Granular Computing and Incremental Classification

Xenia Naidenova[1] and Vladimir Parkhomenko[2]

[1] Military Medical Academy, Saint-Petersburg, Russia
ksennaidd@gmail.com,
ORCID 0000-0003-2377-7093

[2] Peter the Great St. Petersburg Polytechnic University, Saint-Petersburg, Russia
parhomenko.v@gmail.com,
ORCID 0000-0001-7757-377X

Abstract. A problem of incremental granular computing is considered in the tasks of classification reasoning. Objects, values of attributes, partitions of objects (classifications) and Good Maximally Redundant Tests (GMRTs) (special kind of formal concepts) are considered as granules. The paper deals with inferring GMRTs. They are good tests because they cover the largest possible number of objects w. r. t. inclusion relation on the set of all object subsets. Two kinds of classification subcontexts are defined: attributive and object ones. The context decomposition leads to a mode of incremental learning GMRTs. Four cases of incremental learning are proposed: adding a new object (attribute value) and deleting an object (attribute value). Some illustrative examples of four cases of incremental learning are given too.

Keywords: Granular computing · incremental classification · good test · formal concept

1 Introduction

Information granules are becoming important entities in data processing at the different levels of data abstraction. Information granules have also contributed to increasing the precision in data processing [9]. Application of information granules is one of the problem-solving methods based on decomposing a big problem into subtasks.

Several studies devoted to evolving information granules to adapt to changes in the streams of data are described in [12]. The process of forming information granules is often associated with the removal of some element of data or dealing with incomplete data [1]. Generally, we consider object, property, class of objects, and classification as the main granules of human classification reasoning.

The paper deals with inferring good classification (diagnostic) tests. Tests are good because they cover the largest possible number of objects w. r. t. the inclusion relation on the set of all object subsets.

Two kinds of classification subcontexts are defined: attributive and object ones. The context decomposition leads to a mode of incremental learning good classification tests. Four cases of incremental learning are proposed: adding a new object (attribute value) and deleting an object (attribute value). Some heuristic rules allowing decreasing the computational complexity of inferring good tests are considered too.

In [7], it is considered the link between Good Test Analysis (GTA) and Formal Concept Analysis (FCA) [5]. To give a target classification of objects, we use an additional attribute $KL \notin U$. In Tab. 1, we have classification KL containing two classes: the objects in whose descriptions the target value $k(+)$ appears and all the other objects.

Table 1. Example of classification

Index	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	KL
1	2	3	1	2	1	2	2	2	k(+)
2	2	2	2	2	1	2	0	2	k(+)
3	2	3	1	3	1	2	2	2	k(+)
4	2	2	3	1	1	0	2	2	k(+)
5	2	3	2	2	1	2	2	0	k(+)
6	2	3	1	3	0	0	2	2	k(+)
7	2	3	1	2	1	2	2	0	k(-)
8	1	4	2	1	0	2	0	2	k(-)
9	2	3	1	3	1	0	2	2	k(-)
10	1	4	2	2	1	2	2	2	k(-)
11	2	4	1	2	0	2	2	2	k(-)
12	2	4	2	2	1	2	0	0	k(-)

2 Three interrelated sets of classes, objects, and properties

The “atom” of plausible human reasoning is a concept. The concepts are represented by theirs names. We shall consider the following roles of names in reasonings: a name can be the name of object, the name of class of objects and the name of classification or collection of classes. With respect to the role of name in knowledge representation schemes, it can be the name of attribute or attribute’s value. A class of objects may contain only one object; hence the name of the object is a case of the name of a class. For example, fir-tree can be regarded as the name of a tree or the name of a class of trees. Each attribute genarates a classification of a given set of objects; hence the names of attributes can be the names of classifications and the attribute values can be the names of classes. In the knowledge bases, the sets of names for objects, classes and classifications must not intersect.

