Probabilistic formal concepts with negation

Vityaev E.E.[1,2] and Martinovich V.V.[2]

[1] Sobolev institute of mathematics, Novosibirsk, Russia,
evgenii.vityaev@math.nsc.ru,
[2] Novosibirsk State University, Novosibirsk, Russia,
vilco@ya.ru.

Abstract. The probabilistic generalization of formal concept analysis, as well as it's comparison to standard formal analysis is presented. Construction is resistant to noise in the data and give one an opportunity to consider contexts with negation (object-attribute relation which allows both attribute presence and it's absence). This generalization is obtained from the notion of formal concepts with its definition as fixed points of implications, when implications, possibly with negations, are replaced by probabilistic laws. We prove such fixed points (based on the probabilistic implications) to be consistent and wherefore determine correct probabilistic formal concepts. In the end, the demonstration for the probabilistic formal concepts formation is given together with noise resistance example.

Keywords: formal concept analysis, probability, data mining, association rules, noise

1 Introduction

In the formal concept analysis (FCA) [1, 2], formal concepts are used as classification units. The main task of FCA consists in construction of a complete lattice of formal concepts. But FCA induces a potentially dreadful combinatorial complexity and the structures obtained even from small-sized datasets can become prohibitively huge. Noise in data constitutes a primary factor of complication as it favors the existence of many similar but distinct concepts, which may excessively inflate the lattice with superfluous information that significantly impaired readability. Hence, the translation of empirical data into clean and relatively readable structures remains the most important problem. There are some works where concepts formation considered in the presence of noise [3, 4]. But all these papers base on the complete lattice of formal concepts.

In this paper we consider the problem: is it possible to construct the clean and relatively readable structure of idealized or refined concepts directly without the construction of complete lattice of formal concepts. If we consider the complete lattice of formal concepts as a "photo" of data, then the structure of "idealized" concepts may be considered as a "picture" of data.

For solution of this problem we introduce the probabilistic generalization of formal concepts. The first step was made in [5, 6], where the probabilistic generalization of the formal concepts without negations was developed. Here we introduce the probabilistic generalization of the formal concepts with negations for many-valued contexts. For that purpose we utilize, as in [5, 6], the definition of a formal concept in terms of fix-points of implications. Then we define a probability measure and generalize implications into probabilistic implications, so that they minimize the intent of concepts and eliminate random attributes. After that we define a probabilistic formal concept as a fix-point of probabilistic implications. For that purpose we prove the consistency of these fixed points. Resulting fixed points don't directly depend on data and are defined in pure probabilistic terms and thus produce a "picture" of data. At the end of paper the results of the experiment that illustrate the formation of probabilistic formal concepts are presented.

2 Formal Concept Analysis

Here we give a short review of the formal concept analysis. For details we refer to [1, 2]. FCA examines the set of objects G, which have properties from a fixed set M. We say that "the object g has the property m" by using a relation $I \subseteq G \times M$.

Definition 1. *Formal context is a triple (G, M, I), where G and M are sets of the arbitrary nature and $I \subseteq G \times M$ is a binary relation.*

On the formal context (or simply context) we define the operation $'$ as follows:

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

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

Definition 3. *Pair (A, B) is called a formal concept, if B expresses all common features for objects in A, and A - objects that have all attributes from B. In other words, $A' = B$ and $B' = A$.*

Here and further we delve a bit into the theory of FCA [1, 2, 7], but only just as much as it is necessary for the design of construction, proposed in the article.

Lemma 1. *Suppose $A_1 \subseteq A_2 \subseteq G, B_1 \subseteq B_2 \subseteq M$. Then*

1. $A_2' \subseteq A_1', B_2' \subseteq B_1'$
2. $A \subseteq A'', B \subseteq B''$
3. (A, B) *- concept* $\Rightarrow B'' = B$

In fact, usually objects are not formed from attributes in a completely arbitrary manner. Attributes form numerous relationships, which can be described in terms of implications. In definitions below B, C are subsets of attributes $\subseteq M$.

Definition 4. *Implication is a pair (B, C) which we write as $B \to C$. Implication $B \to C$ is true on $K = (G, M, I)$, if $\forall g \in G(B \nsubseteq g'$ or $C \subseteq g')$. The set of all true implications will be denoted as $Imp(K)$.*

Definition 5. *Implication $B \to C$ is called a non-trivial on K, if $C \nsubseteq B$ and $B' \neq \varnothing$. The set of all non-trivial truth implications on K we denote as $ntImp(K)$.*

Definition 6. *For any set of implications L we can construct the operator of direct inference f_L that add conclusions of all implications to the set-operand*

$$f_L(X) = X \cup \{C \mid B \subseteq X,\ B \to C \in ntImp(K)\}$$

Successively applying the operator of direct inference to any set X, we are gradually approaching it's closure [1, 6].

Definition 7. *Operator cl_L, closing the set X relative to the operator of direct inference is $cl_L(X) = f_L^{\infty}(X)$.*

Theorem 1. *For any set $B \subseteq M$ the following is accomplished [2]:*

1. *$f_{Imp(K)}(B) = B \Leftrightarrow B'' = B$;*
2. *If $B' \neq \varnothing$, then $f_{ntImp(K)}(B) = B \Leftrightarrow B'' = B$.*

3 Many-valued contexts. Formulae on binary contexts

Definitions of the previous section present in a set-theoretical notions about attributes and objects. In variety of practical problems this hardly limits the space of possible models and reality interpretations, and sometimes as well - results [6, 8]. This is a case for combinations of attributes and reasoning in terms of implications [8].

There are several different approaches to extend the I relation. Some classic examples can be found in [7, 8]. In this chapter we enrich contexts, providing for each pair (g, m) the degree of belonging the attribute to the object.

