FCA-based Approach to Machine Learning*

Dmitry V. Vinogradov[1,2]

[1] Federal Research Center for Computer Science and Control,
Russian Academy of Science, Moscow 119333, Russia
vinogradov.d.w@gmail.com
[2] Russian State University for Humanities, Intelligent Robotics Laboratory,
Moscow 125993, Russia
WWW home page: http://isdwiki.rsuh.ru/index.php/User:Vinogradov.Dmitry

Abstract. The main result of the paper provides a lower bound on sufficient number of randomly generated formal concepts to correctly predict all important positive test examples with given confidence level. The technique coincides with modern approach to the famous theorem of V.N. Vapnik and A.Ya. Chervonenkis. However the situation is dual to the classical one: in our case test examples correspond to fixed subsets and probabilistically generated formal concepts must fall into selected areas of sufficient large volume.

Keywords: formal context, formal concept, Boolean hypercube, lower half-space, prediction, confidence

1 Introduction

Formal Concept Analysis (FCA) [1] is a popular means based on lattice theory for formalizing methods of data analysis in case of small samples.

Applicability of FCA to Big Data has several obstacles:

- Exponentially large number of hypotheses with respect to size of the initial formal context in the worst case.
- Many problems of FCA belong to famous classes of NP- and $\#P$-complete problems [3].
- There is a positive probability of "accidental" concepts appearance that correspond of overfitting phenomenon [7].

The paper [6] introduces the Markov chain approach to probabilistic generation of formal concepts (so-called VKF-method). The computer VKF-system uses the coupling Markov chain to generate random sample of concepts. Each run of this chain terminates with probability 1. Since each hypothesis (formal concept) is generated by independent run of the Markov chain, the system makes the induction step in parallel by several threads. Finally the system predicts target class of each test example by the analogy reasoning.

* Partially supported by RFBR grant 17-07-00539A

The key question of the approach is how to determine sufficient number of hypotheses to predict target class with given level of confidence. The paper proposes an answer to this question.

Used technique mostly coincides with modern approach to the famous theorem of V.N. Vapnik and A.Ya. Chervonenkis. However the situation is dual to the classical one: in our case test examples correspond to fixed subsets and probabilistically generated formal concepts must fall into selected areas of sufficient large volume. The general approach of Vapnik-Chervonenkis uses the "Occam razor" principle where no assumption on selected hypothesis made except to its correctness on all training examples. Hence a hypothesis coincides with area of objects space. To reject a bad hypothesis is needed to randomly pick training objects from the corresponding subset.

2 Background

2.1 Basic definitions and facts of FCA

Here we recall some basic definitions and facts of Formal Concept Analysis (FCA) [1].

A **(finite) context** is a triple (G, M, I) where G and M are finite sets and $I \subseteq G \times M$. The elements of G and M are called **objects** and **attributes**, respectively. As usual, we write gIm instead of $\langle g, m\rangle \in I$ to denote that object g has attribute m.

For $A \subseteq G$ and $B \subseteq M$, define

$$A' = \{m \in M | \forall g \in A(gIm)\}, \tag{1}$$
$$B' = \{g \in G | \forall m \in B(gIm)\}; \tag{2}$$

so A' is the set of attributes common to all the objects in A and B' is the set of objects possesing all the attributes in B. The maps $(\cdot)' : A \mapsto A'$ and $(\cdot)' : B \mapsto B'$ are called **derivation operators** (**polars**) of the context (G, M, I).

A **concept** of the context (G, M, I) is defined to be a pair (A, B), where $A \subseteq G$, $B \subseteq M$, $A' = B$, and $B' = A$. The first component A of the concept (A, B) is called the **extent** of the concept, and the second component B is called its **intent**. The set of all concepts of the context (G, M, I) is denoted by $\mathbf{B}(G, M, I)$.

Let (G, M, I) be a context. For concepts (A_1, B_1) and (A_2, B_2) in $\mathbf{B}(G, M, I)$ we write $(A_1, B_1) \leq (A_2, B_2)$, if $A_1 \subseteq A_2$. The relation $\leq$ is a **partial order** on $\mathbf{B}(G, M, I)$.