Let k be the name of an objects' class, c be the name of a property of objects (value of an attribute), and g be the name of an object. Each class or property has only one maximal set of objects as its interpretation that is the set of objects belonging to this class or possessing this property: $k \rightarrow I(k) = \{g : g \leq k\}$, $c \rightarrow I(c) = \{g : \ g \leq c\}$, where the relation '$\leq$' denotes 'is a' relation and has causal nature (the dress is red, an apple is a fruit). Each object has only one corresponding set of all its properties:$C(g) = \{c : g \leq c\}$. We shall say that $C(g)$ is the description of object g. The link $g \rightarrow C(g)$ is also of causal nature. We shall say that $C(k) = \{\cap C(g) : g \leq k\}$ is the description of class k, where $C(k)$ is a collection of properties associated with each object of class k. The link $k \rightarrow C(k)$ is also of causal nature. Figure 1 illustrates the causal links between classes of objects, properties of objects, and objects.

Clearly, each description (a set of properties) has one and only one interpretation (the set of objects possessing this set of properties). But the same set of objects can be the interpretation of different descriptions (equivalent with respect to their interpretations). The equivalent descriptions of the same class are said to be the different names of this class. The task of inferring the equivalence relations between names of classes and properties underlies the processes of plausible reasoning.

The identity has the following logical content: class K is equivalent to property k $((K \leftrightarrow k))$ if and only if the interpretations $I(K), I(k)$ on the set of conceivable objects are equal $I(K) = I(k)$. It is possible to define also the relationship of approximate identity between concepts: k approximates B $(k \leq B)$ if and only if the relation $I(k) \subseteq I(K)$ is satisfied. We can consider instead of one property (concept) any subset of properties joined by the union $\cup$ operation: $(c_1 \cup c_2 \cup \ldots \cup c_i \cup \ldots \cup c_n) \leq B$.

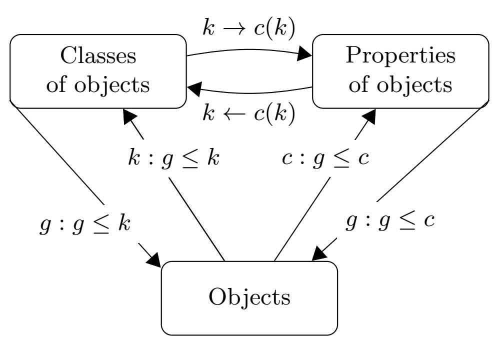

Fig. 1. Links between Objects, Classes, and Properties of Objects

Connection directed from 'Properties of objects' to 'Classes of objects' is constructed by learning from examples of objects and their classes. This connection is also of causal nature, it is expressed via the "if – then" rule: to say that "if tiger, then mammals" means to say that I(tiger) $\subseteq$ I(mammals).

3 Background definitions

.

Let G = {1, 2,. . ., N} be the set of objects' indices (objects, for short) and M = {m_1, m_2, . . ., m_j, . . . m_q} be the set of attributes' values (values, for short). Each object is described by a set of values from M. The object descriptions are represented by rows of a table R the columns of which are associated with the attributes taking their values in M. Denote a description of $g \in G$ by $\delta(g)$. Let D_+ and G_+ ($D_- = \mathrm{D}/D_+$) and $G_- = \mathrm{G}/G_+$) be the sets of positive or negative object descriptions and the set of indices of these objects, respectively. The definition of good tests is based on two mapping $2^{\mathrm{G}} \to 2^{\mathrm{M}}$, $2^{\mathrm{M}} \to 2^{\mathrm{G}}$. Let A $\subseteq$ G, B $\subseteq$ M. Denote by B_i, $B_i \subseteq$ M, i = 1,. . ., N the description of object with index i. The relations $2^{\mathrm{G}} \to 2^{\mathrm{M}}$, $2^{\mathrm{M}} \to 2^{\mathrm{G}}$ are: A' = val(A) = {intersection of all B_i: $B_i \subseteq$ M, i $\in$ A} and B' = obj(B) = {i: i $\in$ G, B $\subseteq$ B_i}.

These mapping are the Galois's correspondences. Operations val(A), obj(B) are reasoning operations (derivation operations). We introduce two generalization operations: generalization_of(B) = B''= val(obj(B); generalization_ of(A) = A'' = obj(val(A)). These operations are the closure operations [8]. A set A is closed if A = obj(val(A)). A set B is closed if B = val(obj(B)). For g $\in$ G and m $\in$ M, g' is called object intent and m' is called value extent. We illustrate the derivation and generalization operations (Tab. 1):

A = {4, 8}, val(A) = {x_4 = 1, x_8 = 2}; A'' = obj({x_4 =1, x_8 = 2}) = {4, 8} = A;

m = {x_8 = 0}, obj ({x_8 = 0}) = {5, 7, 12}; m'' = val({5, 7,12}) = {x_1 = 2, x_4 = 2, x_5 = 1, x_6 = 2, x_8 = 0 }; B = {x_4 = 3, x_5 = 1}, obj({B}) = {3,9}; B'' = val({3, 9}) = {x_1 = 2, x_2 = 3, x_3 = 1, x_4 = 3, x_5 = 1, x_7 = 2, x_8 = 2}.

