

УДК 681.518.2:519.816-616.65-006

Усовершенствованный метод и информационная технология решения задачи классификации состояния элементов сложных систем

И. М. Антонян², В. А. Горячая¹, А. И. Зеленский², Е. М. Угрюмова¹¹Национальный аэрокосмический университет им. Н. Е. Жуковского «ХАИ»,²Харьковская медицинская академия последипломного образования, Украина

Усовершенствован метод решения задачи классификации состояния элементов сложных систем. Предложена системная модель процесса диагностирования элементов медико-биологической системы (пациентов). На основе анализа информативности и агрегирования переменных состояния выполнена редукция размерности пространства переменных состояния пациентов. Разработана компьютерная система поддержки принятия решений при медико-биологическом диагностировании состояния пациентов.

Ключевые слова: регрессионный анализ, редукция размерности, классификация состояния.

Удосконалено метод розв'язання задачі класифікації стану елементів складних систем. Запропоновано системну модель процесу діагностування елементів медико-біологічної системи (пацієнтів). На основі аналізу інформативності та ад'єктивування змінних стану виконана редукція розмірності простору змінних стану пацієнтів. Розроблено комп'ютерну систему підтримки прийняття рішень при медико-біологічному діагностуванні стану пацієнтів.

Ключові слова: регресійний аналіз, редукція розмірності, класифікація стану.

The method for solving the classification condition of elements complex systems has been improved. The system model of the diagnosing medical process and biological elements of the system (patients) has been proposed. Based on the information content analysis and aggregating state variables made the reduction of the dimension variables of the status patients. Computer decision support system for biomedical diagnose condition of patients has been developed.

Keywords: regression analysis, reduction of dimension, classification status.

1. Постановка проблемы и её актуальность

Развитие дефектов в технических или медико-биологических системах – сложный динамический процесс. Эксперты предметной области не всегда могут прогнозировать, насколько быстро они будут развиваться. Не всегда удается прийти к единому мнению на какой стадии развития находятся дефекты и, как следствие, какие методы устранения дефектов применять. Контроль и прогнозирование динамического процесса работы систем помогают экспертам и/или пациентам (если рассматриваются медико-биологические системы) принимать решения, которые приводят к лучшим значениям критериев качества работы систем или выживаемости и качества жизни пациентов.

В качестве примера сложной системы рассмотрим медико-биологическую систему, в состав которой входят следующие элементы: лечащий врач, пациенты и подсистема диагностирования состояния пациентов. Системная модель процесса диагностирования медико-биологической системы представлена на

рис.1, где Database – база данных; S – сумматор входов; Controller – управляющий орган – лечащий врач, разрабатывающий сценарии лечения пациентов; u – управляющие переменные; f – внешние воздействия (возмущения); Object of control – объект управления (пациенты); Z – переменные состояния пациентов; W – критерии качества состояния пациентов; Anamnez – анамнез; Laboratory diagnosis – лабораторная диагностика; Visual diagnosis – визуальная диагностика; The diagnosis – установление диагноза.

Начало и конец этапов жизненного цикла лечения пациентов определим как множество конечных состояний пациентов. Количество состояний, принятых для рассмотрения, устанавливается экспертом в предметной области на основе результатов кластерного анализа. Принимается, в качестве рабочей, гипотеза локального равновесия, согласно которой состояние пациента однозначно определяется фундаментальной системой его переменных состояния.

Будем считать, что критическим на множестве конечных состояний является резистентное состояние, при котором, вследствие прогрессирующего развития дефектов функциональных частей, пациент в процессе лечения теряет управляемость.

Объективная проблема заключается в том, что отсутствует структурированное решающее правило перехода для рассматриваемого пациента в резистентное состояние.

