PRACE IPI PAN • ICS PAS REPORTS

Zdzisław Pawlak

Classification of objects by means of attributes

429

January 1981
WARSZAWA

INSTYTUT PODSTAW INFORMATYKI POLSKIEJ AKADEMII NAUK
INSTITUTE OF COMPUTER SCIENCE POLISH ACADEMY OF SCIENCES
00-901 WARSAW, P.O. Box 22, POLAND

Zdzisław Pawlak

CLASSIFICATION OF OBJECTS BY MEANS OF ATTRIBUTES

An approach to inductive inference

429

Warsaw, January 1981

Pracę zgłosił Andrzej Blikle

Mailing address: prof. Zdzisław Pawlak
Institute of Computer Science PAS
P.O. Box 22
00-901 Warszawa PKiN

ISSN 138 - 0648

Nakład 700 egz. Ark. wyd. 0,70; ark. druk. 1,25. Papier offset. kl. III, 70 g, 70 x 100. Oddano do druku w styczniu 1981 r. WDN zam.99/0/81 n.700

Abstract . Streszczenie . Содержание

Many problems of artificial intelligence are connected with classification of objects. A new approach to classification, based on information systems theory, is given in this paper and application to the automation of inductive inference is outlined.

This approach leads to a new formulation of the notion of fuzzy sets (called here the rough sets). The axioms for such sets are given, which are the same as axioms of topological closure and interior.

Klasyfikacja obiektów za pomocą atrybutów

Wiele problemów sztucznej inteligencji związanych jest z klasyfikowaniem obiektów. Podano tu próbę nowego spojrzenia na problemy klasyfikacji w oparciu o teorię systemów informacyjnych, oraz zastosowanie jej do automatyzacji rozumowania indukcyjnego.

Próba ta prowadzi między innymi do nowego sformułowania pojęcia zbioru rozmytego (zwanego tu zbiorem przybliżonym). Podano aksjomaty takich zbiorów, które są identyczne z aksjomatami topologicznego domknięcia i wnętrza.

Классификация объектов при помощи атрибутов

Подход к индуктивному выводу

Многие проблемы искусственного интелекта связаны с классификацией объектов. В настоящей работе представлена попытка нового взгляда на проблемы классификации на основе теории информационных систем, а также применение его к автоматизации индукционного рассуждения.

Эта попытка между прочим ведет к новой формулировке понятия размытого множества (которые здесь называется приближенным множеством). Даны аксиомы таких множеств. Они являются идентичные с аксиомами топологического замыкания и внутренности.

INTRODUCTION

Many problems of artificial intelligence are based on classification of objects. We propose here somewhat new approach to classification inspired by results of Michalski [2].

Departure point of the approach is the notion of an information system introduced by Marek, Pawlak in [1] and modified by Pawlak [3], which is called here Knowledge Representation System. We show here that if we classify objects by means of attributes exact classification is often impossible. We propose in this case approximate (upper and lower) classification, by means of two relations "x surely belongs to X" ($x \underline{\in} X$) and "x possibly belongs to X" ($x \overline{\in} X$).*) The method leads to very simple classification algorithm, which is outlined at the end of the paper. Application of this method to medical diagnosis is briefly discussed.

1. KNOWLEDGE REPRESENTATION SYSTEM

By a knowledge representation system (KRS) we shall mean a 4-tuple

$$S = \langle X, A, V, \rho \rangle$$

where:

---

*) These two relations $\underline{\in}$, $\overline{\in}$ can be assumed as a basis for "rough sets" theory, the problem however will be discussed elsewhere. In the appendix we give only axioms for such theory.

X - is the set of objects,

A - is the set of attributes,

$V = \bigcup_{a \in A} V_a$ - V_a is the set of values of a (or domain of a)

ϱ - is a knowledge function from $X \times A$ into V, such that $\varrho(x,a) \in V_a$ for every $x \in X$, $a \in A$.

We assume that each attribute has at least two values. Any pair (a,v), $a \in A$, $v \in V_a$ will be called descriptor of attribute a.

For each $x \in X$, ϱ_x is to mean a function from A into V such that $\varrho_x(a) = \varrho(x,a)$. This function will be referred to as knowledge about x in S.

Let $\tilde{a}$, $a \in A$ denote a binary relation on X defined as follows:

$$x \underset{a}{\sim} y \quad \text{iff} \quad \varrho(x,a) = \varrho(y,a).$$

One can easily check that $\tilde{a}$ is an equivalence relation.

