

М.С. Дедовец, О.В. Сенько**МЕТОДЫ АНАЛИЗА НАДЕЖНОСТИ, ОСНОВАННЫЕ
НА ПОИСКЕ ЗАКОНОМЕРНОСТЕЙ В ДАННЫХ**

Рассмотрен новый метод анализа надежности, основанный на построении оптимальных статистически верифицированных разбиений пространства прогностических переменных, а также метод оценивания кривых вероятности отказов для отдельных объектов, основанный на принятии коллективных решений по системам закономерностей. Разработанные методы не требуют априорных предположений о характере зависимости и могут быть использованы при решении широкого круга задач анализа надежности различного технологического оборудования.

Ключевые слова: метод анализа надежности, кривые вероятности отказов, закономерности, надежность технологического оборудования.

Введение

Методы распознавания и прогнозирования, основанные на принятии коллективных решений по ансамблям закономерностям, в настоящее время получили достаточно широкое распространение [1, 7]. Многочисленные эксперименты продемонстрировали их эффективность в задачах с высокой размерностью данных. Целью настоящей работы является разработка новых методов для анализа и прогноза надежности, основанных на поиске эмпирических закономерностей в данных, связанных с надежностью, и построении по ним коллективных решений прогностических решений для отдельных объектов (изделий). При этом в качестве прогнозируемой характеристики надежности используется кривая вероятности безотказной работы до какого-то момента времени. Для поиска статистически верифицированных закономерностей предлагается использовать модификацию метода оптимальных достоверных разбиений [2, 3]. Метод оптимальных верифицированных разбиений. Оптимальные разбиения пространства прогностических переменных ищутся внутри априори фиксированных семейств разбиений по некоторой начальной обучающей информации

$$\tilde{S}_0 = \{(\alpha_i, t_i, x_i), \dots, (\alpha_m, t_m, x_m)\},$$

где t_i – момент последнего наблюдения за объектом, s_i , α_i – бинарная индикаторная функция, показывающая функционировал ли объект s_i в момент t_i или в это время произошел отказ, x_i – вектор значений прогностических переменных для объекта s_i . Целью оптимизации разбиений является достижение максимальных различий уровня надежности в соседних областях, (элементах разбиений). Семейство разбиений определяется как множество разбиений с ограниченным числом элементов, которые строятся с помощью одной и той же процедуры. В ходе исследований рассматривались одномерные и двумерные семейства. Простейшее одномерное семейство 1 включает все разбиения с двумя элементами, разделенными одной граничной точкой; семейство 2 включает все разбиения не более, чем с тремя элементами, разделенными двумя граничными точками. Двумерное семейство 3 включает в себя всевозможные разбиения двумерных областей допустимых значений для пар прогностических перемен-

ных с не более, чем четырьмя элементами, разделенными с помощью двух граничных. Качество разбиений оценивается или расхождение между уровнями надежности в группах объектов, индуцируемых его элементами оценивается с помощью функционала качества $F(R, \tilde{S}_0)$.

Пусть $g_{km}(t, \tilde{S})$ – оценка кривой вероятности отказа, рассчитанных согласно методике Каплан-Мейера [5] (КМ оценка) для группы объектов $\tilde{S}$. Другими словами $g_{km}(t, \tilde{S})$ является оценкой вероятности $\Pr(t' < t)$ где t' – время отказа для объекта из той же самой генеральной совокупности, из которой взяты объекты из $\tilde{S}$. Предположим, что некоторое разбиение R пространства прогностических переменных индуцирует разбиение $\{\tilde{S}_1, \dots, \tilde{S}_k\}$ обучающей выборки $\tilde{S}_0$.

Пусть $\rho[g_{km}(t, \tilde{S}'), g_{km}(t, \tilde{S}'')] = \int_0^T [g_{km}(t, \tilde{S}') - g_{km}(t, \tilde{S}'')]^2 dt$ – расстояние между КМ оценками кривых вероятности отказов, по выборкам $\tilde{S}'$, $\tilde{S}''$, где T – максимальное время наблюдений. Тогда величина функционала качества $F(R, \tilde{S}_0)$ для некоторого разбиения R вычисляется в виде

$$F(R, \tilde{S}_0) = \sum_{i=1}^k \left\{ |\tilde{S}_i| \rho[g_{km}(t, \tilde{S}_i), g_{km}(t, \tilde{S}_0)] \right\},$$

где $|\tilde{S}_i|$ – число объектов в выборке $\tilde{S}_i$.

