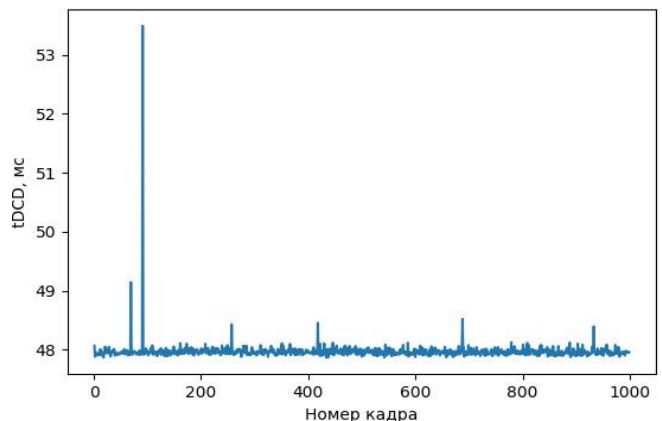


Рис. 2. Тренировочный ряд времени декодирования

Качество предсказания моделями оценивается по метрике MSE (mean squared error – средний квадрат ошибки):

$$MSE = \frac{1}{L} \sum_i \Delta_i^2 = \frac{1}{L} \sum_i (x_i - \hat{x}_i)^2 \quad [12], \quad (10)$$

где x_i – реальные значения, $\hat{x}_i$ – предсказанные значения, L – количество значений.

IV. ЭКСПЕРИМЕНТЫ

Эксперименты показали, что при предсказании времени кодирования и декодирования моделями простых скользящих средних при увеличении q значение MSE падает (рис. 3–4).

Модель простого среднего (рис. 5) показала $MSE = 0,0056$ для временного ряда кодера и $MSE = 0,0346$ для временного ряда декодера.

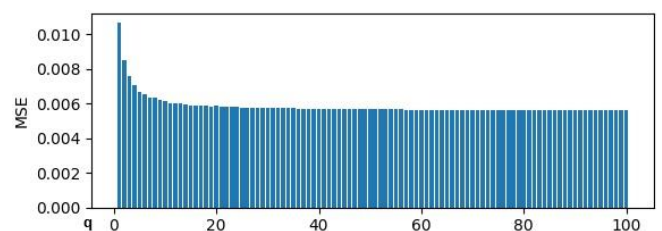


Рис. 3. Зависимость MSE от q для моделей простого скользящего среднего для временного ряда времени кодирования

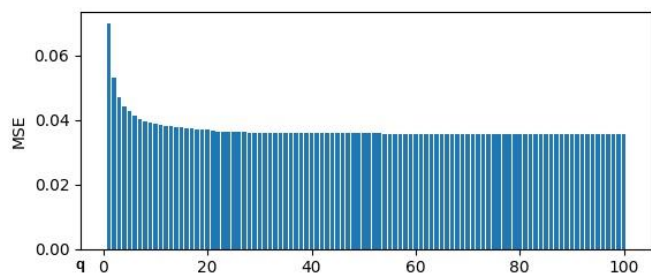


Рис. 4. Зависимость MSE от q для моделей простого скользящего среднего для временного ряда времени декодирования

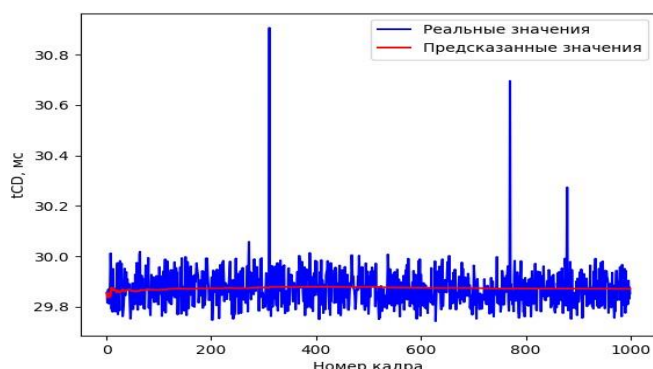


Рис. 5. Предсказание моделью простого среднего для временного ряда времени кодирования

Эксперименты над моделями экспоненциального сглаживания, что при росте a наблюдается сперва понижение (до $a = 0,01$), а затем повышение значений MSE (рис. 6–7).

