ON STATISTICAL ANALYSIS OF EMBEDDING IN BINARY MARKOV CHAIN

E.V. Vecherko
Belarusian State University
Minsk, Belarus
e-mail: vecherko@bsu.by

Abstract

A polynomial algorithm for calculating a likelihood function under the fixed parameters is developed. Maximum likelihood estimators for parameters of embedding and transition matrix are constructed and analyzed.

1 Introduction

Nowadays the models of embedding are used in lots of scientific research areas: intellectual property rights, genetics [2, 3, 5]. It is critical to know if some additional bits are embedded into data sequences. The problem is to make a decision whether a data sequence contains additional bits or not [2,3]. Unfortunately, the most part of existing strategies for solving this problem is based on empirical characteristics. So the decision maker is very dependent on the learning data sets. This article is a step in direction of theoretical analysis of such mathematical models of embedding.

2 Mathematical model of embedding

At first, let us introduce the notations: $V = \{0, 1\}$ is a binary alphabet, V_T – a set of binary T-dimensional vectors, $\mathbb{N}$ – a set of natural numbers, $I\{A\}$ – an indicator function of the event A, $u_{t_1}^{t_2} = (u_{t_1}, \dots, u_{t_2}) \in V_{t_2-t_1+1}$ $(t_1, t_2 \in \mathbb{N},\ t_1 \le t_2)$ – a binary string of $t_2 - t_1 + 1$ bits, $w(\cdot)$ – a Hamming weight.

Let us assume that a cover sequence $x_1^T = (x_1, x_2, \dots, x_T) \in V_T,\ x_t \in V,\ t = 1, \dots, T$, of size T is a stationary binary Markov chain [1] of order 1 with a symmetric transition probabilities matrix $P = P(\varepsilon) = (p_{j_0,j_1}(\varepsilon)),\ j_0, j_1 \in V$:

$$P(\varepsilon) = \frac{1}{2}\begin{pmatrix} 1+\varepsilon & 1-\varepsilon \\ 1-\varepsilon & 1+\varepsilon \end{pmatrix},\ p_{j_0,j_1} = \mathbf{P}\{x_{t+1} = j_1 | x_t = j_0\} = \frac{1}{2}(1 + (-1)^{j_0+j_1}\varepsilon). \quad (1)$$

Here $\varepsilon \in (0, 1)$ is a model parameter: if $\varepsilon = 0$ than x_1^T is a sequence of i.i.d random variables and this situation is investigated in [3]. The stationary probability distribution of x_1^T is equal to $\pi = (1/2, 1/2)$.

A hidden random sequence $\xi_1^M = (\xi_1, \dots, \xi_M) \in V_M,\ M \le T$, is considered to be a sequence of i.i.d. Bernoulli random variables: $\mathbf{P}\{\xi_t = j\} = \theta_j,\ j \in V,\ \theta_1 = 1 - \theta_0,\ t = 1, \dots, M$. As a rule the hidden sequence $\{\xi_t\}$ has a symmetric probability distribution as it is often compressed before embedding: $\theta_1 = \theta_0 = 1/2$.

Let now introduce a special (q, r)-block model of a sequence $\gamma_1^T \in V_T$ which determine the process of embedding. At first, we divide the cover sequence x_1^T into the blocks of the size q: $x_{(1)} = x_1^q, x_{(2)} = x_{q+1}^{2q}, \dots, x_{(K)} = x_{(K-1)q+1}^{Kq}$. Here we assume that $T = qK$. Then we use secondary random variables $\zeta_k \in V,\ k = 1, \dots, T/q$, which are i.i.d. Bernoulli random variables: $\mathbf{P}\{\zeta_k = 1\} = 1 - \mathbf{P}\{\zeta_k = 0\} = \delta$. These new variables are responsible for choosing the blocks of the cover sequence x_1^T for embedding. If $\zeta_k = 1$ than in r randomly chosen bits of the block $x_{(k)}$ we embed r bits of the hidden sequence, if $\zeta_k = 0$ than the embedding operation in the block $x_{(k)}$ is not executed. The sequence γ_1^T is consisted of independent blocks which have the following probability distribution:

$$\mathbf{P}\{\gamma_{(k-1)q+1}^{kq} = u_1^q\} = \begin{cases} 1-\delta, & w(u_1^q) = 0, \\ \delta / C_q^r, & w(u_1^q) = r, \\ 0, & w(u_1^q) \notin \{0, r\}, \end{cases} \qquad k = 1, \dots, K,\ u_1^q \in V_q. \tag{2}$$

We notice that the maximum number of embedding bits is equal to $Tr/q = Kr$ and the power of a set of all possible sequences γ_1^T is $(1 + C_q^r)^{T/q}$ according to its construction. Let remark that if $r = q = 1$ than the power of a set of all possible γ_1^T values is equal to 2^T.

