

О ЦЕЛЕСООБРАЗНОСТИ КЛАССИФИКАЦИИ МНОГОМЕРНЫХ ДАННЫХ*

Э.М. Фархадзаде¹, Ю.З. Фарзалиев¹

¹ Азербайджанский научно-исследовательский и проектно-изыскательский институт энергетики
e-mail: yuszey2002@yahoo.com

Резюме. Классификации многомерных данных по заданным разновидностям признаков целесообразно при условии неслучайного характера расхождения статистических функций распределения построенных по всей совокупности исходных данных $F_{\Sigma}^*(X)$ и выборке $F_V^*(X)$. В качестве критерия оценки принято не превышения оценке показателя надежности, вычисленного по экспериментальным данным выборки $\Pi_3^*(X)$, а критического значения этого показателя Π_k для заданного уровня значимости (ошибки первого рода).

Ключевые слова: Классификация, многомерные данные, надежность, разновидность признаков, выборка, совокупность, энергосистема.

AMS Subject Classification: 62D05.

1. Введение

Повышение эффективности решения эксплуатационных задач в электроэнергетической системах (ЭЭС) требует объективного учета надежности оборудования и устройств (объектов). Традиционно этот учет проводится, в основном, на качественном уровне (опыт работы объектов + интуиция + высокая квалификация персонала). С течением времени:

- доля объектов, срок службы которых превышает расчетный, стала не менее 50% и возрастает. Техническое состояние их ухудшается, возможности снижаются, требуют особого внимания;
- появление новых объектов с иными конструктивными исполнениями и принципами работы, системами управления, возрастающее разнообразие объема и норм испытания и ремонта объектов, снижает значимость накопленного опыта работы и требует повышения качества подготовки специалистов, систематической переподготовки персонала, повышения квалификации;
- автоматизированные системы управления режимами работы объектов ЭЭС, срок службы которых превышает расчетный, недостаточно учитывают изменение энергетических характеристик вследствие

* Работа была представлена на семинаре Института Прикладной Математики 26.04.2019

старения объектов и требуют совершенствования. А системы, предназначенные для непрерывного контроля технического состояния объектов, предоставляя уникальную информацию, к сожалению, не всегда формируют решение по повышению надежности.

Таким образом, методы традиционного учета надежности объектов требуют совершенствования. Одним из наиболее значимых направлений в этом является увеличение составляющей методической поддержки персонала в автоматизированных информационных системах анализа технического состояния объектов [4]. Рекомендации включают:

- ранжировку объектов по надежности и экономичности (эффективности) работы;
- указания на «слабые звенья» объектов, основанные причины, вызывающие ухудшение технического состояния;
- оценку качества:
 - управления режимами работы объектов;
 - восстановления износа в период планового ремонта;
 - консервации в период вынужденного простоя и ряд других

В настоящей статье приводятся метод и алгоритм решения основной составляющей этих рекомендаций – оценки значимости разновидностей признака.

1. Метод и алгоритм оценки значимости разновидности признака.

Известно, что при анализе отказов объектов ЭЭС учитывается большое число информационных признаков и, прежде всего, потому, что трудно с уверенностью утверждать, какие из них окажутся наиболее важными и полезными. Каждый из признаков характеризуется несколькими разновидностями (РП). На основе этой информации и рассчитываются показатели и характеристики надежности (ПН). Однако, при всем этом, вычисляются, как правило, усредненные оценки, а многообразие признаков и их разновидностей при расчетах ПН практически не учитываются. Эти усредненные количественные оценки надежности работы используются, прежде всего, для иллюстрации применения методов расчета ПН, решения отдельных проектных задач. К таким задачам относятся выбор схем распределительных устройств, оценка резерва мощности и др.