We extend context I relation with value-dependent component. Let each attribute m has its own domain V_m. To describe the degree of belonging the attribute to object we need to know the value $v \in V_m$ of the attribute m on object g. For such value we assume that $(g, m, v) \in I$.

Definition 8. *Let G - set of objects and M - set of attributes, and each attribute has a set of possible attribute values V_m. Many-valued context K is a triple*

$$(G, M, \{I_m : G \to V_m \mid m \in M\})$$

In fact $I_m(g)$ maps the object g to value of the attribute m on g. It is not difficult to envision how the many-valued contexts and ordinary contexts are connected. Each attribute with its specific value can be considered as a new independent attribute. That is, for each attribute m, consider the set of pairs (m, v) where $m \in M$ and $v \in V_m$. Relation of the object g to take quality v on m attribute can now be described as $(g, (m, v)) \in I$.

Definition 9. *As a nominally scaled context of* $K = (G, M, \{I_m\})$ *we call* $K^* = (G, M^*, \cup I_m^*)$*, where* $M^* = \{(m, v) \mid m \in M, v \in V_m\}$*, and* $I_m^* = \{(g, (m, v)) \mid I_m(g) = v\}$*. For brevity, we say that* $g \in G$ *has the attribute* m_v*, if* $I_m(g) = v$ *or, equivalently,* $(m, v) \in g'$ *within the* K^**.*

All constructions naturally transeferred to the nominally scaled K^*, as well as statements and theorems, which are presented in classic analysis of formal concepts. Particularly we can talk about formal concepts, defined on many-valued contexts. It is enough to replace occurrences of m by m_v in the relevant cases. Further referring to the classic's structures on K, we mean exactly the same constructions relative to K^*. For example,

Definition 10. *Formal concept on many-valued context* K *is a pair* (A, B)*, where* $A \subseteq G$*,* $B \subseteq M^*$*, such that* (A, B) *- formal concept for* K^**.*

It is natural to consider the proposed structure in the simplest case. In this case, each attribute we interpret in the form of predicate, identifying the a value of 1 with presence of corresponding attribute and 0 - with its absence.

Definition 11. *Binary context is a many-valued context, where* $\forall m (V_m = \{0, 1\})$*. Here and further* m *and* $\overline{m}$ *stay for* $(m, 1)$ *and* $(m, 0)$ *respectively.*

Our immediate task is to build on an arbitrary binary context of a formal system based on the first-order logic.

Definition 12. *For a binary context* $K = (G, M)$ *define a signature* $\sigma = (R, F, \rho)$*:*

1. *Set of predicates* R *- precisely the set of all* I_m*, stating the presence of the corresponding attribute or its negation;*
2. *Empty set of functional symbols* $F = \varnothing$ *(and so does* ρ*);*

All notions, such as atom, term, letter, formula and so on, are determined in a classical manner of formal systems. Formula of defined signature operates with logical connectives $\&, \vee, \rightarrow, \neg$ and predicates. We denote the resulting sets of atoms, letters, formulae and sentences as $At(K)$, $Lit(K)$, $For(K)$, $Sen(K)$, respectively.

Basic set $\mathcal{D} = \{g\}$ together with the predicates forms a model, futher labelled with K_g. The fact of truth of the formula Φ on the model of an object g we denote as follows: $g \vDash \Phi \Leftrightarrow K_g \vDash \Phi$. $G_\Phi \subseteq G = \{g \in G \mid g \vDash \Phi\}$ is called the support for Φ. If $G_\Phi = G$, then Φ - contextual tautology.

Lemma 2. *Note that* $G_{\neg\Phi} = G \setminus G_\Phi$*,* $G_{\Phi \& \Psi} = G_\Phi \cap G_\Psi$*,* $G_{\Phi \vee \Psi} = G_\Phi \cup G_\Psi$

4 Probability and rules on the context

Now we need the definition of probability on binary context.

Definition 13. *Consider probability measure μ on the set G in the Kolmogorov meaning, so G can be interpreted as a set of elementary events. Let us introduce the contextual probability measure:*

$$\nu : For(K) \to [0,1], \nu(\Phi) = \mu(G_\Phi) = \mu(\{g \mid g \vDash \Phi\})$$

Definition 14. *By statistically insignificant objects (subsets) we call $g \in G$ ($A \subset G$) such that $\mu(g) = 0$ ($\mu(A) = 0$). Formula Φ is called a ν-consistent, if $\nu(\Phi) > 0$. Formula Φ is called an almost tautology, if $\nu(\Phi) = 1$.*

Proposition 1. *Context measure ν has the following properties:*

1. *If Φ - is a classical tautology, then Φ – contextual tautology, $\nu(\Phi) = 1$;*
2. *If $\neg(\Phi \& \Psi)$ is almost a tautology on K, then $\nu(\Phi \vee \Psi) = \nu(\Phi) + \nu(\Psi)$;*
3. *$\nu(\Phi \& \Psi) \le \nu(\Phi)$.*

■ 1. Φ - is generally valid, so it is true on any model. In particular, $\forall g \in G(K_g \vDash \Phi)$, so $\Vdash \Phi$. Therefore $\nu(\Phi) = \mu(\{g \mid g \vDash \Phi\}) = \mu(G) = 1$.

2. Remember that $\nu(\Phi \vee \Psi) = \mu(G_{\Phi \vee \Psi}) = \mu(G_\Phi \cup G_\Psi)$. Inclusion-exclusion principle asserts $\mu(G_\Phi \cup G_\Psi) = \mu(G_\Phi) + \mu(G_\Psi) - \mu(G_\Phi \cap G_\Psi)$. The last one is zero due to the fact of $\neg(\Phi \& \Psi)$ is almost a tautology: $\mu(G_\Phi \cap G_\Psi) = \mu(G_{\Phi \& \Psi}) = \nu(\Phi \& \Psi) = 0$. At last, $\nu(\Phi \vee \Psi) = \mu(G_\Phi) + \mu(G_\Psi) = \nu(\Phi) + \nu(\Psi)$.

