

УДК 621.438.003

Т.П. ГРЫЗЛОВА¹, А.О. БАЛЫКИНА²¹*Рыбинская государственная авиационная технологическая академия
им. П.А. Соловьева, Рыбинск, Россия*²*Группа компаний «Тензор», Рыбинск, Россия*

СИСТЕМА ОЦЕНКИ ИНФОРМАТИВНОСТИ ДИАГНОСТИЧЕСКИХ ПРИЗНАКОВ И ПРИЗНАКОВЫХ ПРОСТРАНСТВ

Представлена система для анализа информативности признаков и признаковых пространств, которая позволяет объективно выбрать диагностические признаки и признаковые пространства из заданного набора. Критерием информативности служит отношение среднего расстояния между классами к среднему внутриклассовому расстоянию в признаковом пространстве (пространстве образов). Критерий позволяет отобрать информативные составляющие в спектре сигналов, выбрать метод обработки сигналов для решения задач технической диагностики. Выполнен анализ информативности признаков в различных технических решениях задачи диагностики подшипников трансмиссии ГТД. Показано, что оценки информативности признаков помогают выбрать правильную постановку задачи классификации, метод и алгоритм принятия решений, метод обработки вибросигналов.

Ключевые слова: диагностика, признаки, информативность, набор признаков, внутриклассовое расстояние, межклассовое расстояние, метрики, критерии информативности, поиск информативных признаковых пространств, вибросигнал, подшипники трансмиссии ГТД.

Введение

Задача оценки информативности диагностических признаков является важным этапом разработки любой системы распознавания или автоматического анализа сигналов. Часто в качестве критерия информативности признаков выбирают вероятности правильного распознавания и ошибок первого и второго рода, оцениваемые по испытаниям классификатора в выбранном признаковом пространстве, а также системы распознавания в целом. При неудачном выборе признаков синтез такой системы оказывается напрасной работой.

Таким образом, оценка информативности признаков и объективный выбор алгоритмов вычисления признаков до начала разработки системы распознавания или обработки сигналов имеет очень большое практическое значение.

Кроме того, вопросы оценки информативности признаковых пространств представляют большой теоретический интерес, свидетельством которого является поток публикаций по новым методам и критериям анализа информативности, специальные конференции и гранты на исследования методов анализа информативности разнотипных признаков [1 – 3].

Целью работы была разработка методов анализа информативности заданного набора признаков, которые бы позволили объективно выбрать призна-

ки для решения задачи диагностики состояния сложного источника, а также разработка системы выделения наиболее информативных признаков и пространств из заданного набора.

Если размерность начального набора признаков или искомого пространства велика, то необходима автоматизация отбора информативного пространства признаков.

Решение задачи оценки информативности признаков и признаковых пространств осложняется отсутствием единого критерия информативности. Решающим критерием является, разумеется, величина потерь [4]. Но для ее оценки требуется спроектировать систему обработки сигналов и/или систему распознавания, а это как раз то, чего следует избегать, пока нет оценки информативности выбранных признаков. Известны критерии «трудности распознавания» – среднее значение евклидова расстояния между математическими ожиданиями всех пар сочетаний образов; разность между двумя ожидаемыми величинами логарифмов отношения правдоподобия векторов образов (дивергенция), энтропия распределений плотности вероятности образов в многомерном выборочном пространстве [4, 5]. Для семейства однотипных признаков полезен критерий, введенный в [6] – отношение среднего межклассового расстояния к среднему внутриклассовому расстоянию. Именно он выбран за основу в представляемой работе. Разработано алгоритмическое и программ-

ное обеспечение анализа информативности признаков, которое апробировано на множествах признаков, используемых в различных задачах распознавания, в том числе, и в задаче диагностики подшипников трансмиссии ГТД [7, 8].

1. Формализация задачи отбора информативных признаков

Система отбора информативных признаков в задаче распознавания состояния сложного источника реализует следующий алгоритм.

