

МЕТОДИКА ИДЕНТИФИКАЦИИ ПРОИЗВОДСТВЕННЫХ ХАРАКТЕРИСТИК НА ОСНОВЕ КЛАСТЕРИЗАЦИИ ДАННЫХ

Л.С. Казаринов, О.В. Попова

Основной задачей, стоящей в настоящее время перед промышленными предприятиями, является повышение энергетической эффективности производства. Одним из подходов решения поставленной задачи является выявление резервов повышения эффективности производственных систем на основе идентификации производственных характеристик их элементов по данным эксплуатации.

Реальные производственные характеристики, как правило, зависят от множества внутренних режимных факторов, поэтому не представляется возможным учесть все множество режимных факторов вследствие их большого многообразия для различных составляющих технологических процессов. Предлагается использовать в качестве производственных характеристик элементов сетей не только среднестатистические характеристики, но и мажорантные и минорантные, которые в совокупности описывают рабочую область производственных характеристик. В дальнейшем мажорантные и минорантные характеристики будем называть граничными.

В качестве примера на рис. 1 представлена паровая характеристика турбины электрической станции. Паровая характеристика состоит из базовой зависимости потребления пара (D_0) от выработанной электрической мощности (N_s) и тепловой мощности (Q_T). Реальная диаграмма отличается от базовой на величину поправок по внутренним режимным факторам: температуре и давлению свежего пара, давлению в теплофикационном и конденсационном отборах, температуре подачи сетевой воды и другим. В итоге паровая характеристика в линейном приближении будет иметь вид:

$$D_0 = a_0 + a_1 N_s + a_2 Q_T + \langle \text{поправки} \rangle. \quad (1)$$

Использование точной характеристики (1) при моделировании не представляется возможным, так как невозможно учесть все текущие значения внутренних факторов режимов турбин. Поэтому целесообразно использовать граничную внешнюю характеристику турбины, определенную при оптимальных сочетаниях внутренних факторов, например, по критерию минимума потребления пара:

$$D_0 = b_0 + b_1 N_s + b_2 Q_T.$$

Рассмотрим вопрос определения граничных производственных характеристик узлов на основе данных эксплуатации.

Решение задачи построения граничных характеристик предлагается осуществлять в данной работе на основе двух этапов. На первом этапе на основе решения задачи кластеризации данных из всей совокупности данных эксплуатации выделяются граничные (мажорантные и минорантные).

На втором этапе с использованием граничных данных находятся непосредственно граничные характеристики.

Постановка задачи кластеризации данных на мажорантные и минорантные имеет следующие особенности.

1. Наличие ошибок в измерении значений производственных характеристик элементов сетей и систем.

2. Неполнота информации о производственных характеристиках, содержащихся в данных эксплуатации.

Поэтому решение задачи кластеризации должно включать сглаживание ошибок в данных и регуляризацию постановки задачи.

С учетом сказанного в работе предложена следующая постановка задачи кластеризации, которую для определенности будем рассматривать как выделение мажорантных данных.

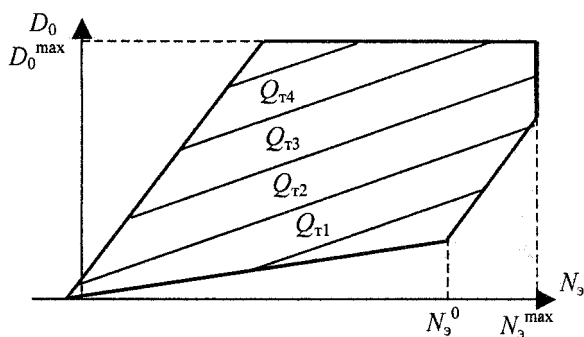


Рис. 1. Паровая характеристика турбины

В основе постановки задачи лежит система неравенств

$$p(\mathbf{a}, x_s) > y_s : \mu_s, s \in I_s; \quad (2)$$

$$p(\mathbf{a}, x_s) > y_s : \mu_s, s \in I_p, \quad (3)$$

где x_s, y_s — статистические данные о производственных характеристиках: x_s — входные данные, y_s — выходные эффекты; $p(\mathbf{a}, x_s)$ — дискриминантная функция, выделяющая мажорантные данные из общей совокупности данных; \mathbf{a} — вектор искомых параметров дискриминантной функции; μ_s — характеристическая функция соответствующего s -го неравенства ($\mu_s = 1$, если неравенство выполняется; $\mu_s = 0$ в противном случае); I_s, I_p — соответственно индексное множество экспериментальных и регуляризующих данных.