3. $G_{\Phi \& \Psi} = G_\Phi \cap G_\Psi \subseteq G_\Phi$; due to the axioms of measure: $\nu(\Phi \& \Psi) = \mu(G_{\Phi \& \Psi}) \le \mu(G_\Phi) = \nu(\Phi)$. □

In this section, we always assume that $L = Lit(K)$, K - binary context and ν - contextual measure on it. We follow the way proposed in [9–11].

Definition 15. *For the set of letters $M \subseteq L$ we construct the composition: $\& M = \underset{P \in M}{\&} P$. For the case of $M = \varnothing$ let $\& M = 1$. Similarly, we construct the negation: $\neg M = \{\neg P \mid P \in M\}$.*

Formulae of the form of simple conjunctions $F = m_{i_1} \& m_{i_2} ... \& m_{i_k}$ have one property that interlinks formulae structure and derivation operator of the classical FCA. In fact, the carrier G_F coincides with $\{m_{i_j}\}'$. In this sense, we can identify the set of letters with their representation in the form of a set of attributes $\{m_{i_j}\}$.

Moreover, the formula $m_{i_1} \& m_{i_2} ... \& m_{i_k} \to m = \& \{m_{i_j}\} \to m$ describes the same process as the implication on the context in classical sense, $(\{m_{i_j}\}, \{m\})$. According to this it is natural to define a class of implications similar to the FCA, but inside the class of formulae. We call them as rules.

Definition 16. *1. Rule is the formula $R = (H_1 \& H_2 ... \& H_k \Rightarrow T)$, where $T, H_i \in L$, $T \notin \{H_1, H_2, ... H_k\}$.*

2. *For the rule R under $\mathrm{head}(R)$ we mean the set $\{H_1, H_2 ..., H_k\}$, and $\mathrm{tail}(R) = T$. If $\mathrm{head}(R_1) = \mathrm{head}(R_2)$ and $\mathrm{tail}(R_1) = \mathrm{tail}(R_2)$, then $R_1 = R_2$.*
3. *The length of the rule is a power of its premise: $\mathrm{len}(R) = |\mathrm{head}(R)|$.*

Definition 17. *The probability of the rule R is the value*

$$\eta(R) = \nu(\mathrm{tail}(R)|\mathrm{head}(R)) = \frac{\nu(\&\mathrm{head}(R)\&\mathrm{tail}(R))}{\nu(\&\mathrm{head}(R))}$$

The rule is global if the expression in the denominator equals to one. If the expression in the denominator is zero, the probability of rule remains undefined.

Definition 18. *Rule R_1 is a sub-rule of R_2, or R_1 is more general then R_2, if* $\mathrm{head}(R_1) \subset \mathrm{head}(R_2)$ *and* $\mathrm{tail}(R_1) = \mathrm{tail}(R_2)$. *This fact we denote as $R_1 \succ R_2$.*

Definition 19. *The rule R_1 is a generalization of the rule R_2, i.e. $R_1 \succeq R_2$, when $R_1 \succ R_2$ or $R_1 = R_2$.*

Definition 20. *The rule R_1 is a refinement of the rule R_2, $R_1 > R_2$, if $R_2 \succ R_1$ and $\eta(R_1) > \eta(R_2)$.*

Theorem 2. *Let R is a non-global rule on the context K with measure ν.*

1. *Probability of R is less or equal of the probability of corresponding implication:*

$$\eta(R) \leq \nu(\mathrm{head}(R) \to \mathrm{tail}(R))$$

2. *R is almost a tautology if $\Leftrightarrow \eta(R) = \nu(R) = 1$.*

■ Let $H = \&\mathrm{head}(R)$, $T = \mathrm{tail}(R)$ and consider the difference $\nu(H)(\eta(R) - \nu(H \to T))$. Note that $H \to T = T \vee \neg H = (T\&H) \vee \neg H$, while $(T\&H)\&\neg H = 0$. Hence, by lemma 4, $\nu(H \to T) = \nu(T\&H) + \nu(\neg H)$. Thus the difference can be transformed as

$$\nu(H)(\eta(R) - \nu(H \to T)) = \nu(H\&T) - \nu(H\&T)\nu(H) - \nu(\neg H)\nu(H) =$$
$$\nu(H\&T)\nu(\neg H) - \nu(H)\nu(\neg H) = -\nu(H\&\neg T)\nu(\neg H) \leq 0$$

Further, equality to 0 is achieved only if $\nu(H\&\neg T) = 0$. However, this is equivalent to the $\nu(H\&T) = \nu(H) - \nu(H\&\neg T) = \nu(H)$ and $\eta(R) = \frac{\nu(H\&T)}{\nu(H)} = 1$. Here we conclude that R is almost a tautology □

Corollary 1. *If the measure μ does not permit insignificant objects, then the set of almost tautologies turns into a set of tautologies, and $\eta(R) = 1 \Leftrightarrow R$ - contextual tautology.*

Definition 21. *R is a probability law, if it is a refinement of every of its sub-rule, i.e. $(R' \succ R) \Rightarrow (R > R')$.*

Now we prove some technical points we need to continue our working with the rules.

