УДК 004.5

РАЗРАБОТКА МАТЕМАТИЧЕСКОЙ МОДЕЛИ РАСПОЗНАВАНИЯ ЗАПРОСОВ-ЗАДАЧ КОММУНИКАЦИОННЫХ УСЛУГ

Аль-Хашеди А.А., Обади А.А., Нуриев Н.К., Печеный Е.А.

ФГБОУ ВО «Казанский национальный исследовательский технологический университет», Казань, e-mail: alhashedi@mail.ru, 19fattah86@mail.ru

В статье рассматривается предметная область с конечным множеством различных объектов. Каждый объект имеет конечное множество себе подобных экземпляров. Любой объект из предметной области однозначно характеризуется конечным множеством признаков. При этом два различных объекта могут иметь общие признаки. Вся предметная область в целом характеризуется набором признаков входящих в нее объектов. В работе рассматриваются две взаимосвязанные задачи: задача 1, задача 2. В задаче 1 требуется разбить всю предметную область на классы объектов, т.е. представить её как объединение классов объектов. Для решения этой задачи разработана математическая модель и соответствующий алгоритм процедуры разбиения, который позволяет исходя из размерности признаков объекта и их содержания выделить отдельные классы объектов в предметной области. После установления всех классов в предметной области рассматривается задача 2, построения математической модели и алгоритма распознавания экземпляра объекта с точки зрения принадлежности его какому-то классу объектов предметной области. Эта задача решается на основе Байесовского подхода, т.е. с использованием формул полной вероятности и Байеса. Для разъяснения процедур классификации и распознавания в работе приводится пример использования предложенных алгоритмов.

Ключевые слова: распознавание образов, коммуникационные услуги, классификация, математическая модель, класс, признак, объект, вероятность, теорема Байеса

DEVELOPMENT OF THE MATHEMATICAL MODEL FOR RECOGNITION TASK REQUESTS OF COMMUNICATIONS SERVICES

Alkhashedi A.A., Obadi A.A., Nuriev N.K., Pechenyy E.A.

Federal state budget educational institution of higher education «Kazan National Research Technological University», Kazan, e-mail: alhashedi@mail.ru, 19fattah86@mail.ru

The article deals with the subject domain with a finite set of different objects. Each object has a finite set of similar copies. Any object from the subject area is, uniquely characterized by a finite set of attributes. Wherein, two different objects may have common attributes. Any object from the domain is, uniquely characterized by a finite set of attributes, which objects include. In article considers two interrelated tasks: task 1, task 2. In task 1 required split the whole subject area to object classes, i.e. present it as a union of object classes. To solve this task developed a mathematical model and corresponding algorithm of the procedure splitting, which allows selecting the individual classes of objects in the subject area on the dimensions of the attributes of the object and their content. After created all classes in the subject area task 2 is considers, constructing a mathematical model and an algorithm for recognizing an object instance in the terms of its belonging to some class of subject area. This task is solved based on the Bayesian approach, i.e. using the full probability and Bayesian formulas. To clarification of classification and recognition procedures, is given an example in this article of using the proposed algorithms.

Keywords: pattern recognition, communications services, classification, mathematical model, class, attribute, object, probability, Bayes' theorem

Цель работы

Решить две взаимосвязанные задачи: ЗАДАЧА 1 – разработать математическую модель и алгоритм вычислительной процедуры, который позволял бы классифицировать объекты из одной предметной области, исходя из их свойств (признаков), при условии, что часть признаков разных объектов могут совпадать. ЗАДАЧА 2 – на основе вероятностного подхода построить модель и алгоритм вычислительной процедуры, позволяющий идентифицировать принадлежность экземпляра объекта определенному классу из предметной области.

Постановка задачи

Данное U множество образов коммуникационных задач с общими признаками $X = \{x_1,\ldots,x_m\}$, m – количество признаков предметной области (в приведенном ниже примере 1 $m = \overline{1,7}$, рис. 1). Это множество необходимо разделить на подмножества признаков, соответствующие классам запросов задач данной предметной области. Система составляет словарь признаков с помощью списка правил [1]. Это множество будем обозначать через X. Пока не поступила первая услуга на систему, существует только один общий класс ω, область предпочтения которого является X. Во множестве образов коммуникационных задач предметной области U нас интересуют некоторые подмножества – класс в зависимости от типов коммуникационных услуг. Множество классов $\Omega = \{\omega_1,\ldots, \omega_m\}$, является конечным (в данной задаче рас-

