

УДК 004.93-1

МЕТОД КЛАССИФИКАЦИИ С ЗАРАНЕЕ НЕИЗВЕСТНЫМ СОСТАВОМ И КОЛИЧЕСТВОМ КЛАССОВ И ЗАДАНЫМ КРИТЕРИЕМ ИХ РАЗДЕЛЕНИЯ

К.С. КУРОЧКА, Н.Н. МАСАЛИТИНА

Гомельский государственный технический университет
пр. Октября, 48, Гомель, 246746, Беларусь

Поступила в редакцию 15 марта 2012

Предложен метод мультиклассовой классификации без учителя, позволяющий использовать качественные критерии классификации в условиях невозможности описания полного алфавита классов и отсутствия объектов, принадлежность которых к искомым классам заранее определена. Метод ориентирован на классификацию объектов, обладающих иерархической структурой управления и основывается на математическом аппарате алгебры логики. Представлен пример классификации промышленных предприятий по критерию принципиального отличия возможностей и целей управления устойчивостью.

Ключевые слова: классификация без учителя, мультиклассовая классификация, качественные критерии классификации, алгебра логики, иерархическая взаимосвязь.

Введение

Методы и модели классификации лежат в основе принятия решений во многих сферах деятельности человека. Качество математического аппарата, позволяющего выделить среди исследуемых объектов относящиеся к различным типам в соответствии с заданным критерием разделения, во многом определяет эффективность управления системой (устранения неисправностей, лечения и т.п.). В связи с этим в современных теоретических и прикладных исследованиях уделяется значительное внимание вопросам классификации объектов различной природы.

Среди существующих подходов к решению задач классификации можно выделить две принципиально различные группы.

Первая группа объединяет интеллектуальные методы, позволяющие формализовать правило классификации на основе выборки, содержащей объекты, принадлежность которых к различным классам заранее задана [1,2]. Такие методы принято объединять под названием классификации с учителем. Они требуют значительного объема исходной информации о классифицируемых объектах: состав классов, на которые нужно разделить объекты; признаки по которым элементы класса отличаются от элементов другого; значительное количество эталонов – объектов, принадлежность которых к заданным классам не вызывает сомнения.

Методы второй группы направлены на выявление среди исследуемых объектов наиболее отличающихся по некоторым количественным признакам на основе вычисления различных расстояний [3,4]. Эти методы носят название методов кластерного анализа или классификации без учителя. Они могут применяться и при отсутствии подробной информации о составе множества классов и не требуют наличия эталонов. Результатом классификации будет разделение множества исследуемых объектов на подмножества, элементы которых схожи по заданным количественным признакам. Однако природа родства между объектами различных классов остается необъяснимой.

Таким образом, задачи классификации по критерию, не поддающемуся количественному описанию, не содержащие достаточно исходной информации для классификации с учителем, не могут быть решены в рамках существующих подходов.

Применение аппарата алгебры логики для классификации объектов, обладающих иерархической структурой управления (ОИСУ), позволило авторам получить метод мультиклассовой классификации без учителя, допускающий использование неколичественного описания критериев разделения классов.

Обобщенная постановка задачи

Формализовать правило, разделяющее множество ОИСУ на n классов b_i ($i = 1 \dots n$, где n – количество классов), таких, что элементы каждого класса b_i отличаются от элементов каждого другого класса b_j ($j = 1 \dots n, j \neq i$) по критерию CR_{ij} – в состав множества M_{i_i} управляющих воздействий, необходимых для повышения устойчивости объекта класса b_i , входят элементы, не входящие в множество M_{tj} управляющих воздействий, необходимых для элементов класса b_j .

Описание метода классификации

Этап 1. Отбор обобщенных дескрипторов для описания ОИСУ.

Этап 1.1. Описание сбоев отдельных подсистем с помощью двоичных показателей I (0 – есть сбой, 1 – подсистема исправна).

Этап 1.2. Описание активности отдельных механизмов управления D (1 – механизм применяется, 0 – механизм не задействован).

Этап 1.3. Описание требований внешнего регулирования EN (1 – объект соответствует некоторым нормам внешнего регулирования, 0 – в противном случае).

Этап 2. Описание всех возможных состояний систем, что равнозначно составлению всех возможных сочетаний p двоичных признаков. В результате будет получено 2^p возможных сочетаний. Значительная часть из них невозможна в реальных условиях. С целью исключения таких сочетаний необходимо выполнить анализ взаимосвязей между показателями I, D и EN .

Этап 3. Описание связей между обобщенными дескрипторами I, D, EN . Результатом данного этапа является формализация множества запрета Z – правил, ограничивающих возможные сочетания значений дескрипторов [5].