Let $B \subseteq A$. By $\tilde{B}$ we shall mean a binary relation on X defined in the following way

$$\tilde{B} = \bigcap_{b \in B} \tilde{b}.$$

Obviously $\tilde{B}$ is also equivalence relation.

If $B = A$ we shall also write instead $\tilde{A}$, $\tilde{S}$, and the relation $\tilde{S}$ will be called classification generated by system S. Equivalence classes of the relation $\tilde{S}$ will be called elementary sets of the system S. If every elementary set of system S has exactly one element then system S will be called selective.

An empty set ϕ and every set being union of some elementary sets of S will be called composed set of S.

2. APPROXIMATIONS IN KRS

Let $S = \langle X,A,V,\varrho \rangle$ be a knowledge representation system and let $Y \subseteq X$. By $\overline{Y}$ we shall mean the smallest composed set of S containing Y, and by $\underline{Y}$ we denote the greatest composed set of S contained in Y. Set $\overline{Y}$ will be referred to as upper approximation of Y in S, and $\underline{Y}$ - as lower approximation of Y in S.

If system S is selective then obviously $\overline{Y} = \underline{Y}$ for any $Y \subseteq X$ in S.

Let $S = \langle X,A,V,\varrho \rangle$ be a knowledge representation system and let $C(X)$ be any classification of X i,e. $C(X) = \{X_1,X_2,\ldots,X_k\}$, $k > 1$ $\bigcup_{i=1}^{k} X_i = X$, and $X_i \cap X_j = \phi$ for $i \neq j$, $i,j = 1,\ldots,k$.

By upper approximation of $C(X)$ in S we shall mean the family $\overline{C}(X) = \{\overline{X}_1,\overline{X}_2,\ldots,\overline{X}_k\}$, and lower approximation of $C(X)$ in S is the family $\underline{C}(X) = \{\underline{X}_1,\underline{X}_2,\ldots,\underline{X}_k\}$. Of course if S is selective then $\overline{C}(X) = \underline{C}(X) = C(X)$. Otherwise $\overline{C}(X) \neq \underline{C}(X)$ and neither $\overline{C}(X)$ nor $\underline{C}(X)$ are classifications.

Our main task is connected with the problem how to express given classification $C(X)$, be means of classification $\tilde{S}$ or in other words - how to express classification $C(X)$ by means of attributes of the system S.

If $\overline{Y}$ is the upper approximation of Y in S then $\eta_{\overline{Y}}$ is to mean accuracy of this approximation which is defined as

$$\eta_{\overline{Y}} = \frac{\text{card}(Y)}{\text{card}(\overline{Y})},$$

and similarly

$$\underline{\eta}_Y = \frac{\text{card}(\underline{Y})}{\text{card}(Y)} .$$

Card(Y) is to mean the number of elements of the set Y.

Accuracy $\eta_{\overline{Y}}$ (or $\eta_{\underline{Y}}$) is the real number from the interval $\langle 0,1\rangle$ and is one for the selective system.

<u>Dispersion</u> of Y in S is the number

$$\delta_Y = \frac{\text{card}(\overline{Y}) - \text{card}(\underline{Y})}{\text{card}(Y)} .$$

For selective systems dispersion of each set $Y \subseteq X$ is always zero.

We extend the notion of accuracy and dispersion for classifications.

Let $S = \langle X,A,V,\varsigma\rangle$ be a knowledge representation system and $C(X)$ a classification. Upper and lower accuracy of the classification $C(X)$ in system S is defined as follows:

$$\eta_{\underline{C}(X)} = \frac{\sum_{i=1}^{k} \text{card}(\underline{X}_i)}{\text{card}(X)} ,$$

$$\eta_{\overline{C}(X)} = \frac{\text{card}(X)}{\sum_{i=1}^{k} \text{card}(\overline{X}_i)} .$$

Dispersion of the classification $C(X)$ in system S is

$$\delta_{C(X)} = \frac{\sum_{i=1}^{k} \text{card}(\overline{X}_i) - \sum_{i=1}^{k} \text{card}(\underline{X}_i)}{\text{card}(X)} .$$

3. REDUCTION OF ATTRIBUTES

Let $S = \langle X,A,V,\varsigma\rangle$ be a knowledge representation system and $C(X) = \{X_1,\dots,X_k\}$ classification on X. The smallest set $B \subseteq A$ will be called $\underline{C}(X)$ - reduct of A in S if lower approximation of every class X_i of $C(X)$ is union of some equivalence classes of the relation $\tilde{B}$. Similarly - $\overline{C}(X)$ - <u>reduct</u> of A in S is the smallest set $B \subseteq A$ such that upper approximation of every class X_i in $C(X)$ is union of some equivalence classes of the relation $\tilde{B}$.