познавания $m=\overline{1,4}$ и равно числу классов услуг: ω_1 – класс услуг интернета, ω_2 – класс услуг тарифов, ω_3 – класс услуг оплаты, ω_4 – класс услуг поддержек), эти классы образуют полную группу подмножеств из U (разбиение множества образов U), т.е. $\bigcup_{i=1}^{m}\omega_i = U$, и $\omega_i \bigcap \omega_j = \varnothing$. для всех $i \neq j$. Классифицировать объект $x \in U$ по классам ω_i значит найти такую индикаторную функцию g: $U \to Y$, которая ставит в соответствие образу $x \in U$ метку $y_i \in Y$ того класса ω_i, которому он принадлежит т.е. $g(x) = y_i$, если $x \in \varpi_i$, где $Y = \{y_i\}$, а y_i признаки объекта.

Система автоматически создает классы с помощью решающих функций при распознавании поступающих на систему неизвестных образов. В системе количество классов увеличивается с разбиением множества признаков предметной области на подмножества классов в зависимости от появляющихся услуг, и эти классы могут пересекаться между собой, обладая общими признаками [5].

Пусть на систему появилась не идентифицированная услуга x, для которой по своим описывающим признакам $\mathbf{x} = (x_1, \ldots, x_n)$, $n \le m$, n – количество признаков задачи, для данной услуги необходимо создать новый класс из множества признаков предметной области $X = \{X_1, X_2\}$. Эта услуга описывается некоторыми признаками в данной предметной области. Эти признаки сравняются с признаками предметной области на совпадение. Если все эти признаки полностью совпадают или с большой вероятностью совпадают с некоторыми признаками предметной области $p(x_j \in \omega_i) > p(x_j \in \omega_r) \forall i \neq r$, то создается новый класс ω_i. Если большинство признаков совпадают, но не все признаки, то в предметной области выводятся новые признаки и создается новый класс. В случае, когда появляется новая задача и в предметной области уже существуют некоторые классы $\omega = \bigcup_{i=1}^{m}\omega_i$, то данная услуга проверяется на принадлежность некоторому из существующих классов, если данные признаки совпадают с признаками некоторого класса, то данная задача относится к данному классу. Но если признаки описывающие данную задачу не совпадают или вероятность совпадения $p(x_j \in \omega_i) = 0$ с признаками существующих классов, то проверяется на совпадения с общими признаками предметной области, чтобы создать новый класс в данной предметной области.

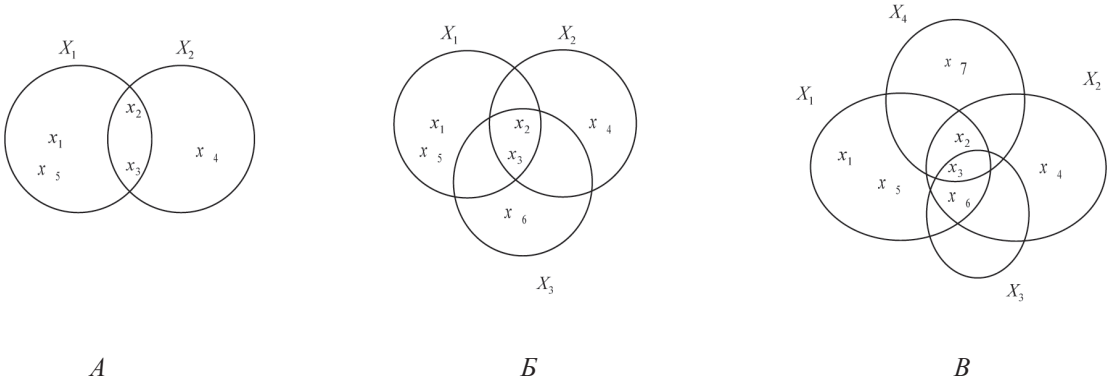
Алгоритм образования новых классов

На рис. 1, блок А, показано, что X_1 – область предпочтения первого класса ω_1, а X_2 – область предпочтения второго класса ω_2. $X_1 \bigcap X_2 = X$ – общие признаки. $(X_1 \bigcup X_2) \setminus (X_1 \bigcap X_2)$ – унифицированные признаки. И так далее при поступлении новых услуг в систему создаются новые классы.

