

MATEMATIKA И МАТЕМАТИЧЕСКО ОБРАЗОВАНИЕ, 2006
MATHEMATICS AND EDUCATION IN MATHEMATICS, 2006
*Proceedings of the Thirty Fifth Spring Conference of
the Union of Bulgarian Mathematicians
Borovets, April 5–8, 2006*

РАЗПОЗНАВАНЕТО КАТО ПОДХОД В ОБУЧЕНИЕТО
ПО ИНФОРМАТИЧНИ ДИСЦИПИЛИНИ*

Маргарита Й. Тодорова, Димитър Й. Биров

В работата се разглежда прилагането на методите за разпознаване на образи в обучението по информатични дисциплини. Показана е обвързаността на дисциплините и практическото приложение на методите, важни за решаването на задачи от различни области. Обучението по компютърни науки се разглежда като конкретна методика за формиране, обобщение и класификация на математически и информатични понятия, класове от обекти и конкретни задачи, т.е. като задача за разпознаване на образи.

1. Терминология. В средата на миналия век се определи обликът на ново научно направление – pattern recognition. Исторически този термин се налага в българския език като взимстван от английската литература. Впоследствие се наложи терминът *разпознаване на образи*. Понятието “образ” (pattern) е многозначно и съдържа в себе си различни семантични значения като например, модел, начин, стил, режим, закономерност, начин на действие. При съвременното му приложение се налага по-широк смисъл, като се има предвид *структурираното описание на разглеждания обект, представляващо принципно свойство на образа*.

Естествено е да се допусне, че образът се състои от две групи символи: описание на *променливите* и описание на *постоянните характеристики* на обекта [2]. Основната цел на въвеждането на понятието е използването му в процеса на установяване на съответствия, т.е. при доказването на идентичността, аналогията, подобие и сходството на обектите по пътя на сравнението (съпоставянето). В различните задачи понятието *образ* има различен смисъл. В класическите схеми по разпознаване това е *признак*, в структурните модели – *създание*, породено от граматика, характеризираща класа, към който разглежданият образ принадлежи и т.н. Терминът *разпознаване* се свързва с процеса на възприятие и познание, присъщи на човека и живия организъм като цяло.

2. Задачи за разпознаването на образи. Основната задача за разпознаването на образи се състои в построяването на ефективни процедури на базата на систематични, теоретични и експериментални изследвания за класифициране на формализирани описания и обекти в съответстващите им класове [2]. Разпознаването на образи се свързва с две функции: *отнасяне на обект към клас от обекти*, който е неизвестен на класификатора, и *идентификация на обекта* като елемент,

* **Ключови думи:** разпознаване на образи, методика на преподаване във ВУЗ

вече известен на класификатора. Първата функция е позната като **процес на от-
деляне** (формиране) на нови класове, а втората като **разпознаване**.

Образът (моделът) сам по себе си не е изолирано наблюдение, а колекция от наблюдения, свързани във времето и пространството. Моделът като цяло демонстрира скритата зад него концепция. Класификацията на образи представя това интуитивно разбиране по-формално в понятието за образ – два различни свята са свързани помежду си – света на физичното наблюдение, върху което могат да се извършават различни по вид измервания и света на концепциите и идеите. Ето защо образът може да се разглежда като двойка променливи: *образ* = [свкупност от наблюдения – v , концепция – ω].

Ако v е свкупността от наблюдения, които принципно биха могли да бъдат направени чрез технически устройства, то ω е концепцията скрита зад тези наблюдения. Ако v може да се представени чрез многомерен континуум от стойности, то ω може да бъде получено от дискретно и крайно множество Ω концепции. За всяка потенциална стойност $\omega \in \Omega$ може да се асоциира различно име. От тази гледна точка отделния образ е едно от многобройните превъплъщения на концепцията ω , представена чрез конкретно разложение на наблюдаваните случаи със стойности описани чрез v . Класификацията на образи е процес на извличане на ω от v , с други думи това е процес на именуване с име ω на наблюденията v .