Lemma 3. *If addition of the letter H into the premise of the rule R reduces it's probability, $\eta(\&\mathrm{head}(R)\ \&\ H \Rightarrow \mathrm{tail}(R)) < \eta(R)$, then $\neg H$ increases it.*

Lemma 4. *For any rule R there exists its generalization R' such that R' is probabilistic law, and $\nu(R') \geq \nu(R)$.*

■ Consider the set $\Pi = \{A \mid \nu(A) \geq \nu(R),\ A \succeq R\}$. So as $R \in \Pi$, then $\Pi \neq \varnothing$. Hence, there is a minimal element in the sense of relation $\succeq$, call it $S = \min \Pi$. Condition 2 of the lemma holds for S by construction of Π.

Suppose S is not a law, i.e. here exists sub-rule S', such that $\nu(S') \geq \nu(S)$ and $S' \succ S$, given $S \succeq R$ we conclude that $S' \succ R$. From the other side, $\nu(S') \geq \nu(S) \geq \nu(R)$, where it follows that $S' \in \Pi$, contradicting the minimality of S. □

5 Refinement theorem

Now we apply the proposed in [10, 12] technics to defined rules.

Definition 22. *Pseudo rule is a formula $R = ((P_1\&...\&P_k)\ \&\ \neg(N_1\&...\&N_s) \Rightarrow T)$; for pseudo rule R,* $\mathrm{head}(R) = (P_1\&...\&P_k)\ \&\ \neg(N_1\&...\&N_s)$ *and* $\mathrm{tail}(R) = T$*; letters P_i we call the positive part of the premise and letters N_j we call the negative part; probability of the pseudo rule R is the value*

$$\eta(R) = \nu(\mathrm{tail}(R)|\mathrm{head}(R)) = \frac{\nu(\&\mathrm{head}(R)\ \&\ \mathrm{tail}(R))}{\nu(\&\mathrm{head}(R))}$$

Theorem 3. *(about refinement) Let $S = ((\&A)\&\neg(\&B)) \Rightarrow T)$ be a pseudo rule, $R = ((\&A) \Rightarrow T)$ the corresponding rule without negative part and moreover $\eta(S) > \eta(R)$. Then for R there is refinement rule $R' > R$ formed with the help of the negative part of pseudo rule S.*

■ For brevity, we denote $\overline{A} = \&A$, $\overline{B} = \&B$. Let us write the probability of pseudo rule S as:

$$\eta(S) = \nu(T \mid \overline{A}\ \&\ \neg\overline{B}) = \nu(T \mid \overline{A}\ \&\ (\neg B_1 \vee ... \vee \neg B_m)) \tag{1}$$

Next we represent the disjunction as disjunction of conjunctions:

$$\neg B_1 \vee ... \vee \neg B_m = \bigvee_{i=(0,...,0)}^{i=(1,...,1,0)} (B_1^{i_1}\ \&\ ...\ \&\ B_m^{i_m})$$

where 0 in multi-index indicates the presence of negation, and 1 - its absence. All multi-indices are included in a lexicographic order except for the last $(1, ..., 1)$, which corresponds to the conjunction $B_1\ \&\ ...\ \&\ B_m$.

Then the conditional probability (1) can be rewritten as

$$\eta(S) = \nu(T \mid \bigvee_{i=(0,...,0)}^{i=(1,...,1,0)} (\overline{A}\ \&\ B_1^{i_1}\ \&\ ...\ \&\ B_m^{i_m})) \tag{2}$$

Suppose the theorem's statement is false and any generalization $R' \succ R$, formed via appending some subset from $\{B_1, ..., B_m\}$ to premise, will fail as a refinement.

This means that all inequalities like $\nu(T \mid \overline{A}\ \&\ B_1^{i_1}\ \&\ ...\ \&\ B_m^{i_m}) \le \nu(T \mid \overline{A})$ are true, if corresponding probabilities are defined. Since $\nu(\overline{A}\ \&\neg\overline{B}) \neq 0$, there is at least one multi-index $(i_1, ..., i_m)$, for which it is true. Then

$$\nu(T\ \&\overline{A}\ \&\ B_1^{i_1}\ \&\ ...\ \&\ B_m^{i_m}) \le \nu(T \mid \overline{A})\nu(\overline{A}\ \&\ B_1^{i_1}\ \&\ ...\ \&\ B_m^{i_m});$$

$$\nu(T \mid \bigvee_{i=(0,...,0)}^{i=(1,...,1,0)} (\overline{A}\ \&\ B_1^{i_1}\ \&\ ...\ \&\ B_m^{i_m})) = \frac{\nu(\vee\ T\ \&\ \overline{A}\ \&\ B_1^{i_1}\ \&\ ...\ \&\ B_m^{i_m})}{\nu(\vee\ \overline{A}\ \&\ B_1^{i_1}\ \&\ ...\ \&\ B_m^{i_m})} =$$

$$\frac{\sum \nu(T\ \&\overline{A}\ \&\ B_1^{i_1}\ \&\ ...\ \&\ B_m^{i_m})}{\sum \nu(\overline{A}\ \&\ B_1^{i_1}\ \&\ ...\ \&\ B_m^{i_m})} \le \frac{\nu(T \mid \overline{A}) \sum \nu(\overline{A}\ \&\ B_1^{i_1}\ \&\ ...\ \&\ B_m^{i_m})}{\sum \nu(\overline{A}\ \&\ B_1^{i_1}\ \&\ ...\ \&\ B_m^{i_m})} = \nu(T \mid \overline{A});$$

The last, according to (2), means that $\eta(S) \le \eta(R)$ – is a contradiction with the theorem premise. So, our assumption is false and for one of the rules we have $\nu(T \mid \overline{A}\ \&\ B_1^{i_1}\ \&\ ...\ \&\ B_m^{i_m}) > \nu(T \mid \overline{A})$. □

6 Semantic probabilistic inference

We define another key concept for this work - the ratio of semantic probabilistic inference on the set of rules [10, 13, 11].