На рис. 2 процедура выполнения данного алгоритма следующая:

1. В блоке 1, А является обучающая выборка объектов, в которой определяются правила выбора признаков и конечное число количество признаков данной предметной области. Множество X является совокупностью признаков предметной области, составляющейся из m признаков (в приведенном ниже примере 1 $m=\overline{1,7}$, рис. 1), $\mathbf{x}$ – вектор признаков распознаваемого образа, образующийся из n признаков и обычно $n < m$.

2. В блоке 2 сравняется m с n, если $m = n$, то множество остается таким и задача (распознаваемый образ) относится к общему классу.



А *Б* *В*

Рис. 1. Разбиение множества признаков на подмножества классов

3. В блоке 3, если $n < m$, то для сравнения определяется начальный признак множества предметной области $i = 1$ и начальный признак распознаваемого образа $j = 1$, и начинается сравнение начального признака распознаваемого образа на совпадение со всеми признаками множества предметной области начинается с первым. Если не совпадают с первым признаком предметной области, то сравнение на совпадение будет с вторым признаком предметной области $i + 1$ и т.д. до того, как найдётся совпадающий элемент в множестве признаков или i станет равным m (i = m), т.е. признак распознаваемого элемента сравняется со всеми признаками предметной области с первого признака $i = 1$ до последнего $i = m$. Затем на сравнение берем следующий признак распознаваемого образа $j + 1$, данный признак будем сравнивать на совпадение со всеми признаками предметной области и т.д. до того момента когда $j = n$ (рис. 2).

4. В блоке 4. После сравнения всех признаков и условие того, что признаки распознаваемого образа совпадают с признаками в данной предметной области, выполняется т.е. $x_j = x_i$, то множество предметной области разбивается и появится новый класс ω_i и $x_j \in \omega_i$. Появление нового класса зависит от значения вероятности совпадения признаков распознаваемого образа с признаками множества предметной области $p\left(x_j \in \omega_i\right)$.

