

Методы восстановления пропущенных значений во временных рядах в системе прогнозирования электропотребления

А.С. Маниковский¹, А.Ю. Мухомов²

¹*Забайкальский институт железнодорожного транспорта, Чита*

²*Иркутский государственный университет путей сообщения, Иркутск*

Аннотация: Основными направлениями энергетической стратегии железнодорожного транспорта являются улучшение структуры управления энергетическим комплексом железных дорог, снижение расходов электрической энергии и сокращение затрат на ее приобретение. Первоначальную информацию для планирования оптимальных режимов работы при управлении энергетическим комплексом предоставляет прогноз электропотребления. Согласно правилам функционирования розничных рынков, потребители обязаны точно планировать объемы электропотребления. При отклонении электропотребления более 5 % от запланированного объема, предприятие несет дополнительные издержки.

Для составления точного прогноза необходимо провести анализ исходных данных – архив электропотребления. На первоначальном этапе анализа данных выявляется проблема пропусков. При наличии пропусков в данных процесс прогнозирования потребления электроэнергии может быть затруднен, а иногда невозможен. Наиболее рациональным решением является заполнение пропусков с использованием современных методов обработки информации. Это позволит ясно представить структуру данных, вычислить необходимые значения и интерпретировать результаты анализа.

Ключевые слова: электропотребление, временной ряд, пропуски данных, восстановление пропущенных значений, прогнозирование электропотребления, тяга поездов, железнодорожный транспорт, нейронные сети.

Введение

Анализ временного ряда (ВР) математическими методами позволяет определить его природу и получить будущее значение рассматриваемой величины. Предшествующей процедурой анализа ВР является первичная обработка. Первичная обработка ВР - важная и актуальная задача, от решения которой зависят результаты последующего анализа данных. Обычно, на этом этапе выявляется проблема пропусков данных [1]. Причины возникновения пропусков данных во ВР электропотребления (ЭП) разнообразны: сбои в оборудовании, устройствах передачи информации, в системе фиксирования данных; изменения в системе отчетности; ошибки персонала; сбои и аварийные ситуации на объектах фиксации данных и др.

При наличии пропусков в данных процесс прогнозирования ЭП может быть затруднен, а при некоторых обстоятельствах даже невозможен [2]. Самым простым решением является исключение измерений, которые содержат пропуски. Такой подход при анализе ВР электропотребления на железнодорожном транспорте приводит к сильному различию статистических выводов, сделанных при наличии в данных пропусков и при их отсутствии.

Рациональным решением является восстановление пропущенных значений перед анализом данных. При таком подходе становится возможным представление структуры данных, вычисление необходимых итоговых значений, интерпретирование результатов анализа.

В качестве объекта исследования использовался временной ряд тягового ЭП тяговой подстанции Белогорск Забайкальской железной дороги. Данный ВР представлен почасовой информацией по потреблению электрической энергии с 2017 по 2021 гг. На рисунке 1 представлен фрагмент графика электропотребления за 2017 год.

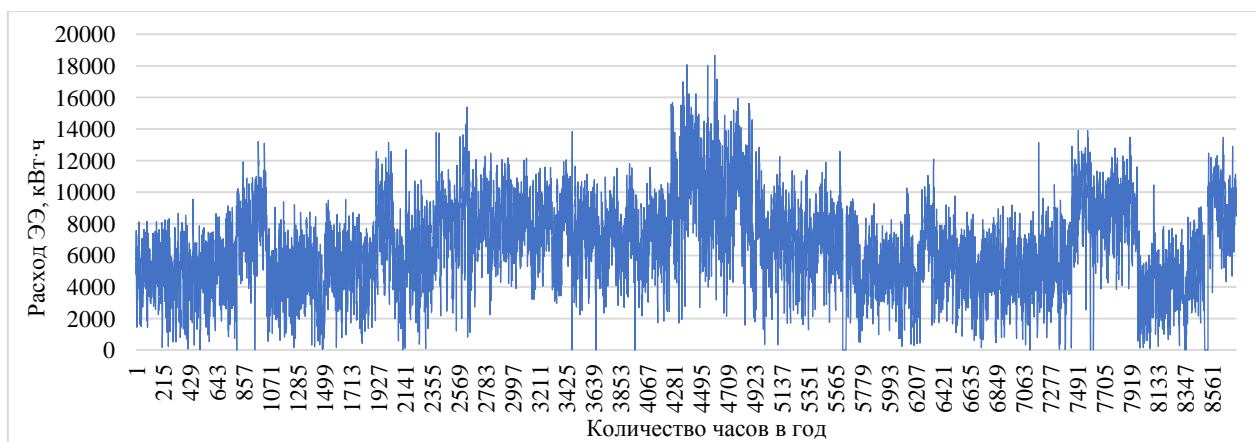


Рис. 1. – Годовой график почасовых значений потребляемой электрической энергии за 2017 год