Definition 23. *The rule R is semantically probabilistic inferred from the rule R', $R' \sqsubset R$ if: R, R' - probabilistic laws,* $\text{len}(R) = \text{len}(R') + 1$*, $R > R'$.*

Definition 24. *Probabilistic law R is the strongest, if $\forall R'\ \neg(R \sqsubset R')$.*

Definition 25. *Semantic Probabilistic Inference (SPI) is a sequence of rules $R_0 \sqsubset R_1 \sqsubset R_2\ ... \sqsubset R_m$, such that:* $\text{len}(R_0) = 0$*, R_m – the strongest probabilistic law.*

In other words, SPI requires the procedure of inference from start to finish.

Definition 26. *A maximal specific law for the predicate T is called as the strongest probabilistic law, if it has the maximal conditional probability among all the other strongest probabilistic laws with the conclusion T.*

The set of maximal specific laws on the context K we denote as MSR_K or MSR, if there is no ambiguity. Designation $\text{MSR}(T)$ stays for those subsets from MSR for which the conclusion is T.

Lemma 5. *For any rule R with* $\text{tail}(R) = T$*, which probability is defined, there always exists a maximal specific law W with the same conclusion T, such that $\eta(R') \ge \eta(R)$.*

■ By lemma 4 there is a generalization R' for the rule R which is a probabilistic law. But for R' there exists the strongest probabilistic law R'' such that $\eta(R'') \ge \eta(R')$. For R'', there is a maximum of the set of the strongest probabilistic laws, i.e. maximal specific law R''' and $\eta(R''') \ge \eta(R'') \ge \eta(R') \ge \eta(R)$. $W = R'''$ still has same $\text{tail}(W) = T$ and so it is the sough for. □

7 Classes of rules

In [10, 14] classes of rules presented. They are used to justify the correctness of the semantic probabilistic inference. Slightly modifying these definitions, we will receive the comparable results.

Definition 27. $R \in M_1(T) \Leftrightarrow ((\varnothing \Rightarrow T) \succ R \ \Rightarrow \ R > (\varnothing \Rightarrow T))$

Definition 28. $R \in M_2(T) \Leftrightarrow R \in M_1(T)$ *and* $(\forall R' \in M_1(T))[R \succ R' \Rightarrow \eta(R') \leq \eta(R)]$

Definition 29. $M_1 = \bigcup\limits_{T \in Lit(K)} M_1(T);\ M_2 = \bigcup\limits_{T \in Lit(K)} M_2(T)$

In other words, class M_1 requires rules to be meaningful, thus enable them to make sense compared with the unconditional approval of T. Class M_2 requires that the rule can not be more specific (no matter how we have expanded the rule R, we can never improve the estimation of its probability). We have the following relationship:

Proposition 2. $\mathrm{MSR} \subset M_2 \subset M_1$.

■ The second inclusion is obvious. Let $R \in \mathrm{MSR}$. There is some SPI for R according to definition 27, starting with the unconditional rule $R' = \varnothing \Rightarrow \mathrm{tail}(R)$. If the premise of R is not empty, then $\varnothing \Rightarrow \mathrm{tail}(R) \succ R$ and from the chain of semantic probabilistic inference relations it follows that $R > R'$ and $R \in M_1$. If the premise R is empty, then the last is automatically fulfilled.

Consider $R \succ R' \in M_1$ and assume that $\eta(R') > \eta(R)$. Lemma 5 implies that there exists $S \in \mathrm{MSR} : \eta(S) \geq \eta(R') > \eta(R)$. This contradicts the maximal specificity of R and therefore $\eta(R') \leq \eta(R)$. Hence $R \in M_2$. □

Definition 30. *As a system of the rules we will call any* $\Pi \subseteq M_2$.

To investigate the formal concept of the binary context of K in the spirit of the approach indicated in [5, 6], it is sufficient to understand the structure of corresponding prediction operator's fixed points on the nominally scaled context K^*. Here we aim to study the fixed points for the probabilistic operator of prediction on K, accounting the availability of negations in the formulae of a special kind. Let $L \subset Lit(K)$ be a arbitrary set of letters from context formal system.

The defenition is completely similar to deterministic one (compare with defenition 6). We strictly follows the generelization idea and the only difference will be the nature of used implications: they are turned in probabilistic entities.

Definition 31. *Operator of direct predictions on the system* Π *works as follows:*

$$\mathrm{Pr}_\Pi(L) = L \cup \{T \mid \exists R \in \Pi : \mathrm{head}(R) \subseteq L, \mathrm{tail}(R) = T\}$$

so Pr_Π *adds conclusions of all the implications, the premise of which is contained in* L *and fullfilled on it, to the operand.*

Definition 32. *Closure of a set of letters* L *is the smallest fixed point of operator of direct prediction:* $\mathrm{PR}_\Pi(L) = \mathrm{Pr}^\infty_\Pi(L)$.

8 Consistency theorem

The correctness for construction proposed in the previous section needs to be proven, as in [10, 14, 15]. Correctness here is understood in two senses: in probabilistic and logical. We show that both they are satisfied.

Definition 33. *Set of letters L is called compatible, if $\&L$ – ν-consistent on K.*

In fact, the set is compatible when there is a set of statistically significant objects of G_K, on which the formula $\&L$ is fulfilled. On "normal" probability contexts K (with no more than countable set of objects G, each of which is statistically significant), the compatible sets L will be those sets (and with them also the corresponding sets of attributes L from M_{K^*}) for which $\mathrm{L}' \neq \varnothing$.

Object can not have both $m \in M_{K^*}$ and $\overline{m} \in M_{K^*}$ simultaniously. It can either possess or lack any attribute accordingly to binary context defenition. Such attribute combinations looks very suspecious and leads to known logical problems, consistency of attribute sets is a desirable property here.