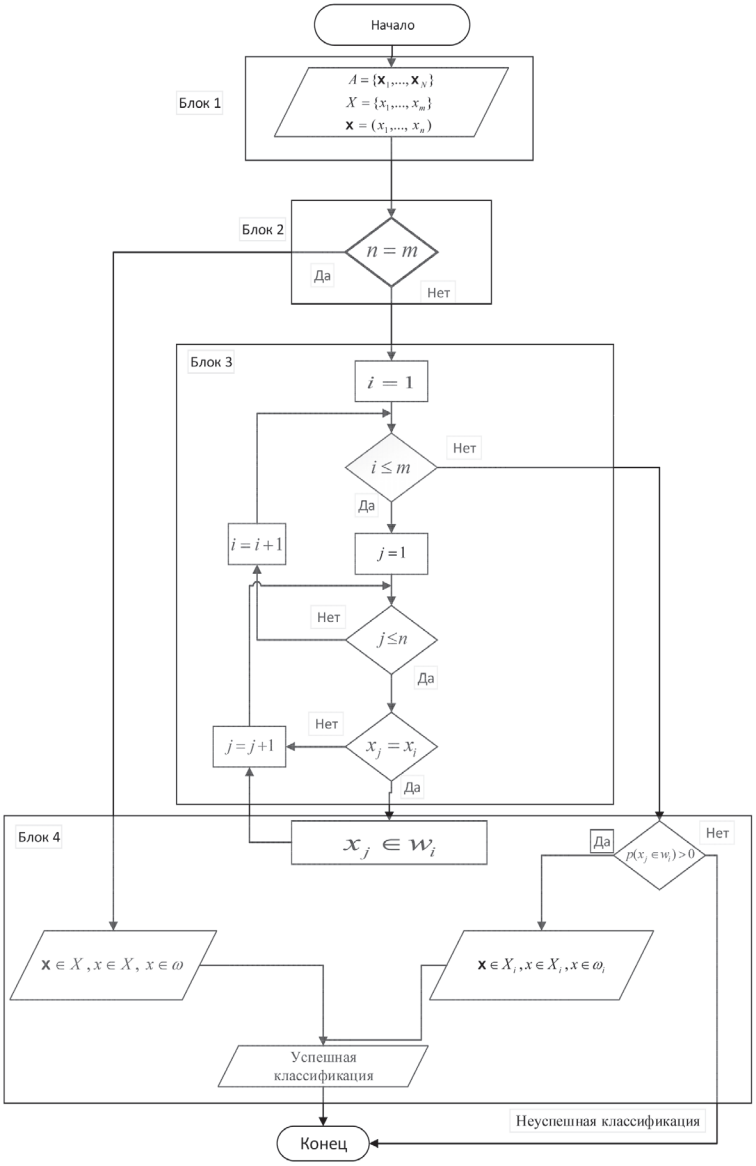


Рис. 2. Логическая блок-схема алгоритма классификации предметной области

Пример классификации с помощью решающих функций

Одной из основных задач распознавания образов является задача описывания классов [1]. Как отметили выше, у нас имеются 4 класса коммуникационных услуг и каждый класс описывается некоторым набором признаков в множестве признаков, например класс интернета описывается следующими признаками: {x_1 – интернет, x_2 – номер телефона, x_3 – название компании и т.д.}. А каждый образ x описывается некоторым набором признаков в множестве признаков – вектором **x**. Все множество признаков X разбивается на $m + 1$ попарно несовместных множеств (полную группу множеств, в зависимости от классов) [1, 3, 4], $X_0, X_1, \ldots, X_m : X_i \bigcap X_j = \varnothing$, для всех $i \neq j$ в случае непересекающихся классов, а в случае пересекающих классов может быть такое равенство $X_i \bigcap X_j \neq \varnothing$ как на рис. 1, $\bigcup_{i=0}^{m} X_i = X$. Каждое множество X_i, является множеством предпочтения одного класса ω_i в множестве X, рис. 1. Таким образом, классами распознаваемых образов будем считать области X_i ($i = \overline{1,4}$).

Автоматическое определение принадлежности объектов к классам – одна из основных задач теории распознавания образов [1, 2, 4]. Принадлежность распознаваемых объектов можно определять с помощью решающей функции.

Будем считать, что пространство признаков является конечным множеством. В этом случае предполагается, что существует $m + 1$ функции $d_j(x)$ так, что $X_j = \{x \in X: d_j(x) > 0\}$. Можно считать, что образ x принадлежит классу ω_i, если выполняется неравенство $d_j(x) < 0$ для всех $j \neq i$ и $d_i(x) > 0$. Решающей называют функцию вида

$$d(x) = d(x_1, x_2, \ldots, x_n) =$$
$$= w_1 x_1 + w_2 x_2 + \ldots + w_n x_n + w_{n+1} = (w, x),$$

где $x = (x_1, x_2, \ldots, x_n, 0)$ вектор признаков, а $(w_1, w_2, \ldots, w_n, w_{n+1})$ вектор весовых коэффициентов, линейных решающих функций $\sum_{i=1}^{n} w_i = 1$ [1–3].

Пример 1: пусть на систему поступила такая задача (Set up mobile internet, phone number: +967774541452, model: Apple iPhone 4S, phone carrier: GSM, operating system: IOS). Данную задачу будем считать объектом со следующим вектором признаков

x_1 – Set up, x_2 – phone number, x_3 – model,
x_4 – operating system, x_5 – mobile,
x_6 – internet, x_7 – phone carrier).

Для того чтобы определить решающую функцию каждого класса, будем считать предикаты всех признаков способом сравнения этих признаков с признаками пространства предметной области или с помощью булевых значений таблицы истинности, таким образом если $x_j \in \omega_i$ т.е. признак объекта соответствует какому-то признаку класса ω_i, $x_j = x_i$, то $P(x_j \in \omega_i) = 1$, а если $x_j \notin \omega_i$ т.е. признак распознаваемого объекта не соответствует никакому признаку класса ω_i, то $P(x_j \in \omega_i) = 0$.

На рис. 1 в блоке С множество признаков X разбивается на 4 группы подмножества признаков $X_i, i = \overline{1,4}$.

$$X = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7\},$$

$$X_1 = \{x_1, x_2, x_3, x_5, x_6\},$$

$$X_2 = \{x_2, x_3, x_4, x_6\},$$

$$X_3 = \{x_3, x_6\},$$

$$X_4 = \{x_2, x_3, x_7\}.$$

Результат вычисления предикатов:

$$P(x_1 \in \omega_i) : i = \overline{1,4} \Rightarrow 1;0;0;0,$$

$$P(x_2 \in \omega_i) : i = \overline{1,4} \Rightarrow 1;1;0;1,$$

$$P(x_3 \in \omega_i) : i = \overline{1,4} \Rightarrow 1;1;1;1,$$

$$P(x_4 \in \omega_i) : i = \overline{1,4} \Rightarrow 0;1;0;0,$$

$$P(x_5 \in \omega_i) : i = \overline{1,4} \Rightarrow 1;0;0;0,$$

$$P(x_6 \in \omega_i) : i = \overline{1,4} \Rightarrow 1;1;1;0,$$

$$P(x_7 \in \omega_i) : i = \overline{1,4} \Rightarrow 0;0;0;1.$$

Поставим $w = \frac{1}{7}$, тогда решающая функция будет иметь такой вид:

$$d(x) = d\left(\frac{x_1}{7}, \frac{x_2}{7}, \ldots, \frac{x_n}{7}\right) = \frac{x_1}{7}, + \frac{x_2}{7} + \ldots + \frac{x_n}{7};$$