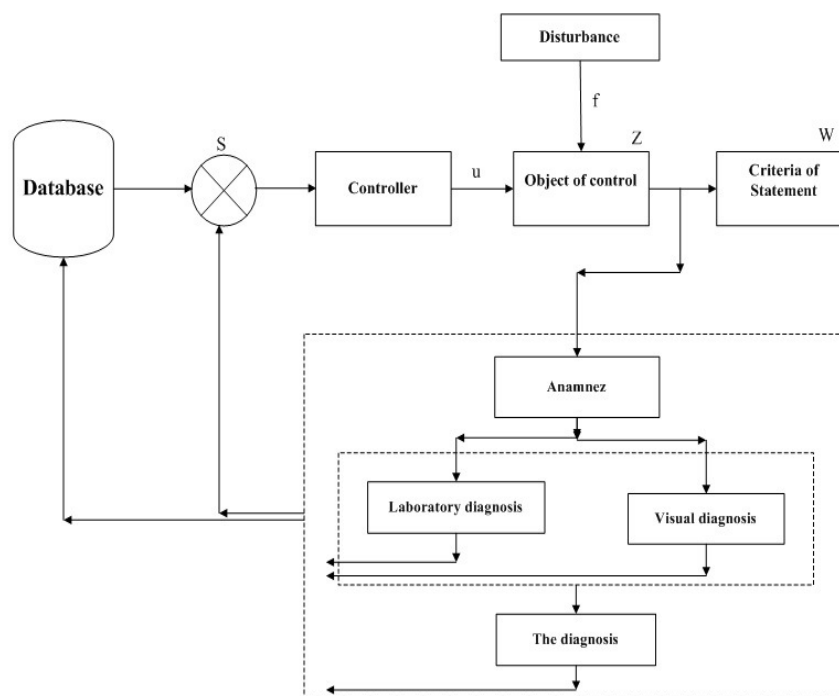


Рис. 1. Системная модель процесса диагностирования медико-биологической системы.

Рассмотрению задач теории и практики диагностирования состояния сложных систем уделяется большое внимание как учеными на Украине, так и за её пределами. К настоящему времени опубликовано множество работ, посвященных описанию методов решения задач диагностирования технических и медико-биологических систем (см., например [1-10]). В публикациях [1-3] достаточно подробно описаны общие положения теории обучаемых искусственных нейронных сетей (ИНС), которые широко применяются для построения формальных математических моделей (диагностических моделей) в форме уравнений регрессии. В работах [4-10] рассмотрены перспективы применения и развития методов классификации состояния (распознавания образов): формирование информативных переменных состояния; классификация на основе обучающих выборок; учет динамики переменных состояния объектов управления.

Следует отметить, в большинстве работ, посвященных решению задач диагностирования и прогнозирования на базе ИНС, отсутствует анализ информативности переменных диагностических моделей.

Анализ существующих литературных источников показывает, при разработке методов решения задач диагностирования, таких как построение диагностической модели, классификация и прогнозирование состояния, возникает ряд математических проблем:

- большая размерность задач;
- переменные состояния являются коррелируемыми величинами (мультиколлинеарность);
- низкая информативность переменных состояния;
- существует неопределенность в формировании и структурировании решающих правил классификации состояния элементов системы.

Для обеспечения достоверности полученной с помощью регрессионных моделей информации, эти модели должны обладать свойством стабильности в условиях априорной неопределенности входных данных, а также достаточной с практической точки зрения точностью.

Разработанное на сегодняшний день информационное обеспечение не позволяют с достаточно высоким уровнем достоверности решать задачи классификации состояния элементов сложных систем.

Таким образом, возникает необходимость в совершенствовании моделей, математических методов и средств информационной технологии обработки статистических данных для анализа состояния элементов динамических систем.

Целью данного исследования является анализ статистической зависимости между переменными, определяющими состояние элементов медико-биологической системы (пациентов); определение принадлежности пациента к определенному классу на основе данных регистрируемых переменных его состояния.

В процессе исследования решались следующие научные задачи: построение диагностической модели; редукция размерности вектора входных данных; агрегирование переменных состояния в комплексы, которые попарно были бы слабо коррелируемыми величинами (ортогонализация); классификация состояния пациентов.

2. Постановка и метод решения задачи построения диагностической модели

На основе системного анализа процесса диагностирования было выявлена иерархия этапов диагностирования: лабораторная диагностика (биохимические анализы крови и т.п.), визуальная диагностика (УЗИ, МРТ и т.п.) и соответствующие каждому этапу регистрируемые переменные состояния пациентов. Первоначальная размерность пространства переменных состояния была равна 24.

