A Markov Chain Approach to Random Generation of Formal Concepts

Dmitry V. Vinogradov[1,2]

[1] All-Russia Institute for Scientific and Technical Information (VINITI), Intelligent Information Systems Laboratory, Moscow 125190, Russia
vin@viniti.ru

[2] Russian State University for Humanities, Intelligent Robotics Laboratory, Moscow 125993, Russia
http://isdwiki.rsuh.ru/index.php/User:Vinogradov.Dmitry

Abstract. A Monte Carlo approach using Markov Chains for random generation of concepts of a finite context is proposed. An algorithm similar to CbO is used. We discuss three Markov chains: non-monotonic, monotonic, and coupling ones. The coupling algorithm terminates with probability 1. These algorithms can be used for solving various information retrieval tasks in large datasets.

Keywords: formal context, formal concept, Markov chain, termination with probability 1

1 Introduction

In many natural problems of information retrieval the piece of information which we are looking for is not contained in few documents. The query generates a huge amount of relevant documents and the task is to generates a cluster of related documents together with the set of common terms describing its common meaning. There are many choices for such cluster. So, the user has to look for plausible answers to his query.

JSM-method is a logical device to provide plausible reasoning for generation, verification and falsification of such clusters and for explanation of whole collection of all relevant documents by means of accepted clusters. The first variant of JSM method was presented by Prof. V.K. Finn in 1983 [1] (in Russian). FCA corresponds to the generation ("induction") step of JSM method. See [5] for details. This correspondence allows to use FCA algorithms in JSM-method and vice versa. For example, the well-known algorithm "Close-by-One" (CbO) was initially introduced by S.O. Kuznetsov in [4] for JSM-method and later translated into FCA framework. The state of art for JSM-method is represented in [2].

In our opinion, the main drawback of 'old-fashioned' JSM-method is the computational complexity of JSM algorithms, especially for the induction step. Paper [7] presents a results of comparison between various (partially improved by the survey's authors) variants of famous deterministic algorithms of FCA.

Paper [6] provides theoretical bounds on computational complexities of various JSM tasks.

The development of JSM-method has resulted in intelligent systems of JSM type that were applied in various domains such as sociology, pharmacology, medicine, information retrieval, etc. In practice there were situations when a JSM system generates more than 10,000 formal concepts (JSM similarities) from a context with about 100 objects. In our opinion the importance of all generated concepts is doubtful, because when experts manually select important JSM causes they reject majority of generated JSM similarities.

In this paper we propose Monte Carlo algorithms using Markov Chain approach for random generation of concepts of a finite context. In other words, we replace the lattice of all concepts by small number of its random elements.

2 Background

2.1 Basic definitions and facts of FCA

Here we recall some basic definitions and facts of Formal Concept Analysis (FCA) [3].

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

Example 1 (Boolean cube with n atoms). Consider the context (G, M, I), where $G = \{g_1, \ldots, g_n\}$, $M = \{m_1, \ldots, m_n\}$, and

$$g_j I m_k \Leftrightarrow j \neq k. \tag{3}$$

Then $(\{g_{j_1}, \ldots, g_{j_k}\})' = M \setminus \{m_{j_1}, \ldots, m_{j_k}\}$, $(\{m_{j_1}, \ldots, m_{j_k}\})' = G \setminus \{g_{j_1}, \ldots, g_{j_k}\}$, $A'' = A$ for all $A \subseteq G$, and $B'' = B$ for all $B \subseteq M$. Hence $\mathbf{B}(G, M, I)$ has element $(\{g_{j_1}, \ldots, g_{j_k}\}, M \setminus \{m_{j_1}, \ldots, m_{j_k}\})$ for every $\{g_{j_1}, \ldots, g_{j_k}\} \subseteq G$.