Этап 3.1. Формализация запретов физических противоречий ZP – т.е. запретов, определяемых физическими законами существования и взаимодействия подсистем классифицируемых объектов.

Этап 3.2. Формализация запретов иерархии ZH – т.е. запретов, определяемых иерархическими связями между подсистемами классифицируемых объектов.

Этап 3.3. Формализация запретов рационального управления ZRM , т.е. запретов, описывающих логику управления, нацеленного на повышение устойчивости и применение для этого наиболее эффективных средств.

Этап 3.4. Формализация запретов внешнего регулирования ZEN , т.е. ограничений возможностей функционирования и управления классифицируемыми объектами со стороны внешних сил (законодательной системы, морально-этических норм, охраны труда и проч.).

Этап 4. Исключение из множества возможных состояний Sb ОИСУ таких сочетаний дескрипторов, которые соответствуют запретам Z .

Этап 5. Анализ множества оставшихся непротиворечивых состояний с целью объединение их в группы, наиболее соответствующие требованиям критерия CR_{ij} .

Этап 6. Представление лицу, принимающему решение (ЛПР), полученного на этапе 5 состава классов ОИСУ. Если результат удовлетворителен, следует перейти к этапу 7. В противном случае следует рассматривать другие варианты классификации, пока не будет найден удовлетворительный вариант или пока не будут рассмотрены все возможные варианты классификации. Если ни один из предложенных вариантов не удовлетворяет требованиям ЛПР, следует повторить этапы 1-5, изменив состав обобщенных дескрипторов.

Этап 7. Построение классификатора в виде дизъюнктивной нормальной формы [5] от показателей, определяющих отличие элементов различных классов $DNF (I', D', EN')$, где $I' \in I, D' \in D, EN' \in EN$.

Классификация предприятий в соответствии с критерием различия целей и возможностей управления устойчивостью

Возможности применения предложенного метода классификации представим на примере распознавания нарушений различных стадий кризисного процесса промышленных предприятий. Предприятия характеризуются многоуровневой системой управления и, при этом, отдельные подсистемы связаны отношениями иерархической подчиненности. Они существенно различны в отношении тяжести кризисных изменений. Инструменты управления, эффективные на одних стадиях кризиса, неэффективны (и часто вредны) на других. Вместе с тем, сложно четко сформулировать признаки, отличающие предприятия, требующие различных приемов управления.

Таким образом, для данного класса систем обобщенная постановка задачи примет вид: разработать классификатор, позволяющий разделить предприятия на классы, такие, что экземпляры каждого класса схожи между собой и два любых объекта, принадлежащих различным классам, принципиально отличаются по составу целей и возможностей управления устойчивостью.

Этап 1. Отбор обобщенных дескрипторов для описания состояния предприятия.

В соответствии с разработанной авторами моделью устойчивости производственной системы [6], состояние sb объекта b может быть описано с помощью следующих обобщенных дескрипторов:

- исправность I его структурных элементов: материальных объектов Mb , материальных Pb и финансовых потоков Fb ;
- активность D механизмов устойчивости и антикризисного управления.

Требования системы внешнего регулирования в сфере управления устойчивостью предприятий EN определяют законодательно установленные критерии диагностики экономической несостоятельности.

Таким образом, состояние sb объекта b в каждый момент может быть описано вектором вида:

$$sb = \begin{bmatrix} I(Mb) \\ I(Pb) \\ I(Fb) \\ D(U_{Mb}) \\ D(U_{Pb}) \\ D(U_{Fb}) \\ D(A_{Mb}) \\ D(A_{Pb}) \\ D(A_{Fb}) \\ D(A_{sb}) \\ EN \end{bmatrix},$$

где Mb – множество материальных объектов предприятия; Pb – множество материальных потоков предприятия; Fb – множество финансовых потоков предприятия; U_{Mb}, U_{Pb}, U_{Fb} – множества механизмов устойчивости, действующих на уровне материальных объектов, материальных потоков и финансовых потоков соответственно; A_{Mb}, A_{Pb}, A_{Fb} – множества механизмов антикризисного управления, действующих на уровне материальных объектов, материальных потоков и финансовых потоков соответственно; A_{sb} – множество механизмов антикризисного управления, воздействующих одновременно на несколько подсистем предприятия; I – свойство исправности подсистем (материальных объектов, материальных потоков, финансовых потоков) предприятия; D – свойство активности механизмов устойчивости и антикризисного управления; EN – обобщенный показатель соответствия предприятия требованиям внешнего регулирования вопросов экономической несостоятельности.

