Formalization of "Natural" Classification and "Natural" Concepts by Probabilistic Generalization of Formal Concepts Analysis

Evgenii E. Vityaev[a], Vladislav Degtiarev[b], Bayar Pak[b] and Yuri Meister[b]

[a]*Sobolev Institute of Mathematic, Novosibirsk, Russia*
[b]*Novosibirsk State University, Novosibirsk, Russia*

Abstract

In the previous works, a probabilistic generalization of the formal concepts analysis was developed. This generalization is induced by the problem of formal concepts determining under noise conditions, when the lattice of formal concepts exponentially grows. In this paper, probabilistic formal concepts with negation are determined, as well as a statistical method for detecting these probabilistic formal concepts. The purpose of this paper is to show that probabilistic formal concepts have a deeper meaning. It is argued that probabilistic formal concepts formalize the "natural" concepts described in cognitive sciences by "causal models", which are characterized by a highly correlated structure of attributes. The same structure is specific for the "natural" classification of objects of the external world. The definition of "natural" classification given by J. Stuart Mill is fairly accurately formalized by probabilistic formal concepts.

1. Introduction

In previous works [Vityaev et al. 2012], [Vityaev et al. 2015], [Martynovich et al. 2016], a probabilistic generalization of the Formal Concepts Analysis (FCA) was developed. This generalization solves the problem of the formal concepts discovery in noise conditions associated with the exponential growth of the lattice of formal concepts in the presence of noise. This issue is discussed in a more detail in section 2.

Probabilistic generalization of formal concepts was obtained by generalizing the definition of formal concepts as fix-points of implication. For this purpose, the probabilistic measure on the objects of context and the most specific probabilistic implication were defined. Consistency of fix-points based on maximally specific probabilistic implications is proved and then probabilistic formal concepts are defined as fix-points of predictions based on maximally specific probabilistic implications. Figuratively speaking, the obtained probabilistic formal concepts not display data in all details as in photographs, but give an "idealized" representation of the data.

In this work, probabilistic formal concepts with negation are defined, as well as a statistical method of detecting these probabilistic formal concepts based on the works [Vityaev et al. 2014], [Vityaev et al. 2015], [Martynovich et al. 2016].

The purpose of this work is to show that probabilistic formal concepts have deeper meaning than their restoration in noise conditions. In [Vityaev et al. 2014], it was shown that the fixpoints of anticipation underlying probabilistic formal concepts model the process of perception in exact accordance with the psychology of perception described in [Smirnov 1985], [Naiser 1981].

In addition, this paper shows that probabilistic formal concepts simulate "natural" concepts explored in cognitive sciences. "Natural" concepts are concepts about directly perceived objects, not abstract objects like production, plant and object. The following section, based on research in cognitive sciences, provides principles of categorization and description of "natural" concepts. It follows that "natural" concepts reflect a highly correlated structure of features of objects of the external world, which can be represented by "causal models". It is argued that causal models are rather precisely modeled by probabilistic formal concepts.

The highly correlated structure of features of objects of the external world is also manifested in the "natural" classification (see section 3). J. Stuart Mill has observed that "natural" classes of animals or plants are described by a potentially infinite number of properties [Mill 2011]. The naturalists who built the "natural" classifications also noted that the more similar the compared objects are in the greater number of significant features, the more likely their similarity in other features [Rutkovsky 1884], and that the construction of the "natural" classification consists in "indication" - from the infinite number of features we need to move to a limited number of features that would replace all other features [Smirnov 1938]. This means that in the "natural" classes the features are strongly correlated, for example, if classes 128 and features are binary, the independent "indicator" features among them will be only 7 features, because 2^7 =128, and the remaining features, which can be tens and hundreds (denoted by N) can be predicted by these 7 features, which means the presence of N-7 regularities. Since the "indicator" features can be different 7 features from N, the choice of which is equal to C_N^7, the number of regularities that predict non-indicator features based on the indicator features is not much less than $(N - 7) \cdot C_N^7$. Highly correlated structure of features is well detected by the probabilistic formal concepts; moreover, the following statement by J.S. Mill almost precisely defines the meaning of probabilistic formal concepts: "Natural groups ... are defined by features ... However, at the same time not only features which are certainly common for all the objects included into the group but the whole set of those features is taken into account, of which all are found in the majority of these objects and the majority - in all" [Mill 2011].

2. "Natural" concepts

In the works of Eleanor Rosch [Rosch 1973, 1978], [Rosch et al. 1975, 1976, 1978] the principles of categorization of "natural" categories [Rosch 1973, 1978] are formulated on the basis of the conducted experiments. Directly perceived objects (basic objects) – information rich bundles of observed and functional properties that form a natural partition that creates a categorization. These bundles form "prototypes" of objects classes: Categories can be viewed in terms

of their clear cases if the perceiver places emphasis on the correlational structure of perceived attributes ... By prototypes of categories we have generally meant the clearest cases of category membership [Rosch et al. 1978], [Rosch 1978]. Rosch and Mervis (1975) have shown that the more prototypical of a category a member is rated, the more attributes it has in common with other members of the category and the fewer attributes in common with members of the contrasting categories [Rosch et al. 1975, 1976].