Исходные данные представлены в виде таблицы $\mathbf{Th}$, строки которой соответствуют образам $\mathbf{x}_k$, $k = 1 \dots K$, а столбцы – признакам $\mathbf{x}_i$, $i = 1 \dots n$. К таблице добавляется еще один столбец, его элементами являются индексы классов m , $i = 1 \dots M$, к которым относятся образы обучающей выборки $\mathbf{x}_k \leftrightarrow m_k$. $m_k \in [1..M]$ – номер класса, к которому относится образ $\mathbf{x}_k$.

Итак, K – количество образов $\mathbf{x}_k$, $k = 1 \dots K$ в обучающей выборке, а n – количество признаков, размерность пространства образов. Индекс $i = 1 \dots n$ – номер признака или координаты пространства образов. Элементами таблицы исходных данных являются значения признаков $\mathbf{x}_{ki}$

$$\mathbf{Th} = \begin{bmatrix} m_1 & x_{11} & \dots & x_{1i} & \dots & x_{1n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ m_k & x_{k1} & \dots & x_{ki} & \dots & x_{kn} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ m_K & x_{K1} & \dots & x_{Ki} & \dots & x_{Kn} \end{bmatrix} = \begin{bmatrix} \kappa & \mathbf{X} \end{bmatrix}, \quad (1)$$

где $\kappa: [1..K] \rightarrow [1..M]$ – характеристический вектор обучающей последовательности образов $\mathbf{x}_k$:

$$\kappa = \begin{bmatrix} \kappa_1 & \dots & \kappa_k & \dots & \kappa_K \end{bmatrix}^T = \begin{bmatrix} m_1 & \dots & m_k & \dots & m_K \end{bmatrix}^T; \quad (2)$$

$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 & \dots & \mathbf{x}_k & \dots & \mathbf{x}_K \end{bmatrix}^T$ – матрица образов.

Рассчитывается матрица $K \times K$ расстояний между образами, на основе этой матрицы можно оценить потенциальные ошибки распознавания по методу ближайшего соседа (МБС) – случай, когда минимальное расстояние от рассматриваемого образа класса m относится к образу, не принадлежащему классу m .

Расстояние или функция расстояния понимается как отображение $\rho: \mathbf{X} \times \mathbf{X} \rightarrow \mathbf{R}_+$, удовлетворяющее условиям:

- 1) $\rho(\mathbf{x}, \mathbf{x}) = 0, \forall \mathbf{x} \in \mathbf{X}$;
- 2) $\rho(\mathbf{x}_i, \mathbf{x}_j) = 0, \forall \mathbf{x}_i, \mathbf{x}_j \in \mathbf{X}$.

Для индексации функций пар образов используются индексы $ij = [1..K] \times [1..K], i \neq j$.

Далее рассчитываются внутриклассовые расстояния ρ_m и расстояния между классами ρ_{IJ} , $I, J \in [1..M] \times [1..M], I \neq J$, которые могут служить критериями информативности, поскольку характеризуют компактность расположения образов одного класса и удаленность образов разных классов в пространстве образов (признаковом пространстве).

Наконец, рассчитывается основной критерий информативности признакового пространства $\mathbf{F}$ заданной размерности как отношение среднего межклассового расстояния к среднему внутриклассовому:

$$\mathbf{I}(\mathbf{F}) = \frac{\bar{\rho}_{IJ}}{\bar{\rho}_m}. \quad (3)$$

Таким образом, оценка информативности признака f_j , $j = 1 \dots n$ получается с помощью алгоритма, реализующего отображение

$$\mathbf{I}(f_j): \mathbf{Th}^{<j>}, \mathbf{Th}^{<0>} \rightarrow \mathbf{R}_+$$

с учетом количества образов в каждом классе – K_m .

Очевидно, что $K = \sum_{m=1}^M K_m$.

Информативность набора признаков $s = (f_1, f_q, \dots, f_p)$, $l \neq q \neq \dots \neq p$, $l, q, p \in [1..n]$ оценивается алгоритмом

$$\mathbf{I}(\mathbf{F}_s): \mathbf{Th}^{<s>}, \mathbf{Th}^{<0>} \rightarrow \mathbf{R}_+,$$

где $\mathbf{Th}^{<s>}$ – столбцы таблицы исходных данных, соответствующие оцениваемому набору.

Информативные признаки и пространства отбираются по результатам вычисления информативности пространств образов, получаемых прямым перебором.

