Accidental formal concepts in the presence of counterexamples

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. An accidental formal concept corresponds to a subset of attributes that accidentally appear in several objects (of the concept extent) if every such an object belongs to different "real" formal concept extent. There are two standard techniques to forbid such concepts: counterexamples to exclude involved concepts and the lower bound on the size of extent. Both are insufficient. Here we define random formal context with attributes (different from elements of "real" formal concepts) generated by independent Bernoulli variables. The main result has asymptotic form: If number n of the random attributes tends to infinity, the probability of success equals to $\sqrt{\frac{a}{n}}$, and there are $m = b \cdot \sqrt{n}$ counterexamples, then the probability of an accidental formal concept with 2 parents is $1 - e^{-a} - a \cdot e^{-a} \cdot [1 - e^{-b \cdot \sqrt{a}}]$ in the limit.

Keywords: formal context, formal concept, Bernoulli variable, counterexample, overfitting

1 Introduction

Formal Concept Analysis (FCA) [2] is one of the most popular means to analyse small sample data by tools of lattice theory.

Applicability of FCA to Big Data has several obstacles:

- Exponentialy large number of hypotheses with respect to size of the initial formal context appears in the worst case.
- In practice of sociology data analysis 100×100 training sample has generated more than $11,500$ formal concepts.
- Many problems of FCA belong to famous NP and $\#P$ completeness complexity classes [3].

The author adds the overfitting phenomenon to this list, when so-called "accidental" formal concepts appear. The following example demonstrates an accidental concept:

Let

$$O = \{o_1 = B737, o_2 = SSJ100, o_3 = IL76, o_4 = A320\}$$

be a set of damaged planes, described by set

$$F = \{f_1 = empennage\ damage, f_2 = motor\ damage, f_3 = curse\ scratch\}$$

of attributes by formal context

$O \mid F$	f_1	f_2	f_3
o_1	1	0	0
o_2	1	0	1
o_3	0	1	1
o_4	0	1	0

Formal concepts

$$\langle \{o_1, o_2\}, \{f_1\} \rangle$$

"empennage damage is a cause" and

$$\langle \{o_3, o_4\}, \{f_2\} \rangle$$

"motor damage is a cause" are plausible ones, but

$$\langle \{o_2, o_3\}, \{f_3\} \rangle$$

"curse scratch is a cause of damage" is doubtful. The last one occurs since the attribute $\{f_3\}$ accidentally appears in 2 training objects o_2 and o_3 with different "real" causes.

Of couse, separation of formal concepts into "accidental" and "real" ones heavily depends on experts' intuition about research domain. In the previous example we know enough to do this ourselves. However in more complex domain there are no information to reject "accidental" concepts automatically.

Counterexamples give a mean to delete such "accidental" formal concepts by using so-called "counterexamples forbidden condition" procedure. A counterexample is an object (described by combinations of the same attributes as training objects) without the target property. Target property (or goal attribute) is a binary attribute of the objects. We suppose that formal concept contains training objects with the target property and there exists an additional list of counterexamples without this property. This assumption converts FCA into a machine learning method (see, for example, [3]). In the previous example "to be damaged" was the target property of training examples.

We develop a probabilistic model for (retriction of) formal context (on the subset of unessential attributes). Attribute is essential if it occurs as a part of intent of some "real" formal concept. The goal is to estimate of probability of appearance of subset of unessential attributes as a formal concept intent without inclusion into some counterexample. Counterexample intent (or description) corresponds to a random subset of unessential attributes. Inclusion of the intent of a formal concept into some counterexample intent is suspicious because the cause (this formal concept) is present and the target property is absent.

Main result of the paper states unsufficiency of the counterexamples forbidden condition. If number n of the random (unessential) attributes tends to infinity, the probability of success equals to $\sqrt{\frac{a}{n}}$, and there are $m = c \cdot \sqrt{n}$ counterexamples, then the probability of an accidental formal concept with 2 elements extent and without any counterexample is $1 - e^{-a} - a \cdot e^{-a} \cdot [1 - e^{-c \cdot \sqrt{a}}]$ in the limit.

