Applied Mathematics Letters 23 (2010) 1125–1128

Contents lists available at ScienceDirect

Applied Mathematics Letters

journal homepage: www.elsevier.com/locate/aml

# A stopping rule for symbolic dynamic filtering☆

Yicheng Wen, Asok Ray*

*Mechanical Engineering Department, Pennsylvania State University, University Park, PA 16802, United States*

ARTICLE INFO

*Article history:*
Received 22 February 2010
Received in revised form 26 April 2010
Accepted 29 April 2010

*Keywords:*
Stopping rule
Symbolic analysis
Markov chain

ABSTRACT

One of the key issues in symbolic dynamic filtering (SDF) is how to obtain a lower bound on the length of symbol blocks for computing the state probability vectors of probabilistic finite-state automata (PFSA). Having specified an absolute error bound at a confidence level, this short work formulates a stopping rule by making use of Markov chain Monte Carlo (MCMC) computations.



## 1. Introduction

Symbolic dynamic filtering (SDF) has been used as a tool for anomaly detection [1] and also for feature extraction in pattern classification problems [2]. The theory of SDF makes the following assumptions on the underlying dynamical system.

- The system behavior is statistically quasi-stationary at the fast scale of time series data acquisition.
- An observable non-stationary behavior can be associated with changes evolving at a slow time scale.

While time series of sensor data are converted to symbol blocks, a key issue in SDF is how to decide the length of the symbol block for construction of probabilistic finite-state automata (PFSA) models so that these models capture significant statistical properties of the dynamical system. Usually, the longer the symbol block is, the more accurate the state probability vector is expected to be. However, very long symbol blocks may not always be admissible because of the assumption of quasi-stationarity in the SDF analysis.

Anomaly detection usually consists of two phases, namely the offline learning phase and the online monitoring phase. In the learning phase, some features of the system are extracted under a nominal condition. In the monitoring phase, the same features are computed and compared to the features obtained in the learning phase by using an appropriate diversity measure or a statistical test to detect and quantify the anomaly.

Using a stopping rule, referred to as the relative error method in the sequel, has been proposed by Ray [3]; this is based on the Perron–Frobenius theorem of irreducible matrices and the ergodic theory of finite-state Markov chains. The stopping time is computed as the ratio of the number of Markov states and the tolerance $\eta$ (that is a free parameter to be chosen). Although this stopping rule [3] is computationally efficient, it does not specify how to choose the parameter $\eta$.

Flegal and Haran [4] proposed an asymptotic stopping rule, called the fixed-width method, based on the central limit theorem. Having specified a confidence level, an absolute error bound $\varepsilon$ is computed to construct the stopping criterion. However, there is no guarantee that the process will stop at a finite step and it does not provide an explicit relation between the stopping point and the absolute error bound $\varepsilon$. Although this method might work for offline learning, it is not deemed suitable for online monitoring.

☆ This work was supported in part by the US Office of Naval Research under Grant No. N00014-09-1-0688, and by the US Army Research Laboratory and the US Army Research Office (ARO) under Grant No. W911NF-07-1-0376. Any opinions, findings and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the sponsoring agencies.

* Corresponding author.
*E-mail addresses:* yxw167@psu.edu (Y. Wen), axr2@psu.edu (A. Ray).

0893-9659/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.aml.2010.04.048

This short work formulates and validates a stopping rule based on Markov chain Monte Carlo (MCMC) computations. The stopping criterion is obtained via a relation between the tolerance $\eta$ and the absolute error bound $\varepsilon$, which is generated offline by the learning algorithm. Subsequently, the stopping rule is executed online to obtain an adaptive confidence interval of the state probability vector $\boldsymbol{p}$ of the PFSA for anomaly detection.

## 2. Statement of the problem

Let a statistically stationary dynamical system be modeled as a stationary Markov chain $\mathbb{S} = \{s_0, s_1, s_2, \ldots\}$ of finite order $D$, where $D$ is a positive integer. Let the symbols $s_i \in \mathbb{S}$ belong to a (finite) alphabet $\Sigma$ and let $\Sigma^*$ be the set of all finite-length strings of symbols including the null string $\epsilon$. Each state $q$ of $\mathbb{S}$ is labeled with a symbol block of length $D$ belonging to $\Sigma^*$ and we let $q_j$ to be the unique label of the $j$th state. For example, if $\Sigma = \{0, 1\}$ and $D = 2$, then the possible states are $q_1 = 00$, $q_2 = 01$, $q_3 = 10$, and $q_4 = 11$.

