УДК 519.6

А.С. Виноградов, О.Л. Селяничев

АЛГОРИТМ ВОССТАНОВЛЕНИЯ ПРОПУЩЕННЫХ ЗНАЧЕНИЙ ВО ВРЕМЕННЫХ РЯДАХ В СИСТЕМЕ ПРОГНОЗИРОВАНИЯ ЭЛЕКТРОПОТРЕБЛЕНИЯ НА АГЛОМЕРАЦИОННОМ ПРОИЗВОДСТВЕ

В статье рассматривается метод обработки данных в системе прогнозирования электропотребления на агломерационном производстве, восстанавливающий пропущенные значения во временных рядах. Метод заключается в том, что пропуски классифицируются на одиночные и групповые, а затем к ним применяются различные алгоритмы восстановления информации в зависимости от класса пропуска.

Прогнозирование электропотребления, восстановление данных, интерполяция сплайнами, Zet-алгоритм.

The article considers the method of data processing in the system of power consumption forecasting at agglomerative production restoring the missing values in time series. The method consists in the fact that the gaps are classified into individual and group; then various algorithms of data recovery are applied to them depending on the miss.

Forecasting energy consumption, data recovery, spline interpolation, Zet-algorithm.

Система прогнозирования электропотребления на агломерационном производстве, в основе которой лежит нейро-нечеткий модуль управления, использует при получении прогнозных значений ретроспективную информацию. Совокупность данных отражает зависимость различных факторов производства (температура горна, температура шихты, основность агломерата) от времени, т.е. образует временные ряды. Показатели на производстве снимаются в автоматическом режиме, поэтому в связи с несовершенством техники в реальной ситуации имеет место быть потеря массивов информации. Например, при кратковременной потере связи между датчиком и сервером базы данных могут затеряться одиночные значения, а во время сбоя на датчике – целая группа показаний. Для корректной работы системы прогнозирования электропотребления необходимо учесть все нежелательные ситуации, которые могут ухудшить степень соответствия прогнозного и реального значений. Поэтому в систему введен модуль обработки поступающих значений снимаемых показателей.

Перед тем, как рассматривать методы восстановления информации, необходимо классифицировать пропуски в массивах данных [3]. Пропуски в данных считаются полностью случайными, если условная вероятность пропуска P не зависит ни от пропущенных элементов, ни от прочих, т.е. эта вероятность постоянна для всех наблюдений. Этот вид пропусков называется MCAR (data are missing completely at random). К данным такого типа возможно применение методов восстановления пропущенных значений. Если вероятность P не зависит от пропущенных значений, но зависит от других элементов, то это пропуски типа MAR (missing at random). В этих случаях механизм пропусков несущественен, поэтому к данным все еще применимы методы восстановления пропусков. Если вероятность пропуска зависит от самого пропущенного элемента, то механизм пропусков является существенным, поэтому для того, чтобы восстановить пропущенный элемент, необходимо анализировать механизм, из-за которого произошел дефект.

Так как на аглопроизводстве предполагается наличие массивов информации, содержащих пропуски, которые связаны с техническими проблемами, то такие пропуски можно отнести к классу MCAR. Действительно, вероятность появления пробелов в информации не зависит от пропущенных показателей, а также она не зависит от присутствующих значений в ряду. Первый подход исключения пропусков в массивах данных – это смещение массива данных на количество пропущенных элементов. Метод прост в реализации, но необходимым условием его применения является следование данных требованию MCAR, а количество пропусков должно быть небольшим. Требование выполняется, однако, значения одного ряда показаний могут зависеть от значений другого ряда в один и тот же момент времени, что накладывает ограничения на операции с массивами данных. Следовательно, метод не подходит для использования в системе прогнозирования электропотребления на аглопроизводстве.

Второй подход – заполнение пропусков значениями, вычисляемыми некоторым из методов: заполнения средними, заполнения подбором, заполнения по регрессии, метод интерполяции сплайнами, Zet-алгоритм и др. Подход применим к пропускам, удовлетворяющим требованию MCAR, поэтому в системе прогнозирования электропотребления возможно применение алгоритмов исключения пропущенных значений, относящиеся к данному подходу.