Остановимся на случае, когда образом объекта является сигнал. Матрица $\mathbf{X}$ теперь содержит отсчеты сигналов, N – количество отсчетов:

$$\mathbf{X} = \begin{bmatrix} s_{10} & \dots & s_{1(N-1)} \\ \dots & \dots & \dots \\ s_{K0} & \dots & s_{K(N-1)} \end{bmatrix}. \quad (4)$$

Если признаковое пространство обычно имеет размер 2 – 20, то длина выборки сигнала составляет несколько десятков тысяч. Отбор признаков в такой постановке нецелесообразен, но общая оценка информативности сигналов и методов их обработки может оказаться полезной.

Например, образом может быть спектральная плотность мощности, полученная с помощью преобразования Фурье по выборке сигнала длиной N на

частотах, кратных коинтервалу Найквиста ω_1 . Тогда матрица образов равна

$$\mathbf{X} = \begin{vmatrix} S_1(0) & S_1(\omega_1) & \dots & S_1\left(\omega_1 \cdot \frac{N}{2}\right) \\ \dots & \dots & \dots & \dots \\ S_K(0) & S_K(\omega_1) & \dots & S_K\left(\omega_1 \cdot \frac{N}{2}\right) \end{vmatrix}. \quad (5)$$

На основе преобразования Фурье может быть построен образ фаз колебаний с частотами, кратными коинтервалу Найквиста

$$\mathbf{X} = \begin{vmatrix} Ph_1(0) & Ph_1(\omega_1) & \dots & Ph_1\left(\omega_1 \cdot \frac{N}{2}\right) \\ \dots & \dots & \dots & \dots \\ Ph_K(0) & Ph_K(\omega_1) & \dots & Ph_K\left(\omega_1 \cdot \frac{N}{2}\right) \end{vmatrix}. \quad (6)$$

Очевидно, оценка информативности отдельных спектральных составляющих в спектре сигнала полезна для разработки методики диагностики. Выбор составляющих с максимальной плотностью мощности, как это принято делать в настоящее время, оказывается не самым удачным, что показано ниже.

Этап предобработки данных важен для разработки системы распознавания, поскольку позволяет, заранее оценив исходные данные, значительно упростить классификатор или отказаться от построения системы распознавания в выбранном пространстве признаков. Важнейшим элементом предобработки данных является выбор признакового пространства. Формирование информативного пространства минимальной размерности позволяет упростить вычисления и повысить точность результатов классификации.

Признаковое пространство системы распознавания должно соответствовать критериям полноты описания объекта и минимальной размерности пространства.

2. Расстояния в пространстве образов

В теории распознавания большое распространение получили метрические алгоритмы классификации, которые основаны на гипотезе компактности – предположении, что схожие объекты, как правило, отображаются в одну область признакового пространства. Для формализации понятия «сходства» вводится функция расстояния $\rho(\mathbf{x}, \mathbf{x}_0)$ в пространстве образов $\mathbf{X}$. Расстояние должно быть введено так, чтобы объекты одного класса были компактно расположены, а объекты разных классов были удалены друг от друга (расхождение).

Метод основан на понятии хорошего размеще-

ния образов в признаковом пространстве, введенном в теории распознавания [1, 2, 4, 5]. Мы говорим, что образы хорошо размещены, если образы, характеризующие объекты одного класса расположены компактно, расстояния между ними малы. Образы, характеризующие объекты разных классов, должны быть удалены друг от друга, расстояния между ними должны быть велики. Естественно вводится относительный критерий – отношение межклассового расстояния к внутриклассовому.

Система анализа информативности включает в себя алгоритм расчета внутриклассовых (внутримножественных) расстояний, алгоритм расчета расстояний между классами (множествами), алгоритм вычисления критерия информативности. Параметром системы является указание на способ вычисления расстояния: евклидово, Чебышева и др. Входными данными является таблица $\mathbf{Th}$ (1) – таблица «класс – образ», выходными – значения внутримножественных (внутриклассовых) расстояний, расстояний между классами и значения критерия информативности. Подсистема отбора информативных признаков формирует списки признаков и признаковых пространств с заданными пороговыми значениями критерия информативности.