Definition 34. *Set of letters L – consistent, if it does not contain any atom T simultaneously with its negation $\neg T$.*

Proposition 3. *If L – compatible, then L – consistent.*

■ Otherwise $\exists T :\ T \in L$ and $\neg T \in L$, so $\nu(\&L) \leq \nu(T \ \&\ \neg T) = 0$. □

Let Π be any rule system and Pr be according prediction operator Pr_Π. We first show that the direct prediction retains the property of compatibility.

Theorem 4. *(Compatibility) If L is compatible, then* $\mathrm{Pr}(L)$ *is also compatible.*

■ The proof is easy to obtain by looking at the refinement theorem. Consider all rules that contribute to the formation of the direct prediction based on L:

$T = \{R \in \Pi \mid \mathrm{head}(R) \subseteq L\}$. We enumerate all the elements of T in an arbitrary manner, $T = \{T_1, ...T_m\}$, and consider the sequence of sets $U_i = U_{i-1} \cup \{\mathrm{tail}(T_i)\}, U_0 = L$. We show that each U_i is compatible.

$U_0 = L$ is obviously compatible by the premise of the theorem.

Let U_i is compatible. For brevity, let $U = U_i, W = U_{i+1}, R = R_{i+1}$ and $T = \mathrm{tail}(R), H = \mathrm{head}(R), N = U \setminus H$. Suppose that W is inconsistent, i.e. $\nu(\&W) = 0$. Similarly to the refinement theorem, consider pseudo rule $F = (\&H \ \&\ \neg(\&N)) \Rightarrow T$. There are two cases:

1. case: $\nu(\&\mathrm{head}(F)) \neq 0$. Then the probability of F is defined and

$$\eta(F) = \frac{\nu(\&H \ \&\ \neg(\&N) \ \&\ T)}{\nu(\&H \ \&\ \neg(\&N))} = \frac{\nu(\&H \ \&\ T) - \nu(\&H \ \&\ (\&N) \ \&\ T)}{\nu(\&H) - \nu(\&H \ \&\ (\&N))} =$$
$$\frac{\nu(\&H \ \&\ T) - \nu(\&W)}{\nu(\&H) - \nu(\&U)} = \frac{\nu(\&H \ \&\ T)}{\nu(\&H) - \nu(\&U)} > \frac{\nu(\&H \ \&\ T)}{\nu(\&H)} = \eta(R) > 0.$$

   According to the refinement theorem, there is a rule S such that $S > R$, which contradicts that the R is non-refineable (i.e., the fact that $R \in M_2$).

2. case: $\nu(\&\text{head}(F)) = 0$. Then

   $$\nu(\&\text{head}(F)) = \nu(\&H \ \& \ \neg(\&N)) = 0 \Rightarrow \nu(\&H \ \& \ \neg(\&N) \ \& \ T) = 0;$$
   $$0 = \nu(\&H \ \& \ (\&N) \ \& \ T) = \nu(\&H \ \& \ T) - \nu(\&H \ \& \ \neg(\&N) \ \& \ T) = \nu(\&H \ \& \ T).$$

   The last means that $\eta(R) = 0$, which contradicts $R \in M_1$ ($0 = \eta(R) > \eta(\varnothing \Rightarrow T) \geq 0$). □

Corollary 2. *If L – compatible, then* PR(L) *– also compatible.*

Corollary 3. *If L is compatible, then* PR(L) *– consistent.*

9 About incompatible sets

The situation is quite clear for compatible sets. Direct prediction on the compatible set L and the closure of this set are compatible and consistent.

Let's try to understand the structure of incompatible sets L. We will start with a fairly trivial statement, which is the opposite to the compatibility theorem.

Proposition 4. *If L – incompatible, then* PR(L) *– is also incompatible.*

■ Assuming compatibility PR(L) we find that any subset, and in particular L, is compatible. □

Somewhat more difficult is the question of the inconsistency of such closures. For a more detailed study of the structure of incompatible systems of letters we need the following concept.

Definition 35. *We say that M is ν-maximal in L, $M \underset{\nu}{\subseteq} L$, when M is maximal by inclusion subset of L and M is compatible.*

Definition 36. *System of rules Π is called complete, if* MSR $\subset \Pi$.

The following discussion focuses only on complete systems of rules. Requirement of completeness can be slightly relaxed, as it can be seen from the theorem below, but we restrict ourselves to the most specific rules in this article. This means we consider only $PR = PR_{\Pi}$ operators, where Π is complete system. It should be noted, that according to proposition 2 the system containing M_2 are complete.

Theorem 5. *Let $M \underset{\nu}{\subseteq} L$. Then $M \cup \neg(L \setminus M) \subseteq$* PR$(M)$.

■ Let x belong to the left side of the formula. Case $x \in M$ is obvious. Here $x \in$ PR(M) according to the definition of the closure.

Now let $x \in L \setminus M$. By definition, ν-maximal subsets of the set $M \cup \{x\}$ is incompatible (otherwise obtain a new maximal set by inclusion). This means that

$$\nu(\&M \ \& \ x) = 0;$$
$$\nu(\&M \ \& \ \neg x) = \nu(\&M) - \nu(\&M \ \& \ x) = \nu(\&M);$$

Let $R = (\&M \Rightarrow \neg x)$. From the relations above it is easy to calculate the probability of rule R:

$$\eta(R) = \frac{\nu(\&M \ \& \ \neg x)}{\nu(\&M)} = 1;$$

Lemma 11 asserts there is a rule $S \in \mathrm{MSR} \subset \mathit{\Pi}$, which is MSR-rule with conclusion equal $\neg x$, so that for S is fulfilled $\nu(S) \geq 1$. Thus, the rule S inevitably add $\neg x$ into direct prediction of $\Pr(L)$. $\square$

Theorem 6. *Consider* $M \underset{\nu}{\subseteq} L$, $N \underset{\nu}{\subseteq} L$ *and* $M \neq N$. *Then*

1. $\exists x \ : \ x \in \mathrm{PR}(M)$ *and* $\neg x \in \mathrm{PR}(N)$*;*
2. $\mathrm{PR}(M) \supseteq \mathrm{PR}(M \cap N) \subseteq \mathrm{PR}(N)$ *and* $\mathrm{PR}(M) \neq \mathrm{PR}(N)$.