Метод интерполяции сплайнами позволяет восстанавливать одиночные пробелы в массиве информации до значений, наиболее приближенных к истинным показаниям, по сравнению с другими методами. Но в случае групповых пропусков результаты применения данного метода ухудшаются пропор-

ционально увеличению протяженности группы, что объясняется тем, что метод заключается в построении наилучшей аппроксимирующей поверхности для существующих наблюдений.

Задача интерполирования состоит в том, чтобы по значениям функции *f(x)* в нескольких точках отрезка восстановить ее значения в остальных точках этого отрезка. В методе используются сплайны третьей степени (кубические), имеющие на отрезке [*a, b*] непрерывную, первую производную. На отрезке времени [*a, b*] заданы значения некоторого показателя агломерационного производства, т.е. существует функция $f_i = f(x_i)$, где i = 0, ..., n – точки времени снятия показаний.

Интерполяционным кубическим сплайном $S_3(x)$, соответствующим зависимости снятых показаний от времени, будем считать сплайн

$$S_3(x) = a_{i0} + a_{i1} \cdot (x - x_i) + \\ + a_{i2} \cdot (x - x_i)^2 + a_{i3} \cdot (x - x_i)^3, \quad (1)$$

где $x \in [x_i, x_{i+1}]$, удовлетворяющий условиям $S_3(x_i) = f(x_i)$, i = 0, ..., n.

Сплайн (1) на каждом из отрезков $[x_i, x_{i+1}]$, i = 0, ..., n – 1 определяется четырьмя коэффициентами, и поэтому для его построения на всем промежутке [*a, b*] необходимо определить 4*n* коэффициентов.

Сплайн (1) после преобразований [1] примет следующий вид:

$$S_3(x) = \frac{x_i - x}{h_i} \cdot y_{i-1} + \frac{x - x_{i-1}}{h_i} \cdot y_i + \\ + \frac{(x_i - x)^3 - h_i^2 \cdot (x_i - x)}{6h_i} \cdot m_{i-1} + \\ + \frac{(x - x_{i-1})^3 - h_i^2 \cdot (x - x_{i-1})}{6h_i} \cdot m_i, \quad (2)$$

где $h_i = x_i - x_{i-1}$, $m_i = S_3''(x_i)$.

Чтобы восстановить функцию, необходимо найти m_i, для i = 1, ..., n – 1. Составляем n – 1 уравнение согласно условию $m_0 = m_n = 0$:

$$\frac{h_i}{6} \cdot m_{i-1} + \frac{h_i + h_{i+1}}{3} \cdot m_i + \frac{h_{i+1}}{6} \cdot m_{i+1} = \frac{y_{i+1} - y_i}{h_{i+1}} - \frac{y_i - y_{i-1}}{h_i} \quad (3)$$

Система линейных алгебраических уравнений (3) имеет трехдиагональную матрицу с диагональным преобладанием. Такие матрицы являются неособенными. Поэтому неизвестные m_i находятся из системы (3) однозначно. После чего $S_3(x)$ восстанавливается по формуле (2), откуда системой прогнозирования электропотребления находятся пропущенные значения показаний, снятых на аглопроизводстве.

Для восстановления групповых пропусков целесообразно применять Zet-алгоритм. В случае одиночных пропусков данный метод немного уступает методу интерполяции сплайнами, но в случае наличия в данных групповых пропусков качество восстановления пропущенных значений при помощи Zet-алгоритма лучше остальных методов. Это связано с тем, что алгоритм при восстановлении учитывает закономерности ряда, а также может работать как с одномерными рядами данных, так и с таблицами данных, состоящих из множества взаимосвязанных рядов. В основе функционирования Zet-алгоритма лежат три предположения [2]:

1. Гипотеза избыточности: предполагается, что в таблице экспериментальных данных присутствует избыточность в строках и столбцах.

2. Гипотеза аналогичности: предполагается, что если два объекта похожи по значениям n – 1 свойств, то они похожи и по n-му свойству.