Далее была сформирована экспериментальная выборка регистрируемых переменных, характеризующих состояние наблюдаемых пациентов. В дальнейшем выборка была разбита на 3 класса: 50 человек, 45 человек и 52 человека. В качестве классификационного признака при делении общей выборки на классы был выбран уровень прогрессирования заболевания.

Пусть имеется многомерная матрица наблюдений $X = \{x_{i,j}\}$ ($i = 1..I, j = 1..J$), где I – количество наблюдаемых пациентов в выборке, J – количество измеряемых переменных состояния. Традиционно строки этой матрицы называются прецедентами. Выполним центрирование и нормирование данных:

$$x_{ij}^{\circ} = (x_{ij} - \langle X_j \rangle) / y_j, \quad (1)$$

где $\langle X_j \rangle$ – среднее значение j -ой переменной состояния; y_j – ее среднее квадратическое отклонение.

Сформулируем постановку задачи построения диагностической модели. Задана векторная функция набором обучающих пар $(\vec{X}^{(0)}, \vec{d})_p$, $p = 1..P$, где $\vec{X}^{(0)}, \vec{d}$ – векторы входа, размерности N_0 , и выхода, размерности N_{K+1} , соответственно. Необходимо аппроксимировать данную выборку. Результатом решения задачи должен являться некий математический механизм, в результате работы которого можно было бы получить любое значение векторной функции $\vec{Y}^{(K+1)}(\vec{X}^{(0)})$, представленной данной обучающей выборкой, по заданному вектору входа, в диапазоне, ограниченном входными данными.

Для выявления связей и закономерностей в развитии исследуемого процесса использовались методы корреляционного и регрессионного анализа [10-13].

Корреляционный анализ. С целью выявления статистической зависимости между случайными переменными, определяющими состояние пациентов, были получены точечные оценки коэффициентов парных корреляций для каждой из выделенных групп. Критерием качества была выбрана шкала состояния пациентов. На основе анализа коэффициентов парной корреляции были выявлены по каждой из групп переменные, для которых коэффициенты парных корреляций превышают выбранный уровень значимости (> 0.15) [11]. Таким образом, были выделены переменные состояния, влияние которых на шкалу состояния выявилось определяющим.

Регрессионный анализ. Рассматривались первые 3 класса пациентов: 50 человек, 45 человек и 52 человека. На основе нормированных переменных с использованием обобщенного метода наименьших квадратов (МНК) было получено уравнение линейной множественной регрессии в виде [12]:

$Y = \epsilon_0 + \epsilon_1 X_1 + \epsilon_2 X_2 + \dots + \epsilon_J X_J + e$, где β_j ($j=1..J$) – коэффициенты уравнения линейной множественной регрессии.

Оценка информативности переменных диагностической модели в форме коэффициентов значимости вклада переменных состояния в значения критерия

качества – $u_{\epsilon_j} = \frac{\epsilon_j^2}{\sum_{j=0}^J \epsilon_j^2}$, осуществлялась с помощью найденных параметров

уравнения регрессии, согласно критерию условной энтропии [2]:

$$I(Y|X_j, X_j) = \frac{1}{2} \log_2 \left(\frac{y_{Y|X_j}^2}{y_{X_j}^2} \right) = \frac{1}{2} \log_2 \epsilon_j^2.$$

В результате анализа информативности переменных было выявлено, что размерность пространства переменных состояния может быть снижена с 24 до 16 переменных. Коэффициент детерминации составил 0.7722; множественный коэффициент корреляции равен 0.87875, что указывает на адекватность полученной диагностической модели экспериментальным данным.

Таким образом, на основе проведенного анализа выявлено 16 информативных переменных из 24, определяющих состояние пациентов группах согласно выбранному критерию качества шкалы состояния пациентов.

3. Постановка и метод решения задачи классификации состояния элементов динамических систем

Сформулируем постановку задачи классификации. Пусть \vec{X}^* – вектор переменных, описывающих состояние прецедентов, M – множество номеров классов (сценариев). Известно количество возможных сценариев отказов системы в целом, а также для каждого сценария (класса) сформированы подмножества наблюдаемых переменных состояния (симптомов). По значениям проекций вектора \vec{X}^* прецедент относят к одному из возможных множеств R_m , где $m=0..M-1$. Необходимо найти такой m -й сценарий, для которого максимальна плотность распределения условной вероятности появления \vec{X}^* у прецедента при m -ом сценарии:

$$\exists! m^* \in C_m(\rho(\vec{X}_m^* | R_m))(m=0..M-1): \rho(\vec{X}_m^* | R_m) \rightarrow \max, \quad (2)$$

где $\rho(\vec{X}_m^* | R_m)$ – множество m -тых индексов плотность распределения условной вероятности появления \vec{X}^* у прецедента при m -ом сценарии.