When the hidden sequence ξ_1^M is embedded to the Markov cover sequence x_1^T we get a new random sequence $Y_1^T \in V_T$:

$$Y_t = \gamma_t \xi_{\tau_t} + (1 - \gamma_t) x_t = \begin{cases} x_t, & \gamma_t = 0, \\ \xi_{\tau_t}, & \gamma_t = 1. \end{cases} \tag{3}$$

The random sequences $\{x_t\},\ \{\xi_t\},\ \{\gamma_t\}$ are considered to be independent.

3 Statistical parameters estimation

Initially, let divide a set V_t of binary t-dimensional vectors into $t + 1$ disjoint subsets (fig. 1):

$$V_t = \Gamma_0^{(t)} \cup \Gamma_1^{(t)} \cup \ldots \cup \Gamma_t^{(t)}, \tag{4}$$

where

$$\begin{aligned} \Gamma_0^{(t)} &= \{u_1^t \in V_t :\ u_t = 1\}, \\ \Gamma_1^{(t)} &= \{u_1^t \in V_t :\ u_t = u_{t-1} = 0\}, \\ \Gamma_j^{(t)} &= \{u_1^t \in V_t :\ u_t = 0, u_{t-1} = \ldots = u_{t-j-1} = 1, u_{t-j} = 0\},\ 1 < j < t, \\ \Gamma_t^{(t)} &= \{u_1^t \in V_t :\ u_t = 0, u_{t-1} = \ldots = u_1 = 1\}. \end{aligned}$$

Using the division (4) let us define a function of binary variables $u_1^t, y_1^t \in V_t$:

$$\varphi_t(u_1^t, y_1^t) = \begin{cases} \theta_{y_t},\ u_1^t \in \Gamma_0^{(t)}, \\ \frac{1}{2}(1 + (-1)^{y_{t-j} + y_t} \varepsilon^j),\ u_1^t \in \Gamma_j^{(t)},\ 1 \le j < t, \\ \frac{1}{2},\ u_1^t \in \Gamma_t^{(t)}. \end{cases}$$

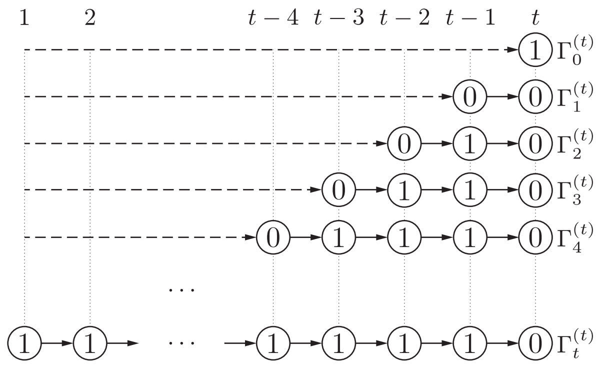

Figure 1: The illustration of the set γ_1^t division; dashed line indicates all possible pathes

Under the introduced notations a likelihood function for an observed sequence $y_1^T \in V_T$ that contains a hidden sequence is

$$L(\varepsilon,\delta) = \mathbf{P}\{Y_1^T = y_1^T\} = \sum_{u_1^T \in V_T^{(q,r)}} (1-\delta)^{b_0(u_1^T)} (\delta / C_q^r)^{b_r(u_1^T)} \prod_{t=1}^{T} \varphi_t(u_1^t, y_1^t). \qquad (5)$$

where a set $V_T^{(q,r)} = \{u_1^T \in V_T : b_h(u_1^T) = 0,\ h \in \{1,\dots,q\}/\{r\}\}$ is needed according to the construction of the sequence $\{\gamma\}$ which determines the points for embedding, $b_h(u_1^T) = \sum_{k=1}^{T/q} I\{w(u_{q(k-1)+1}^{qk}) = h\}$. Calculation of $L(\varepsilon,\delta)$ according to its direct definition (5) involves on the order of $O(T(1+C_q^r)^{T/q})$ calculations.

MLE-estimators $\hat{\varepsilon}, \hat{\delta}$ of the model parameters ε, δ are the solution of problem

$$L(\varepsilon,\delta) = \mathbf{E}\{L_{\gamma_1,\dots,\gamma_T}(\varepsilon)\} \to \max_{\varepsilon \in (-1,1),\ \delta \in [0,1]}.$$

where $L_{u_1,\dots,u_T}(\varepsilon) = \mathbf{P}\{Y_1^T = y_1^T | \gamma_1^T = u_1^T\}$ is a probability of observations y_1^T on condition with $\gamma_1^T = u_1^T$.