3. Гипотеза локальной компетентности: предполагается, что избыточность строк и столбцов носит локальный характер, т.е. для каждого пропущенного значения имеется только некоторое, пусть и большое количество объектов – аналогов объекта с пропуском и свойств – аналогов свойства с пропуском.

Поэтому предлагается использовать для прогнозирования только такие компетентные объекты и свойства, которые выбираются для каждого пропуска отдельно.

Процесс восстановления пропущенных значений в рядах ретроспективной информации с помощью Zet-алгоритма можно условно разделить на три основных этапа [2]:

1) на первом этапе для данного пропуска из исходной матрицы, столбцы которой нормированы по дисперсии, выбирается подмножество компетентных строк и затем для этих строк – компетентных столбцов;

2) на втором этапе автоматически подбираются параметры в формуле, используемой для предсказания пропущенного элемента, при которых ожидаемая ошибка предсказания достигает минимума;

3) на третьем этапе выполняется непосредственно прогнозирование элемента по этой формуле.

Для нахождения пропущенных элементов в массиве показаний, снятых на аглопроизводстве, воспользуемся формулами определения прогнозных величин b_j, порожденной избыточностью, содержащейся в столбцах, и b_i, порожденной избыточностью, содержащейся в строках.

Формула определения b_j имеют следующий вид:

$$b_j = \frac{\sum_{k=1}^{q} b_k \cdot L_{jk}^{\alpha}}{\sum_{k=1}^{q} L_{jk}^{\alpha}}.$$

Здесь *a* – коэффициент, регулирующий влияние компетентности на результат предсказания. Выбор *a*

составляет суть этапа подбора формулы для прогнозирования: все известные элементы *y*-го столбца предсказываются при разных значениях *a* и затем выбирается такое значение *a*, при котором ошибка прогноза была минимальной.

Вычисление b_i происходит по следующей формуле:

$$b_i = \frac{\sum_{l=1}^{s} b_l \cdot L_{il}^{\alpha}}{\sum_{l=1}^{s} L_{il}^{\alpha}}.$$

Для выбора *a* используются все известные элементы *i*-й строки, выбор делается при минимальном значении ошибки их прогнозирования. Окончательное значение пропущенного элемента система прогнозирования считает равным либо прогнозу по строкам, либо прогнозу по столбцам в зависимости от наименьшей ошибки.

Полный алгоритм восстановления пропущенных данных в системе прогнозирования электропотребления на агломерационном производстве будет следующим:

1. В матрице ретроспективных данных, используемых в модуле нейро-нечеткого управления, производится поиск одиночных пропусков.
2. К каждому найденному пропуску применяется метод сплайн-интерполяции.
3. Все остальные пропущенные элементы находятся с помощью Zet-алгоритма.

Таким образом, был получен алгоритм, который использует преимущества двух различных методов восстановления данных. Предварительное применение метода интерполяции сплайнами позволяет не только восстановить пропущенные значения с высокой степенью точности, но и повысить эффективность работы Zet-алгоритма.

Экспериментальные исследования были произведены с использованием ретроспективных данных агломашины № 10 ЧМК ОАО «Северсталь». Было взято 40 значений температуры шихты и температуры в секциях зажигательного горна с интервалом снятия показания в 10 мин. Чтобы проанализировать работу алгоритма, в данные искусственно были введены пропуски в точках снятия показаний 3, 8, 9, 10, 33 для температуры шихты и в точках 20, 21, 22 для температуры в секциях горна.

На рис. 1 представлены графики исходного временного ряда температуры шихты и временного ряда, в котором искусственно введены пропуски значений, обработанного алгоритмом восстановления пропусков. На рис. 2 представлены соответственные графики для температуры в секциях зажигательного горна.

Значения температуры в точках 3, 33 для температуры шихты были восстановлены с помощью метода интерполяции сплайнами, все остальные значения восстановлены с помощью Zet-алгоритма. Алгоритм чувствителен к резким скачкам в изменении показателя. В таких местах наблюдается снижение качества восстановления, где относительная погрешность достигает 3 %. На графиках прослеживается сходство динамик изменения температуры шихты и температуры в секциях зажигательного горна. Zet-алгоритм позволяет отследить эту связь, что снижает погрешность восстановления значений в экстремумах графика. На более гладких участках погрешность не превышает 1 %, что говорит о надежности алгоритма.