Thus if B is $\underline{C}(X)$ - reduct (or $\overline{C}(X)$ - reduct) of A in S it means that we can obtain the same upper and lower approximations of the classification $C(X)$ using only the set B of attributes instead the original set A. So some attributes in the system are superflous from that point of view.

$\underline{C}(X)$ - reduct of A will be denoted by $A_{\underline{C}(X)}$, and similarly $\overline{C}(X)$ reduct of A is denoted by $A_{\overline{C}(X)}$.

4. DESCRIPTION LANGUAGE OF KRS

In order to describe knowledge about objects with each system S a description language L_S will be associated.

Expressions (terms) of the language L_S are built up from constants 0,1 descriptors (a,v), $a \in A$, $v \in V_a$ combined by symbols of boolean operations $+, \cdot, \sim$ in the usual way.

Terms of the language L_S are denoting subsets of objects of the system S. Constants 0,1 are denoting the empty set ϕ, and the whole set X of objects in the system S respectively. Descriptor (a,v) is to mean the set of all objects x in S such that $\varsigma_x(a) = v$. Boolean operarations $+, \cdot, \sim$ are interpreted as set theoretical operations union, intersection and complement respectively.

If t is a term in L_S, then the set of objects denoted by t will be written as $\sigma_S(t)$.

Term t will be called description of the set $\sigma_S(t)$.

If $Y \subseteq X$ and there exists a term t in L_S such that t is description of Y in S (i.e. $\sigma_S(t) = Y$) then Y will be called describable set in S. Two terms t, s in L_S are semantically equivalent in S if $\sigma_S(t) = \sigma_S(s)$.

Term t will be called elementary in L_S (or short elementary) if it is of the form $(a_1,v_{i_1})\cdot(a_2,v_{i_2})\cdot\ldots\cdot(a_n,v_{i_n})$ where $A = \{a_1,\ldots,a_n\}$, $v_{i_j} \in V_{a_j}$.

If t is an elementary term in L_S then $\sigma_S(t)$ is an elementary set in S. So elementary terms in L_S are "names" of elementary sets in S.

Term t is in normal form if $t = t_1 + t_2 + \ldots + t_k$, where t_i are elementary terms. For every term t in L_S there exists term s in L_S in normal form semantically equivalent to t. So describable set in S are union of same elementary sets in S. So the notion of an elementary set and the notion of composed set are exactly the same.

To transform terms of the language L_S preserving its semantics we can use axioms of boolean algebra and the following specific axioms:

A1. $\sim(a,v) = \sum_{\substack{w \neq v \\ w \in V_a}} (a,w)$,

A2. $\sum_{v \in V_a} (a,v) = 1$

A3. $(a,v) \cap (a,w) = 0$ for $v,w \in V_a$ and $v \neq w$.

Terms t,s are syntactically equivalent in S if one of them can be obtained them another one by means of axioms of boolean algebra and specific axioms of the system. (Rules of transformation).

Terms t, s are semantically equivalent in S if and only if they are syntactically equivalent in S.

This property is known as the completeness properly of the language.

5. MAIN PROBLEMS

We are interested in this paper with the following problems:

(i) Characteristic description. Given knowledge representation system $S = \langle X,A,V,\rho \rangle$ and classification $C(X) = \{X_1,\ldots,X_k\}$. Find description in L_S of each class X_i of classification $C(X)$. Because in general case the system S is not selective then classes of the classification are not describable sets in S. So we are unable to give description of them in L_S. We can have only descriptions of lower and upper approximations of each class X_i of $C(X)$.

More exactly, if $C(X) = \{X_1,\ldots,X_k\}$, then the family of terms $\{\underline{t}_1,\ldots,\underline{t}_k\}$ will be called lower description of $C(X)$ if $\sigma_S(\underline{t}_i) = \underline{X}_i$ for $i = 1,\ldots,k$, and similarly the family $\{\bar{t}_1,\ldots,\bar{t}_k\}$ such that $\sigma_S(\bar{t}_i) = \bar{X}_i$ for $i = 1,\ldots,k$, is called upper description of $C(X)$.

If terms $\underline{t}_i$ ($\bar{t}_i$) are built up from the set of attributes $\underline{A}_{C(X)}$ ($\bar{A}_{C(X)}$) then we shall call corresponding families of terms reduced lower description of $C(X)$ or reduced upper description of $C(X)$ respectively.