Для решения эксплуатационных задач, как правило, необходимы ПН сопоставляемых объектов, т.е. показатели и характеристики индивидуальной надежности (ИПН). Однако, ошибочно было бы думать, что оценки ИПН проводят по статистическим данным об отказах и восстановлениях конкретного генератора, трансформатора или выключателя. Специалистам хорошо известно, что такой информации просто нет. И когда мы говорим о ИПН имеются ввиду ПН, которые рассчитываются для значимых РП объектов. Традиционно, классификация статистических данных по

некоторым РП проводится и не представляет какой-либо трудности. Например, ПН рассчитываются для объектов различного класса напряжения, или различного конструктивного исполнения, или различного срока службы. Изредка ПН рассчитываются для двух РП. Например, оцениваются ПН линейных выключателей с номинальным напряжением 110кВ. При этом вопросы целесообразности классификации статистических данных по этим РП не рассматриваются.

Рассмотрим суть решаемой задачи. Пусть в результате сбора и обработки сведений об отказах и восстановлениях объектов ЭЭС мы располагаем некоторой совокупностью статистических данных, формализованных в виде эмпирической таблицы. Поскольку эта совокупность зависит от целого ряда случайных и неслучайных факторов, она относится классу многомерных и именуется конечная совокупность многомерных данных (КСМД). Для КСМД об отказах объектов ЭЭС характерно отсутствие генеральной совокупности и, как следствие, нецелесообразность применения при анализе разработанных для выборки из генеральной совокупности хорошо известных статистических методов. Итак, по КСМД требуется оценить ПН по некоторой группе РП.

РП задаются или соответствующими классификаторами или назначаются. При этом, как правило, число РП выбирается субъективно (максимально возможным) в соответствии со стремлением уточнить характер изменения ПН.

В свою очередь, лаконичная запись отдельных состояний объектов, например, в диспетчерских журналах, часто ограничивает возможное число признаков и их разновидностей. Уточнив РП, собрав и систематизировав в эмпирической таблице КСМД, перейдем к оценке значимости РП. Рекомендуемый метод основан на имитационном моделировании случайного характера оценок ПН и применении основных положений теории проверки статистических гипотез. Поскольку учет случайного характера оценок ПН требует сотни, а чаще тысячи реализаций, вычисления выполняются по разработанной компьютерной технологии. Эта технология состоит из следующих операций: [4]

1. Определяется усредненный ПН по всей КСМД. Обозначим его как Π_{Σ}^*
2. Для каждого из K_S сочетаний РП определяется выборка непрерывных случайных величин t_{003B}
3. Рассчитываются оценки ПН для каждой из K_S выборок. Обозначим их как $\Pi_{V,i}^*$ с $i=1, K_S$;
4. Проводится проверка предположения (гипотезы) H_1 о случайном характере различия $\Pi_{V,i}^*$ от Π_{Σ}^* для $i=1, K_S$. Методика такой проверки приведена в [5];

5. Выделяются сочетания РП, ПН которых неслучайно расходятся с Π_{Σ}^* ;

6. Определяются группы РП с неслучайно различающимися ПН;

7. Проводится ранжировка ПН этих групп в порядке увеличения ошибки второго рода, т.е. уменьшения мощности критерия. Тем самым, устанавливаются значимые сочетания РП, проводится классификация КСМД по уточненным РП [1].

Практическая реализация этого метода показала, что большое число возможных сочетаний РП вносит громоздкость в проведение вычислений и требует недопустимо большого времени счета. Так, при автоматизированном анализе закономерности изменения средней длительности простоя в аварийном ремонте энергоблоков 300 МВт на газомазутном топливе по месяцам года, когда число РП $K=12$, число сочетаний РП $K_S=4094$, а скорость анализа целесообразности классификации равна 10 сочетаний в минуту, время вычислений, даже при уменьшении числа реализаций N в 25 раз, оказалось неприемлемым.