Като основни задачи, които се решават при процедурите за разпознаване и приложението на методите при изучаване на информатични дисциплини могат да се посочат:

Първата задача е свързана с пълното определяне на признаците (параметрите), които характеризират обекта или явлението. Изборът на признаците представлява евристична процедура, която може да се осъществи като се отчитат общите свойства. Основно в информатичните дисциплини (обектно ориентирано програмиране, структури от данни, абстрактни структури от данни, дискретна математика) се използват два типа признаци: *логически* (характеристики на явления или обекти, представими във вид на елементарни съждения) и *структурни признаци* (*примитиви*, които не могат да се раздробят на по-елементарни структури). На базата на тези примитиви се построява обект, аналогично на изграждането на формален език по пътя на комбинации от думи, в смисъла на математическата лингвистика, които от своя страна се състоят от терминални символи – букви.

На основата на ограничен брой атомарни примитиви (в производни елементи) се построяват и описват разнообразни по структура обекти. От един и същ набор атомарни обекти могат да се получат чрез конструкция и рекомбинация богат набор производни обекти. Два еднакви примитива групирани на базата на различни правила, водят до получаване на различни производни обекти. *Като цяло описанието на произволен обект се извършва с помощта на примитиви (непроизводни елементи), характерни предимно за описвания обекта.*

Втората задача решава въпроса с първоначална класификация на обектите, подлежащи на разпознаване, т.е. *съставяне на априорния списък на класове от обекти*. Много често при методите за решаването ѝ преобладават евристичните, като избора на признаците се свежда до определяне на това, какви решения и от кого се приемат в резултат на процеса на разпознаване, формулиране на изискванията към процедурата за разпознаване въз основа на уточнената цел, съставяне на

предварителен списък на класовете (множеството от класове).

Третата задача е свързана с изграждане на **азбуката** на признаците. Решавайки първата задача намираме всички възможни признаци за разпознаване на даден обект или явление. Това важи и за втората задача – определянето на състава на класовете. Ефективността на избраните признаци може да се оцени след провеждането на изпитания с разработената процедура за разпознаване. На база на метода на *последователните приближения* и създаването на математически модел на процедурата може да се избере азбуката на признаците, обезпечаваща исканото качество на решение.

Четвъртата задача се състои в описание на априорната азбука в термините на признаците. Това е най-трудоемката и творческа задача. При решаването ѝ обикновено се построява изображение на всеки клас върху пространството от логическите и структурните признаци.

Пета задача представлява избор на алгоритъма за класификация. Непосредственото решение на задачата се състои в разбиване на пространството от стойности на признаците на отделни области $d_1, d_2, \dots, d_n$, съответстващи на класовете $\omega_1, \omega_2, \dots, \omega_n$. Разбиването трябва да осигурява минимален брой грешки по отношение на попадането на обектите в “чужди” класове. В резултат на класификацията обекта $X_i = (x_1, x_2, \dots, x_n)$ попада в класа ω_i , ако той принадлежи на съответната област от признаци d_i .

При алгоритмите, основани на *логически признаци*, понятието “мярка за близост” на обекта до класа няма смисъл. Достатъчно е да заместим със стойностите на признаците в булевите изрази за съответните класове и да изчислим резултат за булевата функция. За алгоритмите, основани на структурни признаци понятието “мярка за близост” е по-специфично. Тъй като всеки клас се описва с редици от примитиви, характеризиращи структурните особености на обектите, разпознаването на неизвестен обект се извършва чрез идентификация (определя се най-голямото сходство между редицата, описваща разпознавания обект с една от редиците описваща даден клас) на редицата от примитиви по отношение на принадлежността ѝ към определен клас. Процедурата за разпознаване се реализира чрез последователни приближения. В процеса на усъвършенствуването ѝ използваният набор от признаци се попълва, а неефективните признаци се премахват.

Шеста задача е свързана с определяне на *работните множество класове и азбука на признаците*, т.е. на такъв **работен списък** от класове и азбука на признаците, които да обезпечат максималната ефективност на разпознаването.

3. Процедури за разпознаване. За да отразим същността на задачата (процедурата) ще въведем следните означения: $\mathbf{x}$ – входна променлива (или набор от променливи), наречена образ и приемаща стойности от множеството $\mathbf{V}$. Нека $\mathbf{X}$ е множество от променливи, които се изчисляват от разпознаващата процедура в процеса на анализ на входните данни $\mathbf{x}$. Част от тези данни участва във формирането на крайния резултат. Друга част представлява междинни променливи, изчислявани в процеса на работата на процедурата, но не оказващи пряко влияние върху крайния резултат.