A subset $A \subseteq G$ is the extent of some concept if and only if $A'' = A$ in which case the unique concept of which A is the extent is (A, A'). Similarly, a subset B of M is the intent of some concept if and only if $B'' = B$ and then the unique concept with intent B is (B', B).

Proposition 1. *[1] Let (G, M, I) be a context. Then $(\boldsymbol{B}(G, M, I), \leq)$ is a lattice with join and meet given by*

$$\bigvee_{j\in J}(A_j, B_j) = ((\bigcup_{j\in J} A_j)'', \bigcap_{j\in J} B_j), \tag{3}$$

$$\bigwedge_{j\in J}(A_j, B_j) = (\bigcap_{j\in J} A_j, (\bigcup_{j\in J} B_j)''); \tag{4}$$

□

Corollary 1. *For context (G, M, I) the lattice $(\boldsymbol{B}(G, M, I), \leq)$ has (M', M) as the bottom element and (G, G') as the top element. In other words, for all $(A, B) \in \boldsymbol{B}(G, M, I)$ the following inequalities hold:*

$$(M', M) \leq (A, B) \leq (G, G'). \tag{5}$$

□

Definition 1. *For $(A, B) \in \boldsymbol{B}(G, M, I)$, $g \in G$, and $m \in M$ define*

$$CbO((A, B), g) = ((A \cup \{g\})'', B \cap \{g\}'), \tag{6}$$

$$CbO((A, B), m) = (A \cap \{m\}', (B \cup \{m\})''). \tag{7}$$

so $CbO((A, B), g)$ is equal to $(A, B) \vee (\{g\}'', \{g\}')$ and $CbO((A, B), m)$ is equal to $(A, B) \wedge (\{m\}', \{m\}'')$.

We call these operations CbO because the first one is used in Close-by-One (CbO) Algorithm to generate all the elements of $\mathbf{B}(G, M, I)$, see [2] for details.

Useful properties of introduced operations are summarized in the following Lemmas.

Lemma 1. *Let (G, M, I) be a context, $(A, B) \in \boldsymbol{B}(G, M, I)$, $g \in G$, and $m \in M$. Then*

$$g \in A \Rightarrow CbO((A, B), g) = (A, B), \tag{8}$$

$$m \in B \Rightarrow CbO((A, B), m) = (A, B), \tag{9}$$

$$g \notin A \Rightarrow (A, B) < CbO((A, B), g), \tag{10}$$

$$m \notin B \Rightarrow CbO((A, B), m) < (A, B). \tag{11}$$

Lemma 2. *Let (G, M, I) be a context, $(A_1, B_1), (A_2, B_2) \in \boldsymbol{B}(G, M, I)$, $g \in G$, and $m \in M$. Then*

$$(A_1, B_1) \leq (A_2, B_2) \Rightarrow CbO((A_1, B_1), g) \leq CbO((A_2, B_2), g), \tag{12}$$

$$(A_1, B_1) \leq (A_2, B_2) \Rightarrow CbO((A_1, B_1), m) \leq CbO((A_2, B_2), m). \tag{13}$$

Now we represent the coupling Markov chain algorithm that is a core of probabilistic approach to machine learning based on FCA (VKF-method).

Data: context (G, M, I), external function $CbO(\, , \,)$
Result: random concept $(A, B) \in \mathbf{B}(G, M, I)$
$X := G \sqcup M$; $(A, B) := (M', M)$; $(C, D) = (G, G')$;
while $((A \neq C) \vee (B \neq D))$ **do**
  select random element $x \in X$;
  $(A, B) := CbO((A, B), x)$; $(C, D) := CbO((C, D), x)$;
end

Algorithm 1: Coupling Markov chain

The order on two concepts $(A, B) \leq (C, D)$ at any intermediate step of the while loop of Algorithm 1 follows from Lemma 2.

2.2 Probabilistic algorithms for FCA-based machine learning

Now we represent the general scheme of machine learning based on FCA (VKF-method). The reader can learn the classical deterministic FCA-based approach to machine learning from Kuznetsov [4]. Our technique uses probabilistic Algorithm 1 for computing a random subset of formal concepts.

