ALGORITHMS OF SOLVING PATTERN RECOGNITION TASKS WITH UNCERTAIN DATA

Ryabtsev A.

Belarussian State University, Skoryny prospekt, 4, Minsk, Belarus.

Abstract. The extensions of inductive resolution algorithms set are proposed. The extensions allow operating with uncertain (or missing) values of attributes. It is established that algorithms are monotonnous relatively to the volume of known data.

Introduction

The data for pattern recognitions tasks are often characterized by uncertainty. That is why the most adequate way to operate with uncertain data is to use technique of fuzzy mathematics. The key issue of such technique is a method for calculating the set of membership degree. This paper proposes modifications of inductive resolution algorithms set [1, 2]. These modifications allow operating with unknown values of attributes in the pattern recognition tasks.

The base algorithm

Assume that every object is described by the set of attributes J; the members of finite set D_j are possible values of attribute $j \in J$. Without loss of generality it can be assumed that items of $D_j, j \in J$ are the integer numbers from 0 to $|D_j| - 1$.

Suppose the next conditions are hold.

The set X^0 is divided into subsets X_i^0 corresponding to the classes X_i.

The matrix $\| a_{ikj} \|$, $i \in \{1,\ldots,l\}, j \in J$, $k \in \left\{0,\ldots,\max\limits_{j \in J}|D_j| - 1\right\}, a_{ij} \in R$ is assigned to the set $\{1,\ldots,l\} \times J \times \left\{0,\ldots,\max\limits_{j \in J}|D_j| - 1\right\}$. Further assume that:

$$\forall i,j,k \;\; (a_{ikj} \geq 0),\; \forall i \;\; (\sum_{J} a_{ikj} > 0) \tag{1}$$

Now the set $\alpha(a)$ of algorithms can be defined as the following sequence:

For every object $x \in X$ to be recognized the following steps are performed:

1. For each $i = \overline{1,l}$ и $x^u \in X_i^0$ the similarity function is calculated:

$$\mu^{i,u}(x, x^u) = \max\left\{0, \left(\sum_{j \in J} (-1)^{\lambda_j} a_{ix_j j}\right) \cdot \left(\sum_{j \in J} a_{ix_j^u j}\right)^{-1}\right\},$$

where $\lambda_j = \begin{cases} 1, x_j \neq x_j^u \\ 0, \textit{in other case} \end{cases}$

2. For each $i=\overline{1,l}$ do: $P_i^A(x)=\max_{x^u\in X_i^0}\{\mu^{i,u}(x,x^u)\}$

Fig. 1 Decision-making procedure for data with nominal attributes.

Let us describe one of possible schemes that determine parameters $a=\|a_{ikj}\|$.

Assume that J and X_i^0 are fixed. The following sequence of steps is performed:

1. For each $j\in J$ and $i\in\{1,\dots,l\}$ do

$$b_{ix_j j}=\left(\sum_{\substack{x^u\in X_i^0,\\ x_j^u=x_j}}|x_j^u|\right)\cdot(|X_i^0|)^{-1}.$$

2. $b_{kj}=(\sum_{i=1}^{l}b_{ikj})\cdot l^{-1}$ $a_{ikj}=|\,b_{ikj}-b_{kj}\,|$, $i=\overline{1,l}$, $j\in J$, $k\in\left\{1,\dots,\max_{j=\overline{1,n}}|D_j|\right\}$.

Fig. 2 Procedure of determining parameters for data with nominal attributes.

The key problem faced during investigating of learning algorithms is whether the algorithm can operate with partially defined data. It involves both the possibilities to recognize partial defined objects and to use partial descriptions of objects during learning as well. Let us consider the behavior of set $\alpha(a)$ of algorithms for every mentioned case.

Recognizing partially defined data

Consider the object to be recognized that contains some attributes with missing values. The values of such attributes are called uncertain (or unknown). The unknown values is marked with '?'. Let us define the similarity function in the following way:

$$\mu^{i,u}(x,x^u)=\max\left\{0,\left(\sum_{j\in J}(-1)^{\lambda_j}\delta_{ix_j j}\right)\cdot\left(\sum_{j\in J}a_{ix_j^u j}\right)^{-1}\right\} \tag{2}$$

$$\text{where } \delta_{ix_j j}=\begin{cases}a_{ix_j j},x_j\neq\text{'?'},\\ \max_{k\in D_j}(a_{ikj}),x_j=\text{'?'}\end{cases} \tag{3}$$

$$\lambda_j=\begin{cases}1,x_j\neq x_j^u\ \vee x_j=\text{'?'},\\ 0,\textit{in other case}\end{cases} \tag{4}$$

This particular form for function δ was chosen as the most appropriate due to the fact that the unknown (or missed) value of attribute j is assumed to be maximally different from any other possible value of attribute j.