$(x_1, x_2, x_3, x_4, x_5, x_6, x_7) \in \omega_1$, тогда получаем решающую функцию первого класса в таком виде:

$$d_1(x) = \left(\frac{1}{7}x_1 + \frac{1}{7}x_2 + \frac{1}{7}x_3 + \frac{1}{7}x_5 + \frac{1}{7}x_6\right) = \frac{5}{7};$$

$(x_1, x_2, x_3, x_4, x_5, x_6, x_7) \in \omega_2$, $d_2(x) = \left(\frac{1}{7}x_2 + \frac{1}{7}x_3 + \frac{1}{7}x_4 + \frac{1}{7}x_6\right) = \frac{4}{7}$; $d_2(x)-$ решающая функция второго класса;

$(x_1, x_2, x_3, x_4, x_5, x_6, x_7) \in \omega_3$, $d_3(x) = \left(\frac{1}{7}x_3 + \frac{1}{7}x_6\right) = \frac{2}{7}$; $d_3(x)-$ решающая функция третьего класса;

$(x_1, x_2, x_3, x_4, x_5, x_6, x_7) \in \omega_4$, $d_4(x) = \left(\frac{1}{7}x_2 + \frac{1}{7}x_3 + \frac{1}{7}x_7\right) = \frac{3}{7}$; $d_4(x)-$ решающая функция четвертого класса.

Будем считать, что данный объект x принадлежит классу ω_i в таком случае, если $d_i(x) > 0 = z$ и $0 < d_j(x) < z$ для всех $i \neq j$. Другими словами если $d_i(x) > d_j(x)$, то x принадлежит классу ω_i.

Однако можно считать вероятность принадлежности объекта одному классу условной вероятностью. Будем считать, что x принадлежит классу ω_i, если выполняется следующее условие:

$$P(x_j \in \omega_i \,|\, X_j = X_i) > P(x_j \in \omega_r \,|\, X_j = X_r),$$

для этого вычисляется общая вероятность принадлежности признаков всем существующим классам, как показано в приведенном ниже примере 2.

Пример 2. H_i событие, что $Y_i = X_i$, где Y_i признаки образа x, а X_i – признаки i-ого класса ω_i, т.е. признаки распознаваемого объекта соответствуют признакам i-ого класса, A событие, что $x \in \omega_i$ (объект x принадлежит i-тому классу ω_i). Событие A может произойти только при выполнении события $H_i, i = \overline{1,4}$. Вероятность $P(H_i)$ того, что $Y_i = X_i$, равна произведению вероятности $P(A)$ того, что $x \in \omega_i$, на вероятность $P(H_i \,|\, A)$ того, что причиной события стало именно, что $Y_i = X_i$ (событие H_i), а не другая причина. В данном примере:

$P(H_i) = 0{,}25$, вероятность того что данный образец принадлежит i-тому классу, так как у нас имеются 4 класса, т.е. вероятность что данный объект принадлежит одному классу

$$p(H_1) = p(H_2) = p(H_3) = p(H_4) = 0{,}25.$$

Через полученные значения предикатов в примере 1, вероятности выполнения события $H_i, i = \overline{1,4}$, при условии A будут равны следующим значениям:

$$P(A|H_1) = \frac{5}{7} = 0{,}7;$$

$$P(A|H_2) = \frac{4}{7} = 0{,}57;$$

$$P(A|H_3) = \frac{2}{7} = 0{,}28;$$

$$P(A|H_4) = \frac{3}{7} = 0{,}4;$$

$P(A)$ – полную вероятность наступления события A вычислим по формуле

$$P(A) = \sum_{i=1}^{N} P(H_i) P(A \,|\, H_i),\ N = 4,$$

тогда

$$P(A) = \left(P(H_1) * P(A|H_1) + P(H_2) * P(A|H_2) + \right.$$

$$\left. + P(H_3) * P(A|H_3) + P(H_4) * P(A|H_4)\right) =$$

$$= (0{,}25 * 0{,}7 + 0{,}25 * 0{,}57 +$$

$$+ 0{,}25 * 0{,}28 + 0{,}25 * 0{,}4) = 0{,}487.$$

По формуле Байеса получим

$$P(H_1|A) = \frac{P(A \,|\, H_1) * P(H_1)}{\sum_{i=1}^{N} P(H_i) P(A \,|\, H_i)} =$$

$$= \frac{0{,}175}{0{,}487} = 0{,}36;$$

$P(H_1|A)$ – вероятность того, что объект x принадлежит первому классу, именно потому, что его признаки соответствуют признакам данного класса.