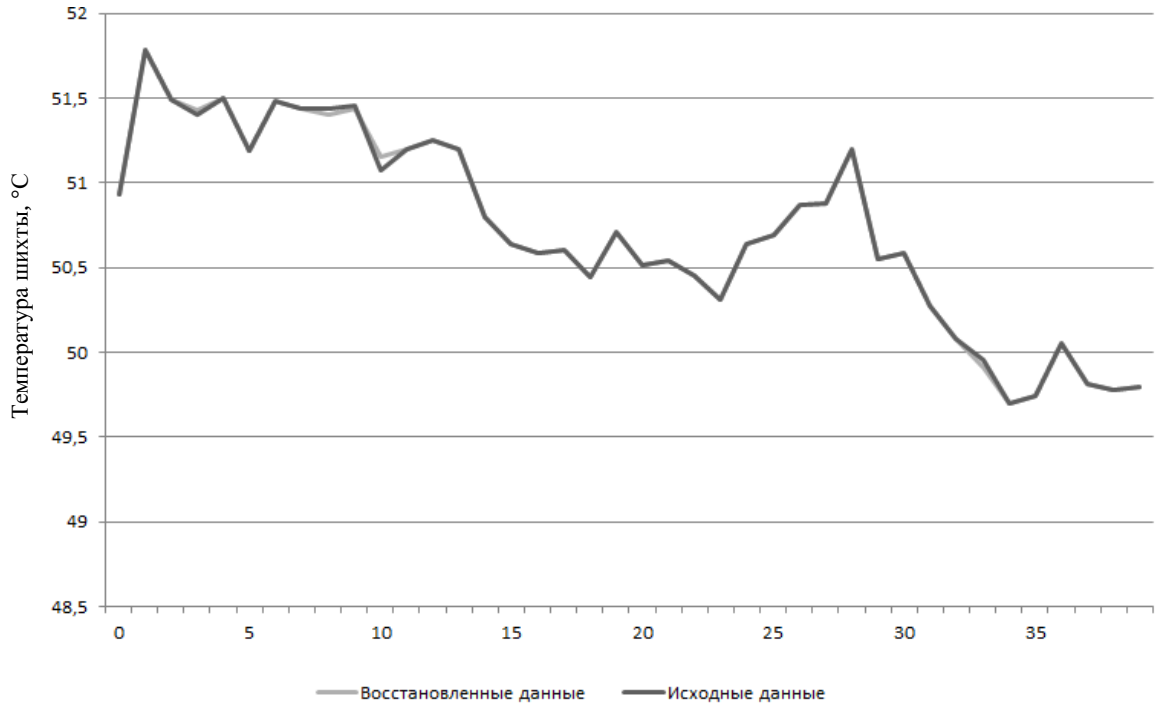


Рис. 1. Графики изменения температуры шихты на агломашине № 10

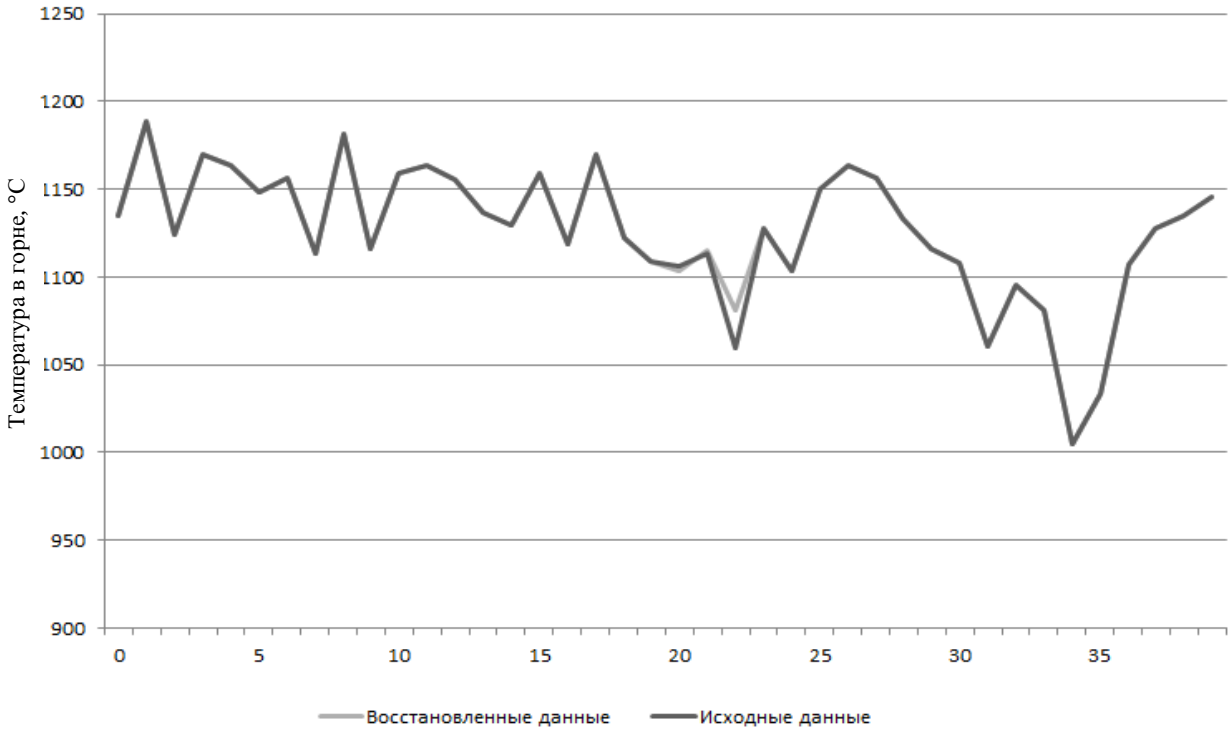


Рис. 2. Графики изменения температуры в секциях зажигательного горна

Таким образом, в системе прогнозирования электропотребления на агломерационном производстве целесообразно использовать алгоритм восстановления пропущенных значений, который повышает достоверность прогнозных значений. В основе алгоритма восстановления информации лежит использование метода интерполяции сплайнами для одиночных пропусков и Zet-алгоритма для групповых пропусков, причем первыми восстанавливаются одиночные пропущенные элементы, а затем полученная выборка обрабатывается с помощью Zet-алгоритма. В статье проанализированы результаты применения описанного метода, сделан вывод, что предложенный алгоритм чувствителен к резким скачкам в показателях, но на участках, где функция изменения значений во времени имеет гладкий график, элементы восстанавливаются с высокой точностью.

Литература

1. *Де Бор, К.М.* Практическое руководство по сплайнам / К.М. Де Бор. – М., 1985.