Later on, the theory of "natural" concepts of Eleanor Rosch was called a "prototype" theory. Its main features are described in the following way: The prototype view (or probabilistic view) keeps the attractive assumption that there is some underlying common set of features for category members but relaxes the requirement that every member have all the features [Ross et al. 2008].

Further research has found that models based on features, similarities and prototypes are not sufficient to describe classes. Theoretical, causal and ontological knowledge relating to class objects must be taken into account. For example, people not only know that birds have wings, can fly and nest in trees, but also that birds nest in trees because they can fly and fly because they have wings.

Based on this research, Bob Rehder, in his work "Categorization as causal reasoning", formulated a causal-model theory: people's intuitive theories about categories of objects consist of a model of the category in which both a category's features and the causal mechanisms among those features are explicitly represented [Rehder 2003a].

In the theory of causal models, the relationship of an object to a category is no longer based on a set of features and proximity based on features, but on the similarity of the generating causal mechanism: Specifically, a to-be-classified object is considered a category member to the extent that its features were likely to have been generated by the category's causal laws, such that combinations of features that are likely to be produced by a category's causal mechanisms are viewed as good category members and those unlikely to be produced by those mechanisms are viewed as poor category members [Rehder 2003a].

Some researchers have used Bayesian networks or causal models to represent causal knowledge [Cheng 1997], [Gopnik et al. 2004], [Griffiths et al. 2009]. However, these models cannot model cyclical causal relationships because Bayesian networks do not support cycles.

In [Rehder et al. 2011], Bob Rehder proposed a model of causal cycles based on the "disclosure" of causal graphical models (CGMs). The "disclosure" is done by creating a Bayesian network that deploys over time. Such a network is already quite well suited for modeling "natural" concepts reflecting the highly correlated structure of the outside world. However, it does not explicitly formalize causal link cycles.

The model we propose further is directly based on cyclic causal links represented by probabilistic formal concepts.

3. "Natural" classification

The first sufficiently detailed philosophical analysis of the "natural" classification belongs to J.St. Mill [Mill 2011]. This analysis revealed all the main properties of the "natural" classifications, which were later confirmed by naturalists who built the "natural" classifications.

According to J.St. Mill [Mill 2011], "artificial" classifications differ from "natural" classifications in that they can be based on any one or more attributes, so that different classes differ only in that they include objects with different values of these attributes. But if we consider classes of "animals" or "plants", they differ in such a large (potentially infinite) number of properties that they cannot be enumerated" [Mill 2011].

J.St. Mill gives the following definition of the "natural" classification: it is such a classification which unites objects into groups concerning which it is possible to express the greatest number of general propositions and it is based on such properties which serve as reasons for many others or at least make up their true attributes. It also defines the notion of the "image" of the class, which is the forerunner of "natural" concepts: "our concept of class – the image by which this class is represented in our mind – is the concept of a certain specimen that possesses all the features of this class ... to the highest degree".

The properties of "natural" classes according to J. St. Mill were confirmed by naturalists. For example, Wavell W.: "The more general statements about objects give an opportunity to make classification, the more natural it is" [Mayen et al. 1976]. L. Rutkovsky writes about the inexhaustible number of general properties of "natural" classes. [Rutkovsky 1884]: "The greater the number of essential features similar to comparable objects, the more likely their similarity in other respects". A similar statement is made by E.S. Smirnov: "The taxonomic problem lies in 'indication': from an infinite number of features we need to move to a limited number of them, which would replace all other features" [Smirnov 1985].

As a result, the problem of defining "natural" classifications has been formulated and is still discussed in the literature [Zabrodin 1981], [The Nature of Classification 2013]. However, from our point of view, the formalization of the "natural" classification is not yet sufficiently adequate.

4. The problem of formal concepts discovery in noise conditions

Let us consider the known problem of formal concepts analysis [Ganter 2003]: standard methods of concept lattice construction on noisy data turn out to be unproductive, because they are designed for accurate data, and in the presence of noise (data distortions) the concept lattice turns out to be filled with noisy clones of initial concept classes [Klimushkin et al. 2010], [Frequent Itemset Mining 2004]. Attempts to solve this problem "on the forehead" do not look promising [Klimushkin et al. 2010], [Kuznetsov 2007]. The main idea of these works is to take a full lattice of concepts, and get rid of noise, "side" concepts by filtering. A widespread evaluation of the "side" concept is the Stability Index [Kuznetsov 2007], [Buzmakov 2014]. The Stability Index and other approaches based on the filtration of lattices have significant disadvantage – the filtered lattices of concepts often have exponential sizes [Emilion et al. 2009], which is unacceptable for large contexts.

5. Basics of formal concepts analysis

Let's start with the main definitions of formal concepts analysis. The majority of definitions here start with classical works on FCA [Ganter 2003], [Ganter et al. 1999], others are taken from [Ganter et al. 2004].