В целях корректности решения задачи классификации был проведен факторный анализ переменных состояния пациентов.

Факторный анализ. Факторный анализ на основе метода главных компонент (ГК) был использован как метод агрегирования переменных состояния в комплексы, которые попарно были бы слабо коррелируемыми величинами (ортогонализация) с целью последующее редукции размерности пространства переменных состояния.

Как известно дисперсия случайных многомерных величин характеризуется корреляционной матрицей Σ .

Суть метода ГК состоит в декомпозиции матрицы значений переменных состояния X° (1):

$$X^\circ = TP^T = \sum_{\beta=1}^A t_\beta p_\beta^T, \quad (3)$$

где T – матрица счетов (scores) с размерностью $I \times A$; P – матрица нагрузок (loadings) с размерностью $J \times A$; A – число главных компонент.

Матрица нагрузок P является матрицей перехода из исходного пространства переменных состояния $X_1..X_J$ (J -мерного) в пространство главных компонент (A -мерное).

В методе ГК используются новые, агрегированные переменные t_α ($\alpha=1..A$), являющиеся линейной комбинацией исходных переменных X_j ($j=1..J$) [13]:

$$T = X^\circ (P^T)^{-1}.$$

Пусть Λ – матрица собственных значений корреляционной матрицы Σ , $\Lambda = \text{diag}\{\lambda_1, \dots, \lambda_A\}$.

Матрица искомым переменных (ГК) F размерностью $I \times A$, где $A < J$, определяется как линейная комбинация:

$$F = X^\circ L, \quad (4)$$

где L – матрица, состоящая из проекций собственных векторов матрицы Σ , соответствующих собственным числам λ_α матрицы Λ , на оси исходной системы координат пространства переменных состояния $X_1..X_J$.

При этом сама матрица L при $A=J$ по построению является ортогональной: $L^T L = LL^T = E$, $L^T = L^{-1}$, где E – единичная матрица.

Матрица факторных нагрузок определяется соотношением:

$$P = L \Lambda^{\frac{1}{2}}. \quad (5)$$

Строки матрицы P соответствуют исходным переменным, столбцы – факторам (или ГК). Для матрицы нагрузок имеет место следующие соотношение:

$$P^T P = \Lambda = \text{diag}\{\lambda_1, \dots, \lambda_A\}, \quad (6)$$

где величины $\lambda_1 \geq \dots \geq \lambda_A \geq 0$ – собственные значения ковариационной матрицы Σ . Выражение (3) с учетом (5) можно переписать в виде:

$$X^\circ = TP^T = T \left(L \Lambda^{\frac{1}{2}} \right)^T. \quad (7)$$

Элементы матрицы счетов T представляют собой нормированные ГК, получающиеся из обычных ГК умножением на $\Lambda^{-\frac{1}{2}}$: $T = F \Lambda^{-\frac{1}{2}}$, $T = X^\circ L \Lambda^{-\frac{1}{2}}$, откуда с учетом (4-7) имеем:

$$F = T \Lambda^{\frac{1}{2}} = X^\circ P \Lambda^{-\frac{1}{2}}. \quad (8)$$

Распознавание состояния. Задача распознавания состояния сводится к задаче классификации (2). В целях исключения мультиколлинеарности применим процедуру ортогонализации переменных состояния (8). Далее для решения задачи классификации состояния объекта использовалась вероятностная нейронная сеть, следующей структуры (рис. 2):

- входной слой $F_1^* \dots F_A^*$ – входными элементами являются значения проекций вектора наблюдаемых симптомов \vec{F}^* прецедента;
- слой образцов $\rho_{1m} \dots \rho_{Mm}$ – центры классов обучающей выборки.

Количество образцов равно количеству классов в обучающей выборке.