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

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

It is easy to check that $A_1 \subseteq A_2$ implies $A_1' \supseteq A_2'$ and for concepts (A_1, A_1') and (A_2, A_2') reverse implication is valid too, because $A_1 = A_1'' \subseteq A_2'' = A_2$. Hence, for (A_1, B_1) and (A_2, B_2) in $\mathbf{B}(G, M, I)$

$$(A_1, B_1) \leq (A_2, B_2) \Leftrightarrow A_1 \subseteq A_2 \Leftrightarrow B_2 \subseteq B_1. \tag{4}$$

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. *[3] 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*

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

$$A_1 \subseteq A_2 \Rightarrow A_1' \supseteq A_2' \quad \text{and} \quad B_1 \subseteq B_2 \Rightarrow B_1' \supseteq B_2', \tag{6}$$

$$A' = A''' \quad \text{and} \quad B' = B''', \tag{7}$$

$$\Big(\bigcup_{j\in J} A_j\Big)' = \bigcap_{j\in J} A_j' \quad \text{and} \quad \Big(\bigcup_{j\in J} B_j\Big)' = \bigcap_{j\in J} B_j', \tag{8}$$

$$A \subseteq B' \;\Leftrightarrow\; A' \supseteq B. \tag{9}$$

Proposition 1. *[3] Let (G, M, I) be a context. Then $(\boldsymbol{B}(G, M, I), \leq)$ is a lattice with join and meet given by*

$$\bigvee_{j\in J} (A_j, B_j) = \Big(\big(\bigcup_{j\in J} A_j\big)'', \bigcap_{j\in J} B_j\Big), \tag{10}$$

$$\bigwedge_{j\in J} (A_j, B_j) = \Big(\bigcap_{j\in J} A_j, \big(\bigcup_{j\in J} B_j\big)''\Big); \tag{11}$$

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

$$(M', M) \leq (A, B) \leq (G, G'). \tag{12}$$

2.2 The "Close-by-One" operations: definition and properties

With the help of expressions for the infinum and the supremum operations in $\mathbf{B}(G, M, I)$ given by Proposition 1 we can introduce local steps of our Markov chains:

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

$$CbO((A, B), g) = ((A \cup \{g\})'', B \cap \{g\}'), \tag{13}$$
$$CbO((A, B), m) = (A \cap \{m\}', (B \cup \{m\})''). \tag{14}$$

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

We call these operations CbO because the first one is used in Close-by-One (CbO) Algorithm to generate all the elements of $\mathbf{B}(G, M, I)$, see [4] for details.

Lemma 2. *Assume that* (G, M, I) *is a context and let* $(A, B) \in \boldsymbol{B}(G, M, I)$, $g \in G$, *and* $m \in M$. *Then*

$$g \in A \Rightarrow CbO((A, B), g) = (A, B), \tag{15}$$
$$m \in B \Rightarrow CbO((A, B), m) = (A, B), \tag{16}$$
$$g \notin A \Rightarrow (A, B) < CbO((A, B), g), \tag{17}$$
$$m \notin B \Rightarrow CbO((A, B), m) < (A, B). \tag{18}$$

Proof. If $g \notin A$ then $A \subset A \cup \{g\} \subseteq (A \cup \{g\})''$ by (5). By definition of the order between concepts this inclusion and (13) imply (17). Relation (18) is proved in the same way, the rest is obvious.

Lemma 3. *Assume that* (G, M, I) *is a context and let* $(A_1, B_1), (A_2, B_2) \in \boldsymbol{B}(G, M, I)$, $g \in G$, *and* $m \in M$. *Then*

$$(A_1, B_1) \leq (A_2, B_2) \Rightarrow CbO((A_1, B_1), g) \leq CbO((A_2, B_2), g), \tag{19}$$
$$(A_1, B_1) \leq (A_2, B_2) \Rightarrow CbO((A_1, B_1), m) \leq CbO((A_2, B_2), m). \tag{20}$$

Proof. If $A_1 \subseteq A_2$ then $A_1 \cup \{g\} \subseteq A_2 \cup \{g\}$. Hence (6) implies $(A_2 \cup \{g\})' \subseteq (A_1 \cup \{g\})'$. Second part of (6) implies $(A_1 \cup \{g\})'' \subseteq (A_2 \cup \{g\})''$. By definition of the order between concepts this is (19). Relation (20) is proved in the same way by using (4).

