

УДК 615.47

ИСПОЛЬЗОВАНИЕ ИНТЕРАКТИВНЫХ МЕТОДОВ КЛАССИФИКАЦИИ ДЛЯ РЕШЕНИЯ ЗАДАЧ МЕДИЦИНСКОГО ПРОГНОЗИРОВАНИЯ

Самаха Башир Аббас, Шевякин В.Н., Разумова К.В., Корневская С.Н.

ФГБОУ ВПО «Юго-Западный государственный университет», Курск, e-mail: kstu-bmi@yandex.ru

В работе рассматривается возможность использования диалоговых методов классификации для решения задач прогнозирования и диагностики заболеваний при плохо формализуемой структуре классов с нечетко определяемыми границами. Основная идея таких методов классификации заключается в отображении многомерных данных в двумерные пространства, в которых человек, используя свои интеллектуальные возможности по обработке зрительных образов, оценивает структуру исследуемых классов состояний человека и под неё подбирает наилучшую классификационную модель. С учетом особенностей структурных данных в задачах медицинского прогнозирования в отображающем пространстве классификацию целесообразно осуществлять, используя методологию нечеткого принятия решений, в соответствии с которой решение о классификации принимается по максимальным величинам функций принадлежности к исследуемым классам состояний. В качестве базовых переменных для функций принадлежности используются величины расстояний от границ отображений классов до координат исследуемого объекта в отображающем пространстве. Приводятся практические результаты применения рассмотренного метода на примере прогнозирования послеоперационных осложнений при урологических заболеваниях и ранней дифференциальной диагностики вибрационной болезни.

Ключевые слова: прогнозирование, распознавание образов, отображение, нечеткая логика

USING OF INTERACTIVE CLASSIFICATION METHODS FOR SOLVING PROBLEMS OF MEDICAL PREDICTION

Samaha Bashir Abbas, Shevyakin V.N., Razumova K.V., Korenevskaya S.N.

Southwest state university, Kursk, e-mail: kstu-bmi@yandex.ru

Dialog classification methods for solving problems of prediction and diagnosis of diseases in poorly formalized structure of classes with fuzzy boundaries are discussed in this article. The basic idea of these methods is to display classification dimensional data into two-dimensional spaces. In these spaces people study evaluates the structure of classes and selects the best model. Researchers use their intellectual abilities to process visual images. Classification is carried out using the methodology of fuzzy decision-making. The decision is taken on the maximum values of the membership functions to the class under consideration states using this methodology. The distances from the boundaries of class maps to coordinate the object under study maps the space used as reference variables for the membership functions. Practical application of the results of this method for the prediction of postoperative complications in urological diseases and early differential diagnosis of vibration disease listed in this article.

Keywords: prediction, pattern recognition, display, fuzzy logic

Задача прогнозирования состояния организма человека относится к классу плохо формализуемых задач, поскольку речь идет о сложной и нерешенной в настоящее время задаче определения границы между здоровьем и болезнью. На практике для решения плохо формализуемых задач часто привлекают аппарат, обеспечивающий изучение структуры данных с выдвиганием соответствующих гипотез на языке специалистов исследуемой предметной области. В анализе данных таким аппаратом является вычислительный эксперимент. Для решения задач классификации вычислительный эксперимент реализуется с помощью диалоговых интерактивных систем распознавания образов (ДСР) [1, 2].

Для привлечения ДСР к решению задач прогнозирования рассмотрим такой ее двухальтернативный вариант: класс ω_0 – обследуемый в течение заданного времени T_0 не переходит в состояние болезни с именем ℓ (не переходит в класс ω_ℓ); класс ω_ℓ – обследуемый в течение времени T_0 переходит в состояние ω_ℓ . Таким образом, задача прогнозирования

превращается в задачу классификации двух состояний ω_0 и ω_ℓ , которая имеет ряд особенностей, значительно осложняющих соответствующую задачу обучения в терминологии распознавания образов. Основными из этих особенностей являются: отсутствие полного статистического материала, одновременно охватывающего различные стороны функционирования исследуемого объекта; наличие казуистических ситуаций; присутствие в обучающем материале объектов с неточными или ошибочными описаниями; часто «учитель» не имеет возможности точно указать класс объекта на обучающей выборке и не дает информации о наличии переходных зон между классами, хотя в обучающей выборке эти объекты имеются; объекты одного и того же класса в исходном пространстве могут быть представлены различными группами, которые, в свою очередь, распределены среди групп других классов и т.д. [1, 2, 4, 5].