As usual, there are two sets of objects called the training and test samples, respectively.

From positive examples of the training sample the program generates a formal context (G^{+}, M, I). The negative examples form the set G^{-} of counter-examples (**obstacles**).

Set G^{τ} of examples to predict the target class contains all test objects.

After that the program applies the coupling Markov chain algorithm 1 to generate a random formal concept $(A, B) \in \mathbf{B}(G^{+}, M, I)$. The program saves the concept (A, B), if there is no obstacle $o \in G^{-}$ such that $B \subseteq o'$.

Data: number N of concepts to generate
Result: random sample S of formal concepts without obstacles
G^{+} := (+)-examples, M := attributes; $I \subseteq G^{+} \times M$ is a formal context for (+)-examples;
G^{-} := (-)-examples; $S := \emptyset$; $i := 0$;
while $(i < N)$ **do**
  Generate concept $\langle A, B \rangle$ by Algorithm 1; *hasObstacle* := **false**;
  for $(o \in G^{-})$ **do**
    if $(B \subseteq o')$ **then**
      hasObstacle := **true**;
    end
  end
  if (*hasObstacle* = ***false***) **then**
    $S := S \cup \{\langle A, B \rangle\}$;
    $i := i + 1$;
  end
end

Algorithm 2: Inductive generalization

Condition ($B \subseteq o'$) of Algorithm 2 means the inclusion of intent B of concept $\langle A, B\rangle$ into the fragment (attributes subset) of counter-example o.

If a concept avoids all such obstacles it is added to the result set of all the concepts without obstacles.

We replace a time-consuming deterministic algorithm (for instance, "Close-by-One") for generation of all concepts by the probabilistic one to randomly generate the prescribed number of concepts.

The goal of Markov chain approach is to select a random sample of formal concepts without computation of the (possibly exponential size) set $\mathbf{B}(G, M, I)$ of all the concepts.

Finally, machine learning program predicts the target class of test examples and compares the results of prediction with the original target value.

Data: random sample S of concepts, list of (τ)-objects
Result: prediction of target class of (τ)-examples
$X := (\tau)$-examples;
for $(o \in X)$ **do**
    $PredictPositively(o) :=$ **false**;
    for $(\langle A, B\rangle \in S^+)$ **do**
        if $(B \subseteq o')$ **then**
            $PredictPositively(o) :=$ **true**;
        end
    end
end

Algorithm 3: Prediction of target class by analogy

3 Main result

Algorithm 3 gives the following

Definition 2. *Object o with fragment (attributes subset) $o' \subseteq M$ is **positively predicted** by concept $\langle A, B\rangle$ if $B \subseteq o'$.*

If there are $n = |M|$ attributes then intent B of any concept $\langle A, B\rangle$ is a point of n-hypercube $\{0, 1\}^n$.

Definition 3. ***Lower half-space** $H^{\downarrow}(o)$ corresponding to object o with fragment $o' \subseteq M$ is defined by linear inequality $x_{j_1} + \ldots + x_{j_k} < \frac{1}{2}$, where $M \setminus o' = \{m_{j_1}, \ldots, m_{j_k}\}$. The empty lower half-space $0 < \frac{1}{2}$ (equals to $\{0, 1\}^n$) is allowed too and corresponds to $o' = M$.*

Remark that cardinality of all possible lower half-spaces is equal to 2^n.

Key observation is

Lemma 3. *Object o is positively predicted if and only if lower half-space $H^{\downarrow}(o)$ contains a fragment B of at least one concept $\langle A, B\rangle$.*

Definition 4. *Object o is called ε-**important** if probability of occurrence of random concept $\langle A, B\rangle$ with $B \in H^{\downarrow}(o)$ is greater than ε.*

*A family of concepts is called ε-**net** if for each ε-important object o there is at least one its member $\langle A, B\rangle$ with $B \in H^{\downarrow}(o)$.*

Now we are interested only in 1-st type error probability (positive prediction fails): we need to determine a number N (depending on ε and δ) such that a random sample of cardinality N forms ε-net with probability greater than $1 - \delta$.