If $\mathbb{S}$ is assumed to be an ergodic Markov chain (i.e., the associated $(n \times n)$ state transition matrix $\boldsymbol{\Pi}$ is irreducible, where $n \leq |\Sigma|^D$), then it follows from the Perron–Frobenius theorem [5] that there exists a unique probability vector $\boldsymbol{p} = [p_1, p_2, \ldots, p_n]$ such that $\boldsymbol{p}\boldsymbol{\Pi} = \boldsymbol{p}$ with the constraints $\sum_i p_i = 1$ and $p_i > 0\ \forall i$. Then, $\boldsymbol{p}$ is called the state probability vector of the Markov chain $\mathbb{S}$. In the setting of SDF, the quasi-stationary state probability vector of the PFSA could be selected as the feature which captures the statistical property of the dynamical system.

For a particular symbol block $\boldsymbol{s}^r = \{s_i\}_{i=1}^r$ generated by $\mathbb{S}$, let $\hat{\boldsymbol{p}}(r) \triangleq [\hat{p}_1(r)\ \hat{p}_2(r)\ \cdots\ \hat{p}_n(r)]$ be the estimated state probability vector. Each element of $\hat{\boldsymbol{p}}(r)$ is defined as

$$\hat{p}_i(r) = \frac{1}{r}\sum_{j=1}^{r} \mathbb{J}_{q_i} \circ T^j(\boldsymbol{s}^r) = \frac{N_i^r}{r}, \quad i = 1, 2, \ldots, n \tag{1}$$

where $\mathbb{J}_{q_i}(x)$ is an indicator function, i.e.,

$$\mathbb{J}_{q_i}(x) = \begin{cases} 1 & \text{if } q_i \text{ is a prefix of the string } x \\ 0 & \text{otherwise.} \end{cases}$$

$T$ is the left shift operator, i.e., $T(s_1 s_2 s_3 \ldots) = s_2 s_3 \ldots$, and $N_i^r$ the number of times the block $\boldsymbol{s}^r$ visits the state $q_i$.

By the ergodic theorem for a Markov chain, it is guaranteed that $\hat{\boldsymbol{p}}(r) \to \boldsymbol{p}$ as $r \to \infty$. The problem is finding a *minimal* stopping point $r_{stop}$ such that $\hat{\boldsymbol{p}}(r_{stop})$ computed from a symbol block of length $r_{stop}$ satisfies the following condition:

$$\|\hat{\boldsymbol{p}}(r_{stop}) - \boldsymbol{p}\|_\infty < \varepsilon \quad \text{with a confidence level } (1 - \alpha) \tag{2}$$

where $\| \bullet \|_\infty$ is the max norm of the finite-dimensional vector $\bullet$, and $\varepsilon$ is the absolute error bound of the estimation.

## 3. Stopping rules

The difficulty encountered in the above stopping problem is that the limit point $\boldsymbol{p}$ is unknown in Eq. (2). Without loss of generality, the block length of each state is set to $D = 1$, because it is possible to rename the alphabet $\Sigma$ as the labels of the states.

### 3.1. Relative error method

The absolute error between successive iterations is obtained as

$$\|\hat{\boldsymbol{p}}(r) - \hat{\boldsymbol{p}}(r+1)\|_\infty = \max_i \left| \frac{N_i^r}{r} - \frac{N_i^{r+1}}{r+1} \right| < \frac{1}{r} \tag{3}$$

because $N_i^{r+1}$ can only take two values $N_i^r$ or $N_i^r + 1$ depending on which state is visited next.