Показатели I принимают значения 1 в случае, если все элементы соответствующего множества исправны:

$$I(Mb) = 1, \text{ если } \forall mb \in Mb | I(mb) = 1;$$

$I(Pb) = 1$, если $\forall pb \in Pb | I(pb) = 1$;

$I(Fb) = 1$, если $\forall fb \in Fb | I(fb) = 1$.

Свойство активности принимает значение 1, если хотя бы один элемент множества активен:

$D(U_{Mb}) = 1$, если $\exists u_{mb} \in U_{Mb} | D(u_{mb}) = 1$;

$D(U_{Pb}) = 1$, если $\exists u_{pb} \in U_{Pb} | D(u_{pb}) = 1$;

$D(U_{Fb}) = 1$, если $\exists u_{fb} \in U_{Fb} | D(u_{fb}) = 1$;

$D(A_{Mb}) = 1$, если $\exists a_{mb} \in A_{Mb} | D(a_{mb}) = 1$;

$D(A_{Pb}) = 1$, если $\exists a_{pb} \in A_{Pb} | D(a_{pb}) = 1$;

$D(A_{Fb}) = 1$, если $\exists a_{fb} \in A_{Fb} | D(a_{fb}) = 1$;

$D(A_{sb}) = 1$, если $\exists a_{sb} \in A_{sb} | D(a_{sb}) = 1$.

В противном случае перечисленные величины принимают нулевое значение.

Показатель EN принимает значение 1, в случае если предприятие соответствует критериям диагностики неплатежеспособности [7], 0 – в противном случае.

Результатом *этапа 2* является матрица размерностью 14×214 , описывающая возможные сочетания 11 выбранных обобщенных дескрипторов. Каждый столбец матрицы которой соответствует отдельному состоянию объекта (т.е. аналогичен вектору sb), строки матрицы соответствуют элементам вектора sb с теми же номерами.

Этап 3. Для исключения из полученной матрицы столбцов, содержащих недопустимые сочетания дескрипторов, заданы следующие запреты.

Запреты физических противоречий ZP :

– в случае отсутствия деструктивных воздействий на предприятие, сохраняется исправность его материальных объектов Mb , материальных потоков Pb и финансовых потоков Fb :

$$I(Mb) = 1 \wedge I(Pb) = 1 \wedge I(Fb) = 1 \quad (1)$$

– без применения каких-либо управляющих воздействий:

$$D(U_{Mb}) = 0 \wedge D(U_{Pb}) = 0 \wedge D(U_{Fb}) = 0 \wedge D(A_{Mb}) = 0 \wedge D(A_{Pb}) = 0 \wedge D(A_{Fb}) = 0 \wedge D(A_{sb}) = 0; \quad (2)$$

– если на предприятие оказывается деструктивное воздействие, но исправность какой-либо подсистемы сохранена, т.е.

$$I(x) = 1, \text{ для } x \in X, X = Mb \cup Pb \cup Fb; \quad (3)$$

то это достигается либо за счет механизмов устойчивости и тогда

$$D(u) = 1 \text{ для } u \in U, U = U_{Mb} \cup U_{Pb} \cup U_{Fb}; \quad (4)$$

либо за счет действия антикризисных мер, тогда

$$D(a) = 1 \text{ для } a \in A, A = A_{Mb} \cup A_{Pb} \cup A_{Fb}. \quad (5)$$

Запреты иерархии Z^H определяются иерархической подчиненностью между материальными объектами, материальными потоками, финансовыми потоками, а также элементами устойчивости и антикризисного управления на уровнях соответствующих подсистем.

Движение материальных потоков предприятия Pb (ритмичность и качество отгрузки продукции, поставка сырья и материалов) определяется состоянием его материальных объектов Mb (качеством работы персонала, состоянием производственной системы и проч.), следовательно,

$$\text{если } I(Mb) = 1, \text{ то } I(Pb) = 1. \quad (6)$$

В результате, сбой, возникшие на уровне материальных объектов Mb , распространяются на уровень материальных потоков Pb , что автоматически приводит к активизации механизмов устойчивости соответствующего уровня U_{Pb} :

$$\text{если } I(Mb)=0, \text{ то } D(U_{Pb})=1. \quad (7)$$

Аналогичная закономерность существует и в отношении финансовых потоков Fb и материальных потоков Pb :

$$\text{если } I(Pb)=1, \text{ то } I(Fb)=1; \quad (8)$$

$$\text{если } I(Pb)=0, \text{ то } D(U_{Fb})=1. \quad (9)$$

Запреты рационального управления ZR :