Однако эти вычисления позволили установить:

1. Сочетания РП, включающие незначимые РП незначимы. Так, в соответствии с таблицей 1, если из восьми РП лишь три незначимы, величина K_S уменьшается в 8,5 раза;

2. Число значимых сочетаний РП не превышает число незначимых сочетаний РП. Следовательно оценку целесообразности классификации КСМД по заданным РП необходимо проводить путем поиска значимых сочетаний РП;

3. С увеличением абсолютной величины относительного отклонения ПН $\Pi_{V,i}^*$ от Π_{Σ}^* с $i=1, K_S$ ошибка первого рода (α) результата сравнения $\Pi_{V,i}^*$ и Π_{Σ}^* уменьшается, а ошибка второго рода (β) – возрастает. Это заключение определяет способ ранжирования ПН выборок, которые предполагается сравнивать с Π_{Σ}^* ;

4. Расчетными следует считать такие сочетания РП, для которых относительные отклонения ПН РП имеют одинаковый знак. Так если для некоторого j_1 –го значимого РП ПН равен Π_{V,j_1}^* , а для j_2 –го значимого РП

ПН равен Π_{V,j_2}^* , то объединение j_1 и j_2 возможно, если

$$\delta\Pi_{V,j_2} = (\Pi_{\Sigma}^* - \Pi_{V,j_2}^*) / \Pi_{\Sigma}^* \text{ и } \delta\Pi_{V,j_1} = (\Pi_{\Sigma}^* - \Pi_{V,j_1}^*) / \Pi_{\Sigma}^* \text{ совпадают;}$$

Количественные оценки эксплуатационных показателей КУ энергоблоков 300 МВт Таблица 1.

| N п/п | Эксплуатационный показатель | | | Условный номер энергоблоков | | | | | | | |
|----------|----------------------------------|--------------------|-------------|-----------------------------|-------|-------|-------|-------|-------|-------|-------|
| | Наименование | Усл.об. | Ед.изм. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 1 | Средняя нагрузка | H_{CP} | т/ч. | 757 | 739 | 752 | 797 | 958 | 804 | 804 | 820 |
| 2 | Удельный расход питательной воды | B | т/ч. | 780 | 758 | 769 | 815 | 973 | 814 | 820 | 832 |
| 3 | Расход тепла «брутто» | Q_b | гДж/ч. | 2061 | 2019 | 2052 | 2161 | 2554 | 2178 | 2178 | 2215 |
| 4 | Температура воздуха после РВП | T_B | $^{\circ}C$ | 261 | 293 | 294 | 305 | 293 | 305 | 325 | 291 |
| 5 | Температура уходящих газов | $T_{y.g.}$ | $^{\circ}C$ | 113 | 109 | 117 | 154 | 110 | 157 | 151 | 146 |
| 6 | Коэффициент избытка воздуха | K_B | о.е. | 1,94 | 1,16 | 1,17 | 1,16 | 1,11 | 1,15 | 1,15 | 1,18 |
| 7 | Присос воздуха на тракте | ΔS | % | 43,3 | 41 | 43 | 42,6 | 39,5 | 41,3 | 42,8 | 44,5 |
| 8 | Потери тепла с уходящими газами | $Q_{y.g.}$ | % | 6,51 | 6,36 | 6,84 | 8,68 | 6,32 | 8,81 | 8,72 | 8,45 |
| 9 | КПД «брутто» | η_b | % | 91,55 | 90,95 | 91,21 | 87,88 | 92,34 | 89,45 | 90,29 | 88,22 |
| 10 | Удельный расход ЭЭ на СН | \mathcal{E}_{CH} | % | 2,58 | 2,56 | 2,39 | 2,34 | 2,11 | 2,46 | 2,18 | 2,17 |
| 11 | Удельный расход ТЭ на СН | Q_{CH} | % | 2,45 | 1,88 | 1,74 | 2,62 | 1,03 | 2,73 | 1,98 | 2,65 |
| 12 | КПД «нетто» | η_n | % | 83 | 84,5 | 85,1 | 83,7 | 86,1 | 83 | 83,6 | 83,8 |