Система неравенств (2)–(3) соответствует двум наборам данных, используемых при решении задачи. Набор данных эксплуатации (I_s) отражает результаты наблюдений реальных значений про-

изводственных характеристик элементов сетей и систем в процессе эксплуатации. Регуляризирующий набор данных (I_p) является дополнительным и используется для восполнения недостающей информации, которой не содержится в данных реальных наблюдений. В качестве регуляризирующих наборов данных могут быть использованы нормативные данные, экспертные оценки и др.

Выбор значений вектора параметров дискриминантной функции (\mathbf{a}) осуществляется по критерию максимального взвешенного числа выполненных неравенств (2)–(3):

$$N_{\mathbf{b}} = (1-\lambda) \sum_{s \in I_s} \mu_s + \lambda \sum_{s \in I_p} \mu_s, \quad \lambda \in [0, 1],$$

при ограничении

$$N_{\text{маж}} = \sum_{s \in I_s \cup I_p} (1-\mu_s) \geq N_{\text{об}}^{\text{мин}} + N_{\text{пов}} = N_{\text{доп}}. \quad (4)$$

Здесь λ – вес регуляризирующих данных; $N_{\text{об}}^{\text{мин}}$ – минимально допустимый размер обучающей выборки данных, необходимых для определения неизвестных значений вектора \mathbf{a} ($N_{\text{об}}^{\text{мин}} = \dim \mathbf{a} + 1$); $N_{\text{пов}}$ – размер поверочной выборки данных, на которых оценивается точность построенной дискриминантной функции ($N_{\text{пов}} \geq 1$).

Вес регуляризирующих данных λ выбирается из условия корректности постановки задачи кластеризации. При $\lambda = 0$ задача решается только на экспериментальных данных и в общем случае является некорректно поставленной. При $\lambda = 1$ задача решается только на регуляризирующих данных и по определению является корректно поставленной. Однако при этом получаемое решение не отражает реальные экспериментальные данные. Поэтому при выборе значения коэффициента λ необходимо задавать минимально возможное значение, при котором сохраняется корректность постановки задачи.

Алгоритм решения задачи кластеризации следующий.

1. Задаются начальные значения веса регуляризирующих данных λ и весов данных $\alpha_s = 1$.

2. Методом взвешенных наименьших квадратов решается система уравнений

$$p(\mathbf{a}, x_s) = y_s : (1-\lambda)\alpha_s, \quad s \in I_s;$$

$$p(\mathbf{a}, x_s) = y_s : \lambda\alpha_s, \quad s \in I_p.$$

3. Вычисляется превышение выходного эффекта y_s над значением дискриминантной функции при входных факторах x_s

$$\Delta y_s = y_s - p(\mathbf{a}, x_s).$$

4. Определяются веса данных

$$\alpha_s = \begin{cases} 1, & \text{если } \Delta y_s > 0; \\ \gamma, & \text{если } \Delta y_s < 0, \end{cases}$$

где γ определяет скорость сходимости алгоритма. При $\gamma = 0$ – максимальная сходимость, однако при этом на первой итерации решения может выполняться неравенство $N_{\text{маж}} < N_{\text{доп}}$. При $\gamma = 1$ решение заканчивается на первой итерации, и в качестве мажорантных точек являются те, что лежат выше средней характеристики. Величина γ в общем случае выбирается из условия выполнения ограничения (4).

5. Данные, которые имеют вес $\alpha_s = 1$ являются мажорантными. Подсчитывается их количество $N_{\text{маж}}$.

6. Число мажорантных данных $N_{\text{маж}}$ сравнивается с допустимым числом данных $N_{\text{доп}}$.

Если на первой итерации решения $N_{\text{маж}} < N_{\text{доп}}$, то постановка задачи некорректна и необходимо дополнить число регуляризирующих данных. Если данное неравенство выполняется на итерации $\neq 1$, то необходимо вернуться к значениям переменных на предыдущей итерации.

Если $N_{\text{маж}} > N_{\text{доп}}$, то осуществляется переход к решению задачи взвешенных уравнений (к п. 2 алгоритма).

Решением задачи кластеризации является множество мажорантных данных, по которым затем методом наименьших квадратов строится непосредственно граничная характеристика.

Литература

1. Дюрбан Б. Кластерный анализ. – М.: Статистика, 1977. – 270 с.

2. Крянев А.В., Лукин Г.В. Математические методы обработки неопределённых данных. – М: ФИЗМАТЛИТ, 2003. – 216 с.