Оптимальное разбиение R_{opt} внутри некоторого семейства ищется путем перебора всевозможных значений функционала $F(R, \tilde{S}_0)$ и выбора разбиения соответствующего его максимальной величине. Найденные закономерности (оптимальные разбиения) статистически верифицируются с помощью перестановочного теста (РТ) проверяющего нулевую гипотезу о полной независимости уровня надежности от объясняющих переменных. Оптимальное значение $F(R, \tilde{S}_0)$ по сути является статистикой перестановочного теста. Предположим, что оптимальное разбиение было найдено по выборке $\tilde{S}_0 = \{(\alpha_i, t_i, x_i), \dots, (\alpha_m, t_m, x_m)\}$, и $F(R_{opt}, \tilde{S}_0)$ – оптимальное значение функционала качества. Для оценки статистической достоверности выявленной закономерности множество из N случайных перестановок $\{\pi_1, \dots, \pi_N\}$ генерируется с помощью датчика случайных чисел. Исходная таблица (выборка) и перестановки $\{\pi_1, \dots, \pi_N\}$ порождают случайные таблицы $\{\tilde{S}_1^r, \dots, \tilde{S}_N^r\}$, где $\tilde{S}_j^r = \{(\alpha_{\pi_j(1)}, t_{\pi_j(1)}, x_1), \dots, (\alpha_{\pi_j(m)}, t_{\pi_j(m)}, x_m)\}$. Для каждой выборки $\tilde{S}_{\pi_j}^r$. Из $\{\tilde{S}_1^r, \dots, \tilde{S}_N^r\}$, оптимальное разбиение ищется внутри того же самого семейства и для тех же самых переменных, что и при поиске оптимального разбиения для обучающей выборки $\tilde{S}_0$. Пусть $N_{gt}[F(R_{opt}, \tilde{S}_0)]$ – число выборок из $\{\tilde{S}_1^r, \dots, \tilde{S}_N^r\}$, для которых $F(R_{opt}, \tilde{S}_j^r) > F(R_{opt}, \tilde{S}_0)$. Отношение $N_{gt}[F(R_{opt}, \tilde{S}_0)] / N$ используется в качестве оценки р-значения. Обычно для оценки р-значений с помощью перестановочных тестов используется от 1000 до 5000 случайных перестановок. Важными преимуществами перестановочных тестов является отсутствие предположений о вероятностных распределениях, а также применимость при произвольных размерах выборок.

Коллективные методы при прогнозировании надежности для отдельных объектов

Часто бывает важным не только оценить влияние на вероятность отказа отдельных прогностических переменных, но также оценить возможность отказа для отдельных изделий. В настоящее время для решения подобных задач суще-

стствует целый ряд подходов, включая использование распространенной модели Кокса [6]. Однако существующие методы обычно требуют существенных предположений о вероятностных распределениях. В этом разделе предлагается подход к построению кривых вероятностей отказов для кривых, основанный на принятии коллективных решений по системам подобластей пространства прогностических переменных, найденных с помощью описанного ранее метода ОВР. Пусть $\tilde{Q}$ – система подобластей пространства прогностических переменных, являющихся элементами оптимальных разбиений. Предположим, что нам требуется оценить кривую вероятности отказов для объекта s^* по его векторному описанию x^* . Пусть x^* принадлежит к пересечению подобластей $q_1, \dots, q_r$ из системы $\tilde{Q}$. Мы предлагаем оценивать вероятность $\Pr(t > t_0)$ для объекта s^* с помощью взвешенной суммы

$$\Pr(t > t_0) = \frac{\sum_{i=1}^r wei_i g_{km}(t, \tilde{S}_i)}{\sum_{i=1}^r wei_i},$$

где wei_i – так называемый «вес» подобласти q_i .

По аналогии с методом Статистически взвешенные синдромы [6] «вес» вычисляется по формуле

$$wei(\tilde{S}_i) = \frac{|\tilde{S}_i|}{|\tilde{S}_i| + 1} \cdot \frac{1}{\hat{d}_i},$$

где $\hat{d}_i = \sqrt{\frac{\sum_{s \in \tilde{S}_i} \rho[g_{km}(t, \tilde{S}_i), g_{km}(t, s)]}{|\tilde{S}_i|}}$, где $\rho[g_{km}(t, \tilde{S}_i), g_{km}(t, s)]$ – расстояние между

кривой вероятности отказов, рассчитанной по группе объектов $\tilde{S}_i$, и кривой вероятности отказов для отдельного объекта s из этой группы. Для случая s с известным моментом «отказа» t^* «кривая» вероятности отказов представляет собой ступенчатую функцию равную 1 на полуинтервале $[0, t^*]$ и 0 на полуинтервале $[t^*, T]$. Но мы сталкиваемся с проблемой оценки «кривой» для случая цензурирования, когда в последний момент наблюдений объект функционировал и действительный момент отказа неизвестен. Оценку ступенчатых функций в этом случае предлагается свести к задаче распознавания с двумя классами. При этом классы задаются с помощью граничных точек на временной оси. Верификация результатов распознавания производится с помощью техники скользящего контроля, а также на множестве цензурированных объектов. Состояние цензурированного объекта (функционирование или отказ), а также соответствующая ступенчатая функция определяются по результатам распознавания.