Definition 1. Formal context is a triple (G, M, I), where G and M – arbitrary sets of objects and attributes, and $I \subseteq G \times M$ – binary relation, expressing the belonging of an object to an attribute.

In formal context, the operators of the derivative play a key role.

Definition 2. $A \subseteq G, B \subseteq M$. Then:

1. $A^{\uparrow} = \{m \in M \mid \forall g \in A, (g, m) \in I\}$
2. $B^{\downarrow} = \{g \in G \mid \forall m \in B, (g, m) \in I\}$

Operators of a derivative link a subset of objects and attributes of a context. In the future we will also refer to both operators as a derivative with a single symbol $\prime$.

Definition 3. A pair (A, B) – is a formal concept, if $A^{\uparrow} = B$ and $B^{\downarrow} = A$.

Definition 4. An implication is a pair of attribute sets (B, C), $B, C \subseteq M$ that is recorded as $B \rightarrow C$. Implication $B \rightarrow C$ is truth on the context $K = (G, M, I)$, if $\forall g \in G(B \nsubseteq g' \vee C \subseteq g')$. A set of all true implications on context K well be denoted as $Imp(K)$.

Definition 5. For any set L of implications we define a direct inference operator f_L that add all possible conclusions of implications to the original set of attributes X:

$$f_L(X) = X \cup \{C \mid B \subseteq X, \ B \rightarrow C \in L\}$$

Theorem 1. [Ganter et al. 1999] For any subset $B \subseteq M$, $f_{Imp(K)}(B) = B \Longleftrightarrow B'' = B$.

Let us illustrate Theorem 1 in the following simple context.

Table 1
[Ganter et al. 1999] Very simple formal context K_0

m_1	m_2	m_3
1	1	0
1	1	1
0	1	1
0	0	0

We can reformulate the task of constructing a concept lattice in terms of implication and output operator. Looking at Table 1, it is not difficult to notice that m_1 and m_3 attributes completely define the object class. In fact, m_1 infer m_2, and the same is true for m_3, from these it follows that $m_1 \rightarrow m_2$ and $m_3 \rightarrow m_2$. The sets $\{m_1, m_2\}, \{m_2, m_3\}$ and $\{m_1, m_2, m_3\}$ will be formal concepts. According to the theorem, they will also be fix-points of the direct inference operator. For example,

$$\{m_1\} \xrightarrow{f_{Imp(K)}} \{m_1, m_2\} \xrightarrow{f_{Imp(K)}} \{m_1, m_2\}.$$

It is therefore $\{m_1, m_2\}$ both a fix-point and a formal concept.

Now let's add some noise to the original context K_0 from Table 1. In order to keep the noise level at a fairly low level, let's also expand the context by adding duplicates of objects:

Table 2
K_0 with duplicates.

m_1	m_2	m_3
1	1	0
1	1	1
0	1	1
1	1	0
1	1	1
0	1	1
1	1	0
1	1	1
0	1	1
0	0	0

Table 3
K_{noise} a little noise

m_1	m_2	m_3
0	1	0
1	1	1
0	1	1
1	1	0
1	0	1
0	1	1
1	1	0
1	1	1
0	1	1
0	0	0

Every small change made to the context will have a dramatic impact on the conceptual lattice. The first context in Table 2 is equivalent to the original context K_0 and generates the same conceptual lattice. However, the lattice with noise of the second context in Table 3 will contain some fictitious notions {m₂} and {m₁, m₃} generated by noise, which will also be formal notions. The number of fictitious notions will increase with the noise level, and the qualitative estimation of speed is exponential [Emilion et al. 2009].

6. Probabilistic logic in a formal context

To avoid the exponential growth of the lattice, a stability index is introduced [Klimushkin et al. 2010], [Kuznetsov 2007],[Buzmakov 2014] to assess the degree of stability of the concept and to select only sustainable concepts.

We will present a fundamentally different way of the lattice construction, based on probabilistic implications. For this purpose, we redefine the context within the framework of logic. Further we will consider only finite contexts.

Definition 6. For the finite context $K = (G, M, I)$ we define a signature Ω_K that contains predicate symbols m(x) for each $m \in M$. For the signature Ω_K and the K context, considered as a model, we define the interpretation of predicate symbols as follows $K \vDash m(x) \Longleftrightarrow (x, m) \in I$.

Definition 7. For the signature Ω_K let's define the following variant of the first order logic:

1. X_K – set of variables;
2. At_K – set of atomic formulas (atoms) $m(x)$ where $m \in \Omega_K$ and $x \in X_K$;
3. L_K – set of literals, including all atoms $m(t)$ and their negations $\neg m(t)$;
4. Φ_K – set of formulas defined inductively: every atom is a formula, for any $\Phi, \Psi \in \Phi_K$ following syntactic constructions $\Phi \wedge \Psi, \ \Phi \vee \Psi, \ \Phi \longrightarrow \Psi, \ \neg \Phi$ are formulas as well.