Definition 1. A Diagnostic Test (DT) for G_+ is a pair (A, B) such that $B \subseteq M, A = obj(B) \neq \emptyset$, $A \subseteq G_+$, and $obj(B) \cap \delta(g) = \emptyset, (\forall g) g \in G_-$.

Definition 2. A Diagnostic Test (DT) for G_+ is maximally redundant if $obj(B \cup m) \subset A$ for all $m \in M \setminus B$.

Definition 3. A Diagnostic Test (DT) for G_+ is good iff any extension $A^* = A \cup i$, $i \in G_+ \setminus A$, implies that $(A^*, val(A^*))$ is not a test for G_+.

Note that the definition of tests for G_+ does not differ from the definition of positive hypotheses given in [6] and [4] in the language of predicates. Definitions 2, 3, 4 remain true if G_+ is replaced by G_-. In what follows, we are interested in inferring GMRTs for positive class of objects. As far as Formal Concept Analysis development and application, the following surveys [11,10,3,2] can be seen

Some examples of formal concepts are in Tab. 1. Let us check if a pair (A, B) = ((1,5, 6,7,9), (x_1=2, x_2 = 3, x_7 = 2)) is a concept or not. It is a concept, because obj((x_1=2, x_2 = 3, x_7 = 2) = (1,5,6,7,9) = A. However, this concept does not distinguish the classes of objects. Pair ((11,12), (x_1 = 2, x_2 = 4, x_4 =

2)) is a concept and a test for k(−), but not a good one, because there is pair ((8, 10, 11, 12), ($x_2 = 4$)) such that (11, 12) ⊂ (8, 10, 11 12).

4 Incremental learning GMRTs

Define two kinds of subtasks:

1. to find all GMRTs intents of which are included in the description of an object;
2. to find all GMRTs into intents of which a given set of values is included.

To solve these subtasks, we need to form subcontexts (projections) of a given classification context.

Definition 4. Let $B \subseteq M$. The object projection $proj(B, G_+)$ on G_+ is $proj(B, G_+) = \{\delta(g) \cap B | g \in G_+, \delta(g) \cap B \neq \emptyset$, and $(obj(\delta(g) \cap B), \delta(g) \cap B)$ is a test for $G_+\}$.

An example of object projection $proj(d_2)$ on G_+ is in Tab. 2.

Table 2. Example of object 2 projection

Index	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	KL	Test?
1	2			2	1	2		2	k(+)	no
2	2	2	2	2	1	2	0	2	k(+)	yes
3	2				1	2		2	k(+)	yes
4	2	2			1			2	k(+)	yes
5	2		2	2	1	2			k(+)	no
6	2							2	k(+)	no

Definition 5. Let $X \subseteq M$. The value projection $proj$ (X, G_+) on G_+ is $proj(X, G_+) = \{\delta(g) | g \in G_+, X \subseteq \delta(g)\}$.

An example of value projection $proj(x_5 = 1, x_6 = 2)$ on G_+ can be presented as descriptions of 1, 2, 3 and 5 objects.

5 Four cases of incremental learning of classification context

We propose four cases of modifying classification contexts: adding/removing objects and adding/removing values of attributes. Modification of GMRTs is based on a decomposition of classification contexts into value and object subcontexts and inferring GMRTs in them. Let $STGOOD_+$ and $STGOOD_-$ be the current sets of extents of GMRTs for positive and negative class of objects, respectively.

We mean that the process in which a change of the classification context implies only updating the sets $STGOOD_+$ and $STGOOD_-$. The classification context can be changed as follows: a new object is added with the indication of its class membership; an object is deleted from G+ or G-; a value is introduced into the classification context; a value is deleted from the classification context. Updating $STGOOD_+$ and $STGOOD_-$ is performed with the use of only subcontexts associated with added (deleted) object or value.