Для расчета информативности используются следующие способы вычисления расстояний. Обозначим евклидово расстояние между точками признакового пространства, которые соответствуют образам k и l классов I и J как $d(\mathbf{x}_{I,k}, \mathbf{x}_{J,l})$. Евклидово расстояние в многомерном пространстве равно

$$d(\mathbf{x}_{I,k}, \mathbf{x}_{J,l}) = \sqrt{\sum_{i=1..n} (\mathbf{x}_{I,k}(i) - \mathbf{x}_{J,l}(i))^2}. \quad (7)$$

Здесь i – номер признака, n – количество признаков, размерность пространства образов.

Расстоянием Чебышева между многомерными числовыми векторами называется максимум модуля разности компонент этих векторов:

$$\rho_{ch}(\mathbf{x}_{I,k}, \mathbf{x}_{J,l}) = \max_{i \in 1..n} |\mathbf{x}_{I,k}(i) - \mathbf{x}_{J,l}(i)|. \quad (8)$$

Расстояние Камберра является относительным, компоненты вектора разности нормированы:

$$\rho(\mathbf{x}_{I,k}, \mathbf{x}_{J,l}) = \sum_{i=1}^n \frac{|\mathbf{x}_{I,k}(i) - \mathbf{x}_{J,l}(i)|}{|\mathbf{x}_{I,k}(b) + \mathbf{x}_{J,l}(i)|}. \quad (9)$$

Внутриклассовое Евклидово расстояние

$$D_m = \frac{2 \sum_{k,l=1..m, k \neq l} d(\mathbf{x}_{m,k}, \mathbf{x}_{m,l})}{K_m(K_m - 1)}, \quad (10)$$

где K_m – количество образов в классе m .

Евклидово расстояние между классами равно:

$$D_{I,J} = \frac{\sum_{k=1}^{K_I} \sum_{l=1}^{K_J} d(\mathbf{x}_{I,k}, \mathbf{x}_{J,l})}{K_I \cdot K_J}. \quad (11)$$

Отношение среднего межклассового Евклидова расстояния в признаковом пространстве к среднему внутриклассовому расстоянию в признаковом пространстве $\mathbf{F}$ выбрано критерием информативности:

$$\mathbf{I}(\mathbf{F}) = \frac{\sum_{I,J=1; I \neq J}^M D_{I,J}}{(M-1) \sum_{m=1}^M D_m}. \quad (12)$$

Особенности представляют собой случаи двух и нескольких классов. Так, образы первого и второго класса могут быть отделены от третьего, а между собой пересекаться. В этом случае целесообразно использовать минимальное отношение межклассового Евклидова расстояния к сумме их внутриклассового расстояния:

$$\mathbf{I}(\mathbf{F})_{\min} = \min_{I,J} \frac{D_{I,J}}{D_I + D_J}. \quad (13)$$

3. Использование системы оценки информативности на примере задачи диагностики состояния подшипников трансмиссии ГТД

Продemonстрируем полезность оценки информативности признаков на примере признаков, используемых в задаче диагностики состояния подшипников трансмиссии ГТД. Множество классов обучающей выборки может быть определено как $\mathbf{C}$ – кондиционный, $\mathbf{B}$ – неисправный, $\mathbf{C}_0$ – кондиционный, правильно распознанный с помощью прибора ИВУ-1 и $\mathbf{N}$ – кондиционный, ошибочно распознанный как плохой подшипник прибором ИВУ-1 и необоснованно снятый. Задача диагностики может ставиться как задача распознавания двух классов –

$$\{\mathbf{C}, \mathbf{B}\}, \mathbf{C} = \mathbf{C}_0 \cup \mathbf{N}$$

или трех классов –

$$\{\mathbf{C}, \mathbf{N}, \mathbf{B}\}, \mathbf{C} = \mathbf{C}_0.$$

Результаты оценки информативности при этом будут различны, поэтому ниже приводятся оценки информативности для случая двух и трех классов подшипников.

Обучающая выборка состоит из 7 подшипников типа Р12: два подшипника класса $\mathbf{C}_0$, два подшипника класса $\mathbf{N}$ и три подшипника класса $\mathbf{B}$.