It's clear that the last term is greater than 0. It means that there is a positive probability of appearance of formal concept with intent consisting of only inessential attributes such that there is no inclusion into some counterexample. "Accidental" formal concepts correspond to the overfitting phenomenon since we try to use all available information on one hand, and "accidental" formal concepts do not correspond to any "real" cause and hence worsen the target property prediction on test examples on the other hand.

2 Basic definitions and facts of FCA

Here we recall some basic definitions and facts of Formal Concept Analysis (FCA) [2].

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 $\mathbf{B}(G, M, I)$.

Fix a context (G, M, I). In the following, let J be an index set. We assume that $A_j \subseteq G$ and $B_j \subseteq M$, for all $j \in J$.

Lemma 1. *[2] Assume that (G, M, I) is a context and let $A \subseteq G$, $B \subseteq M$ and $A_j \subseteq G$ and $B_j \subseteq M$, for all $j \in J$. Then*

$$A \subseteq A'' \quad and \quad B \subseteq B'', \tag{3}$$

$$A_1 \subseteq A_2 \Rightarrow A_1' \supseteq A_2' \quad and \quad B_1 \subseteq B_2 \Rightarrow B_1' \supseteq B_2', \tag{4}$$

$$A' = A''' \quad and \quad B' = B''', \tag{5}$$

$$(\bigcup_{j \in J} A_j)' = \bigcap_{j \in J} A_j' \quad and \quad (\bigcup_{j \in J} B_j)' = \bigcap_{j \in J} B_j', \tag{6}$$

$$A \subseteq B' \;\Leftrightarrow\; A' \supseteq B. \tag{7}$$

□

Lemma 1 implies that for (A_1, B_1) and (A_2, B_2) in $\mathbf{B}(G, M, I)$

$$(A_1, B_1) \leq (A_2, B_2) \Leftrightarrow A_1 \subseteq A_2 \Leftrightarrow B_2 \subseteq B_1. \tag{8}$$

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). Again Lemma 1 is maim tools in proofs of these propositions.

Proposition 1. *[2] Let (G, M, I) be a context. Then $(\boldsymbol{B}(G, M, I), \leq)$ is a lattice with join and meet given by*

$$\bigvee_{j \in J} (A_j, B_j) = ((\bigcup_{j \in J} A_j)'', \bigcap_{j \in J} B_j), \tag{9}$$

$$\bigwedge_{j \in J} (A_j, B_j) = (\bigcap_{j \in J} A_j, (\bigcup_{j \in J} B_j)''); \tag{10}$$

□

A **counterexample** is an object without the target property. If the intent B of formal concept $\langle A, B \rangle$ is a subset of intent $\{c\}'$ of the counterexample c then cause $\langle A, B \rangle$ presents and the property absents for c, and the formal concept must be ignored. This procedure is called **counterexample forbidden condition** check.

3 Probabilistic model for accidental formal concepts

An attribute is called **essential**, if it occurs in some “real” cause. Assume that 2 “real” causes have no common attributes. Other attributes are called **accompanying**.

Essential attributes of counterexamples are absent, and essential attributes of training objects correspond to some of “real” cause in such a way that any of 2 “real” causes appear in some training object.

Denote the number of training objects by k, the number of counterexamples by m, and the number of accompanying attributes by n. It’s clear that accompanying attributes of training objects and counterexamples form a $(k + m) \times n$ binary matrix. It contains $N = (k + m) \cdot n$ bites. Accompanying attributes are generated by Bernoulli series of N tests.

Definition 1. *Bernoulli series of N tests is the probability distribution on $\{0,1\}^N$ with*

$$\boldsymbol{P}(x_1 = \delta_1, \ldots, x_N = \delta_N) = \prod_{j=1}^{N} p^{\delta_j} \cdot (1-p)^{1-\delta_j},$$

where $0 < p < 1$. The number $p > 0$ is called success probability $x_j = 1$ in test j.

