Classification of objects by means of attributes

Zdzisław Pawlak
Zdzisław Pawlak

CLASSIFICATION OF OBJECTS BY MEANS OF ATTRIBUTES

An approach to inductive inference

429

Warsaw, January 1981
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. Poddane tu próbie nowego spojrzenia na problemy klasyfikacji w oparciu o teorie systemów informacyjnych, oraz zastosowanie jej do automatyzacji rozumowania indukcyjnego.

Próba ta prowadzi między innymi do nowego sformułowania pojęcia zbioru rozmytego (zwane tu zbiorem przybliżonym). Podano aksjomaty takich zbiorów, które są identyczne z aksjomatami topologicznego domknięcia i wnętrza.
Many problems of artificial intelligence are based on classification of objects. We propose here a new approach to classification inspired by remote or distant fields of mathematics 


1. KNOWLEDGE REPRESENTATION SYSTEM

By a knowledge representation system (KRS) we shall mean:

$$s = <x_i, y_i, z_i>$$

where:

- $x_i$: set of relations on $x$,
- $y_i$: set of relations on $y$,
- $z_i$: set of relations on $z$.

In the appendix we give only a sketch for such a theory.
2. **APPROXIMATIONS IN KRS**

Let \( S = \langle X, A, V, \mathcal{G} \rangle \) 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 the **upper approximation** of \( Y \) in \( S \), and \( \underline{Y} \) as the **lower approximation** of \( Y \) in \( S \).

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

Let \( S = \langle X, A, V, \mathcal{G} \rangle \) be a knowledge representation system and let \( C(X) \) be any classification of \( X \), i.e.,

\[
C(X) = \left\{ X_1, X_2, \ldots, X_k \right\}, \quad k > 1
\]

\[
X_1 \cap X_j = \emptyset \quad \text{for} \quad i \neq j, \quad i, j = 1, \ldots, k.
\]

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 \( \overline{C(X)} = C(X) = \underline{C(X)} \). Otherwise, \( \overline{C(X)} \neq C(X) \) 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 \( \mathcal{G} \) 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

\[
\eta_{\overline{Y}} \text{ is to mean accuracy of this approximation which is defined as}
\]

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

and similarly.
3. REDUCTION OF ATTRIBUTES

Let $S = \langle X, A, V, \mathcal{S} \rangle$ be a knowledge representation system and $O(X) = \{X_1, \ldots, X_k\}$ classification on $X$. The smallest set $B \subseteq A$ will be called $\bar{O}(X)$ - reduct of $A$ in $B$ if lower approximation of every class $X_i$ of $\bar{O}(X)$ in $O(X)$ is union of some equivalence classes of the relation $\mathcal{S}$. Similarly - $\bar{O}(X)$ - reduct of $A$ in $S$ is the smallest set $B \subseteq A$ such that upper approximation of every class $X_i$ in $O(X)$ is union of some equivalence classes of the relation $\mathcal{S}$.

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

$O(X)$ - reduct of $A$ will be denoted by $A_O(X)$, and similarly $\bar{O}(X)$ reduct of $A$ is denoted by $A_{\bar{O}}(X)$.

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 $\{a, v\}$, $a \in A$, $v \in V$, combined by symbols of boolean operations $+, \cdot$, $\neg$ 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 $\phi$, 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 $\alpha_x(a) = v$. Boolean operations $+, \cdot$, $\neg$ are interpreted as set theoretical operations union, intersection and complement respectively.
If \( t \) is a term in \( L_0 \), then the set of objects denoted by \( t \) will be written as \( \mathcal{G}_t(x) \).

Term \( t \) will be called a **description** of the set \( \mathcal{G}_t(x) \).

If \( Y \subseteq X \) and there exists a term \( t \) in \( L_0 \) such that \( \mathcal{G}_t(x) = Y \) then \( Y \) will be called a **describable set** in \( S \). Two terms \( t, s \) in \( L_0 \) are semantically equivalent in \( S \) if \( \mathcal{G}_t(x) = \mathcal{G}_s(x) \).

Term \( t \) will be called **elementary** in \( L_0 \) (or short elementary) if it is of the form \( (a_1 \cdot y_1) \cdot (a_2 \cdot y_2) \cdot \ldots \cdot (a_n \cdot y_n) \) where \( A = \langle a_1, \ldots, a_n \rangle \), \( Y_1, \ldots, Y_n \).