Case 1 The following actions are necessary:

1. Checking whether it is possible to extend the extents of some existing GMRTs for the class to which a new object belongs (a class of positive objects, for certainty).
2. Inferring all GMRTs, intents of which are included into the new object description; for this goal, the first kind subtask is used.
3. Deleting GMRTs for positive class extents of which are included in the extent of any new GMRTs.
4. Checking the validity of GMRTs for negative objects, and, if it is necessary, modifying invalid GMRTs (test for negative objects is invalid if its intent is included in a new (positive) object description); for this goal, the second kind subtask is used.

Let $s \in STGOOD_-$ and $Y = val(s)$. If $Y \subseteq t_{new}(+)$, then s should be deleted from $STGOOD_-$ because (s, Y) is invalid test for G_-.

Proposition 1. $obj(Y)$ forms the subcontext for finding corrected tests for G_-.

Proof. $Y \subseteq X \leftrightarrow obj(X) \subseteq obj(Y)$. Assume that there exists a GMRT (with intent Z) for G_- such that $obj(Z) \not\subset$ and $\neq obj(Y)$. Then $obj(Z)$ contains some objects not belonging to $obj(Y)$ and Z will be included in some descriptions of objects not belonging to $obj(Y)$ and, consequently, Z has been obtained at the previous steps of the incremental algorithm for finding all GMRTs for G_-.

Consider an example of adding object in the process of inferring GMRTs for the data in Tab. 1. Let us fix the classification context with 3 first objects from G_+ and all the objects of G_-. For this current situation we have one GMRT for G_+, namely, ((1,2,3) (x_1=2, $x_5 = 1$, $x_6 = 2$, $x_8 = 2$)) and one GMRT for G_-, namely, ((8,10, 11, 12), ($x_2 = 4$, $x_6 = 2$)). As a result of adding object 4 into positive class of objects, we obtain a new GMRT for positive class of objects ((2,4), ($x_1 = 2$, $x_2 = 2$, $x_5 = 1$, $x_8 = 2$)). GMRTs for negative class of objects do not changed.

Case 2 Suppose that an object g is deleted from G_+ (G_-) The following actions are necessary:

1. $\forall s$, $s \in STGOOD_+$ ($STGOOD_-$), $g \in s$, delete g from s; in this connection, we observe that ($s \backslash g$, $val(s \backslash g)$) remains to be the test for G_+ (G_-).
2. We denote modified test ($s \backslash g$, $val((s \backslash g)$) by MT. We have the following possibilities:

– the intent of MT has not changed; then MT is a GMRT for $G_+(G_-)$ in the modified context;
– the intent of MT has changed and the extent of MT is included in the extent of an existing GMRT for $G_+(G_-)$, then MT must be deleted; otherwise MT is a GMRT for $G_+(G_-)$.

An example of deleting object. Let us fix the whole classification context (Tab. 1). We have one GMRT for G- obtained in this context: ((8,10,11,12), (x_2 = 4, x_6 = 2). Delete object 8. We have one modified test: ((10, 11, 12), (x_2 =4, x_4 = 2, x_6 = 2)) and it is the GMRT.

Case 3 Suppose that a new value m^* is added to the classification context: m^* appears in the descriptions of some positive and negative objects and $M ::= m^* \cup M$. The task of finding all GMRTs for $G_+(G_-)$ whose intents contain m^* is reduced to the task of the second kind. The subcontext for this task is determined by the set of all objects whose descriptions contain m^*. As result, we obtain all the GMRTs $(obj(Y), Y)$ for $G_+(G_-)$ such that $m^* \in Y$. We can add a set of values if we want that all these values will be included simultaneously in the intents of GMRTs.

Case 4 Suppose that some value m is deleted from the classification context. Let a GMRT ($obj(X)$, X) for $G_+(G_-)$ be transformed into $(obj(X\backslash\, m), X\backslash\, m)$. Then we have $((X\backslash m) \subset X) \leftrightarrow (obj(X) \subseteq obj(X\backslash m))$.