2. *Загоруйко, Н.Г.* Алгоритм заполнения пропусков в эмпирических таблицах (алгоритм Zet) / Н.Г. Загоруйко, В.Н. Елкина, В.С. Тимеркаев // Эмпирическое предсказание и распознавание образов. – Новосибирск, 1975. – Вып. 61: Вычислительные системы. – С. 3 – 27.
3. *Литтл, Р.Дж.А.* Статистический анализ данных с пропусками / Р.Дж.А. Литтл, Д.Б. Рубин. – М., 1990.

УДК 621.313.2:681.513.7:004.896

В.Н. Волков, А.В. Кожевников

ПОСТРОЕНИЕ ЭТАЛОННОЙ МОДЕЛИ ЭЛЕКТРОПРИВОДА ПОСТОЯННОГО ТОКА С ИСПОЛЬЗОВАНИЕМ РЕКУРРЕНТНЫХ НЕЙРОННЫХ СЕТЕЙ

В статье рассматривается использование искусственной нейронной сети для идентификации электропривода постоянного тока. Модель может быть использована для построения эталонных моделей в адаптивных системах.

Электропривод, самонастройка, эталонная модель, нейронные сети, адаптивные системы.

The use of an artificial neural network for identification of DC drive is considered in the article. The model can be used for constructing the reference models in adaptive systems.

Electric drive, self-adjustment, reference model, neural networks, adaptive systems.