Lemma 1. *Under the assumptions (1)-(3), if $q > r$ and $t > 2r+1$ than*

$$\mathbf{P}\{\gamma_1^t \in \bigcup_{j=2r+2}^{t} \Gamma_j^{(t)}\} = 0,$$

$$\mathbf{P}\{Y_t = y_t | Y_1^{t-1} = y_1^{t-1}, \gamma_1^t = u_1^t\} = \psi_t(u_{t-2r-1}^t, y_{t-2r-1}^t) = \\ = \begin{cases} \theta_{y_t}, & u_1^t \in \Gamma_0^{(t)}, \\ \frac{1}{2}(1 + (-1)^{y_{t-j}+y_t} \varepsilon^j), & u_1^t \in \Gamma_j^{(t)},\ 1 \le j \le 2r+1 \end{cases} \qquad (6)$$

and a random sequence $\{Y_t\}$ with a constant sequence $\{\gamma_t\}$ is a supervised Markov chain of conditional order. The conditional order $s_t \in \{0,\dots,2r+1\}$ is dependent on a sequence $\{\gamma_t\}$: $s_t = j$, if $u_1^t \in \Gamma_j^{(t)}$.

The lemma 1 provides a developing of a polynomial algorithm for calculation the likelihood function $L(\varepsilon, \delta)$ which is based on the algorithm "Forward" [4].

We denote by $s \in \mathbb{N}$ a secondary parameter of the algorithm and by $\alpha_t(u_0, \ldots, u_{r-1}) = \mathbf{P}\{Y_1 = y_1, \ldots, Y_t = y_t, \gamma_{t-s+1} = u_0, \ldots, \gamma_t = u_{s-1}\}$, $t > s$, the probability of the partial observations y_1^t and states u_0^{s-1} at times $t - s + 1, \ldots, t$ of the sequence $\{\gamma_t\}$.

Theorem 1. *Under the assumptions (1)-(3), $q > r$, $s > 2r + 1$ the probabilities $\alpha_t(u_0, \ldots, u_{s-1})$, $t = s + 1, s + 2, \ldots, T$, can be calculated recurrently:*

$$\alpha_t(u_0, \ldots, u_{s-1}) = c_{t,u_{s-2},u_{s-1}} \sum_{u_{-1} \in V} \alpha_{t-1}(u_{-1}, \ldots, u_{s-2}) \psi_t(u_{s-2r-2}^{s-1}, y_{t-2r-1}^{t}), \qquad (7)$$

where ψ_t is defined in (6), the probabilities $c_{t,u_{s-2},u_{s-1}} = \mathbf{P}\{\gamma_t = u_{s-1} | \gamma_{t-1} = u_{s-2}\}$.

The probabilities c_{t,j_0,j_1}, $j_0, j_1 \in V$, can be calculated according to the construction of the $\{\gamma_t\}$ sequence. The initial probabilities $\alpha_t(u_0, \ldots, u_{t-1})$, $t = 1, \ldots, s$ are

$$\alpha_1(u_0) = q_{1,0,u_0} \varphi_1(u_0, y_1),$$
$$\alpha_t(u_0, \ldots, u_{t-1}) = \alpha_{t-1}(u_1, \ldots, u_{t-1}) c_{t,u_{t-2},u_{t-1}} \varphi_t(u_0^{t-1}, y_0^{t-1}), \ 2 \leq t \leq s.$$

The likelihood function $L(\varepsilon, \delta)$ is equal to $\sum_{u_0^{s-1} \in V_s} \alpha_T(u_0, \ldots, u_{s-1})$. The proposed algorithm for calculating $L(\varepsilon, \delta)$ based on (7) involves on the order of $O(T 2^{2r})$ calculations if we set a parameter s to its minimum possible value $2r + 2$.

To compute a likelihood function (e.g. using gradient-search procedure) we also need to perform the initial statistical estimation of the model parameters.

References

[1] Billingsley. P. (1961) Statistical methods in Markov chains. *Ann. Math. Statistics*. Vol. **32**, N. 1, pp. 12–40.

[2] Kharin Yu.S., Vecherko E.V. (2010). On statistical hypotheses testing of embedding. *Proceedings of Computer Data Analysis and Modeling Conference*. Vol. **2**. pp. 26-29.

[3] Ponomarev K.I. (2010). A parametric model of embedding and its statistical analysis. *Discrete Mathematics and Applications*. Vol, **19**, pp. 587-596.

[4] Rabiner L.R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*. Vol. **77**, N. 2, pp. 257-286.

[5] Waterman M.S. (1989). *Mathematical methods for DNA sequences*. CRC Press.