■ 1. $M \neq N$ means, that $\exists x \in M \setminus N$ (indeed $M \subset N$ would be contrary to the maximality of M). $x \in M \Rightarrow \mathrm{PR}(M)$ and similarly to Theorem 6 $\neg x \in \mathrm{PR}(N)$.

2. $\mathrm{PR}(N)$ – is compatible and consistent, and $\neg x \in \mathrm{PR}(N)$; it means $x \notin \mathrm{PR}(N)$ and $x \in \mathrm{PR}(M) \setminus \mathrm{PR}(N)$. Then, $M \cap N \subset M$, so $\mathrm{PR}(M \cap N) \subseteq \mathrm{PR}(M)$. $\square$

The last two theorems conclude that there exists an injective mapping from ν-maximal subsets of L to fixed points set, each completely covering the entire set of atoms in L (containing them or their negations).

Inconsistency and compatibility of fixed points for compatible sets proved in section above. For incompatible sets following theorem tends to be an answer.

Theorem 7. *If* L *is incompatible, then* $\mathrm{PR}(L)$ *– inconsistent.*

■ Find ν-maximal subset of L and denote it as M. $M \neq L$, otherwise L would have been compatible. Therefore, there exists $x \in L \setminus M$. Set $\{x\}$ is extended to a maximal compatible $N \underset{\nu}{\subseteq} L$. By construction, $x \in N \setminus M \Rightarrow M \neq N$. By Theorem 6, there exists y, such that $y \in \mathrm{PR}(M)$ and $\neg y \in \mathrm{PR}(N)$:

$$\left.\begin{array}{c} M \subseteq L \\ N \subseteq L \end{array}\right\} \Rightarrow \left.\begin{array}{c} y \in \mathrm{PR}(M) \subseteq \mathrm{PR}(L) \\ \neg y \in \mathrm{PR}(N) \subseteq \mathrm{PR}(L) \end{array}\right\} \Rightarrow \mathrm{PR}(L) \text{ - contradictory. } \square$$

10 Probabilistic formal concepts

The fixed points of PR operator are rather interesting and promising. However, the purpose for their consideration was the motive of introduction of the probabilistic analogous of formal concepts. Using the idea of Theorem 1, it is easy to offer as candidates [5, 6] for inclusion in intent of such concepts. We mean exactly the fixed points of PR.

Selection for concept extent is a bit more complicated. But since all the sets of letters, such that $\mathrm{PR}(M) = B$, have a real connection to the closure, it is logical to propose to collect all objects falling under them. That is:

Definition 37. *By probabilistic formal concept on K we denote (A, B), such that:*

$$\mathrm{PR}(B) = B,\ A = \bigcup_{\mathrm{PR}(C)=B} G_C$$

To distinguish probabilistic concepts from the usual ones in the sense of the context of K^*, the last ones we call strict formal concepts. Our selection justified by the following statement, relating probabilistic and strict formal concepts on the same context.

Theorem 8. *Let K be a binary context.*

1. *If (A, B) is strict concept on K, then there is a probabilistic concept (N, M) such that $A \subseteq N$, and $B \subseteq M$.*
2. *If (N, M) is the probabilistic concept on K, then there is a set of strict notions $\mathcal{C}$, such that*

$$\forall (A, B) \in \mathcal{C}\ (\mathrm{PR}(B) = M),$$
$$N = \bigcup_{(A,B)\in\mathcal{C}} A.$$

■ Suppose $\mathcal{S} = \{S \mid \mathrm{PR}(S) = M\}$.

1. Let $M = \mathrm{PR}(B)$. Then $B \in S$ and $A = G_B \subseteq \bigcup_{\mathcal{S}} G_S = N$. Hence (A, B) is desired.

2. On $\mathcal{S}$ we make a set of strict concepts $\mathcal{C} = \{(S''', S'') \mid S \in \mathcal{S}\}$. From lemma 1 it is easy to understand that $B''' = B'$, that is $\mathcal{C} = \{(S', S'') \mid S \in \mathcal{S}\}$ and all $(A, B) \in \mathcal{C}$ - are strict concepts. Hence $N = \bigcup_{\mathcal{S}} S' = \bigcup_{(A,B)\in\mathcal{C}} A$ It should be added that $M \in \mathcal{S}$ and hence $\mathcal{C} \neq \varnothing$. □

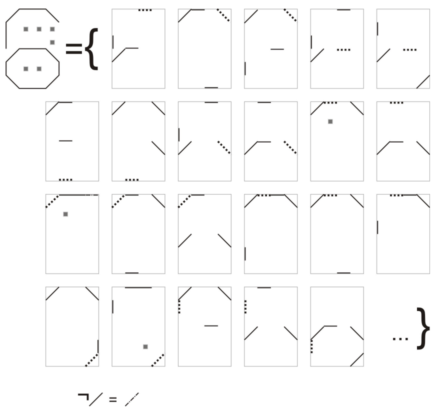
Probabilistic concept is like cluster unifying set of poorly distinguishable strict concepts in terms of a system of rules Π.

