

УДК 004.032.26

**РЕКУРРЕНТНЫЕ НЕЙРОННЫЕ СЕТИ С УПРАВЛЯЕМЫМИ
РЕКУРРЕНТНЫМИ БЛОКАМИ ДЛЯ МНОГОМЕРНЫХ ВРЕМЕННЫХ
РЯДОВ С ПРОПУЩЕННЫМИ ЗНАЧЕНИЯМИ****Тронов Кирилл Александрович**студент КФ МГТУ им. Н.Э. Баумана,
248000, РФ, г. Калуга, ул. Баженова, дом № 2
e-mail: kirtron@mail.ru**Федоров Виктор Олегович**кандидат технических наук, доцент КФ МГТУ им. Н.Э. Баумана,
248000, РФ, г. Калуга, ул. Баженова, дом № 2
e-mail: fedorov_vo@bmstu.ru**Аннотация**

Многомерные данные временных рядов в практических приложениях, таких как здравоохранение, геология и биология, характеризуются множеством пропущенных значений. В прогнозировании временных рядов и других сопутствующих задачах было отмечено, что пропущенные значения и их пропущенные шаблоны часто коррелируют с целевыми метками, также известными как информативное отсутствие. Довольно мало работ посвящено применению отсутствующих шаблонов для эффективного вменения и повышения эффективности прогнозирования. В данной статье предлагается новая модель глубокого обучения, а именно GRU-D. Данная модель основана на Gated Recurrent Unit (GRU), современной рекуррентной нейронной сети и использует два представления отсутствующих паттернов, т. е. маскирование и временной интервал, и эффективно включает их в глубокую архитектуру модели, чтобы не только фиксировать долгосрочные временные зависимости во временных рядах, но также использовать отсутствующие паттерны для достижения лучших результатов прогнозирования.

Ключевые слова: рекуррентные нейронные сети, глубокое обучение, затухающий градиент, GRU, многомерные временные ряды.

**RECURRENT NEURAL NETWORKS WITH GATED RECURRENT UNITS
FOR MULTIVARIATE TIME SERIES WITH MISSING VALUES****Kirill A. Tronov**Student of Bauman Moscow State Technical University (Kaluga Branch),
248000, Russian Federation, Kaluga, Bazhenova st., 2.
kirtron@mail.ru

Viktor O. Fedorov

Candidate of Technical Sciences, Docent of Bauman Moscow State Technical University (Kaluga Branch),

248000, Russian Federation, Kaluga, Bazhenova st., 2.

fedorov_vo@bmstu.ru

ABSTRACT

Multivariate time series data in practical applications such as health care, geology and biology are characterized by many missing values. In time series forecasting and other related tasks, it has been observed that missing values and their missing patterns often correlate with target labels, also known as informative absence. Quite a few works are devoted to the application of missing templates for effective imputation and improving the efficiency of forecasting. This article proposes a new deep learning model, namely GRU-D. This model is based on the Gated Recurrent Unit (GRU), a modern recurrent neural network, and uses two representations of missing patterns, i.e. masking and time interval, and effectively incorporates them into the deep architecture of the model to not only capture long-term time dependencies in time series, but also use missing patterns to achieve better prediction results.

Keywords: recurrent neural networks, deep learning, decaying gradient, GRU, multivariate time series.

Введение

Данные многомерных временных рядов повсеместно используются во многих практических приложениях, начиная от здравоохранения, наук о Земле, астрономии, заканчивая биологией и другими. Они часто неизбежно содержат пропущенные наблюдения по разным причинам, таким как медицинские события, экономия средств, аномалии, неудобства и так далее. Было отмечено, что эти пропущенные значения обычно являются информативными пропущенными, т. е. пропущенные значения и шаблоны предоставляют обширную информацию о целевых метках в контролируемых задачах обучения (например, классификация временных рядов) [1]. Чтобы проиллюстрировать эту идею, покажем несколько примеров из реального набора данных MIMIC-III-2 о здравоохранении, на рисунке 1. построен коэффициент корреляции Пирсона между показателями пропуска переменных, который показывает, как часто переменная отсутствует во временном ряду, и ярлыки, которые представляют интерес, а именно категории смертности и диагнозы по ICD-9. Видно, что значение частоты пропуска коррелирует с метками, а частота пропуска переменных с низким уровнем пропуска обычно сильно (как положительно, так и отрицательно) коррелирует с метками. Другими словами, недостающая частота переменных для каждого пациента полезна, и эта информация более полезна для переменных, которые чаще наблюдаются в наборе данных. Эти результаты демонстрируют полезность паттернов пропуска при решении задачи прогнозирования.