Входной слой и слой образцов образуют полносвязную структуру. Активность элемента слоя образцов определялась зависимостью, соответствующей плотности распределения вероятностей согласно t -закону Стьюдента (что уместно и для ограниченных выборок):

$$\rho_{lm} = \rho(\vec{F}_m^* | R_l) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n} \Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t_{lm}^2}{n}\right)^{-\frac{n+1}{2}}, \text{ где } \Gamma - \text{ гамма-функция Эйлера с } n$$

степенями свободы ($n = K_l + K_m - 2$; K_l, K_m – число прецедентов в соответствующих классах, $l, m = 1..M$);

t_{lm} – статистика Стьюдента:

$$t_{lm} = \sqrt{\frac{MD_{lm}^2}{\frac{1}{K_l} + \frac{1}{K_m}}}, \quad (9)$$

MD_{lm}^2 – расстояние Махаланобиса от неизвестного прецедента (полагая, что он относится к l -тому классу) до m -ого образца –

$MD_{lm}^2 = \frac{1}{A} (\vec{F}^* - \langle \vec{F}_m \rangle)^T \Sigma_{pooled}^{-1} (\vec{F}^* - \langle \vec{F}_m \rangle)$, где \vec{F}^* – значения проекций вектора ГК наблюдаемых симптомов неизвестного прецедента;

$\langle \vec{F}_m \rangle$ – средние значения проекций вектора ГК наблюдаемых симптомов элемента слоя образцов;

Σ_{pooled} – объединенная корреляционная матрица для рассматриваемых сценариев (классов) – определяется с учетом принадлежности прецедента к

тому, либо иному классу: $\Sigma_{pooled} = \frac{1}{K_l + K_m - 2} ((K_l - 1)\Sigma_l + (K_m - 1)\Sigma_m)$;

- выходной слой $m^*, \rho(\vec{F}_m^* | R_{m^*})$ (выходной элемент) представляет собой дискриминатор пороговой величины, указывающий элемент слоя образцов с максимальным значением активности (т.е. указывает класс, к которому принадлежит неизвестный прецедент).

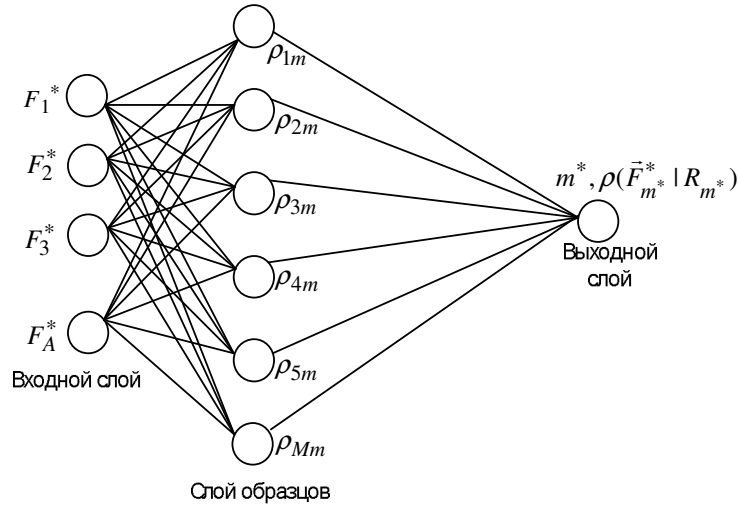


Рис. 2. Архитектура вероятностной нейронной сети.

Следует обратить внимание, что значения статистики Стьюдента (9) зависят от выбора базиса, в котором оценивается степень близости между прецедентом и образцами (при решении задачи классификации), а также между образцами (при анализе значимости расстояний между классами). Поэтому возникает необходимость в структурировании дополнительного статистического решающего правила выбора единственного (опорного) базиса. Согласно принципу максимального правдоподобия в качестве решающего правила выбора опорного базиса m^* было принято следующее:

$$\exists! m^* \in C_m(t_{lm}) (l, m = 0..M-1): \min_l t_{lm} \rightarrow \max, \quad (10)$$

где $C_m(t_{lm})$ – множество m -тых индексов статистик.