The problem of finding lower and upper descriptions of $C(X)$ is rather simple and the corresponding algorithms will be given in the next section. The algorithm for computing reducts of the set of attributes is somewhat more difficult and will be not discussed here.

(ii) <u>Classification</u>. The classification problem is formulated as follows:

Given system $S = \langle X,A,V,\varrho \rangle$, classification $C(X) = \{X_1,\dots,X_k\}$, lower and upper description of $C(X)$ and an elementary term t in L_S. Find term $\underline{t}_i$ such that $\sigma_S(t) \subsetneq \sigma_S(\underline{t}_i)$; If such term does not exist, - find all terms $\overline{t}_j$ such that $\sigma_S(t) \subseteq \sigma_S(\overline{t}_j)$.

The classification algorithm is very simple and it is based on the fact that in order to check whether $\sigma_S(t)$ (t is elementary) is included in $\sigma_S(s)$ or not, it is enough to translate s to normal form and check whether s contains t as an elementary term or not.

The classification algorithm says how to classify objects, given by description (corresponding elementary term) - according to assumed classification and knowledge representation. If the system S is selective we can classify object exactly, i.e. to each object we can find exactly one class to which it belongs. If the system S is not selective exact classification of objects is impossible and we can find only approximate classification, i.e. we are able to find only lower or upper approximations of class to which considered objects belongs.

We can reformulate the situation also in different manner. If we know that $x \in \underline{X}$ we shall say "x belongs <u>surely</u> to X" ($x \underline{\in} X$); if we know that $x \in \overline{X}$, we shall say "x belongs <u>possibly</u> to X" ($x \overline{\in} X$). So if the system S is not selective

first we ask whether given object surely belongs to some class. If the answer is positive the classification process is finished; if not - we ask about classes to which the object belongs possibly, obtaining classes to which considered object may belong.

(iii) <u>Characteristic sets (Samples).</u> In many areas of artificial intelligence e.g. learning, inductive intervence, automatic hypotheses generation etc., we want to infer some general properties of objects from finite sample (finite number of examples). The question how to find a sample of a given set of objects is of main importance for this kind of problems. Solution of this problem in our model is very simple.

Let t_Y, t_Z denote description of describable sets Y,Z in system S. If $Y \subseteq Z$ and $t_Y = t_Z$, then Y will be called <u>sample</u> (or <u>characteristic set</u>) of Z in S. If Y is minimal set such that Y is sample of Z in S then Y will be called <u>proper sample</u> (or <u>proper characteristic set</u>) of Z in S. Thus if $Z = \{Z_1,\dots,Z_k\}$, is a describable set in S and $Z_1,\dots,Z_k$ are elementary sets, then any set consisting of some elements of all elementary sets $Z_1,\dots,Z_k$ is sample of Z. If we take one element from each elementary sets $Z_1,\dots,Z_k$ we obtain proper sample of Z. Thus if we have system $S = \langle X,A,V,\varrho \rangle$ and classification $C(X) = \{X_1,\dots,X_n\}$ we can find proper samples of each describable class X_i as shown before. If class X_i is not describable in S we can find proper samples only of it upper or lower approximations (which are describable) in the same manner as before. Conversely, if we are given set Y supposed to be a sample of some set Z in system S, we have to be sure that Y contains representatives of all elementary sets occuring in Z. Otherwise we do not obtain description of set Z on the basis of set Y.

6. THE ALGORITHM

As mentioned before main problem connected with classification in the proposed model is to find upper and lower description of each class of the classification. We shall outline in this paragraph an algorithm which gives upper and lower description of each class of the classification.

We assume that elementary sets of system S are represented in computer memory as records structured as shown in fig. 1.

OC	CC	...	OC	CC	ET	CC	SC	...	CC	SC

$\underbrace{\text{OC CC ... OC CC}}_{\text{objects}}$ $\underbrace{\text{ET CC SC ... CC SC}}_{\text{classes}}$

OC - object code

CC - class code

ET - elementary term

SC - sort of class

Fig. 1.

Each object of a given elementary set is represented in the record by its code (object code - OC) and class code (CC) i.e. code of the class to which this object belongs. Next we have in the record an elementary term describing considered elementary set, and then each class of the classification is represented in the record by its code and sort of the class. Sort of the class may have values 0,1,2. Sort 0 of the class i is to mean that none of the object in the record is in the class i; 1 - is to denote that all object in the record belong to class i; 2 - is to mean that there are some object in the

record belonging to class i.