5. Рассмотренный выше алгоритм позволяет разделить РП на три группы. Первая группа РП имеет ПН равные Π_{Σ}^* , вторая группа РП имеет ПН равные $\Pi_V^* > \Pi_{\Sigma}^*$, а третья группа РП имеет ПН $\Pi_V^* < \Pi_{\Sigma}^*$. С учетом этих результатов рекомендуются следующие преобразования метода оценки значимости РП:

1. Для каждого из заданных К РП определяется выборка непрерывных случайных величин τ (см. п.2 алгоритма);

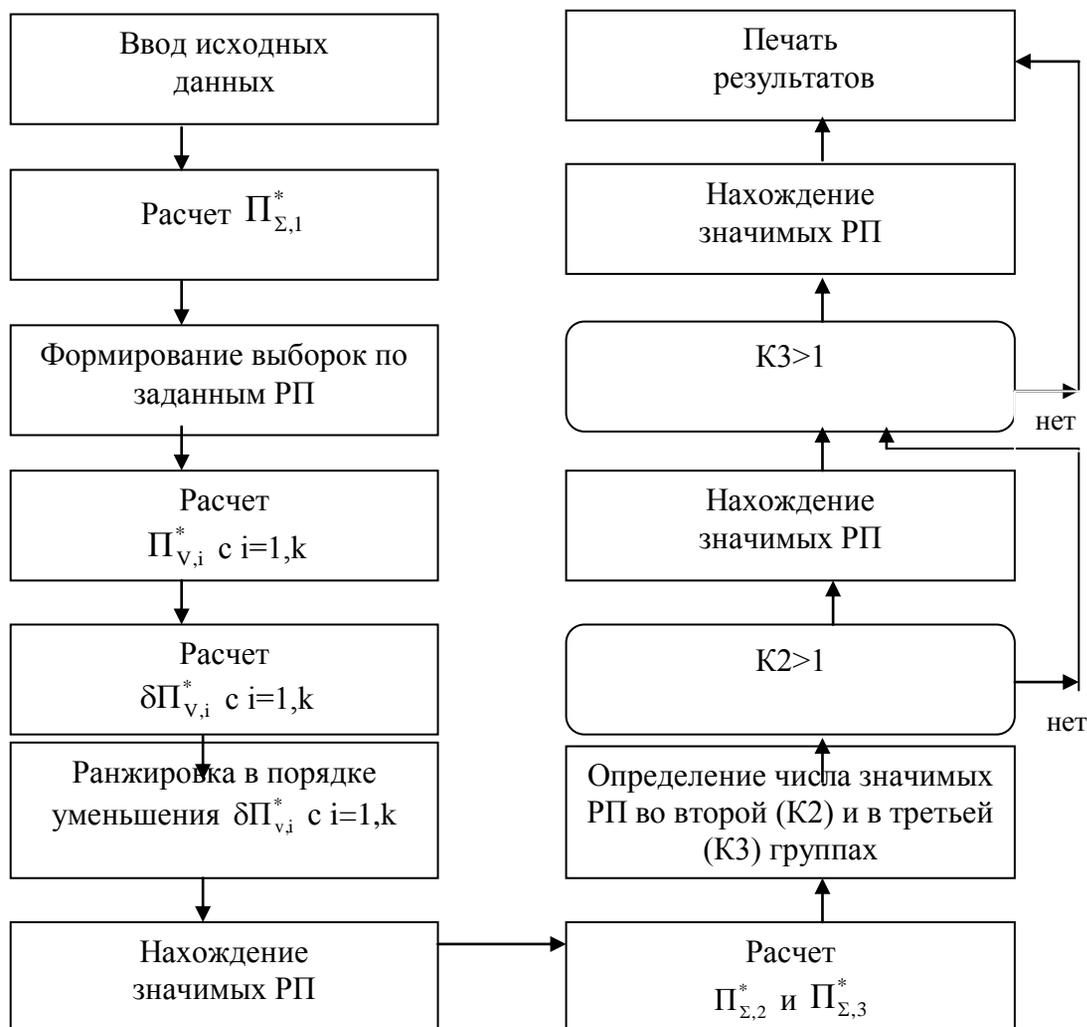
2. Рассчитываются оценки ПН для каждой из К выборок (см. п.3 алгоритма). Определяются относительные изменения каждой из $i=1, K$ оценок ПН по формуле $\delta\Pi_{V,i}^* = |(\Pi_{\Sigma}^* - \Pi_{V,i}^*)| / \Pi_{\Sigma}^*$. Проводится ранжировка абсолютных значений $\delta\Pi_{V,i}^*$ с $i=1, K$ в порядке их убывания. Выделяется наибольшее (первое) значение абсолютных величин $\delta\Pi_{V,\max}^*$;