Ако с $\mathcal{K}$ означим множеството от всевъзможните стойности на $\mathbf{X}$, то реализацията на процедурата се свежда до реализация на изображението: $\mathbf{f} : \mathbf{V} \rightarrow \mathcal{K}$, т.е.

алгоритъмът представлява реализация на отношение, зададено върху множеството $V \times K$. Ако отношението е зададено формално, например с помощта на алгоритъма $F : V \times K \rightarrow \{0, 1\}$, то F е формална постановка на решаваната задача. Решението y е такъв алгоритъм $f : V \rightarrow K$, за който равенството $F(x, f(x)) = 1$ е вярно за произволно x [2]. По този начин се разглеждат два типа алгоритми – единият е свързан с постановката на задачата, а другият с решението y . Те се различават, както се различават функциите зададен явно и неявно. Решението на задачата се свежда до намиране на алгоритъм, реализиращ в явен вид изображението $f : V \rightarrow K$, което се задава неявно с алгоритъма: $F : V \times K \rightarrow \{0, 1\}$.

При задачите за разпознаване се използва понятието сходство (подобие) или различие при вземането на решения за разпознаване към кой клас принадлежи разглеждания обект. Необходимата информация се съдържа както в отделните признаци, които обектите притежават, така и в различните им комбинации. Именно такава комбинация е *неприводимия дескриптор* [1, 3]. Всеки дескриптор може да бъде записан като поредица от стойностите на примитивите, описващи обекта или да се представи във вид на конюнкции. Всички конюнкции, построени за даден обект могат да бъдат свързани с дизюнкции. Така множеството от примитиви, построено за всеки обект може да бъде представено като булева формула в ДНФ. Свързвайки булевите формули с дизюнкции на обектите от даден клас получаваме булева форма, описваща класа. В [2, 3, 4] формулираната по-горе задача, се свежда до построяване на всички примитиви, породени от обектите на даден клас до задача за намирането на съкратена ДНФ на частична булева функция със стойност 1 за обектите на дадения клас и 0 за обектите от останалите класове.

4. Области на приложение в обучението по информатични дисциплини. Естествено и логично превъплъщение на задачите споменати по-горе е обектно-ориентирания подход в моделирането, в който се определят признаците на моделираното явление или обект, класифицират се обектите от предметната област (съставя се априорният списък на класовете от обекти), дефинират се признаците за разграничаване на отделните обекти, което заема съществено място в курсовете по обектно-ориентирано програмиране и абстрактни структури от данни.

Друго съществено практическо приложение на подхода по разпознаване е в процеса на софтуерното производство, където с помощта на така наречените софтуерни модели (design pattern), се класифицират често повтарящи се софтуерни практики в три основни направления: – създаване на обекти, конструкции, библиотеки; изграждане на архитектурата на софтуерни системи; описание на поведение на обектите и софтуера.

Известна теория, която се основава на връзките, взаимодействията и подходите при разпознаване на образи е т.н. теория на класификацията, която се базира на строга математическа основа – λ – смятане от първи, втори и по-висок ред, което позволява процеса на класификация на обектите да се извършва автоматизирано от страна на компилатора. Класификацията на обектите в този случай се извършва на основата на строги правила, като от първа до четвърта задачи се извършват автоматично от страна компилатора, а при пета задача алгоритъма за класификация е формално изведен от теорията на λ – смятането. Интересното е, че подхода на разпознаване на образи добре описва явлението наследяване и параметричен

полиморфизъм, които са едни от основните механизми за абстракция в обектно-ориентирания подход.

Отличаването на даден обект (явление) от друг, се прави на база различни качества, признаци или особености на обектите (и явленията). Сред различните свойства на изучаваните обекти, се отделят: *единични (индивидуални) свойства и общи свойства*. Обобщаването на обектите и явленията посредством **понятията** увеличава познавателната ценност на мисленето: по-общите понятия дават възможност мислено да се обхванат и изучат по-обширни множества от обекти, а отхвърляйки индивидуалните признаци на обекта ние подчертаваме общите, по-устойчиви признаци, които в рамките на по-тясното понятие остават неразкрити. Друг начин на обобщение позволява да се образуват т.н. *конкретни понятия*. Особеността се състои в това, че обобщението тук произтича не само по пътя на отделянето на общите свойства на обектите, но и по пътя на съхраняване на понятието в особени, единични признаци.