Согласно принципу максимального правдоподобия в качестве критерия перехода от одного к другому состоянию пациентов (памятуя, что они упорядочены по уровню прогрессирования заболевания) может быть использовано решающее байесовское правило: $\forall m = 0..M-1: \frac{\rho(\bar{F}_{m+1}^* | R_{m+1})}{\rho(\bar{F}_m^* | R_m)} \geq 1$,

которое справедливо при условии, что $\frac{P(R_m)}{P(R_{m+1})} \cdot \frac{\rho(\bar{X}_{m+1}^*)}{\rho(\bar{X}_m^*)} \approx 1$, где $P(R_m)$ – априорные вероятности реализации классов.

4. Информационная технология решения задачи классификации состояния элементов медико-биологической системы

Системная модель информационно-аналитического обеспечения процесса принятия решений лицом принимающим решения (ЛПР) при медико-биологическом диагностировании состояния пациентов представлена на рис. 3.

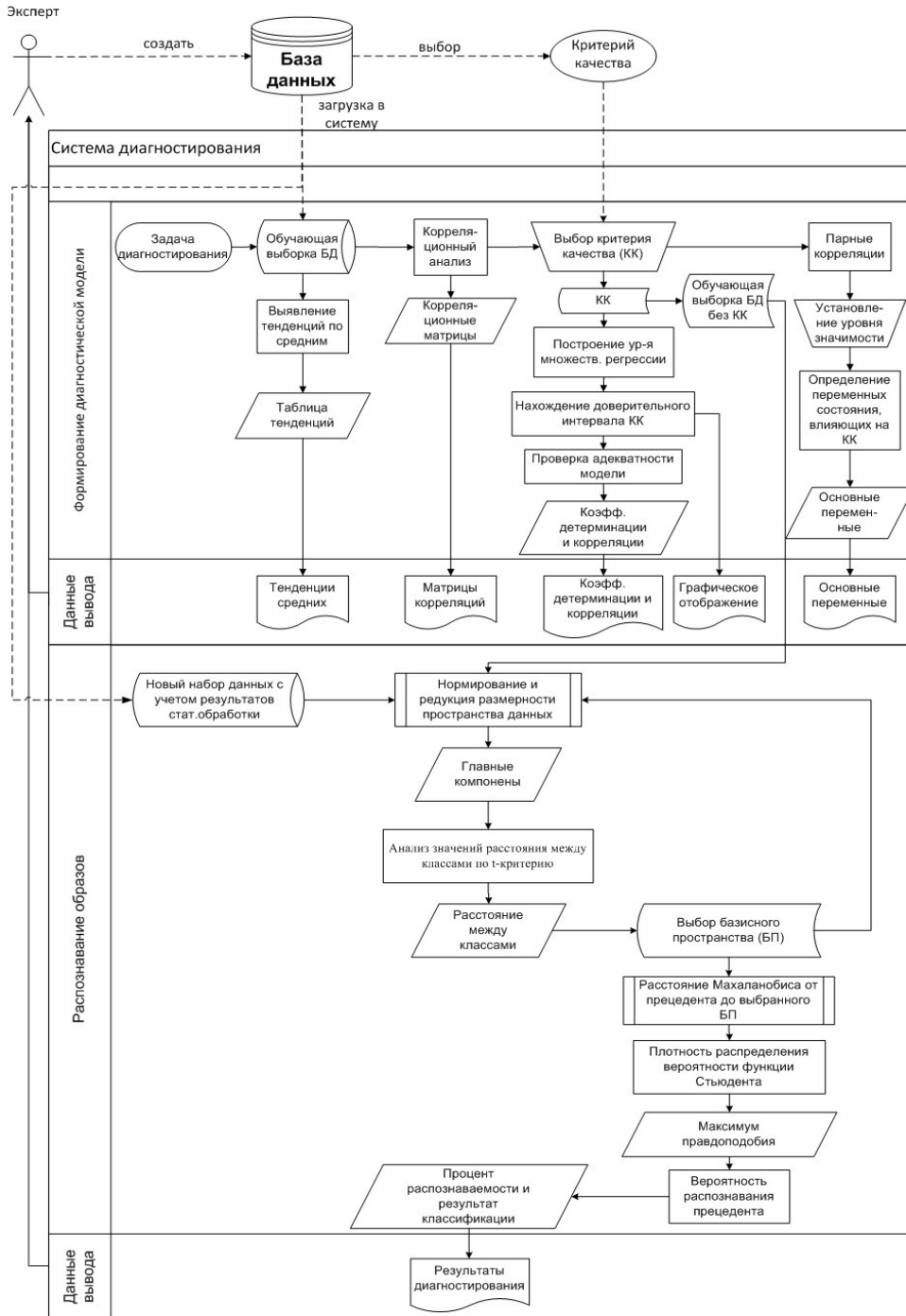


Рис. 3. Системная модель информационно-аналитического обеспечения процесса принятия решений лицом принимающим решения (ЛПР) при медико-биологическом диагностировании состояния пациентов.