For a set of literals $L \subseteq L_K$ we define their conjunctions as $\wedge L = \underset{P \in L}{\wedge} P$. In the same way $\neg L = \{\neg P \mid P \in L\}$.

Definition 8. A single element set $\{g\}$, together with signature Ω_K predicates, forms a model K_g of this object. The fact of truth of a formula ϕ on an object model K_g is written as $g \vDash \phi \Longleftrightarrow K_g \vDash \phi$.

Definition 9. Let us define a probabilistic measure μ on the set G, in the Kolmogorov sense. Then we can define a probabilistic measure on the set of formulas as:

$$\nu : \Phi_K \longrightarrow [0, 1], \ \nu(\phi) = \mu(\{g \mid g \vDash \phi\}).$$

Let us assume that the context has no statistically insignificant objects such as $\mu(\{g\}) = 0, \ g \in G$.

Definition 10. The set of literals M will be called ν-joint if $\nu(\wedge M) > 0$.

In the future, the compatibility of the set of literals will be considered within a probabilistic measure of a context, and in the absence of ambiguity the symbol of the measure will be omitted.

Let us consider some subset of atoms $L \subseteq At(K)$. The formula $m_1 \wedge m_2 \ldots \wedge m_k \longrightarrow m = \wedge\{m_i\} \longrightarrow m$ defines implication $(\{m_i\}, \{m\})$ on the context. The presence of a probabilistic measure allows to determine the probabilistic implication.

Let's define the rules on the context.

Definition 11. Let $\{H_1, H_2, \ldots, H_k, C\} \in L_K, \ C \notin \{H_1, H_2, \ldots H_k\}, \ k \geq 0$.

1. The rule $R = (H_1, H_2., H_k \longrightarrow C)$ is called implication $(H_1 \wedge H_2 \ldots \wedge H_k \longrightarrow C)$;
2. The premise $R^{\leftarrow}$ of rule R is a set of literals $\{H_1, H_2, \ldots, H_k\}$;
3. The conclusion of the rule is $R^{\rightarrow} = C$;
4. The length of the rule is called the premise power $|R^{\leftarrow}|$;

5. If $R_1^{\leftarrow} = R_2^{\leftarrow}$ and $R_1^{\rightarrow} = R_2^{\rightarrow}$, then $R_1 = R_2$.

Definition 12. The probability η of the rule R is the value

$$\eta(R) = v(R^{\rightarrow}|R^{\leftarrow}) = \frac{v(R^{\leftarrow} \wedge R^{\rightarrow})}{v(R^{\leftarrow})}.$$

If the rule denominator $v(\wedge R^{\leftarrow})$ is equal to 0, then the probability is not defined.

7. Rule classes and consistency of predictions

The following sections are a combination of works [Ganter et al. 1999], [Vityayev 1992], [Vityaev 2006], [Vityayev et al. 2012], [Vityaev et al. 2015], [Martynovich et al. 2016].

Definition 13. Let us define a prediction operator by a set of rules $\mathcal{R}$ as

$$\Pi_{\mathcal{R}}(L) = L \cup \{C \mid \exists R \in \mathcal{R} : R^{\leftarrow} \subseteq L, R^{\rightarrow} = C\}.$$

Definition 14. A closure $\bar{L}$ of the set of literals L is the smallest fix-point of a prediction operator containing L

$$\bar{L} = \Pi_{\mathcal{R}}(\bar{L}) = \Pi_{\mathcal{R}}^{\infty}(L) = \bigcup_{k \in \mathbb{N}} \Pi_{\mathcal{R}}^k(L).$$

Definition 15. Rule R_1 is a subrule of rule R_2 if $R_1^{\rightarrow} = R_2^{\rightarrow}$, $R_1^{\leftarrow} \subset R_2^{\leftarrow}$. We denote this fact as $R_1 \sqsubset R_2$.

Definition 16. Rule R_1 specifies rule R_2, or $R_1 > R_2$ if $R_2 \sqsubset R_1$ and $\eta(R_1) > \eta(R_2)$.

Example 2. A rule $m_1 \rightarrow m_2$ from the context K_{noise} (Table 3) has only one, unconditional sub-rule $(\emptyset \rightarrow m_2) \sqsubset (m_1 \rightarrow m_2)$ but it is wrong that the first rule specifies the second $\eta(\emptyset \rightarrow m_2) = \frac{8}{10} = \frac{4}{5} = \eta(m_1 \rightarrow m_2)$. However, $(m_3 \rightarrow m_2) > (\emptyset \rightarrow m_2)$ since $\eta(\emptyset \rightarrow m_2) = \frac{8}{10} < \frac{5}{6} = \eta(m_3 \rightarrow m_2)$.

To achieve consistency and resistance to noise, some restrictions on the rule classes used will be required [Vityayev 1992], [Speransky 2013].

Class M_1 defined below contains only those rules whose conditional probability is strictly greater than the probability of an unconditional rule.