Для разрешения этих и ряда других проблем, связанных с особенностями задач медицинского прогнозирования, в данной работе предлагается использовать

модификацию известных ДСР, определенную в работах [2, 4] как метод интерактивного конструирования двумерных классификационных пространств (ИКДКП), который способен сохранять работоспособность при плохо формализованной структуре классов с нарушениями гипотезы о компактности, при большом количестве артефактов в обучающей выборке, в условиях не полностью определенных «учителем» классов, возможной «вложенности» их друг в друга и др. [2].

Основная идея предлагаемого метода заключается в том, что обучение и классификация ведутся в двумерном отображающем пространстве $\Phi = Y_1 \times Y_2$, куда

$$R(A, B) = \int_{x\Omega} S(X, A, B, \Omega) P(\Omega) P(\Omega / X) L(X, \Omega) dX d\Omega, \quad (1)$$

где $P(\Omega)$ – априорная вероятность появления класса Ω ; $P(\Omega/X)$ – условная вероятность появления Ω ; $L(X, \Omega)$ – функция потерь от ошибочной классификации, например, из-за недостаточной классификации «учителя»; $S(X, A, B, \Omega)$ – площадь наложения классов в Φ .

Задача классификации заключается в определении местоположения отображения объектов $X = (x_1, \dots, x_n)$ относительно границ различных классов, полученных в Φ .

В качестве функций отображения могут применяться как простейшие линейные формы, так и нелинейные модели.

Известно несколько способов динамического конструирования отображающего пространства. Например, один из его вариантов основан на том, что в исходном пространстве признаков выбирается эталонная структура и создается механизм ее скольжения по некоторой средней (в смысле выбранного критерия усреднения) линии, представляющей каждый из исследуемых классов. Получающуюся в результате такого скольжения линию называют «скелетной», поскольку она является основой, вокруг которой формируется вся гиперконфигурация класса, подобно тому, как вокруг скелета формируется конфигурация тела [3].

В практических приложениях исследуемые классы состояний ω_0 и ω_ℓ достаточно сильно пересекаются и имеют размытые границы в исходном пространстве признаков, что, естественно, передается и в отображающее пространство. В таком варианте целесообразно использовать нечеткое описание отображений классов с выделением соответствующих областей уверенности в классификации, коэффициентов уверенности и (или) функций принадлежности [2, 3, 5].

многомерные объекты $X = (x_1, x_2, \dots, x_n)$ отображаются с помощью пары отображающих функций $Y_1 = \varphi_1(A, X)$, $Y_2 = \varphi_2(B, X)$, где $A = (a_1, a_2, \dots, a_n)$ и $B = (b_1, b_2, \dots, b_n)$ – векторы настраиваемых параметров. Задача обучения состоит в нахождении таких значений параметров векторов A и B и видов функциональных зависимостей φ_1 и φ_2 , при которых непересекающиеся образы в многомерном пространстве N при отображении в Φ будут разделимы или будут иметь минимальное (допустимое) наложение классов ω_0 и ω_ℓ . С математической точки зрения задача обучения заключается в минимизации функционала качества вида:

Для изображения в Φ зон с различной уверенностью в классификации удобно использовать линии равных значений коэффициентов уверенности (значений функций принадлежности) или соответствующие цветовые гаммы. Для отображающего пространства Φ функция принадлежности к классу ω_ℓ может быть определена на базовой переменной, соответствующей расстоянию d_k от отображения исследуемого объекта с координатами $\{Y_1^0, Y_2^0\}$ до опорных конфигураций L_k класса ω_ℓ .