Современные условия работы железнодорожного транспорта отличаются неравномерностью загрузки и неполнотой исходных данных, которые формируют и характеризуют рассматриваемую систему [3]. Эти и

другие факторы, приведенные ранее, приводят к образованию пропущенных значений и, как следствие, осложняют или делают невозможным процесс прогнозирования ЭП.

В рассматриваемом временном ряду присутствуют пропущенные значения, имеющие не эпизодический характер. Для использования данного ВР в последующем анализе необходимо восстановить пропущенные значения.

На сегодняшний день существует множество разнообразных методов восстановления пропущенных значений. Все эти методы можно обобщить в три группы в зависимости от степени эффективности и сложности реализации:

- простые методы, основанные на линейной интерполяции;
- методы прогнозирования при помощи одиночных моделей;
- нейросетевые методы заполнения пропусков.

Для оценки эффективности анализируемых методов в экспериментальный ВР был искусственно введен пропуск длиной в 48 значений (Рисунок 2). Тем самым была смоделирована ситуация, когда фиксация величины ЭП была нарушена в течение двух суток.



Рисунок 2. – График электропотребления с искусственно введенными пропусками

Метод линейной интерполяции

Методы, основанные на линейной интерполяции, характеризуются простой и быстрой реализацией [4]. При данном методе заданные точки с координатами x_i, y_i при $i=0, 1, 2, \dots, n$ соединяются отрезками. Уравнение прямой, проходящей через две точки, используется, как уравнение интерполяционного многочлена. Для i -го интервала уравнение прямой, проходящей через точки (x_{i-1}, y_{i-1}) и (x_i, y_i) , выглядит следующим образом:

$$\frac{y - y_{i-1}}{y_i - y_{i-1}} = \frac{x - x_{i-1}}{x_i - x_{i-1}}, \quad (1)$$

отсюда

$$y = a_i x + b_i, \quad x_{i-1} \leq x \leq x_i. \quad (2)$$

Коэффициенты уравнения линейной интерполяции определяются согласно следующих формул:

$$a_i = \frac{y_i - y_{i-1}}{x_i - x_{i-1}}, \quad (3)$$

$$b_i = y_{i-1} - a_i x_{i-1}. \quad (4)$$

На рисунке 3 представлен результат заполнения пропусков.



Рисунок 3. – График электропотребления с заполнением пропусков методом линейной интерполяции

Графически этот метод сводится к соединению точки непосредственно перед отсутствующими данными с точкой сразу после отсутствующих данных прямой линией. Метод линейной интерполяции точен только в том

случае, когда значения переменной близки и изгиб кривой между ними невелик.

Метод прогнозирования

Методы второй группы – методы итеративного прогнозирования при помощи одиночных моделей [5].

На начальном этапе данного метода реализуется прогноз на один шаг вперед. Далее, на основе полученных данных, строится новый вектор в пространстве задержек:

$$\hat{x}_{L+1} = \begin{pmatrix} \hat{x}_{L+1} \\ x_L \\ \vdots \\ x_{L-p+2} \end{pmatrix}, \quad (5)$$

x_L – значение временного ряда.

Полученный вектор подставляется в правую часть выражения (6) и реализуется прогноз на еще один шаг

$$\hat{x}_{L+1} = f(x_L, \hat{a}), \quad (6)$$

где \hat{a} – параметр модели.

Поскольку параметры модели выбираются те же, что и на первом этапе, то каждое последующее вычисляемое значение выражается через предыдущее. Результат заполнения пропусков представлен на рисунке 4.



Рисунок 4. – График электропотребления с заполнением пропусков методом итеративного прогнозирования при помощи одиночных моделей

К преимуществу данного способа можно отнести простоту прогнозирования следующих значений, которая достигается за счет оценки параметров модели только на первом этапе. Однако, данный метод характеризуется невысокой точностью. Это связано с тем, что соседями новых векторов задержек являются векторы, которые были соседями стартового. Поэтому применение данного метода для восстановления пропущенных значений временного ряда тягового электропотребления является не предпочтительным.