Proposition 1 *Suppose* $A\in\alpha(a)$ *uses the similarity function* $\mu^{i,u}(x,x^u)$ *calculated by formulas (2)-(4). Assume that some object* $x\in X$ *contains unknown values for any*

$j \in \widetilde{J} \subset J$. *Let* $\widetilde{X}$ *be a set of objects such that for any object* $\widetilde{x} \in \widetilde{X}$ *and for all* $j \in J \setminus \widetilde{J}$ *the next condition is hold:* $\widetilde{x}_j = x_j$. *Then for any* $x' \in \widetilde{X}$ *and any object* x^u *of training set that belongs to the arbitrary class* X_i^o *the next condition is hold:*

$$\mu^{i,u}(x, x^u) \leq \mu^{i,u}(x', x^u) \quad (5)$$

Proof. Let A be defined and some procedure for calculating of parameters $\|a_{ikj}\|$ be fixed (it is not necessary that this procedure coincides with the procedure defined at Fig.2). Then,

$$\left(\sum_{j\in J} a_{ix_j^u j}\right) \cdot \mu^{i,u}(x, x^u) = \sum_{j\in J} (-1)^{\lambda_j} \delta_{ix_j j} = \sum_{j\in J\setminus\widetilde{J}} (-1)^{\lambda_j} \delta_{x_j j} + \sum_{j\in\widetilde{J}} (-1)^{\lambda_j} \delta_{ix_j j} = \sum_{j\in J\setminus\widetilde{J}} (-1)^{\lambda_j} a_{ix_j j} -$$

$$- \sum_{j\in\widetilde{J}} \max(a_{ikj});$$

$$\left(\sum_{j\in J} a_{ix_j^u j}\right) \cdot \mu^{i,u}(x', x^u) = \sum_{j\in J} (-1)^{\lambda_j} \delta_{ix'_j j} = \sum_{j\in J\setminus\widetilde{J}} (-1)^{\lambda_j} \delta_{ix'_j j} + \sum_{j\in\widetilde{J}} (-1)^{\lambda_j} \delta_{ix'_j j} = \sum_{j\in J\setminus\widetilde{J}} (-1)^{\lambda_j} a_{ix'_j j} +$$

$$+ \sum_{j\in\widetilde{J}} (-1)^{\lambda_j} a_{ix'_j j} = \sum_{j\in J\setminus\widetilde{J}} (-1)^{\lambda_j} a_{ix_j j} + \sum_{j\in\widetilde{J}} (-1)^{\lambda_j} a_{ix'_j j} \geq \sum_{j\in J\setminus\widetilde{J}} (-1)^{\lambda_j} a_{ix_j j} - \sum_{j\in\widetilde{J}} \max(a_{ikj}) =$$

$$= \left(\sum_{j\in J} a_{ix_j^u j}\right) \cdot \mu^{i,u}(x, x^u).$$

By construction $\left(\sum_{j\in J} a_{ix_j^u j}\right) > 0$; dividing both sides by $\left(\sum_{j\in J} a_{ix_j^u j}\right)$, we get (5).

Learning partially defined data

It is a common case that a training set has objects with unknown values of some attributes. In many cases such objects are of key role in description of classes, and skipping these objects during learning process can cause falsification of algorithm result obtained during recognition of new objects. Let us consider modification of inductive resolution algorithms set for the case when the training set contains objects with some unknown (missing) values of attributes.

The array $\|a_{ikj}\|$ of dimension $l \cdot \max_{j\in J}(|D_j|) \cdot n$ is replaced by the array of dimension $l \cdot \max_{j\in J}(|D_j| + 1) \cdot n$. It allows calculating parameters $\|a_{ikj}\|$ for unknown (missing) attribute values of training set objects. Without loss of generality it can be assumed that in the array $\|a_{ikj}\|$ the layer with coordinates $\left(i; \max_{j\in J}(|D_j| + 1); j\right)$ corresponds to the unknown values.