Definition 17. $R \in M_1(C) \iff \eta(R) > v(R^{\rightarrow}), R^{\rightarrow} = C, R^{\leftarrow} = \emptyset$.

The next class M_2 requires the property of maximum specificity – the impossibility to increase the conditional probability by refining the rule [Vityaev 2006].

Definition 18. $R \in M_2(C) \iff (R \in M_1(C)) \& (R \sqsubset \tilde{R} \implies \eta(\tilde{R}) \le \eta(R))$.

We have proved that maximally specific rules of M2 solve the problem of statistical ambiguity [Vityaev 2006], [Vityaev et al. 2015].

Example 3. A rule $(m_3 \rightarrow m_2)$ can be specified to the rule $(\neg m_1 \wedge m_3 \rightarrow m_2)$, because of $\eta(m_3 \rightarrow m_2) = \frac{5}{6} < 1 = \frac{3}{3} = \eta(\neg m_1 \wedge m_3 \rightarrow m_2)$. The last rule meets all the requirements of class M_2, and therefore $(\neg m_1 \wedge m_3 \rightarrow m_2) \in M_2(m_2)$.

The next class Imp contains all precise implication and corresponds to the set $Imp(K)$ of definition 4.

Definition 19. $R \in Imp(C) \iff (R^{\rightarrow} = C)$.

Definition 20. Define classes M_1, M_2, *Imp* by joining over all literals

$$M_1 = \underset{C \in Lit(K)}{\cup} M_1(C); M_2 = \underset{C \in Lit(K)}{\cup} M_2(C); Imp = \underset{C \in Lit(K)}{\cup} Imp(C).$$

Definition 21. A set of rules $\mathcal{R}$ will be called exact, if $Imp \subset \mathcal{R}$.

Definition 22. By the system of rules we will call any subset $\mathcal{R} \subseteq M_2$.

A set of literals L is consistent if it does not simultaneously contain some atom C and its negation $\neg C$.

Theorem 2. (Joint predictions [Vityaev et al. 2015]). If L is joint, then $\Pi_{\mathcal{R}}(L)$ is also joint and consistent for any system of rules $\mathcal{R} \subseteq M_2$.

8. Probabilistic formal concepts

Fix-points of the prediction operator are obvious candidates for the role of the content of probabilistic concepts. To determine the volume of notions let's refer to [Kuznetsov 2007],[Buzmakov 2014]. The idea is to use all possible prototypes of the closing operator, i.e. all sets of M such that $\Pi_{\mathcal{R}}(M) = B$ to compose their derivatives into a single volume of notion A. This will allow to restore the initial correspondence of objects-concept using the prediction operator and also to include all objects of the same class into a single volume of the notion.

Definition 23. A probabilistic formal concept on the context K is a pair (A,B) that meets the conditions:

$$\Pi_{\mathcal{R}}(B) = B, \ A = \underset{\Pi_{\mathcal{R}}(C)=B}{\cup} C', \mathcal{R} \subseteq M_2 \ (def. \ 22).$$

.

Example 4. Let's consider a $K_{squares}$ context consisting of two non-intersecting squares that give two independent formal concepts. We will change the values of some cells to make the task a bit more complicated.

Table 4
Context of Ksquares with 2 notions and superimposed noise.

G	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8
g_1	1	1	1	1	1	0	0	0
g_2	1	1	1	1	0	0	0	0
g_3	1	1	1	1	0	0	0	0
g_4	1	1	1	1	0	0	0	0
g_5	0	0	0	0	1	1	1	1
g_6	0	0	0	0	1	1	0	1
g_7	0	0	0	0	1	1	1	1
g_8	0	0	0	0	1	1	1	1

Note that the most specific rules of M_2 will be $(m_{i=1...4} \rightarrow \neg m_{j=5...8})$, but not the rule $m_5 \rightarrow \neg m_{i=1...4}$ for which you can find the following clarification $\eta(m_5 \wedge m_6 \rightarrow \neg m_1) = 1 > \frac{4}{5} =$

$\eta(m_5 \wedge \rightarrow \neg m_1)$. This example shows how the noise is processed and invested in probability in the process of rule refinement.

To calculate a concept containing the feature m_1, we first calculate its closing $\Pi_{\mathcal{R}}(\{m_1\}) = \{m_1, m_2, m_3, m_4, \neg m_5, \neg m_6, \neg m_7, \neg m_8\}$. Then all the objects whose closing turns out to be exactly the same are gathered in a single volume of the concept $\{g_2, g_3, g_4\}$.

To distinguish probabilistic formal concepts from classical formal concepts on the context K, we will call the latter simply formal concepts. The definition of set A is also based on the following theorem linking probabilistic and simple formal concepts on context K.

Theorem 3. Let K be a formal context.

1. If (A,B) is a formal concept in K, then there is a probabilistic formal concept (N,M) in the same context, such as $A \subseteq N$, .