Нейросетевой метод

Перспективным представляется метод восстановления пропусков в данных с использованием искусственных нейронных сетей (ИНС) [6, 7]. ИНС способны обрабатывать большой объем данных, выделять и обобщать скрытые зависимости между входными и выходными данными. ИНС позволяют моделировать зависимости, в основе которых могут быть линейные, обобщенно линейные и нелинейные модели [8].

В качестве прогнозной модели была использована рекуррентная нейронная сеть (рисунок 5). Данная архитектура ИНС обладает рядом преимуществ и наилучшим образом подходит для задачи прогнозирования [9, 10].

Дискретная модель для прогнозной сети имеет следующий вид:

$$x_i(n+1) = \sum_{j=1}^r d_{ij}y_j(n) + b_i, \quad (7)$$

$$y_i(n+1) = F(x_i(n+1)), \quad (8)$$

где $x_i(n+1)$ – сигнал активации i -го нейрона в момент времени $(n+1)$,

$d_{ij} = 1, 2, \dots, r$ – синаптические веса входных сигналов i -го нейрона,

b_i – вес смещения i -го нейрона,

F – функция активации,

$y_i(n+1)$ – выходной сигнал i -го нейрона в момент времени $(n+1)$.

Соответствующая непрерывная модель прогнозной модели представляется в виде:

$$\frac{dx_i(t)}{dt} = \sum_{j=1}^r d_{ij}y_j(t) + b_i,$$

$$y_i(t) = F(x_i(t)),$$

где t – переменная непрерывного времени.

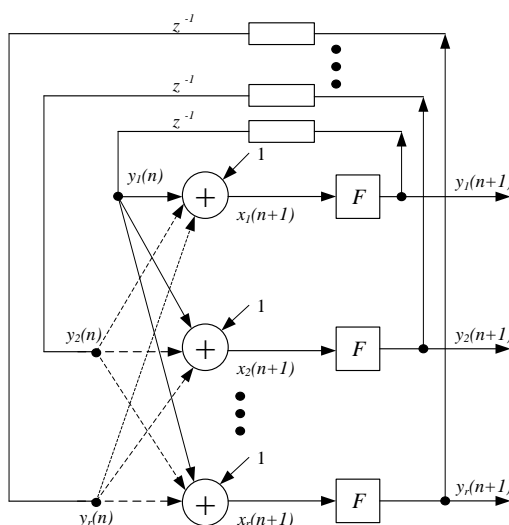


Рисунок 5 – Структура прогнозной сети

Автором предлагается использовать двусторонний прогноз для эффективного восстановления пропусков во временном ряду тягового ЭП.

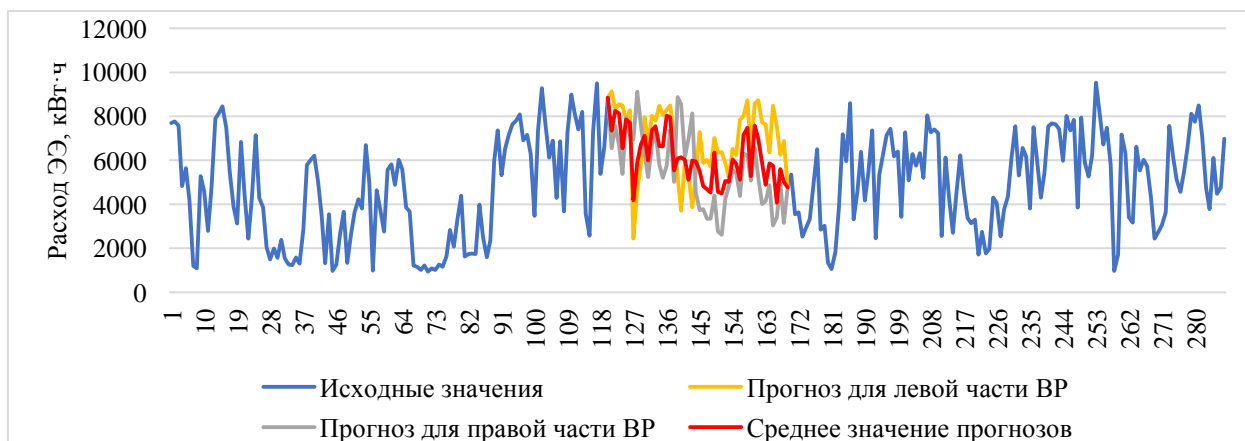


Рисунок 6. – График электропотребления с заполнением пропусков методом двустороннего прогноза временного ряда

Первым этапом используется часть ВР, расположенная слева от пропуска. На этом участке ВР обучается модель, формирующая прогноз на такое количество элементов, сколько их есть в пропуске. Затем выполняется прогноз для правой относительно пропуска части. Для этого известная часть ВР инвертируется, затем обучается модель и формируется прогноз, который также инвертируется. После этого осуществляется комбинация двух прогнозов при помощи взвешенного среднего (Рисунок 6).