Сформированы состав и структура компьютерной интерактивной системы поддержки принятия решений (КИ СППР) при медико-биологическом диагностировании состояния пациентов. Разработана информационная технология расчетов в среде разработанной КИ СППР.

5. Пример решения задачи классификации состояния элементов (пациентов) медико-биологической системы

В соответствии с мнением экспертов существует проблема выявления стадии перехода от гормон-чувствительного к гормон-резистентному состоянию. Согласно решению экспертов к анализируемой ранее обучающей выборке, которая состояла из трех классов, был добавлен класс пациентов (33 человека), соответствующих гормон-резистентному состоянию.

На этапе факторного анализа была проведена редукция размерности пространства ГК на основе критерия Кайзера [13]. Было принято решение оставить 16 ГК из 24 ГК, что позволило снизить размерность пространства переменных состояния.

Проведен анализ значимости расстояний между классами с учетом выбранного количества переменных состояния и количества прецедентов в классах. Расстояния между центрами классов в каждом из выбранных базисных пространств определялись на основе статистик Стьюдента $\{t_{lm}\}$. Использовалось статистическое решающее правило Стьюдента – гипотеза о равенстве средних [2, 11]. Из анализа данных в Табл. 1 видно, что расстояния между классами значимы, о чем свидетельствует превышение значений статистик Стьюдента критических.

Следуя решающему правилу (10) в качестве опорного был выбран базис, соответствующий классу '0'. На рис. 4 представлено расположение классов '0', '1', '2' и '3' в выбранном базисном пространстве.

Табл. 1. Результаты анализа значимости расстояний между классами

Класс '0'		Класс '1'		Класс '2'		Класс '3'	
Значения статистик Стьюдента $\{t_{lm}\}$, соответствующих расстояниям между центрами классов при числе ГК=16 {критические значения – $t_{crit}(n_{min} = 70; P=0,99)=2,6$ }							
0	0	0	12,58686	0	9,84595	0	9,61469
1	7,19796	1	0	1	3,84722	1	3,98223
2	9,54973	2	2,84275	2	0	2	2,74201
3	6,56994	3	3,23714	3	3,36595	3	0

Результаты решения задачи классификации на основе оценок вероятностей принадлежности прецедентов тому, либо иному классу приведены в Табл. 2. Вероятность распознавания класса самого себя превышает 80%.

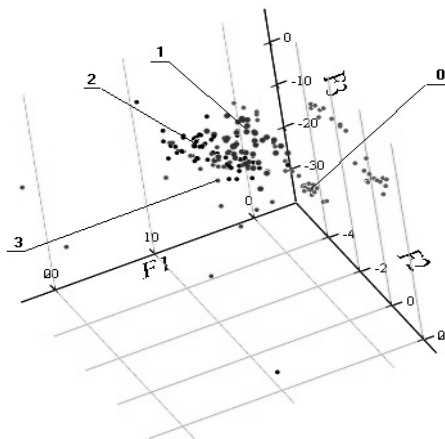


Рис. 4. Расположение прецедентов обучающей выборки в пространстве переменных состояния относительно опорного базиса.

Таблица 2. Результаты решения задачи классификации

Вероятность распознавания класса, %							
Класс '0'		Класс '1'		Класс '2'		Класс '3'	
ГК=16	98	ГК=16	84,4	ГК=16	82,2	ГК=16	100

6. Результаты и выводы

Предложена системная модель процесса диагностирования элементов медико-биологической системы (пациентов).