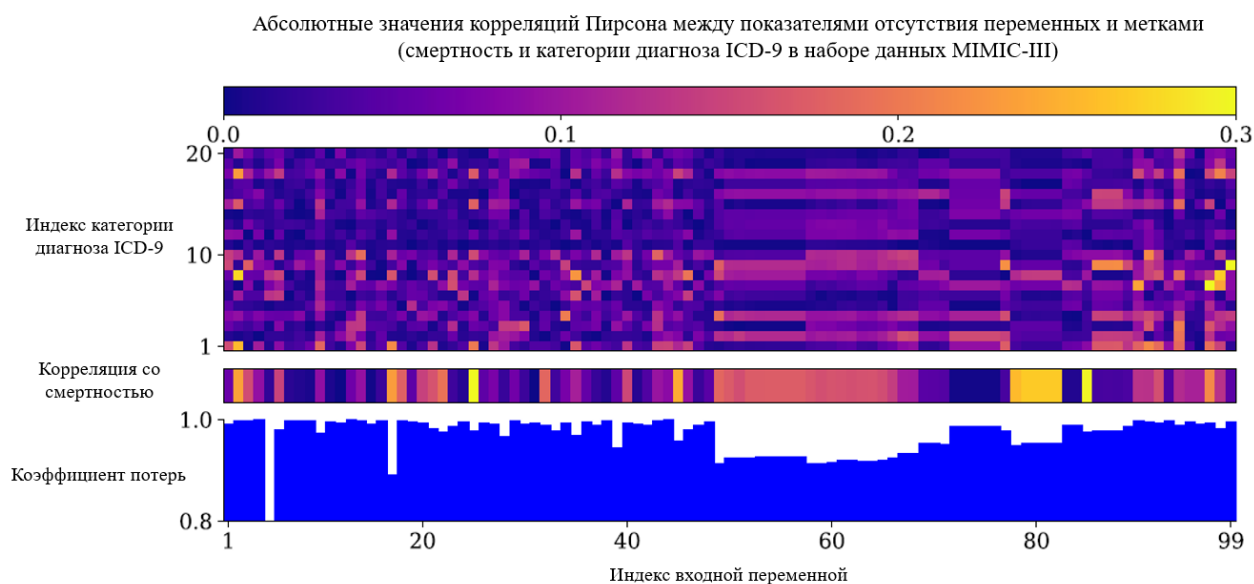


Рисунок 1. Демонстрация отсутствия информации в наборе данных MIMIC-III.

В последние десятилетия были разработаны различные подходы к устранению пропущенных значений во временных рядах. Простое решение состоит в том, чтобы опустить отсутствующие данные и выполнить анализ только наблюдаемых данных, но это не обеспечивает хорошей производительности, когда процент отсутствующих данных высок и хранятся неадекватные выборки [2]. Другое решение состоит в том, чтобы заполнить пропущенные значения замещенными значениями, что известно как вменение данных [3]. Методы сглаживания, интерполяции и сплайнов просты и эффективны, поэтому широко применяются на практике. Однако эти методы не фиксируют корреляции переменных и могут не фиксировать сложные закономерности для выполнения вменения. В последнее время рекуррентные нейронные сети (RNN), такие как Long Short-Term Memory (LSTM) и Gated Recurrent Unit (GRU), продемонстрировали достижение наилучших результатов во многих задачах с временными рядами или последовательными данными [4]. RNN обладают несколькими важными свойствами, такими как высокая производительность прогнозирования, а также способность фиксировать долгосрочные временные зависимости и наблюдения переменной длины [5]. В этой статье предлагается новая модель глубокого обучения на основе GRU, а именно GRU-D, для эффективного использования двух представлений информативных шаблонов отсутствия, то есть маскировки и временного интервала. Маскирование информирует модель о том, какие входные данные наблюдаются (или отсутствуют), а временной интервал инкапсулирует входные шаблоны наблюдения.