3. Проводится проверка предположения (гипотезы H_1) о случайном характере различия $\Pi_{V,\max}^*$ от Π_{Σ}^* (см. п.4 алгоритма). Метод сравнения Π_{Σ}^* и Π_V^* зависит от типа ПН. Если, например, Π^* есть модель распределения случайной величины длительности простоя в аварийном ремонте, то сравниваются статистические функции распределения (с.ф.р.) КСМД. $F_{\Sigma}^*(\tau_{ав})$ и с.ф.р. выборки $F_V^*(\tau_{ав})$ в соответствии с [6]. Если наблюдаемое расхождение Π_{Σ}^* и Π_V^* случайно, то расхождение с Π_{Σ}^* ПН, вычисленных для остальных РП также будет случайным. Иначе говоря, классификация КСМД по рассматриваемым РП нецелесообразна. В противном случае, когда $\Pi_{V,\max}^*$ неслучайно отличается от Π_{Σ}^* , переходим к ПН следующего в вариационном ряду РП и проверяем характер его расхождения с Π_{Σ}^* . Этот процесс продолжается до тех пор, пока расхождение Π_{Σ}^* и Π_V^* не окажется случайным.

4. Формируются три группы выборок из КСМД. В первую (I) группу входят РП, оценки ПН которых случайно различаются от Π_{Σ}^* . Эти K_1 РП изымаются из полного перечня РП, как незначимые РП. Во вторую (II) группу входят РП, оценки ПН которых неслучайно больше Π_{Σ}^* . Если число РП второй группы (K_2) более одного то, в качестве упрощения, ПН этой группы вычисляются как среднее арифметическое оценок ПН РП второй группы. Этой оценке ПН сопоставляется укрупненная РП. Например, если во вторую группу вошли объекты с номинальным напряжением 110 и 220кВ, то укрупненной РП будет (110 – 220) кВ. Для РП третьей (III) группы ПН $\Pi_{V,III}^*$ вычисляется как среднее

арифметическое выбороч случайных величин $\tau_{ав}$, для которых $\Pi_{\Sigma}^* > \Pi_{V,i}^*$, а Π_V^* не случайно отличается от Π_{Σ}^* . ПН $\Pi_{V,III}^*$ также сопоставляется укрупненная РП [2;3].

На рисунке 1 приведена укрупненная блок-схема алгоритма решения задачи о значимости РП



Заключение. 1. Разработан метод, алгоритм и программа оценки значимости разновидностей признаков;
2. Существенным преимуществом этого метода является возможность повысить объективность решения многих эксплуатационных задач по располагаемым статистическим данным;
3. Результаты исследований позволяют перейти от традиционного анализа статистических данных эксплуатации оборудования и устройств электроэнергетических систем как представительной выборки из генеральной совокупности к методам анализа, учитывающих многомерный характер этих данных.