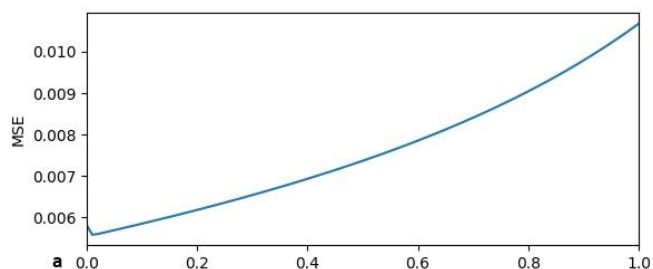


Рис. 6. Зависимость MSE от q для моделей экспоненциального сглаживания для временного ряда времени декодирования

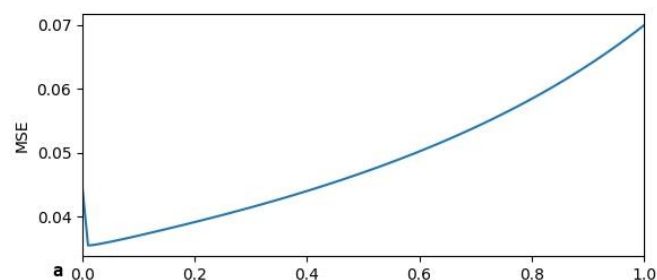


Рис. 7. Зависимость MSE от q для моделей экспоненциального сглаживания для временного ряда времени декодирования

ТАБЛИЦА I. ЛУЧШИЕ ПРЕДСКАЗАТЕЛЬНЫЕ МОДЕЛИ ДЛЯ ВРЕМЕНИ КОДИРОВАНИЯ

Модель	Параметр	MSE
Экспоненциальное сглаживание	$a = 0,01$	0,00558
Модель простой скользящей средней	$q = 100$	0,00559
Модель простой средней	-	0,00560

ТАБЛИЦА II. ЛУЧШИЕ ПРЕДСКАЗАТЕЛЬНЫЕ МОДЕЛИ ДЛЯ ВРЕМЕНИ КОДИРОВАНИЯ

Модель	Параметр	MSE
Модель простой средней	-	0,03464
Модель простой скользящей средней	$q = 100$	0,03543
Экспоненциальное сглаживание	$a = 0,01$	0,03553

V. ЗАКЛЮЧЕНИЕ

В данном исследовании было исследовано применение простых динамических предсказательных моделей для предсказания среднего времени кодирования и декодирования вариационным автокодировщиками KL-f4.

Как показали эксперименты, лучшими моделями являются: модель простой средней, модель скользящей средней при $q = 100$ (и более) и модель экспоненциального сглаживания с $a = 0,01$ (табл. 1–2). Модели дают близкие результаты и в одних случаях (время кодирования) лучше себя показывает экспоненциальное сглаживание, а в других – модель простой средней.

Это вызвано тем, что экспоненциальное сглаживание хуже реагирует на резкие выбросы. В связи с этим рекомендуется использовать либо модель простой скользящей средней, либо модель простой средней.

СПИСОК ЛИТЕРАТУРЫ

- [1] Anwer M.S., Guy C. A survey of VANET technologies // Journal of Emerging Trends in Computing and Information Sciences. 2014. Т. 5. №. 9. С. 661-671.
- [2] Evensen K., Kaspar D., Griwodz C., Halvorsen P., Hansen A.F., Engelstad P. Using bandwidth aggregation to improve the performance of quality-adaptive streaming // Signal Processing: Image Communication. 2012. Т. 27. №. 4.
- [3] Березкин А.А., Ченский А.А., Киричек Р.В., Захаров А.А. Исследование конфигураций нейросетевых кодеков для адаптивной системы сжатия кадров FPV-видеопотока при управлении беспилотными системами. Часть I. Методика // Электросвязь. Санкт-Петербург, 2024. № 9. С. 42-51.
- [4] Березкин А.А., Ченский А.А., Киричек Р.В., Захаров А.А. Исследование конфигураций нейросетевых кодеков для адаптивной системы сжатия кадров FPV-видеопотока при управлении беспилотными системами. Часть II. Эксперимент // Электросвязь. Санкт-Петербург, 2024. № 10. С. 59-69.
- [5] Choquette J., Gandhi W., Giroux O., Stam N., Krashinsky R. Nvidia A100 tensor core GPU: Performance and innovation // IEEE Micro. 2021. Т. 41. №. 2. С. 29-35.
- [6] Polino A., Pascanu R., Alistarh D. Model compression via distillation and quantization // arXiv preprint arXiv:1802.05668. 2018.
- [7] Zhu M., Gupta S. To prune, or not to prune: exploring the efficacy of pruning for model compression // arXiv preprint arXiv:1710.01878. 2017.
- [8] Шевченко И.В. Некоторые модели анализа и прогнозирования временных рядов // Системная информатика. 2013. №. 2. С. 23-40.
- [9] Shumway R.H., Stoffer D.S., Shumway, R.H., & Stoffer, D. S. ARIMA models. Time series analysis and its applications: with R examples. 2017.
- [10] Бардасов С.А. Эконометрика: учебное пособие. 2-е издание, переработанное и дополненное. Тюмень: Изд-во Тюменского государственного университета, 2010. 264 с.
- [11] Stable Diffusion // Github. URL: <https://github.com/pesser/stable-diffusion> (дата обращения: 11.12.2024).
- [12] Saxena A., Celaya J., Balaban E., Goebel K., Saha B., Saha S., Schwabacher M. Metrics for evaluating performance of prognostic techniques // 2008 international conference on prognostics and health management. IEEE, 2008. С.1-17.