2. If (N,M) is a probabilistic formal concept on K, then there is a family $\mathcal{C}$ of formal concepts in the same context, such that

$$\forall(A, B) \in \mathcal{C} \ (\Pi_{\mathcal{R}}(B) = M),$$
$$N = \bigcup_{(A,B)\in\mathcal{C}} A$$

9. Statistical method for detection of probabilistic formal concepts

The procedure for obtaining all M_2 rules by exhaustive search is exponential, because the $O(2^{|M|})$ candidate rules (each atom or its negation can be included in the rule) are evaluated at $|G|$. Of course, this estimation is rather rough and the procedure can be improved. Instead a more radical simplification of the problem by finding a statistical approximation of M_2-rules will be defined.

Definition 24. R is a probabilistic law, if $(\tilde{R} \sqsubset R) \implies (\tilde{R} < R)$.

Definition 25. A rule $\tilde{R}$ is semantically derived from the rule R, we denote as $R \triangleright \tilde{R}$, if R and $\tilde{R}$ – probabilistic law and $\tilde{R} > R$.

Definition 26. The probabilistic law R will be called the strongest if there is no probabilistic law $\tilde{R}$ such as $(\tilde{R} > R)$.

Definition 27. We define the Semantic Probabilistic Inference (SPI) of some strongest probabilistic law R_m as such a sequence of probabilistic laws $R_0 \triangleright R_1 \triangleright R_2. \triangleright R_m$ that $R_0^{\leftarrow} = \varnothing$ and R_m is the strongest probabilistic law.

Let us denote by $\mathcal{R}$ the set of all the strongest probabilistic laws obtained by various semantic probabilistic inferences. Specification check R1 > R2 requires an inequality $\eta(R_1) > \eta(R_2)$ check.

In general, from a practical point of view, we cannot assume that a probability measure is defined as in definition 9, so we will consider it unknown. The probabilistic inequalities used in the SPI definition we will determine using some statistical criterion, e.g. the exact Fisher independence criterion with a confidence level α. The resulting set of rules $\mathcal{R}_\alpha$ will already make contradicting predictions. Therefore, to approximate the prediction operator $\Pi_{\mathcal{R}}(L)$, it is necessary to introduce an additional criterion of mutual consistency of predictions by rules $\mathcal{R}_\alpha$ on the set L.

Definition 29. The rule $R \in \mathcal{R}_\alpha$ is confirmed on the set of letters L, if $R^\leftarrow \subset L$ and $R^\rightarrow \in L$. In this case, we write that $R \in Sat(L) \subseteq \mathcal{R}_\alpha$.

Definition 30. The rule $R \in \mathcal{R}_\alpha$ is refuted on the set of letters L, if $R^\leftarrow \subset L$ and $R^\rightarrow \in \neg L$. In this case, we write that $R \in Fal(L) \subseteq \mathcal{R}_\alpha$.

The difference between the quantity and quality of rules confirmed and refuted on the concept-hypothesis L is the determinant for adding or deleting literals to the content of the concept.

Definition 31. The consistency criterion for the set L is the value

$$Int(L) = \sum_{R \in Sat(L)} \gamma(R) - \sum_{R \in Fal(L)} \gamma(R).$$

The choice of rule evaluation γ may depend on additional task specifics. In the experiments below, we are guided by Shannon's considerations as

$$\gamma(R) = -log(1 + \epsilon - \eta(R)).$$

Definition 32. Define a consistent prediction operator $Y(L)$ that changes a set of literals L by one element so as to strictly increase the consistency criterion:

1. For all $G \in L_K \backslash L$ calculate the effect of adding G: $\triangle^+ = Int(L \cup \{G\}) - Int(L)$;
2. For all $G \in L$ calculate the effect of removing G: $\triangle^- = Int(L \backslash \{G\}) - Int(L)$;
3. The operator $Y(L)$ adds a letter G to the L set if $\triangle^+ > 0$ and $\triangle^+ > \triangle^-$; the operator delete a letter G from L if $\triangle^- > 0$ and $\triangle^- > \triangle^+$. If $\triangle^- = \triangle^+$ and $\triangle^- > 0$, the operator will delete literal G;
4. If $\triangle^+ \le 0$ and $\triangle^- \le 0$, the operator returns L and we have a fix-point.

Definition 33. A probabilistic formal concept discovered by a statistical method is a probabilistic formal concept in which $\mathcal{R}_\alpha$ is a set of all strongest probabilistic laws (def. 26) obtained by various semantic probabilistic inferences with a confidence level α in the context K.

10. Experiments in the detection of probabilistic formal concepts by statistical method

Let us illustrate the detection of probabilistic formal concepts by statistical method on the example of classes (concepts) detection of digits. Let's encode the digits with 24 features, located as shown in fig. 1. Each feature has 7 values (see fig. 2). Some lines of numbers are encoded twice, for example, the vertical line at number 4 and 1 is encoded both by the vertical right line (feature 3) in cells 3,7,11,15,19,23 and by the vertical left line (feature 5) in cells 4,8,12,16,20,24.