If \( t \) is an elementary term in \( L_0 \) then \( \mathcal{G}_t(x) \) is an elementary set in \( S \). So elementary terms in \( L_0 \) 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 \( L_0 \) there exists term \( s \) in \( L_0 \) in normal form semantically equivalent to \( t \). So describable sets 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 \( L_0 \) preserving its semantics we can use axioms of boolean algebra and the following specific axioms:

A1. \( \neg(s, y) = \sum_{x \in V_a} (s, w) \)

A2. \( \sum_{x \in V_a} (s, y) = 1 \)

A3. \( (s, y) \wedge (s, w) = 0 \) for \( y, w \in V_a \) and \( y \neq w \).

Terms \( t, s \) are syntactically equivalent in \( S \) if one of them can be obtained from 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 property of the language.

5. MAIN PROBLEMS

We are interested in this paper with the following problems:

(i) **Characteristic description.** Given knowledge representation system \( S = \langle X, A_i, Y_i, \mathcal{G}_i \rangle \) and classification \( C(X) = \langle X_1, \ldots, X_k \rangle \). Find description in \( L_0 \) 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 \( L_0 \). We can have only descriptions of lower and upper approximations of each class \( X_i \) of \( C(X) \).

More exactly, if \( C(X) = \langle X_1, \ldots, X_k \rangle \), then the family of terms \( \langle \mathcal{G}_t(x) = X_i \rangle \) will be called **lower description** of \( C(X) \) if \( \mathcal{G}_t(x_i) = X_i \) for \( i = 1, \ldots, k \), and similarly the family \( \langle \mathcal{G}_t(x_i) = \overline{X}_i \rangle \) such that \( \mathcal{G}_t(x_i) = \overline{X}_i \) for \( i = 1, \ldots, k \), is called **upper description** of \( C(X) \).

If terms \( t_i \) are built up from the set of attributes \( A_0(X) (A_C(X)) \) then we shall call corresponding families of terms **reduced lower description** of \( C(X) \) or **reduced upper description** of \( C(X) \) respectively.
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) Characteristic sets (Samples). In many areas of artificial intelligence e.g. learning, inductive inference, 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 \( Y \), \( t_y \) denote description of describable sets \( Y \), \( Z \) in system \( S \). If \( Y \subseteq Z \) and \( t_y = t_z \), then \( Y \) will be called proper sample (or characteristic set) of \( Z \) in \( S \). If \( Y \) is minimal set such that \( Y \) is sample of \( Z \) in \( S \) then \( Y \) will be called proper sample (or proper characteristic set) of \( Z \) in \( S \).

Thus if \( Z = \{ z_1, \ldots, z_k \} \) is a describable set in \( S \) and \( z_1, \ldots, z_k \) are elementary sets, then any set consisting of some elements of all elementary sets \( z_1, \ldots, z_k \) is sample of \( Z \). If we take one element from each elementary sets \( z_1, \ldots, z_k \) we obtain proper sample of \( Z \). Thus if we have system \( S = \langle X, A, V, G \rangle \) and classification \( C(X) = \{ x_1, \ldots, x_l \} \) 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 occurring 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 | OC | SC | ... | CC | SC
```

objects  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 superfluous 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 \( S = \langle X, A, V, S \rangle \) 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 occurring in records having in \( i \)-th class sort 1 or 2 ( \( \neq 0 \)).

Thus one run through the list of records gives 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:

\[
\begin{array}{ccc}
C & C & \text{old SC} & \text{new SC} \\
0 & 0 & 0 & 0 \\
0 & 1 & 2 & 2 \\
0 & 2 & 2 & 2 \\
1 & 0 & 2 & 2 \\
1 & 1 & 1 & 1 \\
1 & 2 & 2 & 2 \\
\end{array}
\]

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 univocally we propose here two membership functions $\bar{C}$ (surely belongs), and $\tilde{C}$ (possibly belongs). This can be considered as an alternative approach to fuzzy set theory introduced by Zadeh.

Let $U$ be a fixed set. Subset of set $U$ are denoted by $X, Y, Z$ etc., $\emptyset$ 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 $\bar{C}, \tilde{C}$ are defined as

\[
\forall x \in \overline{X} \iff x \in X,
\]
\[
\forall x \in \underline{X} \iff x \in \overline{X}.
\]

We assume the following axioms for approximations:

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

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