– механизмы антикризисного управления разрабатываются и применяются в случае нарушения исправности некоторой одной или нескольких подсистем предприятия, а значит

$$\text{если } I(Mb)=0, \text{ то } D(A_{Mb})=1; \quad (10)$$

$$\text{если } I(Pb)=0, \text{ то } D(A_{Pb})=1; \quad (11)$$

$$\text{если } I(Fb)=0, \text{ то } D(A_{Fb})=1; \quad (12)$$

– проект санации разрабатывается и реализуется в случае, если кризисный процесс не удалось остановить с помощью антикризисных мер, действующих на отдельные подсистемы предприятия, т.е.

$$D(A_{sb})=1, \text{ если } I(Mb)=0 \wedge I(Pb)=0 \wedge I(Fb)=0 \wedge D(A_{Mb})=1 \wedge D(A_{Pb})=1 \wedge D(A_{Fb})=1. \quad (13)$$

Запреты внешнего регулирования ZEN . В силу того, что экономическая несостоятельность предприятия оценивается на основе показателей, характеризующих финансовые потоки [7], признаки этого состояния могут возникнуть только в случае нарушения исправности элементов множества финансовых потоков:

$$EN=1 \text{ тогда и только тогда, когда } I(Fb)=0 \quad (14)$$

Этан 4. С учетом рассмотренных взаимосвязей между дескрипторами состояние sb предприятия b под действием кризисного процесса в течение одного цикла производства-реализации продукции может характеризоваться одним из 10 значений:

$$Sb = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Каждый столбец Sb_i ($i=1..10$) матрицы Sb соответствует отдельному состоянию объекта (т.е. аналогичен вектору sb), строки матрицы соответствуют элементам вектора sb с теми же номерами.

Представленные в матрице значения получены посредством исключения сочетаний, которые противоречат выражениям (1)–(14), из множества возможных сочетаний двоичных величин I , D и EN .

Этан 5. Сравнения представленных матрицей Sb состояний предприятия позволило выделить среди них ОИСУ, имеющие принципиальные отличия по заданному критерию CR_{ij} .

1. Предприятия, не требующие антикризисного управления – устойчивые. Объединяет состояния, для которых справедливо следующее условие:

$$I(Mb) = 1 \wedge I(Pb) = 1 \wedge I(Fb) = 1.$$

Обозначим класс таких предприятий b_1 :

$$b_1 = \{Sb^1, Sb^2\}.$$

2. Для управления устойчивостью достаточно сдерживающих мер на уровне отдельных подсистем, т.к. без привлечения приемов глубокого оздоровления, охватывающего различные подсистемы $D(A_{Sb}) = 0$, сохраняется исправность элементов уровня финансовых потоков $I(FB) = 1$, определяющих возобновление резервов управления.

Матрица Sb содержит 5 столбцов, соответствующих таким условиям. Для обозначения данного класса будем использовать символ b_2

$$b_2 = \{Sb^3, Sb^4, Sb^5, Sb^6, Sb^7\}.$$

3. Требуется инициативной санации – состояние самостоятельного антикризисного управления – b_3 . Для таких объектов недостаточно поддерживающих антикризисных мер на уровне отдельных подсистем:

$$D(U_{Mb}) = 1 \wedge D(U_{Pb}) = 1 \wedge D(U_{Fb}) = 1 \wedge D(A_{Mb}) = 1 \wedge D(A_{Pb}) = 1 \wedge D(A_{Fb}) = 1,$$

$$\text{при этом } I(Mb) = 0 \wedge I(Pb) = 0 \wedge I(Fb) = 0$$

Для обозначения данного класса будем использовать символ b_3 :

$$b_3 = \{Sb^8, Sb^9\}.$$

4. Требуется внешнего антикризисного управления – потенциальные банкроты. Для таких предприятий $EN = 1$.

$$b_4 = \{Sb^{10}\}.$$

Элементы выделенных классов характеризуются принципиальной разницей целей и возможностей управления устойчивостью. Поэтому в случае отсутствия дополнительных требований со стороны ЛПР, этап 6 можно считать пройденным с положительным результатом.