В качестве опорных конфигураций в зависимости от целей и типов решаемых задач могут быть выбраны координаты центра отображений классов, опорные точки с координатами $\{Y_1^s, Y_2^s\}$ внутри или вне классов, скелетные (усредняющие) линии классов (прямые линии с уравнениями типа $a_1 Y_1 + a_2 Y_2 + a_0 = 0$, ломаные линии), эталоны различных конфигураций (круги, квадраты, прямоугольники), собственно границы классов, которые в общем случае описываются уравнениями вида $f_k(A, Y_1, Y_2) = 0$.

На этапе обучения специальная компьютерная программа или группа экспертов определяет семейство функций принадлежности $\mu_\ell(d_k)$. Далее уверенность в принятии решения определяется по отношению к ближайшей L_k .

Например, если границы классов в пространстве Φ определяются кусочно-линейной аппроксимацией, то до каждой из k линий базовая переменная d_k определяется выражением

$$d_k = \frac{|a_1 Y_1^0 + a_2 Y_2^0 + a_0|}{\sqrt{a_1^2 + a_2^2}}, \quad (2)$$

а общая уверенность (коэффициент уверенности) в классификации ω_ℓ определяется выражением

$$KY_\ell = \min[\mu_{\omega_\ell}(d_k)]. \quad (3)$$

В качестве практического примера рассмотрим решение задачи синтеза решающего правила для прогнозирования послеоперационных осложнений после трансуретральной резекции предстательной железы класс ω_R [5].

Для получения прогностических решающих правил было отобрано 12 медицинских признаков для построения соответствующих математических моделей: x_1 – наличие в анамнезе острой задержки мочеиспускания, катетеризации мочевого пузыря; x_2 – наличие эпицистостомы; x_3 – степень инфицирования мочи (КОЕ/мл); x_4 – количество лейкоцитов в поле зрения; x_5 – наличие сопутствующих воспалительных заболеваний мочевыделительной системы; x_6 – суммарный балл оценки симптомов ($I^0 - PSS$); x_7 – качество жизни вследствие расстройств

мочеиспускания (L); x_8 – состояние иммунитета по показателю $CD 8$; x_9 – показатель иммунитета по интерлейкину – 4; x_{10} – показатель иммунитета по тесту восстановления нитросинеготетразолия (НСТ – тест); x_{11} – антиокислительная активность (АОА) сыворотки крови (в %); x_{12} – количество церулоплазмينا в сыворотке крови.

В ходе проведения разведочного анализа с использованием пакета прикладных программ, описанного в работах [2, 4], было установлено, что при разделении классов ω_0 и ω_R наибольшей информативностью обладают признаки x_4 , x_6 , x_{11} и x_{12} , причем пара признаков x_4 и x_6 оперативно и легко получается в ходе обследования. На объектах репрезентативной обучающей выборки в системах координат $\{x_4, x_6\}$ были построены образы классов: ω_0 – осложнений не ожидается; и ω_R – ожидаются послеоперационные осложнения (рис. 1).

Таким образом, отображающее пространство Φ в данном варианте построено на координатах исходного пространства признаков: $Y_1 = x_6$; $Y_2 = x_4$.

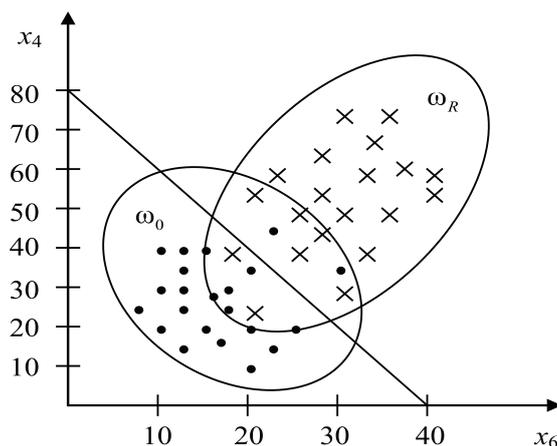


Рис. 1. Расположение классов $\omega_0(x)$ и $\omega_R(x)$ в двумерном пространстве признаков

Анализ изображений полученных образов показал, что между классами ω_0 и ω_R может быть проведена разделяющая линия типа $x_4 + 2x_6 = 80$ с выделением зоны пересечений этих классов.