Because sort of each class is computed from class codes of each object in the record - so they are superflous in the record. However for simplification of the algorithm it is worthwhile to have this information directly in the record. So by means of list of such records we can represent each system $S = \langle X, A, V, \varrho \rangle$ and classification $C(X)$ in computer memory.

In order to find lower description of a class we have to read out elementary terms from all records having in i-th class sort number 1; upper approximation of i-th class is the sum of all elementary terms occuring in records having in i-th class sort 1 or 2 ($\neq 0$).

Thus one run through the list of records give upper and lower description of each class.

We can also compute easily in the same run accuracy and dispersion of the classification.

7. APPLICATION TO MEDICAL DIAGNOSIS

Consider a medical data base concerning some patients. Each single patient in the data base is described diagnostically, pathogenetically, prognostically and therapeutically. Usually this description consist of sentences in ordinary English, however it can be easily replaced by „attribute - value" type description. Thus patient description can be treated as an elementary term in some description language.

Assume we are interested in classifying patients in two classes only, for example having heart disease or not. This information is given in the position CC in the record: 0 - for "not" and 1 - for "yes".

Organizing the data base as a file of records mentioned in the previous section we can easily find characteristic description (lower and upper) of ill patients, and solve the classification problem, which can be formulated as follows: given a data base as mentioned before, concerning some patient investigated for heart disease. So beside the description of the patient in the data base we have the information whether each patient is ill or not.

Now we can ask whether description of the class of ill patient is characteristic for the considered disease or not; possibly with some approximation. If so every new case can be decided on the basis of its description, i.e. having the description of a new patient we have to check only whether this description fits to the class of ill patients (or its lower or upper description).

In the case of approximate classification we can compute the accuracy of lower and upper approximation. If the accuracy is not good enough we can add new examples to the data base and compute whether they improve the accuracy essentially or not. If it is the case we can use now the extended data base as a sample of ill patients - if not, we have to search for new examples improving the accuracy of the system. In this way we obtain a learning algorithm which gives better decisions with increasing number of examples accumulated in the data base.

To this end let us remark that the updating algorithm is very simple. Adding new example to the data base we have to check first whether such example already exists in the data base. If not, we have to add new record to the data base according to the rule given before. If such a case already exist in

the data base we have then to add to the corresponding record, the new case and update "sort of the class" code of this record according to the rule:

C C	old SC	new SC
0	0	0
0	1	2
0	2	2
1	0	2
1	1	1
1	2	2

where

CC = 0 means healthy

CC = 1 heart disease

APPENDIX. ROUGH SETS

In order to deal with situations in which the membership function is not defined univocal we propose here two membership functions $\underline{\in}$; (surely belongs), and $\overline{\in}$ (possibly belongs). This can be consider as an alternative approach to fuzzy set theory introduced by Zadeh.

Let U be a fixed set. Subset of set U are dented be X, Y, Z etc., ϕ is to mean an empty set.

With each set X we associate its upper approximation $\overline{X}$ and lower approximation $\underline{X}$. Then both membership functions $\overline{\in}$, $\underline{\in}$ are defined as

$$x \overline{\in} X \quad \text{iff} \quad x \in \overline{X},$$

$$x \underline{\in} X \quad \text{iff} \quad x \in \underline{X}.$$

We assume the following axioms for approximations:

1. $\overline{X} \supseteq X \supseteq \underline{X}$
2. $\overline{U} = \underline{U} = U$
3. $\underline{\phi} = \overline{\phi} = \phi$
4. $\overline{X \cup Y} = \overline{X} \cup \overline{Y}$
5. $\underline{X \cap Y} = \underline{X} \cap \underline{Y}$
6. $\overline{\overline{X}} = \overline{X}$
7. $\underline{\underline{X}} = \underline{X}$
8. $\overline{X} = -(\underline{-X})$
9. $\underline{X} = -(\overline{-X})$

It is easy to see that an upper approximation of set X satisfy axioms of topological closure, and axioms of lower approximation of X are due to axioms of interior operation. Thus in order to deal with approximate classifications we can use standard topological methods.

References

1. W. Marek, Z. Pawlak, Information Storage and Retrieval Systems: Mathematical Foundation. Theoretical Computer Sciences 1 (1976), 331-354.
2. R. Michalski, Variable-valued logic and its application to pattern recognition and machine learning, (1974).
3. Z. Pawlak, Information System Theory, Mathematical Foundation, Information Systems (to appear).