Consider the modification of similarity function for algorithms $A \in \alpha(a)$:

$$\mu^{i,u}(x, x^u) = \max\left\{0, \left(\sum_{j\in J} (-1)^{\lambda_j} \delta_{ix_j j}\right) \cdot \left(\sum_{j\in J} a_{ix_j^u j}\right)^{-1}\right\} \quad (6)$$

$$\text{where } \delta_{ix_j j} = \begin{cases} a_{ix_j j}, x_j \neq '?' \& x_j^u \neq '?', \\ a_{i(|D_j|+1)j}, x_j = '?' \& x_j^u = '?', \\ \max\limits_{k \in D_j}(a_{ikj}), \textit{in other case.} \end{cases} \tag{7}$$

$$\lambda_j = \begin{cases} 1, x_j \neq x_j^u \ \& \ x_j^u \neq '?', \\ 0, \textit{in other case} \end{cases} \tag{8}$$

Proposition 2 *Suppose* $\mathrm{A} \in \alpha(a)$ *uses the similarity function* $\mu^{i,u}(x, x^u)$ *calculated by formulas (6)-(8). Assume that some object* $x \in X$ *contains unknown values for any* $j \in \widetilde{J} \subset J$ *. Let* $\widetilde{X}$ *be a set of objects such that for any object* $\widetilde{x} \in \widetilde{X}$ *and for all* $j \in J \setminus \widetilde{J}$ *the next condition is hold:* $\widetilde{x}_j = x_j$*. Then for any* $x' \in \widetilde{X}$ *and any object* x^u *of training set that belongs to the arbitrary class* X_i^o *the next condition is hold:*

$$\mu^{i,u}(x, x^u) \leq \mu^{i,u}(x', x^u) \tag{9}$$

Proof. Fix some object x^u that belonging to the training set. Assume that every attribute $j \in J_u \subseteq J$ of the object x^u contains the unknown value.

$$\left(\sum_{j \in J} a_{ix_j^u j}\right) \cdot \mu^{i,u}(x, x_u) = \sum_{j \in J} (-1)^{\lambda_j} \delta_{ix_j j} = \sum_{j \in \widetilde{J} \cap J_u} (-1)^{\lambda_j} \delta_{x_j j} + \sum_{j \in \widetilde{J} \setminus (\widetilde{J} \cap J_u)} (-1)^{\lambda_j} \delta_{ix'_j j} +$$
$$+ \sum_{j \in (J \setminus \widetilde{J}) \cap J_u} (-1)^{\lambda_j} \delta_{ix'_j j} + \sum_{j \in (J \setminus \widetilde{J}) \setminus ((J \setminus \widetilde{J}) \cap J_u)} (-1)^{\lambda_j} \delta_{ix'_j j} = \sum_{j \in \widetilde{J} \cap J_u} a_{i(|D_j|+1)j} - \sum_{j \in \widetilde{J} \setminus (\widetilde{J} \cap J_u)} \max(a_{ikj}) +$$
$$+ \sum_{j \in (J \setminus \widetilde{J}) \cap J_u} \max(a_{ikj}) + \sum_{j \in (J \setminus \widetilde{J}) \cap ((J \setminus \widetilde{J}) \cap J_u)} (-1)^{\lambda_j} a_{ix_j j};$$

$$\left(\sum_{j \in J} a_{ix_j^u j}\right) \cdot \mu^{i,u}(x', x_u) = \sum_{j \in J} (-1)^{\lambda_j} \delta_{ix'_j j} = \sum_{j \in \widetilde{J} \cap J_u} (-1)^{\lambda_j} \delta_{x'_j j} + \sum_{j \in \widetilde{J} \setminus (\widetilde{J} \cap J_u)} (-1)^{\lambda_j} \delta_{ix'_j j} +$$
$$+ \sum_{j \in (J \setminus \widetilde{J}) \cap J_u} (-1)^{\lambda_j} \delta_{ix'_j j} + \sum_{j \in (J \setminus \widetilde{J}) \setminus ((J \setminus \widetilde{J}) \cap J_u)} (-1)^{\lambda_j} \delta_{ix'_j j} \overset{\text{by definition}}{=} \sum_{j \in \widetilde{J} \cap J_u} (-1)^{\lambda_j} \delta_{x'_j j} +$$
$$+ \sum_{j \in \widetilde{J} \setminus (\widetilde{J} \cap J_u)} (-1)^{\lambda_j} \delta_{ix'_j j} + \sum_{j \in (J \setminus \widetilde{J}) \cap J_u} (-1)^{\lambda_j} \delta_{ix_j j} + \sum_{j \in (J \setminus \widetilde{J}) \setminus ((J \setminus \widetilde{J}) \cap J_u)} (-1)^{\lambda_j} \delta_{ix_j j} = \sum_{j \in \widetilde{J} \cap J_u} (-1)^{\lambda_j} \delta_{x'_j j} +$$
$$+ \sum_{j \in \widetilde{J} \setminus (\widetilde{J} \cap J_u)} (-1)^{\lambda_j} \delta_{ix'_j j} + \sum_{j \in (J \setminus \widetilde{J}) \cap J_u} \max(a_{ikj}) + \sum_{j \in (J \setminus \widetilde{J}) \setminus ((J \setminus \widetilde{J}) \cap J_u)} (-1)^{\lambda_j} a_{ix_j j} \tag{10}$$