Рассматривая зону пересечения как область нечеткой классификации и используя переменную $Z = x_4 + 2x_6$ в качестве базовой переменной, была построена пара функций принадлежности к классам ω_0 и ω_R (рис. 2).

Максимальное значение функций принадлежности на уровне 0,92 определяет доверие экспертов к полученному решающему правилу. Решение о прогнозировании принимается по максимальной величине соответствующих функций принадлежности.

Предложенный метод классификации обеспечивает хорошие результаты и при решении задач ранней и дифференциальной диагностики различных заболеваний. В качестве второго примера рассмотрим задачу нечеткой классификации вибрационной болезни по трем классам: здоров (ω_0); ранняя (доклиническая, донозологическая) стадия (класс ω_p); клиническая стадия (класс ω_k) вибрационной болезни.

Для решения классификационной задачи было выбрано 24 признака x_i ; x_1 – профстаж работы с инструментом не менее 8 лет; x_2 – парестезия рук после работы и ночью; x_3 – боли в руках после работы и ночью; x_4 – зябкость рук; x_5 – тугоподвижность пальцев рук по утрам; x_6 –

судорожные стягивания пальцев рук; x_7 – прерывистость сна из-за болей и парестезий в руках; x_8 – побеление или посинение пальцев; x_9 – гиперемированный акроцианоз с мраморностью; x_{10} – акрогипотермия; x_{11} – акрогипергидроз; x_{12} – трофическое изменение кистей; x_{13} – гипалгезия (гипреалгезия) по типу «перчаток»; x_{14} – снижение вибрационной чувствительности; x_{15} – спастический или спастикоатонический тип капилляров; x_{16} – термоасимметрия на кистях

и груди более $0,5^{\circ}\text{C}$; x_{17} – акрацианоз при холодной пробе; x_{18} – акроспазмы (синдром Рейно); x_{19} – замедленное восстановление кожной температуры после холодной пробы; x_{20} – снижение силы рук; x_{21} – преобладание периферических сосудистых нарушений над церебральными по данным РЭГ, ПРГ; x_{22} – костные нарушения кистей и позвоночника; x_{23} – наличие неврита лучевого, локтевого нерва; x_{24} – хроническая коронарная недостаточность.

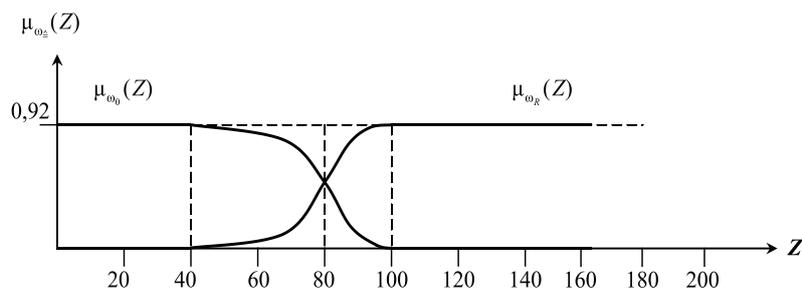


Рис. 2. Графики функций принадлежности к классам ω_0 и ω_R

В ходе обучения была получена пара отображающих функций вида:

$$Y_1 = 1,4x_1 + 0,18x_2 + 0,23x_4 + 0,33x_6 + 1,2x_7 + 1,44x_{11} + 0,9x_{14} + 0,43x_{15} + 0,36x_{16} + 0,9x_{17} + 0,09x_{18} + 0,08x_{22} + 0,2x_{23} + 0,23x_{24};$$