В задаче распознавания на два класса характеристический вектор равен

$$\kappa = [1 \ 1 \ 1 \ 1 \ 2 \ 2 \ 2]^T,$$

в задаче распознавания на три класса –

$$\kappa = [1 \ 1 \ 2 \ 2 \ 3 \ 3 \ 3]^T.$$

Заметим, что из обучающей выборки 30 вибросигналов, представленной «ОАО «Сатурн», исключен ряд подшипников, так как она содержит вибросигналы подшипников с разным числом тел качения, причем в ней преобладают исправные подшипники типа Р12 и неисправные подшипники типа Р7.

В табл. 1 приведена оценка информативности 256 отсчетов вибросигналов s_0^{255} (начало сигнала s_0^{29999}) и s_{25000}^{25255} (конец сигнала), оценка информативности вибросигнала, проквантованного на два уровня, а также количество ошибок распознавания по методу ближайшего соседа (МБС).

Таблица 1

Информативность вибросигналов

Данные	Количество ошибок МБС		Информативность, $n = 256$	
	Три класса	Два класса	Три класса	Два класса
Сигнал (конец)	4	2	0,593	0,561
Сигнал (начало)	3	2	0,608	0,554
Бинарный сигнал	7	6	0,495	0,532

При распознавании на два класса информативность начала сигнала ниже, чем конца; при распознавании на три класса информативность начала сигнала выше (3 ошибки – 4 ошибки). Если сигнал преобразовать к бинарному, то информативность падает, а количество ошибок растет. В целом, информативность сигнала низка. Это означает, что требуется обрабатывать сигнал и вводить диагностические признаки.

Для диагностики состояния подшипников трансмиссии ГТД используются разнообразные признаки, представленные в табл. 2. Требуется объективный выбор технического решения. Объективным критерием потенциального качества распознающей системы является информативность признаков, которую можно оценить с помощью «Системы оценки информативности диагностических признаков и признаковых пространств», описанной выше.

В табл. 2 приведено максимальное значение информативности $\mathbf{I}_{\max}$, в первой строке – в задаче распознавания на два класса, во второй – на три. Видно, что информативность метода, используемого в ИВУ-1 выше (0,692).

Таблица 2

Информативность диагностических признаков в технических решениях
задачи диагностики подшипников трансмиссии ГТД

№	Система	Признаки	$I_{\max}$	Примечание
1	ИБУ-1	СКО вибросигнала	0,692 0,572	Эксплуатируется
2	МІС-200 (ВДК-44)	Спектральные	0,519 0,733	Применяется для исследований
3	На основе расчета информативности	Спектральные	0,511 4,465	Предложение
4	Bearing	Статистические, информационные характеристики, регрессия сигналов и последовательностей Wavelet-коэффициентов	0,827 0,462	НИР
5	На основе метода характерных последовательностей (ХП)	Частоты встречаемости ХП	~ 3,744 Данные отсутствуют	НИР
6	На основе анализа графического представления реализаций вибросигналов	Статистика яркостей, геометрические характеристики, отношения сигнал/шум	0,67 2,182	Исследуются
7	На основе декомпозиции вибросигнала по характеристикам Wavelet-коэфф-ов	Статистика сегментов	1,744	Исследуются
8	На основе НВ-анализа (полувольтной анализ)	Статистика и динамика полувольт	0,841 3,785	Исследуются

Если обратить внимание на динамику поведения стрелки ИБУ-1, которую можно определить вектором

$$\mathbf{x} = \left[\sigma(s_0^{4999}) \quad \sigma(s_{5000}^{9999}) \quad \dots \quad \sigma(s_{25000}^{29999}) \right],$$

то получим следующий вектор информативности:

$$\mathbf{I}(\sigma_i) = [0,547 \quad 0,562 \quad 1,076 \quad 0,658 \quad 0,524 \quad 0,83].$$

Спектральные признаки, определяемые с помощью МІС200 не информативнее, чем СКО при разбиении выборки на два класса, но информативнее при распознавании трех классов. Если спектральные признаки отобрать по критерию информативности, то можно найти признаки с информативностью больше 1 в диапазонах частот 20 – 96 Гц, 370 – 410 Гц и 3760 – 3800 Гц для задачи распознавания двух классов и с информативностью 4,465 в диапазоне частот 20 – 60 Гц для распознавания трех классов. Также есть большой набор информативных пространств ($I > 4$) с дополнительными диапазонами 1560 – 1870 Гц и 2430 – 5000 Гц. Если в качестве образа выбрать фазовую характеристику, то информативность признаков в задаче распознавания двух классов можно повы-

сить до 2,381 (диапазоны 4340 – 4380 Гц и 880 – 920 Гц), а для трех классов результаты почти не меняются.