Анализ результатов

На основании оценки величины среднеквадратичной ошибки был выполнен анализ эффективности методов восстановления пропусков во ВР электропотребления. На рисунке 7 представлен сравнительный анализ методов восстановления пропусков. В таблице 1 приведено сравнение средних ошибок аппроксимации рассмотренных методов.

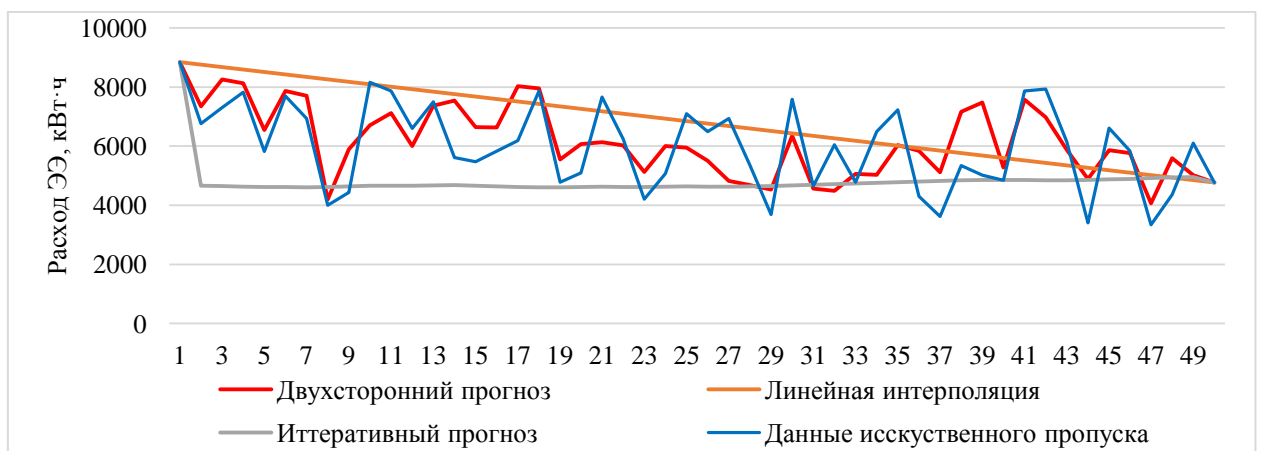


Рисунок 7 – Сравнение методов восстановления пропущенных значений

Таблица № 1

Средние ошибки аппроксимации методов восстановления пропусков.

Метод восстановления пропусков	Средняя ошибка аппроксимации
Линейная интерполяция	30%
Итеративный прогноз	21%
Двухсторонний прогноз	13%

Как видно из приведенных графика и таблицы, наилучшим способом восстановления пропущенных значений является двусторонний прогноз ВР.

Для временного ряда, имеющего восстановленные пропуски, становится возможным:

- определение статистических выводов;
- определение структуры данных;
- анализ исходных данных;
- определение итоговых значений;
- интерпретирование результатов анализа.

Даже при наличии достаточно высокой величины средней ошибки аппроксимации, временной ряд с восстановленными пропущенными значениями становится более пригодным для дальнейшего анализа и прогнозирования, чем ряд с имеющимися пропусками. В свою очередь, получение точных результатов прогнозирования ЭП позволит улучшить структуру управления энергетическим комплексом железнодорожного транспорта, снизить расходы электрической энергии и сократить затраты на ее приобретение.

Литература

1. Камаев В.А., Щербаков М.В., Бребельс А. Интеллектуальные системы автоматизации управления энергосбережением // Открытое образование. - 2011. - № 2. - С. 227-231.
2. Маркин А.В., Щербаков М.В. Метод автоматического восстановления значений в потоках данных на основе взвешенной модели // Прикаспийский журнал: управление и высокие технологии. - 2013. - №3 (23). - С. 49-54.
3. Шумилова, Г.П., Готман Н.Э., Старцева Т.Б. Прогнозирование электрических нагрузок при оперативном управлении электроэнергетическими системами на основе нейросетевых структур:

монография; Под ред. доктора технических наук П. И. Бартоломея. – Сыктывкар: КНЦ УрО РАН, 2008. – 85 с.