Lemma 2. *Accident formal concept with extent of cardinality k can be replaced by Bernoulli series of n tests $\langle a_1, \ldots, a_n \rangle$ with success probability p^k of $a_j = 1$ added to the string of 0s corresponding to essential attributes.*

□

Lemma 2 easily implies the first negative result (insufficiency of lower bound on the size of extent to reject all accidental concepts):

Proposition 2. *Probability of an accidental formal concept with the extent of size k exceeds $\varepsilon > 0$ if the success probability $p \geq (-\ln(1-\varepsilon)/n)^{1/k}$.*

□

In analysis of the counterexample forbidden rejection procedure we restrict ourselves by $k = 2$.

Lemma 3. *The probability of appearance of accidental formal concept without m random counterexamples is*

$$\sum_{j=0}^{m} \binom{m}{j} \cdot (-1)^j \cdot (1 - p^2 + p^{2+j})^n.$$

Proof. Lemma 2 implies that accidental formal concept with 2 elements extent corresponds to Bernoulli series of n tests with success probability p^2.

Fix the cardinality l of intent of accidental formal concept. Then the probability equals to

$$\sum_{l=0}^{n} \binom{n}{l} \cdot \left(p^2\right)^l \cdot \left(1-p^2\right)^{n-l} \cdot \left(1-p^l\right)^m,$$

since p^l is the probability of deleting this formal concept by some fixed counterexample, hence $\left(1-p^l\right)^m$ is the probability that any (of m) counterexample doesn't delete the formal concept, and $\binom{n}{l} \cdot \left(p^2\right)^l \cdot \left(1-p^2\right)^{n-l}$ is the probability of appearance of an accidental formal concept with intent of cardinality l. But

$$\sum_{l=0}^{n} \binom{n}{l} \cdot \left(p^2\right)^l \cdot \left(1-p^2\right)^{n-l} \cdot \left(1-p^l\right)^m =$$

$$= \sum_{l=0}^{n} \binom{n}{l} \cdot \left(p^2\right)^l \cdot \left(1-p^2\right)^{n-l} \cdot \left(\sum_{j=0}^{m} \binom{m}{j} \cdot (-1)^j \cdot p^{lj} \right) =$$

$$= \sum_{j=0}^{m} \binom{m}{j} \cdot (-1)^j \cdot \left(\sum_{l=0}^{n} \binom{n}{l} \cdot \left(p^{2+j}\right)^l \cdot \left(1-p^2\right)^{n-l} \right) =$$

$$= \sum_{j=0}^{m} \binom{m}{j} \cdot (-1)^j \cdot (1-p^2+p^{2+j})^n.$$

□

Also the proof of Lemma 3 can be performed by direct application of "inclusion-exclusion" principle.

4 Main result

Lemma 4. *For any constant c the following holds*

$$\sum_{j=0}^{m} \binom{m}{j} \cdot (-1)^j \cdot c = 0.$$

This is proved by Newton's binomial formula. □

Lemma 5. *For $j > 0$*

$$\left(1 - \frac{a}{n} + \left(\frac{a}{n}\right)^{1+j/2}\right)^n - \left(1 - \frac{a}{n}\right)^n = a \cdot e^{-a} \cdot \left(\frac{a}{n}\right)^{j/2} + o(n^{-j/2})$$

holds as $n \to \infty$.

Proof.

$$\left[1-\frac{a}{n}+\left(\frac{a}{n}\right)^{1+j/2}\right]^n-\left(1-\frac{a}{n}\right)^n=$$

$$=\left[1-\frac{a}{n}\cdot\left(1-\frac{a}{n}\right)^{j/2}\right]^n-\left(1-\frac{a}{n}\right)^n=$$

$$=1-n\cdot\frac{a}{n}\cdot\left(1-\frac{a}{n}\right)^{j/2}+\ldots$$

$$+\frac{n\cdot\ldots\cdot(n-k)}{k!}\cdot\left(\frac{-a}{n}\right)^k\cdot\left(1-k\cdot\left(\frac{a}{n}\right)^{j/2}+o(n^{-j/2})\right)+\ldots$$