11 Probabilistic concepts search

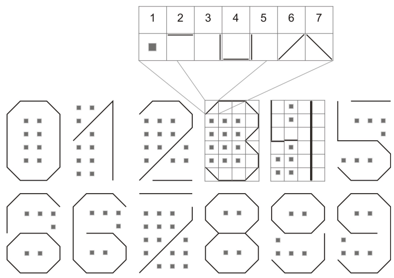
In this section we restrict ourselves to the case of finite context K. From the last one, we can drop out statistically insignificant objects without loss of generality.

Assume that the system of rules Π on context K has already been found by one of the algorithms, for example [11, 13]. Probabilistic concept definition implies the following search procedure.

1. On step $k = 1$ we generate the set $C^{(1)} = \{\mathrm{PR}(\mathrm{head}(R)) \mid R \in \Pi\}$.
2. On step $k > 1$ in case of $C^{(k-1)} = \varnothing$ algorithm finishes its execution and output a list of detected probability concepts.
3. Else on step $k > 1$ the set $A = \{g \in G \mid \mathrm{PR}(g' \cap B) = B\}$ is calculated for each $B \in C^{(k-1)}$. If $A \neq \varnothing$, pair (A, B) is added to list of found concepts.
4. The set $C^{(k)} = \{\mathrm{PR}(B \cup C) \mid B, C \in C^{(k-1)}, \mathrm{PR}(B \cup C) \notin C^{(k-1)}\}$ is generated.
5. Let $k := k + 1$ and go to step 2.



(a) Fix point for the digit 6



(b) Coding of the digits

Finally, we present one of many examples. Real example can be found in [16]. In [17] there is experiment for analyzing of postal envelopes digits. The data contains 12 digits (2 options for each "6" and "9"). The context is based on 24 attributes, each of which has 7 values (for different shapes in the relevant sector of the digit partition). Set G consists of 360 objects (30 copies of each digit) with gap in information (each digit misses one randomly deleted attribute), which are mixed, plus negative sample of 1050 objects with random attributes. There is no attribute that designate which digit is the object representing. On these data, 73458 rules was found. Then the all fixed points were computed using set of all rules, which turn up just 14. From them, 12 digits were exactly our numbers, and for each of "6" and "9" were still two fixed points containing an extra space in the features that distinguishes 2 options of those digits ("6" and "9" are not mixed up due to fixing top-bottom orientation while coding procedure).

12 Conclusion

Negations (and, in general, the values for attributes) in formal contexts, produce a much more expressive system of concepts. This provides such properties (in some sense) as correctness and completeness to proposed method.

Our considerations and algorithm allows us to find consistent probabilistic concepts, and, at the same time, do not lose strict concepts. Moreover, the proposed method preserves a binary noise - the random binary noise imposed into the values of attributes don't change the set of concepts [5, 6].

The concept of fixed points may be rather natural applied for formalization of classtering [17]. Therefore, the fix point theory and probabilistic formal concepts may be used for the new Data Mining method development.

References

1. Ganter, B.: Formal Concept Analysis: Methods, and Applications in Computer Science. TU Dresden (2003)
2. Ganter, B., Wille, R.: Formal concept analysis – Mathematical Foundations. Berlin-Heidelberg-New York, Springer (1999)
3. Klimushkin, M., Obiedkov, S., Roth, C.: Approaches to the Selection of Relevant Concepts in the Case of Noisy Data. In: ICFCA 2010, LNAI 5986, 255266 (2010)
4. Kuznetsov, S.O.: On Stability of a Formal Concept. Annals of Mathematics and Artificial Intelligence. 49, 101-115 (2007)
5. Demin, A., Ponomaryov, D., Vityaev, E.: Probabilistic concepts in formal contexts. In: Proc. ARCOE Workshop at IJCAI'11, pp. 31-35. Barcelona (2011)
6. Vityaev, E.E., Demin, A.V., Ponomaryov, D. K.: Probabilistic Generalization of Formal Concepts. Programming and Computer Software. 38(5), 219-230 (2012)
7. Ganter, B., Obiedkov, S.: Implications in Triadic Formal Contexts. TU Dresden, Springer (2004)
8. Missaoui, R., Kwuida, L.: Implications in Triadic Formal Contexts. In: 9th International Conference, ICFCA 2011. Nicosia, Cyprus, Springer (2011)
9. Vityaev, E., Kovalerchuk, B.: Empirical Theories Discovery based on the Measurement Theory. Mind and Machine 14(4), 551-573 (2004)
10. Vityaev, E.E.: The logic of prediction. In: Mathematical Logic in Asia. Proceedings of the 9th Asian Logic Conference, pp.263-276, World Scientific, Singapore (2006)
11. Vityaev, E.E.: Knowledge discovery. Computational cognition. Cognitive process models. Novosibirsk State University Press, Novosibirsk (2006) (in Russian)
12. Vityaev, E., Smerdov, S.: On the Problem of Prediction In: KONT/KPP 2007, LNAI 6581, pp. 280-296, Springer, Heidelberg, (2011)
13. Kovalerchuk, B., Vityaev, E.: Data Mining in Finance: Advances in Relational and Hybrid methods. Kluwer Academic Publishers (2000)
14. Speransky, S.O.: On the logical consistency of probabilistic predictions. Messenger of NGU. Series: mathematics, mechanics, informatics 11(1), 99-115 (2011) (in Russian)
15. Speransky, S.O.: Logic of probability and probability logic. Novosibirsk State University, Novosibirsk, PhD thesis. (2013) (in Russian)
16. Vityaev, E.E., Lapardin, K.A., Khomicheva, I.V., Proskura, A.,L.: Transcription factor binding site recognition by regularity matrices based on the natural classification method. Intelligent Data Analysis 12(5), 495-512 (2008)
17. Vityaev, E.E., Neupokoev, N.V.: Formal model of perception and pattern as fix point of anticipations. In: Approaches to the thinking modeling, pp.155-172, Moscow, URSS Editorial (2014) (in Russian)