Two experiments were conducted for two different contexts. The first contained 30 copies of each digit (360 in total) randomly mixed together. The program does not know where the digit is and it treats them as different images. The second context contained 10 copies of each of the 10 digits (without repeating sixes and nines, a second copy was taken) and 300 random images in which each of the 24 features took a random value from the seven available. All numbers and images were also randomly mixed. In each context a corresponding set $\mathcal{R}_\alpha$ of strongest
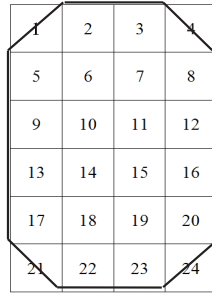
Figure 1: Digits encoding
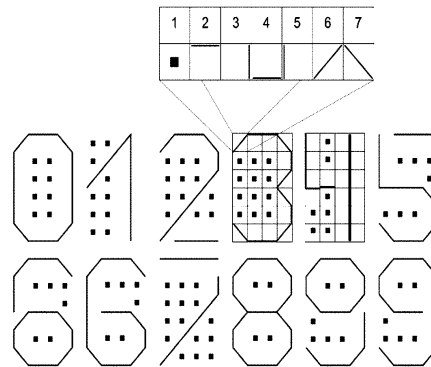
Figure 2: Digits encoding

probabilistic laws with α = 0.05 for the Fisher criterion were found. In the first experiment 66160 regularities with conditional probability greater than 0.9 and α= 0.05 were found. In the second experiment 629494 regularities with conditional probability greater than 0.9 and α= 0.05 were found.

In the first context, the method found as classes all 12 digits – they were fix-points, as well as two fix-points for 6 and 9, which did not contain, respectively, signs 13 and 12, with which two copies of digits differed. In the second context, the method identified 10 digits as fix-points.

In both experiments, the stability of the probabilistic formal concepts found was tested with regard to the lack of information on any feature of a digit. It was checked that if any digit described by the set of literals L removes one of the literals m_i or $\neg m_i$, i.e., removes one of the values of the feature, the digit will still be assigned to the same class and the unknown feature will be correctly identified and added to L, i.e. $\Pi_{\mathcal{R}}(L \backslash m_i) = L$. This means that the removed value of the feature will be correctly predicted by the regularities from $\mathcal{R}_\alpha$ without contradictions on the part of the logical relationship of this feature with other features. In the first experiment, for all 360 digits, a feature was accidentally removed and then restored by the prediction operator. In the second experiment, the removed features were restored in 96% of cases.

Acknowledgments

References

[Vityaev 1992] Vityaev E.E. Semantic approach to creation of knowledge bases. Se-mantic Probabilistic Conclusion of the Best Prediction PROLOGRAMS by Probabilistic Data Model. // Logic and Semantic Programming (Computing Systems, vol. 146), Novosibirsk, 1992, pp. 19-49.

[Vityaev et al. 2014] Vityaev E.E., Neupokoev N.V. Formal model of perception and image as fix-point of anticipation. Approaches to Thinking Modeling. (Collected under the editorship of V.G. Redko, Ph.D.-m.D.). URSS Editorial-Al, Moscow, 2014, pp. 155-172.

[Vityaev et al. 2015] Vityaev E.E., Martynovich V.V. Formalization of "natural" classification and systematics through fixed prediction points // Siberian Electronic Mathematical Reports, Vol. 12, Institute of Mathematics, Moscow, Russia. Siberian Electronic Mathematical Reports, Volume 12, Sobolev Institute of Mathematics, Siberian Branch of the Russian Academy of Sciences, 2015. 1006-1031.

[Vityaev et al. 2012] Vityaev E.E., Demin A.V., Ponomariov D.K. Probabilistic generalization of formal concepts // Programming, T.38, #5, 2012, p.18-34.

[Smirnov 1985] Smirnov S.D. Patterns psychology. Moscow: Moscow State University, 1985, 231s.

[Naiser 1981] Naiser U. Knowledge and reality. M.: "Progress", 1981, 229 c.

[Mill 2011] Mill J.S. Logic system of syllogistic and inductive, 2011, Prel. and adj. V.C. Finn, M.: Lenand, 832 s.

[Rutkovsky 1884] Rutkovsky L. Elementary textbook of logic. St. Petersburg, 1884.

[Mayen et al. 1976] Meyen S.V., Schreider S.A. Methodological aspects of classification theory // Problems of philosophy, 1976, #12.

[Smirnov 1938] Smirnov E.S. Design of a kind of taxonomic point of view // Zool. Zhurn. Vol. 17, No.3, 1938, Pp. 387-418.

[Zabrodin 1981] Zabrodin V.Yu. On criteria of natural classification // NTI, ser.2, 1981, #8.

[Cheng 1997] Cheng, P. From covariation to causation: A causal power theory. Psychological Review, 104, 367-405.

[Gopnik et al. 2004] Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T. A theory of causal learning in children: Causal maps and Bayes nets. Psychological Review, 111, 3-23.