Обозначения

Обозначим многомерный временной ряд с D переменными длины T как $X = (x_1, x_2, \dots, x_t)$, где для каждого $t \in \{1, 2, \dots, T\}$, x_t представляет собой t -е наблюдение (также известное как измерение) всех переменных, а x_t^d обозначает измерение d -й переменной x_t . Пусть s_t обозначает отметку времени, когда получено t -е наблюдение, и мы предполагаем, что первое наблюдение сделано с отметкой времени 0 (т. е. $s_1 = 0$). Временной ряд X может иметь пропущенные значения. Мы вводим маскирующий вектор $m_t \in \{0, 1\}^D$, чтобы обозначить, какие переменные отсутствуют на временном шаге t , а также сохраняем временной интервал δ_t^d для каждой переменной d с момента ее последнего наблюдения. Пример обозначений проиллюстрирован на рисунке 2. В данном случае нас интересует проблема классификации временных рядов, где мы предсказываем метки $l_n \in \{1, \dots, L\}$, учитывая данные временного ряда D , где $X_n = [x_1^{(n)}, \dots, x_{T_n}^{(n)}], S_n = [s_1^{(n)}, \dots, s_{T_n}^{(n)}], M_n = [m_1^{(n)}, \dots, m_{T_n}^{(n)}]$.

$$m_t^d = \begin{cases} 1, & x_t^d - \text{наблюдается} \\ 0, & \text{иначе} \end{cases}$$

$$\delta_t^d = \begin{cases} s_t - s_{t-1} + \delta_{t-1}^d, & t > 1, m_{t-1}^d = 0 \\ s_t - s_{t-1}, & t > 1, m_{t-1}^d = 1 \\ 0, & t = 1 \end{cases}$$

X: Входная последовательность (2 переменные)

M: Маскирование

s: Метки времени

Δ: Временной интервал

$$X = \begin{bmatrix} 47 & 49 & NA & 40 & NA & 43 & 55 \\ NA & 15 & 14 & NA & NA & NA & 15 \end{bmatrix}$$

$$M = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$s = [0 \quad 0.1 \quad 0.6 \quad 1.6 \quad 2.2 \quad 2.5 \quad 3.1]$$

$$\Delta = \begin{bmatrix} 0.0 & 0.1 & 0.5 & 1.5 & 0.6 & 0.9 & 0.6 \\ 0.0 & 0.1 & 0.5 & 1.0 & 1.6 & 1.9 & 2.5 \end{bmatrix}$$

Рисунок 2. Пример векторов измерений x_t , меток времени s_t , маскирования m_t и интервала времени δ_t .

Общий случай GRU

Использование рекуррентных нейронных сетей (RNN) для классификации временных рядов подходит больше всего, поскольку их рекурсивная формулировка позволяет им естественным образом обрабатывать последовательности переменной длины [6]. Более того, RNN использует одни и те же параметры на всех временных шагах, что значительно сокращает общее количество параметров, которые необходимо изучить [7]. Среди различных вариантов RNN особое внимание уделено RNN с закрытыми рекуррентными единицами (GRU), но аналогичные методы и модификации применимы и к другим моделям RNN, таким как Long Short-Term Memory (LSTM). Существует три простых способа обработки пропущенных значений без применения каких-либо подходов к вменению или внесения каких-либо изменений в сетевую архитектуру GRU. Первый подход состоит в том, чтобы просто заменить каждое отсутствующее наблюдение средним значением переменной в обучающих примерах. Второй подход заключается в использовании временной структуры. Например, мы можем предположить, что любое отсутствующее значение совпадает с его последним измерением, и использовать прямое вменение (GRU-Forward), т. е.

$$x_t^d \leftarrow m_t^d x_t^d + (1 - m_t^d) x_{t'}^d,$$

где $t' < t$ – время последнего наблюдения d -й переменной. Третий же подход вместо явного ввода пропущенных значений просто указывает, какие переменные отсутствуют и как долго они отсутствовали как часть входных данных, путем объединения векторов измерений, маскирования и временных интервалов.