ЛИТЕРАТУРА

1. Орлов А.И. Организационно-экономическое моделирование: М.: Изд-во МГТУ им. Н. Баумана, 2012. -624с.
2. Панде П., Холп Л. Что такое «шесть сигм»? Революционный метод управления качеством. / Пер. с англ. – М.: Альпина Бизнес Букс, 2004. -158.
3. Смирнова О.С. Программное обеспечение для статистического анализа и заводская лаборатория. Диагностика материалов. 2008. Т. 74. №5. - с. 68-75.
4. Фархадзаде Э.М., Фарзалиев Ю.З., Мурадалиев А.З. Автоматизированное формирование рекомендаций для повышения надежности и экономичности энергоблоков ТЭС М., Энергетик №2, 2015, 15-17с.
5. Фархадзаде Э.М., Фарзалиев Ю.З., Мурадалиев А.З. Сравнение показателей усредненной и индивидуальной надежности оборудования электроэнергетических систем. Электричество. №12, 2015, 31-37 с.
6. Эфрон Б. Нетрадиционные методы многомерного статистического анализа. М.: Финансы и статистика, 1988. -263с. – ISBN 5 – 279 – 00277 -11.

ABOUT EXPEDIENCY OF CLASSIFICATION OF MULTIDIMENSIONAL DATA

E.M. Farhadzadeh¹, Yu.Z. Farzaliyev¹

¹Azerbaijan Scientific-Research and Design-Prospecting Institute of Energetics
e-mail: yuszey2002@yahoo.com

ABSTRACT

Classifications of multidimensional data on the set kinds of signs expediently on condition of nonrandom character of a divergence of statistical functions of distribution of the basic data constructed on all population $F_{\Sigma}^*(X)$ and selection $F_s^*(X)$. As evaluation criterion it is accepted not excess to an assessment of the indicator of reliability calculated on experimental data of selection $A_e^*(X)$, and critical value of this indicator A_c for the set significance value (an ERROR I Type).

Keywords: Classification, multidimensional data, reliability, type of features, sampling, combination,

power grid.

References

1. Orlov A.I. Organizatsionno-ekonomicheskoe modelirovanie: M.: Izd-vo MGTU im. N. Baumana, 2012. -624s. –(Orlov A.I. Organizational and economic modeling: Moscow: Pub N. Bauman MGTU, 2012, p-624) (In Russian)
2. Pande P., Kholp L. Chto takoe «shest' sigm»? Revolyutsionnyy metod upravleniya kachestvom. / Per. s angl. – M.: Al'pina Biznes Buks, 2004. -158. (Panda P., Holp L. What is “Six Sigma”? A revolutionary quality management method. / Translated from English - Moscow: Alpina Business Books, 2004, p.158) (In Russian)
3. Smirnova O.S. Programmnoe obespechenie dlya statisticheskogo analiza i zavodskaya laboratoriya. Diagnostika materialov. 2008. T. 74. №5. - s. 68-75. (Smirnova O.S. Software for statistical analysis and factory laboratory. Diagnostics of materials. 2008. T. 74. №5. pp.68-75) (In Russian)
4. Farkhadzade E.M., Farzaliyev Yu.Z., Muradaliyev A.Z. Avtomatizirovannoe formirovanie rekomendatsiy dlya povysheniya nadezhnosti i ekonomichnosti energoblokov TES M., Energetik №2, 2015, 15-17 s. (Farhadzadeh E.M., Farzaliyev Yu.Z., Muradaliyev A.Z. Automated generation of recommendations to improve the reliability and efficiency of power units of thermal power plants Moscow, Energetic, №2, 2015, pp.15-17) (In Russian)
5. Farkhadzade E.M., Farzaliyev Yu.Z., Muradaliyev A.Z. Sravnenie pokazateley usrednennoy i individual'noy nadezhnosti oborudovaniya elektroenergeticheskikh sistem. Elektrichestvo. №12, 2015, 31-37 c. (Farhadzadeh E.M., Farzaliyev Yu.Z., Muradaliyev A.Z. Comparison of indices of the average and individual reliability of electric power systems' equipment. Electricity, №12, 2015, pp.31-37) (In Russian)
6. Efron B. Netraditsionnye metody mnogomernogo statisticheskogo analiza. M.: Finansy i statistika, 1988. -263s. – ISBN 5 – 279 – 00277 -11. (Efron B. Nontraditional methods of multidimensional statistical analysis. Moscow: Finance and Statistics, 1988. p263. - ISBN 5 - 279 - 00277-11.) (In Russian)