$$Y_2 = x_1 + 0,7x_2 + 0,03x_3 + 0,1x_4 + 0,07x_5 + 0,1x_6 + x_7 + 0,08x_8 + 0,03x_9 + 0,04x_{10} + x_{11} + 0,015x_{12} + 0,018x_{13} + x_{14} + 0,1x_{15} + 0,7x_{16} + x_{17} + x_{18} + 0,11x_{19} + 0,09x_{20} + 0,011x_{21} + 0,12x_{22} + 0,18x_{23} + 0,1x_{24}.$$

Относительно выделенных экспертами границ классов в пространстве Φ согласно рекомендациям

[1, 3] получены функции принадлежности, графики которых приведены на рис. 3.

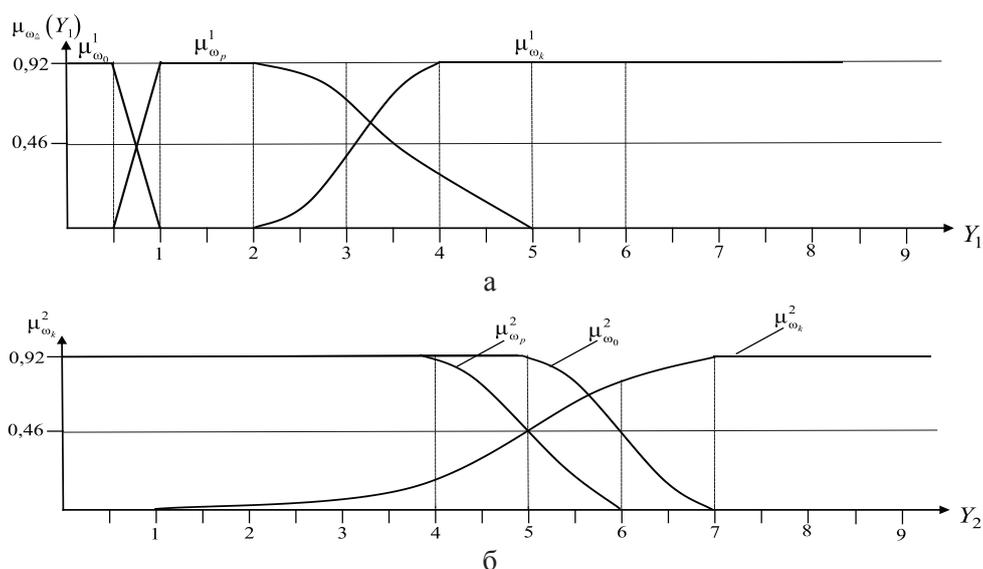


Рис. 3. Графики функций принадлежности к классам вибрационной болезни по шкале: а – Y_1 ; б – Y_2

Решение о принадлежности к каждому из классов ω_ℓ ($\ell = 0, p, k$) принимаются в соответствии с выражениями

$$\mu_{\omega_0} = \min(\mu_{\omega_0}^1, \mu_{\omega_0}^2); \mu_{\omega_p} = \min(\mu_{\omega_p}^1, \mu_{\omega_p}^2);$$

$$\mu_{\omega_k} = \min(\mu_{\omega_k}^1, \mu_{\omega_k}^2).$$

Решение о классификации принимается по максимальной величине функций принадлежности. При равных значениях функций принадлежности предпочтение отдаются в следующем порядке $\omega_k, \omega_p, \omega_0$, начиная с ω_k .

Проверка качества классификации на контрольной выборке объемом 50 человек на каждый класс показала, что полученная диагностическая эффективность решающих правил превышает 0,9, что позволяет рекомендовать их к использованию в медицинской практике.

Аналогичные результаты были получены в задаче диагностики профессиональных заболеваний сварщиков с уверенностью в дифференциальной диагностике таких заболеваний, как интоксикация, пневмокониоз и хронический бронхит на уровне 0,9 при диагностической эффективности не ниже 0,88.