Усовершенствован статистический метод классификации состояния элементов динамических систем на основе аппарата вероятностных нейронных сетей, который содержит процедуры ортогонализации и редукции размерности факторного пространства переменных состояния и, в отличие от существующих, использует в качестве меры близости прецедентов в пространстве переменных состояния статистику Стьюдента в выбранном, согласно принципу максимального правдоподобия, опорном базисе. Высокая надежность результатов распознавания состояния прецедентов достигается в случае, если расстояния между классами являются значимыми.

В качестве примера, получены результаты решения задач построения диагностической модели и классификации состояния элементов медико-биологической системы (пациентов) для конкретного типа заболевания. На основе анализа информативности и агрегирования переменных выполнена редукция размерности пространства переменных состояния.

Разработана КИ СППР при медико-биологическом диагностировании состояния пациентов. Предложена информационная технология расчетов в среде разработанной КИ СППР.

Получено, что с использованием разработанного метода и реализующей его КИ СППР вероятность распознавания класса самого себя превышает 80%.

ЛИТЕРАТУРА

1. Руденко О.Г. Искусственные нейронные сети. Учебное пособие / О.Г.Руденко, Е.В. Бодянский. – Харьков: ООО «Компания СМИТ», 2005. – 408с.
2. Хайкин Нейронные сети: полный курс, 2е издание / Хайкин, Саймон: Пер. с англ. М. Издательский дом "Вильямс", 2006. – 1104 с.
3. Угрюмова Е.М. Обучаемые искусственные нейронные сети в построении формальных математических моделей систем при априорной неопределенности данных / Е.М. Угрюмова // Вісник Харківського національного університету: зб. наук. пр. Сер. Математичне моделювання. Інформаційні технології. Автоматизовані системи управління. – 2010. – Випуск 13 (№890). – С. 237 – 253.
4. Taguchi G. The Mahalanobis–Taguchi Strategy. A pattern technology system / Taguchi G., Jugulum R. New York: John Wiley & Sons, 2002. – 235 p.
5. Высочина О.С. Модель мониторинга телекоммуникационной сети на базе модифицированной вероятностной нейронной сети / О.С. Высочина, С.И. Шматков, А.М. Салман // Восточно-европейский журнал передовых технологий. – Х., 2010. – №5. – С.67-71.
6. Колтунов И.А. Новые статистические методы распознавания и классификации для автоматического дешифрования дистанционных наблюдений / И.А. Колтунов, А.Н. Думин, В.А. Катрич, Р.Р. Наумов // Вестник Харк. нац. ун-та., – 2011. – № 966. Сер. “Радиофізика та електроніка”, вып. 18. – С. 37-49.
7. Крицкий О.Л. Информационная матрица Фишера для многомерного метода динамических условных корреляций DCC-MGARCH(1,1) / О.Л. Крицкий // Вестник Томск. гос. ун-та. – 2009. – № 4. Сер. “Управление, вычислительная техника и информатика”, вып. 9. – С. 67-82.
8. Igor Loboda, Sergiy Yepifanov, Yakov Feldshteyn, 2012, “A more realistic scheme of deviation error representation for gas turbine diagnostics”, Proceedings of ASME Turbo Expo 2012 GT2012, June 11-15, 2012, Copenhagen, Denmark, 10p.
9. Sokolov O., Radyvonenko O., Korchak T, Gololobova O. Epidemics prediction with the use of neuro-fuzzy methods for time series under uncertainty conditions // Journal of Health Sciences, Poland, 2012. P. 64-70.
10. Афанасьевская В.Е. Методология решения задачи классификации технического состояния авиационного газотурбинного двигателя / В.Е. Афанасьевская, Е.М. Угрюмова, В.В. Нерубасский, Т.В. Гайденко // Восточно-Европейский журнал передовых технологий. – 2012. – № 3/10 (57). – С.40-43.
11. Айвазян С.А. Прикладная статистика и основы эконометрики –Учебник для вузов / С.А. Айвазян, В.С. Мхитарян. – М.: ЮНИТИ, 1998. – 1022с.
12. Корреляционный анализ ДОС [Электронный ресурс]. – Режим доступа: <http://www.twirpx.com/file/298707/> – Загл. с экрана.
13. Халафян А.А. Учебник STATISTICA 6. Статистический анализ данных / А.А. Халафян. – М.: ООО «Бином-Пресс», 2007. – 512 с.