3 Markov Chain Algorithms

Now we represent Markov chain algorithms for random generation of formal concepts.

Data: context (G, M, I), external function $CbO(\, , \,)$
Result: random concept $(A, B) \in \mathbf{B}(G, M, I)$
$t := 0$; $(A, B) := (M', M)$;
while $(t < T)$ **do**
  select random element $x \in (G \setminus A) \sqcup (M \setminus B)$;
  $(A, B) := CbO((A, B), x)$;
  $t := t + 1$;
end

Algorithm 1: Non-monotonic Markov chain

Example 2 (Random walk on Boolean cube). Consider the context (G, M, I) of Example 1. Then Non-monotonic Markov chain corresponds to Random Walk on Boolean Cube of all the concepts in $\mathbf{B}(G, M, I)$.

Definition 2. *An* ***(order) ideal*** *of partially odered set (poset)* $(S, \leq)$ *is a subset* J *of* S *such that*

$$\forall s \in S \ \forall r \in J[s \leq r \Rightarrow s \in J]. \tag{21}$$

A Markov chain S_t *with values into poset* $(S, \leq)$ *is called* ***monotonic*** *if for every pair of start states* $a \leq b$ $(a, b \in S)$ *and every order ideal* $J \subseteq S$

$$\mathrm{P}[S_1 \in J | S_0 = a] \geq \mathrm{P}[S_1 \in J | S_0 = b]. \tag{22}$$

Proposition 2. *There exists the context* (G, M, I) *such that Non-monotonic Markov chain for* $(\boldsymbol{B}(G, M, I), \leq)$ *isn't monotonic one.*

See [8] for the proof of Proposition 2. The following Markov chain is always monotonic one.

Data: context (G, M, I), external function $CbO(\, , \,)$
Result: random concept $(A_T, B_T) \in \mathbf{B}(G, M, I)$
$t := 0$; $X := G \sqcup M$; $(A, B) := (M', M)$;
while $(t < T)$ **do**
  select random element $x \in X$;
  $(A, B) := CbO((A, B), x)$;
  $t := t + 1$;
end

Algorithm 2: Monotonic Markov chain

Proposition 3. *For every context* (G, M, I) *the Monotonic Markov chain for* $(\boldsymbol{B}(G, M, I), \leq)$ *is monotonic one.*

The proof of Proposition 3 can be found in [8].The Monotonic Markov chain algorithm has another advantage: random selection of elements of the static set $X := G \sqcup M$. However both previous algorithms have common drawback: the unknown value T for the termination time of the calculation. In the Monte Carlo Markov Chain (MCMC) theory this value corresponds to the **mixing time** of the chain. For some special Markov chains (for instance, for the chain of Example 2) the mixing time is estimated by sophisticated methods. In general case, it is an open problem. The following algorithm has not this problem at all.

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

Algorithm 3: Coupling Markov chain

Intermediate value of quadruple $(A,B) \leq (C,D)$ on step t corresponds to Markov chain state Y_t. The A_t, B_t, C_t and D_t are first, second, third, and fourth components, respectively.

Definition 3. *A* ***coupling length*** *for context* (G,M,I) *is defined by*

$$L = \min(\mid G \mid, \mid M \mid). \tag{23}$$

A ***choice probability*** *of fixed object or attribute in context* (G,M,I) *is equal to*

$$p = \frac{1}{\mid G \mid + \mid M \mid}. \tag{24}$$

Lemma 4. *If* $\mid G \mid < \mid M \mid$ *then for every integer* r *and every pair of start states* $(A,B) \leq (C,D)$ *(*$(A,B),(C,D) \in \boldsymbol{B}(G,M,I)$*)*

$$\mathrm{P}[A_r = A_{r+L} = C_{r+L} \& B_r = B_{r+L} = D_{r+L} | Y_r = (A,B) \leq (C,D)] \geq p^L. \tag{25}$$