Выводы

Рассматриваемый метод интерактивно-го конструирования двумерных классификационных пространств позволяет решать задачи прогнозирования, ранней и дифференциальной диагностики заболеваний в условиях плохой формализации, недостаточной статистики, различной структуры классов и признакового описания, при наличии казуистических ситуаций, отсутствии информации об объективно существующих переходных зонах между классами и отсутствии априорной информации о структуре классов. Кроме того, разработанные методы позволяют изучать структурные особенности исследуемых классов, получать дополнительные сведения как о структуре классов, так и о системе признаков, представляющих эти классы, включая выяснение роли признаков в формировании того или иного класса или зоны перехода между классами.

Практическая апробация метода интерактивного конструирования двумерных классификационных пространств показыва-

ет приемлемые для практической медицины результаты по качеству классификации.

Список литературы

1. Корневский Н.А., Бунаев В.В. Метод синтеза двумерных классификационных пространств // Известия ВУЗов. Приборостроение. – 2005. – Т. 48. – № 2. – С. 35–38.
2. Корневский Н.А., Крупчатников Р.А., Горбатенко С.А. Синтез нечетких сетевых моделей, обучаемых по структуре данных, для медицинских экспертных систем // Медицинская техника. – 2008. – № 2. – С. 18–24.
3. Корневский Н.А., Филлист С.А., Устинов А.Г., Рябкова Е.Б., Геометрический подход к синтезу нечетких решающих правил для решения задач прогнозирования и медицинской диагностики // Биомедицинская радиоэлектроника. – 2012. – № 4. – С. 20–26.
4. Корневский Н.А. Проектирование систем поддержки принятия решений для медико-экологических приложений: монография / Н.А. Корневский, В.С. Титов, И.Е. Черneckая. – Курск. гос. техн. ун-т Курск, 2004. – 180 с.
5. Синтез комбинированных нечетких решающих правил для прогнозирования послеоперационных осложнений в урологии / С.П. Серегин, С.Д. Долженков, С.Н. Корневская, Т.Н. Сапитонова // Известия Юго-Западного государственного университета. Серия Управление, вычислительная техника, информатика. Медицинское приборостроение. – 2012. – № 2. – Ч. 3. – С. 293–297.

References

1. Korenevskij N.A., Bunjaev V.V. Method for the synthesis of two-dimensional classification of spaces / Proceedings of the universities. Instrument. 2005. Vol. 48. no. 2, pp. 35–38.
2. Korenevskij N.A., Krupchatnikov R.A., Gorbatenko S.A. Synthesis of fuzzy network models, trained on the data structure for medical expert systems / Medical Technology. 2008. no. 2. pp. 18–24.
3. Korenevskij N.A., Filist S.A., Ustinov A.G., Rjabkova E.B. The geometric approach to the synthesis of fuzzy decision rules for solving medical diagnosis and prediction / Biomedical electronics. 2012. no. 4. pp. 20–26.
4. Korenevskij N.A., Titov V.S., Cherneckaja I.E. Designing decision support systems for medical and environmental applications : monograph / Kursk.gos. tehn. un-t. Kursk, 2004. 180.
5. Seregin S.P., Dolzhenkov S.D., Korenevskaja S.N., Sapitono-va T.N. Synthesis combined fuzzy decision rules for predicting post-operative complications in urology / Proceedings of the Southwestern State University. Series Management, Computer Science, Computer Science. Medical devices. 2012. no. 2. Ch 3. pp. 293–297.

Рецензенты:

Мишустин В.Н., д.м.н., профессор, декан факультета последипломного образования, ГОУ ВПО «Курский государственный медицинский университет» Росздрава, г. Курск;

Бурмака А.А., д.т.н., профессор, главный научный сотрудник НИЦФГУП 18 ЦНИИ МО РФ, г. Курск.

Работа поступила в редакцию 27.01.2014.