Всяко понятие може да бъде разгледано по съдържание и по обем. *Съдържанието на понятието* определя обема му и обратно обема напълно определя съдържанието. По този начин изменението на съдържанието довежда до изменение на обема му и обратно. В процеса на обобщение обема става по-голям, а съдържанието се стеснява. При процеса на специализация на понятието обема на понятието се свива, но се разширява съдържанието му.

Процеса на разкриване на съдържанието на дадено понятие се състои в изброяването на неговите признаци. Изброяването на необходимите и достатъчни признаци на понятието в свързано изречение (мисъл) е и определение на понятието (математическия обект). Всеки от признаците влизащ в определението трябва да бъде необходим, а заедно с това достатъчен за установяване на даденото понятие. В определението трябва да се разкрива основното съдържание на понятието. *Понятието може да се определи чрез различни начини*: чрез най-близкият род и видово различие; генетически (начин, посочващ произхода на понятието); индуктивно; чрез абстракция. В *процеса на изясняване* на понятието се прави **класификацията** на понятието (разделяне на множествата от обекти, съставляващи обема на родовото понятие) на видове. Това разделение е основано на сходството на обектите от един вид и различието им от обектите на други видове в съществените признаци.

Правилната класификация изисква съблюдаването на определени правила: *класификацията да се провежда на база на определен признак*, оставащ неизменен в процеса на класификация; понятията – резултат на класификацията трябва да бъдат *взаимно независими*, а обемът от понятия, получаващи се при класификация трябва да се равнява по обем на изходното понятие. В процеса на класификация е необходимо да се премине към най-близкият вид на даденото родово понятие. Обучението може да бъде евристично – учителя (преподавателя) не съобщава на обучаваните готови, подлежащи на усвояване сведения, а ги навежда към самостоятелно преоткриване на съответстващите предложения, твърдения и правила. Същността е самостоятелно, планирано в общи линии търсене на решението на поставения проблем. Целта на всяка евристика е да се изследват правила и методи, водещи до открития.

5. Заключение. В процеса на развитие на методиката на преподаване в послед-

но време се набляга на преодоляването на разрыва между училищната информатика и компютърните науки. В определен смисъл можем да разглеждаме методиката на преподаване по информатични дисциплини и като граничен раздел на педагогика с философията, логиката, психологията, кибернетиката, биологията и др., изследващ закономерностите на обучението по компютърни науки на определен етап от нейното развитие, в съответствие с целите на обучение, поставени от обществото и призвана да разкрие закономерностите на процеса на обучение по компютърни науки. В контекста на казаното по-горе можем да разглеждаме обучението по компютърни науки и като конкретна методика за формиране, обобщение и класификация на математически и информатични понятия, класове от обекти и конкретни задачи, т.е. като задача за разпознаване на образи.

ЛИТЕРАТУРА

- [1] В. ВЪЛЕВ. Неприводими дескриптори и някои техни приложения в задачите за разпознаване на образи. *Математика и математическо образование*, **25** (1996), 29–44.
- [2] Ю. И. ЖУРАВЛЕВ, И. Б. ГУРЕВИЧ. Распознавание, классификация, прогноз. Математические методы и применение. Вып. 2. Москва, Наука, 1989, 302 с.
- [3] М. ТОДОРОВА, Н. СИНЯГИНА, Д. БИРОВ. Интегриране на обучението по дисциплините разпознаване на образи с математически и информатични дисциплини. *Математика и математическо образование*, **34** (2005), 382–388.
- [4] М. ТОДОРОВА, Н. СИНЯГИНА. Взаимообвързаност на обучението по дисциплините “Разпознаване на образи” и “Математическа логика”. Ст. Загора, 2004, Том I, 372–377.

Маргарита Йорданова Тодорова
ЮЗУ “Н. Рилски”, ПМФ
катедра “Информатика”
2700 Благоевград
e-mail: todorova@aix.swu.bg

Димитър Йорданов Биров
СУ “Климент Охридски”
Факултет по математика и информатика
Катедра “Изчислителни системи”
ул. Дж. Баучер № 5
1126 София
e-mail: birov@fmi.uni-sofia.bg

THE RECOGNITION AS AN APPROACH IN THE LEARNING OF COMPUTER SCIENCE

Margarita Todorova, Dimitar Birov

The applicability of the approach of teaching informatics in recognition problems is discussed in this paper. The work is an initial step towards the integration the recognition procedures and to applying methods in the methodology of integrated teaching of computer science.