If $\mid G \mid \geq \mid M \mid$ *then for every integer* r *and every pair of start states* $(A,B) \leq (C,D)$ *(*$(A,B),(C,D) \in \boldsymbol{B}(G,M,I)$*)*

$$\mathrm{P}[A_{r+L} = C_{r+L} = C_r \& B_{r+L} = D_{r+L} = D_r | Y_r = (A,B) \leq (C,D)] \geq p^L. \tag{26}$$

Proof. For coupling to $(A,B) \leq (A,B)$ it suffices to get an element from $A \setminus C$, but $\mid A \setminus C \mid \leq \mid G \mid = L$. Relation (26) is proved in a similar way. □

Theorem 1. *The coupling Markov chain has the probability of coupling (termination) before* n *steps with limit 1 when* $n \to \infty$.

Proof. Lemma 3 implies that $(A_{(k-1)\cdot L}, B_{(k-1)\cdot L}) \leq (C_{(k-1)\cdot L}, D_{(k-1)\cdot L})$. Let $r = (k-1) \cdot L$. Then Lemma 4 implies that $\mathrm{P}[A_{k\cdot L} \neq C_{k\cdot L} \vee B_{k\cdot L} \neq D_{k\cdot L} | Y_r = (A_r, B_r) \leq (C_r, D_r)] \leq (1 - p^L)$. After k independent repetitions we have $\mathrm{P}[A_{k\cdot L} \neq C_{k\cdot L} \vee B_{k\cdot L} \neq D_{k\cdot L} | Y_0 = (M', M) \leq (G, G')] \leq (1-p^L)^k$. But when $k \to \infty$ we have $(1-p^L)^k \to 0$. □

Conclusions

In this paper we have described a Monte Carlo approach using Markov Chains for random generation of concepts of a finite context. The basic steps of proposed Markov chains are similar to ones of algorithm CbO. We discuss three Markov chains: non-monotonic, monotonic, and coupling ones. The coupling algorithm terminates with probability 1.

Acknowledgements.

The author would like to thank Prof. Victor K. Finn and Tatyana A. Volkova for helpful discussions. The research was supported by Russian Foundation for Basic Research (project 11-07-00618a) and Presidium of the Russian Academy of Science (Fundamental Research Program 2012-2013).

References

1. V.K. Finn, About Machine-Oriented Formalization of Plausible Reasonings in F. Beckon-J.S. Mill Style. Semiotika I Informatika, 20, 35–101 (1983)
2. V.K. Finn, The Synthesis of Cognitive Procedures and the Problem of Induction. Autom. Doc. Math. Linguist., 43, 149–195 (2009)
3. B. Ganter, R. Wille, Formal Concept Analysis: Mathematical Foundations, Springer-Verlag, 1999
4. S.O. Kuznetsov, A Fast Algorithm for Computing All Intersections of Objects in a Finite Semi-Lattice, Automatic Documentation and Mathematical Linguistics, vol.27, no.5, 11-21, 1993.
5. S.O. Kuznetsov, Mathematical aspects of concept analysis. Journal of Mathematical Science, Vol. 80, Issue 2, pp. 1654-1698, 1996.
6. S.O. Kuznetsov, Complexity of Learning in Concept Lattices from Positive and Negative Examples. Discrete Applied Mathematics, 2004, no. 142(1-3), pp. 111-125.
7. S.O. Kuznetsov and S.A. Obiedkov, Comparing Performance of Algorithms for Generating Concept Lattices. Journal of Experimental and Theoretical Artificial Intelligence, Vol. 14, no. 2-3, pp. 189-216, 2002.
8. D.V. Vinogradov, Random generation of hypotheses in the JSM method using simple Markov chains. Autom. Doc. Math. Linguist., 46, 221–228 (2012)