GRU-D: модель с обучаемыми затуханиями

Чтобы в корне решить проблему отсутствующих значений во временных рядах, мы отмечаем два важных свойства отсутствующих значений временных рядов, особенно в области здравоохранения: во-первых, значение отсутствующей переменной, как правило, близко к некоторому значению по умолчанию, если его последнее наблюдаемое значение было зафиксировано давно. Это свойство присуще медицинским данным о человеческом организме и известно как механизмы гомеостаза и считается критическим для диагностики и лечения заболеваний. Во-вторых, влияние входных переменных со временем исчезнет, если переменная какое-то время отсутствовала. Например, одна медицинская характеристика в электронных медицинских картах имеет значение только в определенном

временном контексте. На основе этих суждений предлагается модель GRU-D, в которой механизм затухания предназначен для входных переменных и скрытых состояний, чтобы зафиксировать вышеупомянутые свойства. В модель вводятся скорости затухания, чтобы управлять механизмом затухания, учитывая следующие важные факторы. Во-первых, каждая входная переменная во временных рядах медицинских данных имеет свое значение и важность в задачах здравоохранения. Скорость затухания должна отличаться от переменной к переменной в зависимости от основных свойств, связанных с переменными [8]. Во-вторых, поскольку наблюдается, что многие отсутствующие шаблоны информативны и потенциально полезны в задачах прогнозирования, но неизвестны и, возможно, сложны, делается упор на изучение скорости затухания на основе обучающих данных, а не фиксированных априори. То моделируется вектор скоростей распада γ :

$$\gamma_t = \exp\{-\max(0, W_\gamma \delta_t + b_\gamma)\}$$

где W_γ и b_γ – параметры модели, которые обучаются совместно со всеми остальными параметрами GRU. Выбран экспоненциальный отрицательный выпрямитель, чтобы скорость каждого затухания монотонно уменьшалась в разумном диапазоне от 0 до 1. Предлагаемая GRU-D включает в себя два различных обучаемых механизма затухания, чтобы использовать отсутствие непосредственно со значениями входных признаков и неявно в состояниях RNN. Во-первых, для отсутствующей переменной используется входное затухание γ_x , чтобы уменьшить его с течением времени до эмпирического среднего (которое принимается в качестве конфигурации по умолчанию), вместо того, чтобы использовать последнее наблюдение как таковое. В этом предположении обучаемая схема затухания может быть легко применена к вектору измерения с помощью:

$$\hat{x}_t^d = m_t^d x_t^d + (1 - m_t^d)(\gamma_{x_t^d}^d x_{t'}^d + (1 - \gamma_{x_t^d}^d) \tilde{x}^d)$$

где $x_{t'}^d$ – последнее наблюдение d-й переменной ($t' < t$), а \tilde{x}^d – эмпирическое среднее значение d-й переменной. При прямом затухании входной переменной мы ограничиваем W_{γ_x} диагональю, что эффективно делает скорость затухания каждой переменной независимой от других.

Иногда входное затухание может не полностью отражать отсутствующие шаблоны, поскольку не вся информация об отсутствии может быть представлена в затухающих входных значениях. Чтобы извлечь более богатые знания из отсутствующих данных, имеется также скрытое затухание состояния γ_h в GRU-D. Интуитивно это приводит к затуханию извлеченных признаков (скрытые состояния GRU), а не к необработанным входным переменным напрямую. Это реализуется путем распада предыдущего скрытого состояния h_{t-1} перед вычислением нового скрытого состояния h_t как

$$\hat{h}_{t-1} = \gamma_{h_t} \square h_{t-1}$$

в этом случае W_{γ_h} не ограничивается диагональностью. Кроме того, маскирующие векторы (mt) подаются непосредственно в модель. Функции обновления GRU-D:

$$r_t = \sigma(W_r \hat{x}_t + U_r \hat{x}_{t-1} + V_r m_t + b_r)$$

$$z_t = \sigma(W_z \hat{x}_t + U_z \hat{x}_{t-1} + V_z m_t + b_z)$$

$$\tilde{h}_t = \tanh(W \hat{x}_t + U(r_t \square \hat{h}_{t-1}) + V m_t + b)$$

$$h_t = (1 - z_t) \square \hat{h}_{t-1} + z_t \square \tilde{h}_t$$

Преимущества модели

Предложенная модель фиксирует наблюдения и их зависимости, применяя маскирование и временной интервал (с использованием термина затухания) к входным данным и состояниям сети GRU, а также совместно обучает все компоненты модели с использованием обратного распространения. Таким образом, модель не только фиксирует долгосрочные временные зависимости наблюдений временных рядов, но также использует недостающие закономерности для улучшения результатов прогнозирования. Кроме того, метод дает полезную информацию о более общих исследовательских задачах анализа временных рядов с отсутствующими данными, помимо задач классификации, включая 1) общую структуру глубокого обучения для обработки временных рядов с отсутствующими данными, 2) эффективное решение для характеристики отсутствующих моделей данные временных рядов, не полностью отсутствующие в случайном порядке, с маскированием и временным интервалом, и 3) проницательный подход к изучению влияния отсутствия переменной на метки прогноза с помощью анализа затухания.

Исследуя условия распада, полученные в модели, врачи могут оценить влияние недостающих данных для каждой переменной и улучшить стратегии сбора данных, чтобы получить более важные переменные. Кроме того, прогноз, полученный на основе модели, можно использовать в качестве заменителя системы оценки. Наконец, предлагаемая модель также может быть использована для изучения оценки риска смертности в режиме реального времени для пациентов и может показать, как состояние здоровья пациента меняется с течением времени.

Ограничения

Предлагаемая модель ориентирована на получение точных и надежных прогнозов для данных многомерных временных рядов с пропущенными значениями. Эта модель опирается на информацию, относящуюся к задачам прогнозирования, которая представлена в отсутствующих шаблонах, для повышения эффективности прогнозирования по сравнению с исходными базовыми показателями GRU-RNN. Если отсутствующие данные вообще неинформативны или неотъемлемая корреляция между отсутствующими паттернами и задачами прогнозирования неясна, модель может получить ограниченный прирост улучшения результата или даже потерпеть неудачу. Предлагаемый метод явно не предназначен для заполнения пропущенных значений в данных и не может быть непосредственно использована для неконтролируемой настройки без меток прогнозирования.

Заключение

Обработка пропущенных значений в данных многомерных временных рядов с использованием рекуррентных нейронных сетей является сложной задачей. Готовые архитектуры RNN с импутацией могут обеспечить производительность, сравнимую только со случайными лесами и SVM, и, кроме того, они не демонстрируют всех преимуществ репрезентативного обучения. Использование популярных методов вменения приводит к длительной процедуре прогнозирования и может снизить качество. Чтобы решить вышеуказанные проблемы, предлагается новая модель на основе GRU, которая фиксирует

отсутствие информации путем включения маскирования и временного интервала непосредственно в архитектуру GRU. Предложенная модель GRU-D с обучаемыми затуханиями имеет такое же время работы и пространственную сложность, что и исходные модели RNN. Хотя в данной статье преимущества модели описаны в рамках данных временных рядов, возникающих в медицинских задачах, предложенные подходы будут широко полезны для различных задач прогнозирования временных рядов за пределами области здравоохранения.