Consider two possibilities: $obj(X\backslash m) = obj(X)$ and $obj(X) \subset obj(X\backslash m)$. In the first case, $(obj(X\backslash m), X\backslash m)$ is a GMRT for $G_+(G_-)$. In the second case, ($obj(X\backslash m)$, $X\backslash m$) is not a test. However, $obj(X\backslash m)$ can contain extents of new GMRTs for $G_+(G_-)$ and these tests can be obtained by using the subtask of the second kind.

An example of deleting value $x_6 = 2$. The GMRT for the negative class ((8, 10, 11, 12), (x_8 = 4)) remains GMRT. The GMRTs ((2, 4), (x_1 = 2, x_2 = 2, x_5 = 1, x_8 = 2)) remains GMRT, but the pair ((1, 2, 3), (x_1 = 2, x_5 =1, x_8 = 2)) is not a test for positive objects after deleting x_6 = 2. It is impossible to find any GMRTs for positive objects 1, 2, 3, 4.

Recognizing the class membership for a new object not belonging to training set is performed as follows:

– If (and only if) description of object contains an intent of GMRT of only one class, then the object can be assigned to this class;
– If description of an object does not contain any intent of GMRTs, then we have the case of uncertainty.

In two last cases, it is necessary to continue learning by adding new objects or to change the classification context.

6 Conclusions

The paper examines the relationship between an incremental model of good classification test inferring with granular computing. In the process of finding good tests, the following granules are highlighted: objects, attribute values, and object classes. The incremental test inferring is carried out as a process in which granules can be added and removed, changing the classification context. The decomposition of classification contexts into subcontexts is considered based on the selection of objects and values of attributes (granules). Thus, granules become active elements of test inferring and allow this process to be made data driving based on data selection.

Acknowledgements

The research was partially supported by Russian Foundation for Basic Research, research project No. 18-07-00098A. This research work was supported by the Academic Excellence Project 5-100 proposed by Peter the Great St. Petersburg Polytechnic University.

References

1. Al-Hmouz, R., Pedrycz, W., Balamash, A.S., Morfeq, A.: Granular description of data in a non-stationary environment. Soft Computing 22(2), 523–540 (Jan 2018), https://doi.org/10.1007/s00500-016-2352-2
2. Buzmakov, A., Kuznetsov, S.O., Napoli, A.: A new approach to classification by means of jumping emerging patterns. In: CEUR Workshop Proceedings. vol. 939, pp. 15–22 (2012)
3. Codocedo, V., Napoli, A.: Formal concept analysis and information retrieval – a survey. In: Baixeries, J., Sacarea, C., Ojeda-Aciego, M. (eds.) Proceedings of the 13th ICFCA. pp. 61–77 (2015)
4. Finn, V.: On machine-oriented formalization of plausible reasoning in the style of F.Bacon–J.S. Mill. Semiotika i Informatika 20, 35–101 (1983), (in Russian)
5. Ganter, B., Wille, R.: Formal concept analysis: mathematical foundations. Springer, Berlin (1999)
6. Ganter, B., Kuznetsov, S.O.: Formalizing hypotheses with concepts. In: Conceptual Structures: Logical, Linguistic, and Computational Issues: Proceedings of the 8th International Conference on Conceptual Structures. pp. 342–356 (2000)
7. Naidenova, X.: Good classification tests as formal concepts. In: Domenach, F., Ignatov, D., Poelmans, J. (eds.) Proceedings of the 10th International Conference on Formal Concept Analysis. vol. 7278, pp. 211–226 (2012)
8. Ore, O.: Galois connections. Trans. Amer. Math. Soc 55, 494–513 (1944)
9. Pedrycz, W.: Allocation of information granularity in optimization and decision-making models: Towards building the foundations of granular computing. European Journal of Operational Research 232(1), 137 – 145 (2014)
10. Poelmans, J., Ignatov, D.I., Kuznetsov, S.O., Dedene, G.: Formal concept analysis in knowledge processing: A survey on applications. Expert Systems with Applications 40(16), 6538 – 6560 (2013)
11. Poelmans, J., Kuznetsov, S.O., Ignatov, D.I., Dedene, G.: Formal Concept Analysis in knowledge processing: A survey on models and techniques. Expert systems with applications 40(16), 6601–6623 (NOV 15 2013)
12. Xu, J., Wang, G., Li, T., Pedrycz, W.: Local-density-based optimal granulation and manifold information granule description. IEEE Transactions on Cybernetics 48(10), 2795–2808 (Oct 2018)