Приложение **Bearing**, описанное в [8], позволило в интерактивном режиме с помощью визуального анализа отобрать признаковое пространство, в котором классификатор, обученный по методу потенциальных функций безошибочно разделял все подшипники на два класса. Это было среднее значение сигнала и нормированная информационная мера конца сигнала $NPIM(s_{25000}^{29999})$.

Как показали расчеты, информативность этого пространства составляет 0,827 в задаче распознавания двух классов, для которой, в общем-то, оно и выбиралось.

Показатель информативности пространства $\left[\bar{M}(s_0^{29999}), NPIM(s_{25000}^{29999}) \right]$ выше, чем у признаков, используемых в настоящее время. Тем не менее, исследования продолжались, было предложено семейство признаков с очень высоким показателем информативности – частоты встречаемости характерных последовательностей (ХП) в сигналах разных классов [6, 7]. Метод МХП имеет хо-

рошие показатели, но является очень сложным и трудоемким.

В настоящее время исследуется ряд принципиально новых для диагностики методов, основанных на анализе сложной структуры сигнала (метод декомпозиции и НВ-анализ) и метод распознавания графиков вибросигналов (изображения). Показатели информативности признаков, используемых в этих методах высоки, а у метода анализа статистик полувольт (НВ-анализ) выше, чем у метода ХП, при том, что он на несколько порядков проще МХП.

Заключение

В работе предложен метод оценки информативности признаков пространств на основе расчета межклассовых и внутриклассовых Евклидовых расстояний. Метод позволяет отбирать наиболее информативные пространства признаков, оценить разделимость или линейную разделимость образов в пространстве признаков. Метод позволяет определить наиболее информативное признаковое пространство из набора признаков при отсутствии разделимости классов.

По значению критерия информативности можно получить рекомендации по построению классификатора, можно судить о возможном типе классификатора: линейный классификатор, классификатор по методу ближайшего соседа, классификаторы на основе потенциальных функций.

Система оценки информативности диагностических признаков и признаков пространств позволяет объективно оценить информативность каждого признака.

Добавление признака с большой информативностью часто не увеличивает, а даже снижает общую информативность по критерию отношения межклассового расстояния к среднему внутриклассовому, поскольку увеличиваются внутриклассовые расстояния, а межклассовые расстояния меняются незначительно.

Система позволяет:

- подобрать набор информативных признаков для разных типов алгоритмов, как метрических, так и логических;
- выявить причины неинформативности признаков и признаков пространств;

- определить, необходима или нет обработка сигнала, выбрать метод обработки;
- определить, необходимо ли вычислять характеристики сигналов;
- выбрать наиболее информативные составляющие в спектре сигналов.

Об эффективности системы можно судить по представленным результатам анализа вариантов решения задачи диагностики подшипников трансмиссии ГТД. Подход к разработке алгоритмов диагностики на основе объективного выбора диагностических признаков и технических решений по испытательным приборам представляется перспективным и необходимым.