Список литературы:

1. Lipton, Z. C., Kale, D., Wetzel, R. Directly modeling missing data in sequences with rnns: Improved classification of clinical time series. // In Machine Learning for Healthcare Conference, pp. 253–270, 2016.
2. Jaques N., Taylor S., Sano A., Picard R., Multimodal autoencoder: A deep learning approach to filling in missing sensor data and enabling better mood prediction. // in Proc. 7th Int. Conf. Affective Comput. Intell. Interact. (ACII), San Antonio, TX, USA, pp. 202–208, 2017.
3. Тронов К.А., Белов Ю. С. Алгоритм скользящего окна переменной длины для генерации выборок временных рядов // В сборнике: Фундаментальные и прикладные исследования. Актуальные проблемы и достижения. Сборник избранных статей Всероссийской (национальной) научной конференции. Санкт-Петербург, 2022. С. 32-34.
4. S. Yang, X. Yu and Y. Zhou, LSTM and GRU Neural Network Performance Comparison Study: Taking Yelp Review Dataset as an Example. // 2020 International Workshop on Electronic Communication and Artificial Intelligence (IWEC AI), pp. 98-101, 2020.
5. Ростовцев, В. С. Искусственные нейронные сети: учебник для вузов / В. С. Ростовцев. – 2-е изд., стер. – Санкт-Петербург: Лань, 2021. – 216 с. – ISBN 978-5-8114-7462-2. – Текст: электронный // Лань: электронно-библиотечная система. – URL: <https://e.lanbook.com/book/160142> (дата обращения: 10.12.2022).
6. Choi E., Bahadori M. T., Schuetz A., Stewart W. F., Sun, J. Doctor ai: Predicting clinical events via recurrent neural networks. // In Machine Learning for Healthcare Conference, pp. 301–318, 2016.
7. Pham T., Tran T., Phung D., Venkatesh S. Deepcare: A deep dynamic memory model for predictive medicine. // In Advances in Knowledge Discovery and Data Mining, pp. 30–41, 2016.
8. Che Z., Purushotham S., Cho K., Sontag D., Liu, Y. Recurrent neural networks for multivariate time series with missing values. // arXiv preprint arXiv:1606.01865, 2016. – URL: <https://arxiv.org/pdf/2107.02248v2.pdf> (дата обращения 09.12.2022)

References:

1. Lipton, Z. C., Kale, D., Wetzel, R. Directly modeling missing data in sequences with rnns: Improved classification of clinical time series. // In Machine Learning for Healthcare Conference, pp. 253–270, 2016.

2. Jaques N., Taylor S., Sano A., Picard R., Multimodal autoencoder: A deep learning approach to filling in missing sensor data and enabling better mood prediction. // in Proc. 7th Int. Conf. Affective Comput. Intell. Interact. (ACII), San Antonio, TX, USA, pp. 202–208, 2017.
3. Tronov K.A., Belov Y.S. Variable length sliding window algorithm for time series samples generation // In the collection: Fundamental and applied research. Actual problems and achievements. Collection of selected articles of the All-Russian (national) scientific conference. St. Petersburg, 2022, pp. 32-34.
4. S. Yang, X. Yu and Y. Zhou, LSTM and GRU Neural Network Performance Comparison Study: Taking Yelp Review Dataset as an Example. // 2020 International Workshop on Electronic Communication and Artificial Intelligence (IWEC AI), pp. 98-101, 2020.
5. Rostovtsev, V. S. Artificial neural networks : textbook for universities / V. S. Rostovtsev. — 2nd ed., revised. — Saint Petersburg: Lan, 2021. — 216 p. — ISBN at 978-5-8114-7462-2. — Text: electronic // Lan: electronic library system. — Address: <https://e.lanbook.com/book/160142> (accessed: 12/10/2022)
6. Choi E., Bahadori M. T., Schuetz A., Stewart W. F., Sun, J. Doctor ai: Predicting clinical events via recurrent neural networks. // In Machine Learning for Healthcare Conference, pp. 301–318, 2016.
7. Pham T., Tran T., Phung D., Venkatesh S. Deepcare: A deep dynamic memory model for predictive medicine. // In Advances in Knowledge Discovery and Data Mining, pp. 30–41, 2016.
8. Che Z., Purushotham S., Cho K., Sontag D., Liu, Y. Recurrent neural networks for multivariate time series with missing values. // arXiv preprint arXiv:1606.01865, 2016. — URL: <https://arxiv.org/pdf/2107.02248v2.pdf> (accessed 12/09/2022)