$$+\frac{n!}{n!}\cdot\left(\frac{-a}{n}\right)^n\cdot\left(1-n\cdot\left(\frac{a}{n}\right)^{j/2}+o(n^{-j/2})\right)-$$

$$-\left(1+\ldots+\frac{n\cdot\ldots\cdot(n-k)}{k!}\cdot\left(\frac{-a}{n}\right)^k+\ldots+\frac{n!}{n!}\cdot\left(\frac{-a}{n}\right)^n\right)=$$

$$=n\cdot\frac{a}{n}\cdot\left(\frac{a}{n}\right)^{j/2}-\ldots-\frac{n\cdot\ldots\cdot(n-k)}{k!}\cdot\left(\frac{-a}{n}\right)^k\cdot\left(k\cdot\left(\frac{a}{n}\right)^{j/2}+o(n^{-j/2})\right)-\ldots$$

$$-\frac{n!}{n!}\cdot\left(\frac{-a}{n}\right)^n\cdot\left(n\cdot\left(\frac{a}{n}\right)^{j/2}+o(n^{-j/2})\right)=$$

$$=a\cdot\left(\frac{a}{n}\right)^{j/2}+\ldots+a\cdot\frac{(-a)^{k-1}}{(k-1)!}\cdot\left(\frac{a}{n}\right)^{j/2}+\ldots+a\cdot\frac{(-a)^{n-1}}{(n-1)!}\cdot\left(\frac{a}{n}\right)^{j/2}+$$

$$+o(n^{-j/2})=a\cdot e^{-a}\cdot\left(\frac{a}{n}\right)^{j/2}+o(n^{-j/2})$$

□

Lemma 6. *Formula*

$$1-\left(1-\frac{a}{n}\right)^n=1-e^{-a}+o(1)$$

holds as $n\to\infty$.

□

Theorem 1. *If number n of the random attributes tends to infinity, the probability of success equals to $\sqrt{\frac{a}{n}}$, and there are $m=b\cdot\sqrt{n}$ counterexamples, then the probability of an accidental formal concept with 2 elements extent and without any counterexample is $1-e^{-a}-a\cdot e^{-a}\cdot[1-e^{-b\cdot\sqrt{a}}]$ in the limit.*

Proof. Lemma 3 reduces the theorem to estimation of

$$\sum_{j=0}^{m}\binom{m}{j}\cdot(-1)^j\cdot(1-p^2+p^{2+j}))^n.$$

Substitute $p = \sqrt{\frac{a}{n}}$, then Lemma 4 implies

$$\sum_{j=0}^{m} \binom{m}{j} \cdot (-1)^j \cdot \left(1 - \frac{a}{n} + \left(\frac{a}{n}\right)^{1+j/2}\right)^n =$$
$$= \sum_{j=0}^{m} \binom{m}{j} \cdot (-1)^j \cdot \left[\left(1 - \frac{a}{n} + \left(\frac{a}{n}\right)^{1+j/2}\right)^n - \left(1 - \frac{a}{n}\right)^n\right].$$

Lemma 5 and Lemma 6 imply

$$\sum_{j=0}^{m} \binom{m}{j} \cdot (-1)^j \cdot \left[\left(1 - \frac{a}{n} + \left(\frac{a}{n}\right)^{1+j/2}\right)^n - \left(1 - \frac{a}{n}\right)^n\right] =$$
$$= \left[1 - \left(1 - \frac{a}{n}\right)^n\right] +$$
$$= \sum_{j=1}^{m} \binom{m}{j} \cdot (-1)^j \cdot \left[\left(1 - \frac{a}{n} + \left(\frac{a}{n}\right)^{1+j/2}\right)^n - \left(1 - \frac{a}{n}\right)^n\right] =$$
$$= \left[1 - e^{-a} + o(1)\right] + \sum_{j=1}^{m} \binom{m}{j} \cdot (-1)^j \cdot \left[a \cdot e^{-a} \cdot \left(\frac{a}{n}\right)^{j/2} + o(n^{-j/2})\right].$$