Этап 7. Для построения классификатора в виде дизъюнктивной нормальной формы $DNF (I', D', EN')$ рассмотрим значения каждого используемого дескриптора для объектов различных классов. Для каждого класса существует один или два дескриптора, значение которых определяет принадлежность объекта к данному классу:

$$- D(U_{Pb}) = 1, \text{ для класса } b_1;$$

- $I(Fb) = 1$ и $I(U_{Pb}) = 1$, либо $I(Fb) = 1$ и $D(A_{Mb}) = 1$ для класса b_2 (в дальнейшем для определения принадлежности классу b_2 будем использовать показатель $I(Fb)$ и $D(UPb)$);

$$- I(Fb) = 0 \text{ для класса } b_3;$$

$$- EN = 1 \text{ для класса } b_4.$$

Следовательно, для однозначного разграничения классов b_1, b_2, b_3, b_4 достаточно трех дескрипторов:

$$\{I', D', EN'\} = \begin{bmatrix} I(Fb) \\ D(U_{Pb}) \\ EN \end{bmatrix}.$$

Для объектов распознаваемых классов b_1, b_2, b_3, b_4 матрица Sb' примет следующий вид:

$$Sb' = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Разграничение классов b_1, b_2, b_3, b_4 может быть выполнено на основе следующей дизъ-

юнктивной формы:

$$DNF = \overline{I(Fb)} + D(U_{pb}) + EN,$$

принимаящей значение 0 – для объектов класса b_1 ; 1– для объектов класса b_2 ; 2– для объектов класса b_3 ; 3– для объектов класса b_4 .

Значительный практический интерес имеет также сокращенная форма классификатора, позволяющего распознать классы b_2 , b_3 и b_4 . в ситуации, когда известно, что классифицируемое множество не содержит элементов класса b_1 :

$$DNF' = \overline{I(Fb)} + EN.$$

DNF' принимает значение 0 – для объектов класса b_2 ; 1– для объектов класса b_3 ; 2– для объектов класса b_4 .

Результаты и их обсуждение

В результате проведенных исследований получен метод классификации с заранее неизвестным составом классов и заданным критерием их разделения, а также математическая модель классификации предприятий по критерию различия целей и возможностей управления устойчивостью.

Полученные метод и математическая модель иллюстрируют возможности решения задач, в равной степени недоступных для классических методов классификации с учителем и без учителя – задач с неполной информацией о составе классов. При этом предложенный инструментарий позволяет разделять объекты не произвольным случайным образом, а в соответствии с заданным критерием классификации.

В отличие от методов классификации с учителем, предложенный математический аппарат более гибок и не требует упрощения исходной задачи с целью обязательного определения полного алфавита классов на начальных этапах синтеза математической модели классификации.

Применение аппарата алгебры логики с целью отбора значимых признаков, используемых при классификации, также имеет ряд преимуществ перед альтернативными методами (корреляционный анализ или вычисление различных расстояний). Предложенный вариант отражает логику взаимодействия между реальными подсистемами классифицируемых объектов и, в отличие от аналогов, экономичен и свободен от ошибок, связанных с неадекватностью характеристик выборочной совокупности, используемой при оценке информативности дескрипторов.

Представление классификатора в дизъюнктивной нормальной форме предоставляет широкие возможности автоматизированной и аппаратной реализации предложенного метода классификации.

THE CATEGORIZATION METHOD WITH UNDEFINED CLASS ALPHABETIC AND GIVEN CRITERION OF THEIR DIVISION

K.S. KURACHKA, N.N. MASALITINA

Abstract

The multiclass categorization method without learning was proposed. It allows to use a qualitative criterions of categorization. The whole class alphabetic and etalons ensemble is not required. The method is orientated on objects with an hierarchy structure of management system. Principles of Boolean algebra are in the base of the method. The example of en industrial enterprises categorization is represented.

Список литературы

1. Головки В.А. Нейросетевые методы обучения и обработки информации в системах управления и прогнозирования: Дис. на соиск. учен. степ. д-ра техн. наук. Минск, 2002.
2. Корноушенко, Е.В. // Материалы межд. научн. конгресса по информатике «Информационные системы и технологии», Минск, 31 окт. – 3 нояб. 2011 г. Минск: БГУ, 2011. С.81–85
3. С.А. Айвазян. Прикладная статистика: Классификация и снижение размерности: справ. изд. М., 1989.
4. Мандель И.Д. Кластерный анализ. М.:, 1988.
5. Закревский Д.А. Логика распознавания. Минск, 1988.
6. Масалитина Н.Н. // Информатика. 2010, №4. С. 57–71.
6. Инструкция по анализу и контролю за финансовым состоянием и платежеспособностью субъектов предпринимательской деятельности: Постановление Министерства финансов Республики Беларусь, Министерства экономики Республики Беларусь, Министерства статистики и анализа Республики Беларусь, 14 мая 2004 г., № 81/128/65 // Эталон-Беларусь [Электрон. ресурс] / Нац. центр правовой информ. Республики Беларусь. – Минск, 2004.

Библиотека БГУМР