4. Воробьев А.В., Воробьева Г.Р. Методы резервирования в задачах восстановления временных рядов геомагнитных данных // Информационные процессы. – 2018. – Т. 18. – № 1. – С. 1-18.

5. Лоскутов А.Ю., Котляров О.Л., Истомин И.А., Журавлев Д.И. Проблемы нелинейной динамики. Локальные методы прогнозирования временных рядов // Вестник МГУ, Сер.3. Физика и Астрономия. - 2002, - №6, С. 3-21.

6. Староверов Б. А., Семенов И.В. Схемы взаимодействия поставщиков и получателей прогнозов электропотребления на основе использования нейросетевой информационной системы // Инженерный вестник Дона, 2018, №1. URL: ivdon.ru/ru/magazine/archive/n1y2018/4784.

7. Кретов Д. А., Рузанов Р.В. Прогнозирование электропотребления энергосбытовой компании с использованием искусственной нейронной сети // Инженерный вестник Дона, 2015, №2. URL: ivdon.ru/ru/magazine/archive/n2y2015/2880.

8. Рудковская Д., Пилиньский М., Рутковский Л. Нейронные сети, генетические алгоритмы и нечеткие системы: перевод с польского И. Д. Рудинского // М.: Горячая линия–Телеком, 2006. – 452 с.

9. Indrajeet Y. Javeri, Mohammadhossein Toutiaee, Ismailcem B. Arpinar, and John A. Miller, Tom W. Miller Improving Neural Networks for Time-Series Forecasting using Data Augmentation and AutoML - University of Georgia, Athens, Georgia 2021. pp. 1-8.

10. Cirstea Razvan-Gabriel, Guo Chenjuan, Yang Bin. Graph Attention Recurrent Neural Networks for Correlated Time Series Forecasting - Full version - Aalborg 2021. pp. 1-6.

References

1. Kamaev V.A., Shherbakov M.V., Brebel's A. Otkrytoe obrazovanie. 2011. № 2. P. 227-231
2. Markin A.V., Shherbakov M.V. Prikaspijskij zhurnal: upravlenie i vysokie tehnologii. 2013. №3 (23). pp. 49-54.
3. Shumilova, G.P., Gotman N.Je., Starceva T.B. Prognozirovanie elektricheskikh nagruzok pri operativnom upravlenii elektroenergeticheskimi sistemami na osnove nejrosetevykh struktur: monografiya [Forecasting of electric loads during operational control of electric power systems on the basis of neural network structures: monograph]. Syktyvkar: KNC UrO RAN, 2008. 85 p.
4. Vorob'ev A.V., Vorob'eva G.R. Informacionnye processy. 2018. T. 18. № 1. pp. 1-18.
5. Loskutov A.Ju., Kotljarov O.L., Istomin I.A., Zhuravlev D.I. Vestnik MGU, Ser.Z. Fizika i Astronomija. 2002, №6, pp. 3-21.
6. Staroverov B. A., Semenov I.V. Inzhenernyj vestnik Dona, 2018, №1. URL: ivdon.ru/ru/magazine/archive/n1y2018/4784.
7. Kretov D. A., Ruzanov R.V. Inzhenernyj vestnik Dona, 2015, №2. URL: ivdon.ru/ru/magazine/archive/n2y2015/2880.
8. Rudkovskaja D., Pilin'skij M., Rutkovskij L. Neironnye seti, geneticheskie algoritmy i nechetkie sistemy [Neural networks, genetic algorithms and fuzzy systems]: perevod s pol'skogo I. D. Rudinskogo Gorjachaja linija. Telekom, 2006. P. 452.
9. Indrajeet Y. Javeri, Mohammadhossein Toutiaee, Ismailcem B. Arpinar, and John A. Miller, Tom W. Miller Improving Neural Networks for Time-Series Forecasting using Data Augmentation and AutoML. University of Georgia, Athens, Georgia 2021. pp. 1-8.
10. Razvan-Gabriel Cirstea, Chenjuan Guo, Bin Yang Graph Attention Recurrent Neural Networks for Correlated Time Series Forecasting. Full version. Aalborg 2021. pp. 1-6.