$$P(H_2|A) = \frac{P(A \,|\, H_2) * P(H_2)}{P(A)} = \frac{0{,}1425}{0{,}487} = 0{,}3;$$

$$P(H_3|A) = \frac{P(A \,|\, H_3) * P(H_3)}{P(A)} = \frac{0{,}07}{0{,}487} = 0{,}14;$$

$$P(H_4|A) = \frac{P(A \,|\, H_4) * P(H_4)}{P(A)} = \frac{0{,}1}{0{,}487} = 0{,}2.$$

Вычисленные вероятности по формуле Байеса нам позволяет сократить перебор классов. Для этого проранжируем по убыванию значения вероятности, т.е.

$$p(H_2|A) = 0{,}3;$$

$$p(H_4|A) = 0{,}2;$$

$$p(H_3|A) = 0{,}14.$$

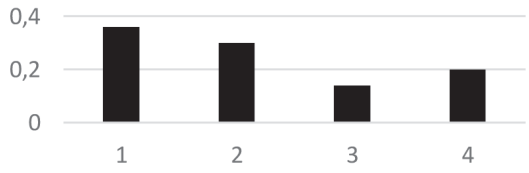


Рис. 3. График вероятностей принадлежности объекта i-тому классу ω_i, $i=1,4$

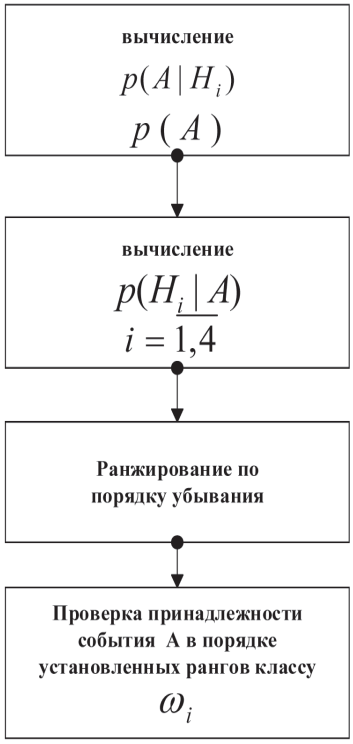


Рис. 4. Блок-схема алгоритма определения порядка идентификации

Идентификацию принадлежности события A определенному классу будем проверять в последовательности ω_1, ω_2, ω_4, ω_3.

Заключение

Решена задача 1, т.е. разработаны математическая модель и алгоритм вычислительной процедуры, который позволяет нам классифицировать объекты из одной предметной области, исходя из их свойств (признаков), при условии, что часть признаков разных объектов могут совпадать.

Решена задача 2, т.е. на основе вероятностного подхода построены модель и алгоритм вычислительной процедуры, позволяющий идентифицировать принадлежность экземпляра объекта определенному классу из предметной области.

В заключение следует отметить, что предложенный алгоритм можно использовать для решения задач большой размерности, что является их преимуществом перед множеством известных простых методов классификации.

Список литературы

1. Лепский А.Е. Математические методы распознавания образов / А.Е. Лепский, А.Г. Броневич. – Таганрог: Курс лекций, 2009. – 155 с.

2. Патрик Э.А. Основы теории распознавания образов / Пер. с англ.; под ред. Б.Р. Левина. – М.: Советское радио, 1980. – 407 с.

3. Потапов А.С. Распознавание образов и машинное восприятие: общий подход на основе принципа минимальной длины описания. – СПб.: Политехника, 2007. – 548 с.

4. Чабан Л.Н. Теория и алгоритмы распознавания образов / Л.Н. Чабан. – М.: Учебное пособие, М.: МИИГАиК, 2004. – 70 с.

5. Goswami J.C. Fundamentals of wavelets. Theory, algorithms, and applications / J.C. Goswami, A.K. Chan. – Wiley, 2000. – 359 p.