Заключение

Таким образом, предложен новый метод анализа надежности, основанный на построении оптимальных статистически верифицированных разбиений пространства прогностических переменных, а также метод оценивания кривых вероятности отказов для отдельных объектов, основанный на принятии коллективных решений по системам закономерностей. Разработанные методы не требуют априорных предположений о характере зависимости и могут быть

использованы при решении широкого круга задач анализа надежности технологического оборудования в горной промышленности: выявление факторов, влияющих на вероятность отказов; оценку надежности отдельных узлов; оптимизации условий эксплуатации и т.д.

СПИСОК ЛИТЕРАТУРЫ

1. Журавлев Ю.И. Рязанов В.В., Сенько О.В. Распознавание. Математические методы. Программная система. Практические применения. – М.: Фазис, 2006.
2. Sen'ko O.V., Kuznetsova A.V. The use of partitions constructions for stochastic dependencies approximation. Proceedings of the International conference on systems and signals in intelligent technologies. – Minsk, 1998. – pp. 291–297.
3. Senko O.V., Kuznetsova A.V., Kropotov D.A. The Methods of Dependencies Description with the Help of Optimal Multistage Partitioning. Proceedings of the 18th International Workshop on Statistical Modelling Leuven, Belgium. 2003. – pp. 397–401.
4. Sen'ko O.V., Kuznetsova A.V. The Optimal Valid Partitioning Procedures. Statistics on the Internet <http://statioumals.net/>. April, 2006.
5. Kaplan E.L., Meier P. Nonparametric estimation from incomplete observations // J. Amer. Stat. Assoc. 1958, v. 53, pp. 457–481.
6. Cox D.R. Regression models and life tables // J. R. Statist. Soc. B., pp. 34–187.
7. Кузнецов В.А., Сенько О.В., Кузнецова А.В. и др. Распознавание нечетких систем по методу статистически взвешенных синдромов и его применение для иммуногематологической нормы и хронической патологии. // Химическая физика. – 1996. – т. 15. – № 1. – С. 81–100. **ГИАЗ**

КОРОТКО ОБ АВТОРАХ

Дедовец М.С. – аналитик, «АрстелКонсалтинг»,
Сенько О.В. – e-mail: senkoov@mail.ru,
Вычислительный центр им. А.А. Дородницына РАН, МГИ НИТУ «МИСиС».

UDC 681.3.019.3

METHODS OF RELIABILITY ANALYSIS BASED ON SEARCHING OF REGULARITIES IN DATA

Dedovets M.S., analyst, «ArstelKonsalting», Russia,
Sen'ko O.V., e-mail: senkoov@mail.ru, Computing center named after A.A. Dorodnicyn RAS,
Moscow Mining Institute, National University of Science and Technology «MISiS», Moscow, Russia.

The article presents a new reliability analysis method based on building optimal, statistically verified partitions of space of predictive variables and a method of estimation of life curves for different objects, based on collective decision-making using sets of objective laws. These methods require no a priori assumption on the character of a functional connection and are of use to solving wide range problems concerned with analysis process equipment reliability.

Key words: reliability analysis, life curves, functional connections.

REFERENCES

1. Zhuravlev Yu.I. Ryazanov V.V., Sen'ko O.V. Raspoznavanie. Matematicheskie metody. Programmnaya sistema. Prakticheskie primeneniya (Identification. Mathematical methods. Program system. Practical applications), Moscow, Fazis, 2006.
2. Sen'ko O.V., Kuznetsova A.V. The use of partitions constructions for stochastic dependencies approximation. Proceedings of the International conference on systems and signals in intelligent technologies. Minsk, 1998. pp. 291–297.
3. Senko O.V., Kuznetsova A.V., Kropotov D.A. The Methods of Dependencies Description with the Help of Optimal Multistage Partitioning. Proceedings of the 18th International Workshop on Statistical Modelling Leuven, Belgium. 2003. pp. 397–401.
4. Sen'ko O.V., Kuznetsova A.V. The Optimal Valid Partitioning Procedures. Statistics on the Internet <http://statioumals.net/>. April, 2006.
5. Kaplan E.L., Meier P. Nonparametric estimation from incomplete observations. J. Amer. Stat. Assoc. 1958, v. 53, pp. 457–481.
6. Cox D.R. Regression models and life tables. J. R. Statist. Soc. B., pp. 34–187.
7. Kuznetsov V.A., Sen'ko O.V., Kuznetsova A.V. Khimicheskaya fizika. 1996, vol. 15, no 1, pp. 81–100.