Литература

1. *Methods of Recognition Based on the Function of Rival Similarity* / N.G. Zagoruiko, I.A. Borisova, V.V. Dubanov, O.A. Kutnenko // *Pattern Recognition and Image Analysis*. – 2008. – V. 18. – P. 1-6.
2. *Сходство и компактность* / И.А. Борисова, В.В. Дюбанов, Н.Г. Загоруйко, О.А. Кутненко // *Математические методы распознавания образов: 14-я Всероссийская конференция*. – М.: МАКС Пресс, 2009.
3. Колесникова С.И. *Методы анализа информативности разнотипных признаков* / С.И. Колесникова // *Вестник Томского государственного университета*. – 2009. – № 1 (6). – С. 69-80.
4. Загоруйко Н.Г. *Методы распознавания и их применение* / Н.Г. Загоруйко. – М.: Советское радио, 1972. – 320 с.
5. Ту Дж. *Принципы распознавания образов* / Дж. Ту, Р. Гонсалес. – М.: Мир, 1978. – 460 с.
6. Gorshkov A.P. *A Family of Effective Features for Condition Diagnostics of Complex Technical Systems from the Example of GTE Transmission Bearings* / A.P. Gorshkov, T.P. Gryzlova // *Pattern Recognition and Image Analysis*. – 2009. – Vol. 19, No. 4. – P. 641-646.
7. *Диагностика состояния подшипников трансмиссии газотурбинных двигателей в пространствах статистик характерных последовательностей вибраций* / А.П. Горшков, Т.П. Грызлова, Б.И. Комаров, В.Т. Шепель // *Авиационно-космическая техника и технология*. – 2006. – № 10 (36). – С. 152 – 156.
8. Шепель В.Т. *Выбор признаков для диагностики технического состояния подшипников трансмиссии ГТД* / В.Т. Шепель, Б.И. Комаров, Т.П. Грызлова // *Авиационно-космическая техника и технология*. – 2005. – № 8 (24). – С. 200-205.

Поступила в редакцию 6.05.2011

Рецензент: д-р техн. наук, проф., начальник КО сертификации авиационных двигателей В.Т. Шепель, «ОАО «Сатурн», Рыбинск, Россия.

СИСТЕМА ОЦІНКИ ІНФОРМАТИВНОСТІ ДІАГНОСТИЧНИХ ОЗНАК І ОЗНАКОВИХ ПРОСТОРІВ

Т.П. Гризлова, А.О. Баликіна

Представлена система для аналізу інформативності ознак і ознакових просторів, яка дозволяє об'єктивно вибрати діагностичні ознаки і ознакові простори із заданого набору. Критерієм інформативності служить відношення середньої відстані між класами до середньої внутрікласової відстані в ознаковому просторі (просторі образів). Критерій дозволяє відібрати інформативні складові в спектрі сигналів, вибрати метод обробки сигналів для вирішення завдань технічної діагностики. Виконаний аналіз інформативності ознак в різних технічних рішеннях задачі діагностики підшипників трансмісії ГТД. Показано, що оцінки інформативності ознак допомагають вибрати правильну постановку завдання класифікації, метод і алгоритм ухвалення рішень, метод обробки вібросигналів.

Ключові слова: діагностика, ознаки, інформативність, набір ознак, внутрікласова відстань, міжкласова відстань, метрики, критерії інформативності, пошук інформативних ознакових просторів, вібросигнал, підшипники трансмісії ГТД.

DIAGNOSTIC FEATURE'S AND PATTERN SPACES' INFORMATIVITY ESTIMATION SYSTEM

T.P. Gryzlova, A.O. Balykina

A methodical solution is proposed for the problem of feature search based on informativity estimations of characteristics of digital signals. The diagnostic feature's and pattern spaces' informativity estimation system makes possible an objective test of feature set. Informativity for evaluation of feature set that is the ratio of the mean interclass Euclidean distance in a feature space to the mean intraclass distance is used. The result is a pattern space set with proper clustering. In the example of the problem of system diagnostics of transmission bearings in a gas_turbine engine, the advisability of change to a new approach for choice of features' set is shown. Informativity helps in choice of spectrum coefficients with high informativity. Also it helps to choose signal processing method to solve diagnostics problems. There was shown that informativity estimations of features are helpful for correct formalization of classification problem and for choice of vibration signal processing method and decision-making behavior.

Key words: diagnostics, features, informativity, feature set, interclass Euclidean distance, intraclass distance, distance function, automatic feature search, vibration, transmission bearings, gas_turbine engine.

Грызлова Татьяна Павловна – канд. техн. наук, доцент кафедры МПО ЭВС РГТА им. П.А. Соловьева, Рыбинск, Россия, e-mail: ktnpgryzlova@mail.ru.

Балыкина Ася Олеговна – магистр техники и технологий, программист группы компаний «ТЕНЗОР», Рыбинск, Россия, e-mail: tsuzuki_asmodei@mail.ru.