Lemma 4. *For all ε with $l > \frac{2}{\varepsilon}$ and for any independent random samples S_1 and S_2 of concepts of cardinality l the following inequality holds:*

$$\boldsymbol{P}^l\{S_1 : \exists H \in (Sub\downarrow)\,[S_1 \cap H = \emptyset, \boldsymbol{P}H > \varepsilon]\} \leq$$
$$\leq 2 \cdot \boldsymbol{P}^{2l}\{S_1S_2 : \exists H \in (Sub\downarrow)\,[S_1 \cap H = \emptyset, |S_2 \cap H| > \varepsilon \cdot l/2]\}.$$

Lemma 5. *For all ε and for any independent random samples S_1 and S_2 of concepts of cardinality l the following inequality holds:*

$$\boldsymbol{P}^{2l}\{S_1S_2 : \exists H \in (Sub\downarrow)\,[S_1 \cap H = \emptyset, |S_2 \cap H| > \varepsilon \cdot l/2]\} \leq$$
$$\leq m^{Sub\downarrow}(2l) \cdot 2^{-\varepsilon l/2}.$$

Theorem 1. *For $n = |M|$ and for any $\varepsilon > 0$ and $1 > \delta > 0$ random sample of concepts of cardinality*

$$N \geq \frac{2 \cdot (n+1) - 2 \cdot \log_2 \delta}{\varepsilon}$$

forms ε-net with probability $> 1 - \delta$.

Proof. Solve inequality $2 \cdot 2^n \cdot 2^{-\varepsilon N/2} \leq \delta$ with respect to N to obtain the estimate.

Conclusions

In this paper we provided a lower bound on sufficient number of randomly generated formal concepts to correctly predict all important positive test examples with given confidence level. The technique mostly coincides with modern approach to the famous theorem of V.N. Vapnik and A.Ya. Chervonenkis, but the situation is dual to the classical one.

Acknowledgements.

The author would like to thank Prof. Victor K. Finn and his colleagues at Federal Research Center for Computer Science and Control of Russian Academy of Science for support and helpful discussions.

The author is grateful to anonymous reviewers for improving the style of presentation.

References

1. Ganter, Bernard and Wille, Rudolf. *Formal Concept Analysis: Mathematical Foundations*, Springer-Verlag, 1999
2. Kuznetsov, S.O.: A Fast Algorithm for Computing all Intersections of Objects in a Finite Semi-Lattice. Autom. Doc. Math. Linguist. 27:5, 11-21 (1993)
3. Kuznetsov, S.O. Complexity of Learning in Concept Lattices from Positive and Negative Examples. *Discrete Applied Mathematics*, no. 142(1-3). – 2004. – pp. 111–125
4. Kuznetsov, S.O. Machine Learning and Formal Concept Analysis. *Proc. 2nd International Conference on Formal Concept Analysis: Springer LNAI*, Vol. 2961. – 2004. – pp. 287-312
5. Makhalova, T.P., Kuznetsov, S.O. On Overfitting of Classifiers Making a Lattice. *Proc. 14th International Conference on Formal Concept Analysis: Springer LNAI*, Vol. 10308. – 2017. – pp. 184-197
6. Vinogradov, D.V. A Markov Chain Approach to Random Generation of Formal Concepts. *Proceedings of the Workshop Formal Concept Analysis Meets Information Retrieval (FCAIR 2013): CEUR Workshop Proceedings*, Vol. 977. – 2013. – p. 127–133
7. Vinogradov, D.V. Accidental Formal Concepts in the Presence of Counterexamples. *Proceedings of International Workshop on Formal Concept Analysis for Knowledge Discovery (FCA4KD 2017): CEUR Workshop Proceedings*, Vol. 1921. – 2017. – p. 104–112
8. Vorontsov, K.V., Ivahnenko, A. Tight Combinatorial Generalization Bounds for Threshold Conjunction Rules. *Proceedings of 4th International Conference on Pattern Recognition and Machine Intelligence.* – 2011. – p. 66-73