[Griffiths et al. 2009] Griffiths, T. L., Tenenbaum, J. B. Theory-based causal induction. Psychological Review, 116, 56.

[Rehder et al. 2001] Rehder, B., Hastie, R. Causal knowledge and categories: The effects of causal beliefs on categorization, induction, and similarity. Journal of Experimental Psychology: General, 130, 323–360.

[Rehder 2003a] Bob Rehder. Categorization as causal reasoning // Cognitive Science 27 (2003) 709–748.

[Rehder 2003b] Rehder, B. A causal-model theory of conceptual representation and categorization. J. of Exper. Psych.: Learning, Memory, and Cognition, 29, 1141-1159.

[Rehder et al. 2011] Bob Rehder, Jay B. Martin. Towards A Generative Model of Causal Cycles // 33rd Annual Meeting of the Cognitive Science Society 2011, (CogSci 2011), Boston, Massachusetts, USA, 20-23 July 2011, V.1 pp. 2944-2949.

[Rosch 1973] Rosch, E.H. Natural categories // Cognitive Psychology 4. 1973. P. 328-350.

[Rosch et al. 1975] Rosch, E., Mervis, C.B. Family resemblances. Studies in the internal structure of categories // Cognitive Psychology, 7. P. 573–605.

[Rosch et al. 1976] Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., Boyes-Braem, P. Basic objects in natural categories // Cognitive Psychology, 8. P. 382–439.

[Rosch et al. 1978] Rosch, Eleanor and Lloyd, Barbara B. (eds) Cognition and categorization. Hillsdale, NJ: Lawrence Erlbaum, 1978, P. 27-48.

[Rosch 1978] Rosch, E. Principles of Categorization // Rosch, E. & Lloyd, B.B. (eds), Cognition and Categorization, Lawrence Erlbaum Associates, Publishers, (Hillsdale), 1978. P. 27–48

[Ross et al. 2008] B. H. Ross, E. G. Taylor, E. L. Middleton, and T. J. Nokes. Concept and Category Learning in Humans // H. L. Roediger, III (Ed.), Cognitive Psychology of Memory. Vol. [2] of Learning and Memory: A Comprehensive Reference, 4 vols. (J.Byrne Editor), Oxford: Elsevier, 2008, P. 535-556.

[The Nature of Classification 2013]   The Nature of Classification. Relationships and Kinds in the Natural Sciences. Palgrave Macmillan. 2013. 208p.

[Ganter 2003] Ganter, B. Formal Concept Analysis: Methods, and Applications in Computer Science. TU Dresden (2003).

[Ganter et al. 1999] Ganter, B., Wille, R. Formal concept analysis – Mathematical Foundations. Berlin-Heidelberg-New York, Springer (1999)

[Ganter et al. 2004] Ganter, B., Obiedkov, S. Implications in Triadic Formal Contexts. TU Dresden, Springer (2004)

[Klimushkin et al. 2010] Klimushkin, M., Obiedkov, S., Roth, C. Approaches to the Selection of Relevant Concepts in the Case of Noisy Data. In: ICFCA 2010, LNAI 5986, 255-266. (2010)

[Frequent Itemset Mining 2004] Bayardo Jr., R., Goethals, B., Zaki, M. (eds.) Proc. of the IEEE ICDM Workshop on Frequent Itemset Mining Implementations (FIMI 2004). CEUR-WS.org (2004)

[Kuznetsov 2007] Kuznetsov, S.O. On Stability of a Formal Concept. Annals of Mathematics and Artificial Intelligence. 49, 101-115 (2007)

[Buzmakov 2014] Buzmakov, A., Kuznetsov, S., Napoli, A. Concept Stability as a Tool for Pattern Selection. CEUR Workshop proceedings, vol. 1257, ECAI 2014, pp. 51-58 (2014)

[Emilion et al. 2009] Emilion R., Levy G. Size of random Galois lattices. Discrete Applied Math. J. 157, 2945-2957 (2009)

[Speransky 2013] Speransky, S.O. Logic of probability and probability logic. Novosibirsk State University, Novosibirsk, PhD thesis. (2013)

[Vityaev 2006] Vityaev, E.E. The logic of prediction. In: Proceedings of the 9th Asian Logic Conference Mathematical Logic in Asia, Novosibirsk, Russia, August 16–19, 2005, pp. 263–276. World Scientific, Singapore (2006)

[Vityaev et al. 2015] Vityaev, E.E., Martynovich, V.V. Probabilistic Formal Concepts with Negation. Perspectives of System Informatics. A. Voronkov, I. Virbitskaite (Eds.), LNCS vol. 8974, pp. 385-399 (2015)

[Martynovich et al. 2016] Vitaliy V. Martynovich, Euvgeniy E. Vityaev. Recovering Noisy Contexts with Probabilistic Formal Concepts // Proceedings of the 2nd International Workshop on Soft Computing Applications and Knowledge Discovery (SCAKD 2016), Moscow, Russia, July 18, 2016. CEUR Workshop Proceedings, v. Vol-1687, pp. 24-35.