Consider expression $\sum_{j \in \widetilde{J} \cap J_u} (-1)^{\lambda_j} \delta_{ix'_j j}$. If $x'_j = ?$ then $\delta_{ix'_j j} = a_{i(|D_j|+1)j}$; in the other case, $\delta_{ix'_j j} = \max(a_{ikj})$. For any x'_j by definition of the algorithm: $\lambda_j = 0$. Taking into account that $\max(a_{ikj}) \geq a_{i(|D_j|+1)j}$, we obtain the next inequality:

$$\sum_{j \in \widetilde{J} \cap J_u} (-1)^{\lambda_j} \delta_{ix'_j j} \geq \sum_{j \in \widetilde{J} \cap J_u} (-1)^{\lambda_j} \delta_{ix_j j} \tag{11}$$

By definition of the algorithm (6)-(8)

$$\sum_{j \in \widetilde{J} \setminus (\widetilde{J} \cap J_u)} (-1)^{\lambda_j} \delta_{ix'_j j} = -\sum_{j \in \widetilde{J} \setminus (\widetilde{J} \cap J_u)} \max(a_{ikj}) = \sum_{j \in \widetilde{J} \setminus (\widetilde{J} \cap J_u)} (-1)^{\lambda_j} \delta_{ix_j j} \tag{12}$$

Combining (11), (12) and (10), we obtain

$$\left(\sum_{j\in J} a_{ix^u_j j}\right)\cdot \mu^{i,u}\left(x',x^u\right)=\sum_{j\in \tilde{J}\cap J_u}(-1)^{\lambda_j}\delta_{ix'_j j}+\sum_{j\in \tilde{J}\setminus(\tilde{J}\cap J_u)}(-1)^{\lambda_j}\delta_{ix'_j j}+\sum_{j\in (J\setminus\tilde{J})\cap J_u}\max\left(a_{ikj}\right)+$$

$$+\sum_{j\in (J\setminus\tilde{J})\setminus((J\setminus\tilde{J})\cap J_u)}(-1)^{\lambda_j}a_{ix_j j}=\sum_{j\in \tilde{J}\cap J_u}(-1)^{\lambda_j}\delta_{ix'_j j}-\sum_{j\in \tilde{J}\setminus(\tilde{J}\cap J_u)}\max\left(a_{ikj}\right)+\sum_{j\in (J\setminus\tilde{J})\cap J_u}\max\left(a_{ikj}\right)+$$

$$+\sum_{j\in (J\setminus\tilde{J})\setminus((J\setminus\tilde{J})\cap J_u)}(-1)^{\lambda_j}a_{ix_j j}\geq\sum_{j\in \tilde{J}\cap J_u}(-1)^{\lambda_j}\delta_{x_j j}-\sum_{j\in \tilde{J}\setminus(\tilde{J}\cap J_u)}\max\left(a_{ikj}\right)+\sum_{j\in (J\setminus\tilde{J})\cap J_u}\max\left(a_{ikj}\right)+$$

$$+\sum_{j\in (J\setminus\tilde{J})\setminus((J\setminus\tilde{J})\cap J_u)}(-1)^{\lambda_j}a_{ix_j j}=\left(\sum_{j\in J} a_{ix^u_j j}\right)\cdot \mu^{i,u}\left(x,x^u\right)$$

By construction $\left(\sum_{j\in J} a_{ix^u_j j}\right)>0$; dividing both sides by $\left(\sum_{j\in J} a_{ix^u_j j}\right)>0$, we get (9).

Conclusion

The paper proposes modifications of the pattern recognition algorithm. These modifications allow operating with unknown values of attributes. It is proved that modifications are monotonous relatively to the volume of known data. It means that as the volume of known information about an object is less then any estimation of membership for the object is also less.

References

1. Bergmans J., Krasnoproshin V., Obraztsov V., Vissia H. //Proceedings of 4th International Conference PRIP97, Minsk, 1997, Vol.1-pp. 305-312.
2. Ryabtsev A. //Proceedings of 4th International Conference PRIP97, Minsk, 1997, Vol.2-pp. 227-231.