Then

$$\sum_{j=1}^{m} \binom{m}{j} \cdot (-1)^j \cdot \left[a \cdot e^{-a} \cdot \left(\frac{a}{n}\right)^{j/2} + o(n^{-j/2})\right] =$$
$$= \sum_{j=1}^{m} \binom{m}{j} \cdot (-1)^j \cdot \left[a \cdot e^{-a} \cdot \left(\frac{b \cdot \sqrt{a}}{m}\right)^{j} + o(m^{-j})\right] =$$
$$= a \cdot e^{-a} \cdot \left[\left(1 - \frac{b \cdot \sqrt{a}}{m}\right)^m - 1\right] + o(1) =$$
$$= a \cdot e^{-a} \cdot \left[e^{-b \cdot \sqrt{a}} - 1\right] + o(1),$$

finishes the proof, since the last equation holds by Lemma 6. □

To prove that

$$1 - e^{-a} - a \cdot e^{-a} \cdot \left[1 - e^{-b \cdot \sqrt{a}}\right] > 0$$

for any $a > 0$ and $b > 0$ use the following reasoning

$$1 - e^{-a} - a \cdot e^{-a} \cdot \left[1 - e^{-b \cdot \sqrt{a}}\right] \geq$$
$$\geq 1 - e^{-a} - a \cdot e^{-a} = e^{-a} \cdot (e^a - 1 - a) \geq e^{-a} \cdot \frac{a^2}{2!} > 0.$$

Our model of random counterexamples is incorrect if the number m of counterexamples is comparable with 2^n, since in this case many pairs of counterexamples coincide. But this situation is also inpractical from data analysts' point of view, since majority of attrubutes absents, when real data are coded by binary attributes. Hence formal context has a small fraction of 1s.

Conclusions

In this paper we demonstrate that Formal Concept Analyses has the overfitting shortcoming similar to other machine learning methods. JSM method [1] has the same drawback. The author [5] developed probabilistic approach to JSM, FCA, and related formalisms. The inspiration for the study was the overfitting phenonenon of classical methods.

Recently Tatiana Makhalova and Prof. Sergey O. Kuznetsov [4] develop an interesting approach to compare of overfitting phenomena between different machine learning algorithms by means of FCA. They suppose that formal context contains errors of prediction of the target property of objects by different algorithms (attributes). Hence Formal Concept Analysis is a means to estimate the overfitting of algorithms when set of all objects randomly separated into two parts (to learn and to test). There are some theoretical results based on "weak probability axiomatization" proposed by Prof. Konstantin V. Vorontsov [6].

In the present paper we discuss the overfitting of FCA itself considered as a machine learning method. We consider "accidental" formal concepts appearence as overfitting because it occurs during computing of all formal concepts (in attempt to use full available information ignoring a possibility of "accidental" coincidence between unessential attributes).

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

References

1. Finn, V.K. The Synthesis of Cognitive Procedures and the Problem of Induction. *Automatic Documentation and Mathematical Linguistics*, Vol. 43. – 2009. – pp. 149–195
2. Ganter, Bernard and Wille, Rudolf. *Formal Concept Analysis: Mathematical Foundations*, Springer-Verlag, 1999
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. Makhalova, T. and Kuznetsov, S.O. On Overfitting of Classifiers Making a Lattice *Proc. 14th International Conference on Formal Concept Analysis (ICFCA 2017), Lecture Notes in Artificial Intelligence (LNAI)*, vol. 10308. – Springer. – 2017. – pp. 184–197.
5. Vinogradov, D.V. VKF-method of Hypotheses Generation. *Communications in Computer and Information Science*, Vol. 436. – 2014. – pp. 237–248
6. Vorontsov, K.V. Combinatorial Probability and Generalization Bounds Tightness. *Pattern Recognition and Image Analysis*, Vol. 18. – no. 2. – 2008. – pp. 243–259