With the objective of identifying the stopping point $r_{stop}$, a tolerance $\eta$ $(0 < \eta \ll 1)$ is specified for the relative error such that

$$\frac{\| \left(\hat{\boldsymbol{p}}(r) - \hat{\boldsymbol{p}}(r+1)\right) \|_\infty}{\| \left(\hat{\boldsymbol{p}}(r)\right) \|_\infty} \leq \eta \quad \forall r \geq r_{stop}. \tag{4}$$

Since the minimum possible value of $\| \left(\hat{\boldsymbol{p}}(r)\right) \|_\infty$ for all $r$ is $\frac{1}{n}$, where $n$ is the number of states (i.e., the dimension of the vector $\hat{\boldsymbol{p}}(r)$), the least of the most conservative values of the stopping point is obtained from Eqs. (3) and (4) as

$$r_{stop} = \max\left( \text{ceil}\left(\frac{n}{\eta}\right),\ r_{min} \right) \tag{5}$$

where ceil$(\bullet)$ rounds the real number $\bullet$ to the nearest integer towards infinity, and $r_{min}$ is the minimum allowable value of $r_{stop}$. The rationale for including $r_{min}$ is to avoid the risk of premature termination possibly due to an erroneous value of $\eta$. Note that a shortcoming of the relative error method in its present form is that the relation between the absolute error $\varepsilon$ and the tolerance $\eta$ is not specified. This might lead to an arbitrary choice of $\eta$.

### 3.2. Markov chain Monte Carlo (MCMC) methodology

The Markov chain Monte Carlo (MCMC) tools have been used to estimate $E_\mu g = \int_\Omega g(x)\mu(\mathrm{d}x)$ where $g(x)$ is a real-valued, $\mu$-integrable function on $\Omega$. Under certain specified conditions, the ergodic theorem [6] implies that

$$\bar{g}_r := \frac{1}{r}\sum_{i=0}^{r-1} g(s_i) \xrightarrow{a.e.} E_\mu g \quad \text{as } r \to \infty. \tag{6}$$

The following statement of the central limit theorem [6] has been used in formulating the MCMC stopping rule.

**Theorem 3.1.** *Given a Markov chain $\mathbb{S} = \{s_0, s_1, \ldots\}$ and a real-valued and $\mu$-integrable function $g$ on $\Omega$, if $\mathbb{S}$ is uniformly ergodic* [6] *and $E_\mu g^2 < \infty$, then the following central limit theorem for $g$ holds:*

$$\sqrt{r}(\bar{g}_r - E_\mu g) \xrightarrow{d} N(0, \sigma_g^2) \quad \text{as } r \to \infty \tag{7}$$

*where $\sigma_g^2 \triangleq Var_\mu\{g(s_0)\} + 2\sum_{i=1}^{\infty} Cov_\mu\{g(s_0), g(s_i)\}$.*

The central limit theorem holds because an ergodic Markov chain defined on a finite state space is uniformly ergodic [7]. Taking $g$ as the indicator function of the state $i$, the limit of the time average $\bar{g}_r$ becomes $p_i$, the $i$th element of the state probability vector $\boldsymbol{p}$.

Fixed-width methodology [4] constructs an asymptotically valid confidence interval of the estimate $\hat{p}_i$ in terms of the $N(0, 1)$ cumulative distribution function $\Phi$, the confidence level $(1 - \alpha)$ and the block length $r$. Given the absolute error bound $\varepsilon$ and the estimated standard deviation $(\hat{\sigma}_g)_i$, the stopping rule for estimated probability $\hat{p}_i$ of the state $q_i$ is

$$\frac{(\hat{\sigma}_g)_i}{\sqrt{r}}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) + \theta(r, r_{min}) \le \varepsilon \quad \forall i \in \{1, 2, \ldots, n\} \tag{8}$$

where $\theta(r, r_{min}) \triangleq \varepsilon\mathbb{I}(r < r_{min})$ and $\mathbb{I}$ is the usual indicator function, namely,

$$\mathbb{I}(\bullet) = \begin{cases} 1 & \text{if the statement } \bullet \text{ is true} \\ 0 & \text{if the statement } \bullet \text{ is false.} \end{cases}$$

The rationale for including the term $\theta(r, r_{min})$ in Eq. (8) is to avoid the risk of premature termination possibly due to inaccuracy in the computation of $(\hat{\sigma}_g)_i$.

There are several ways for obtaining an estimate $\hat{\sigma}_g$. As suggested in [4], the method of batch means requires weak conditions on the function $g$ and is implemented as follows. If the symbol block is of total length $r = ab$, then the batch means estimate of variance of state $q_i$ is calculated as follows:

$$\bar{Y}_{ij} \triangleq \frac{1}{b}\sum_{k=(j-1)b+1}^{jb} \mathbb{I}(s_k = q_i) \tag{9}$$

$$\left(\hat{\sigma}_g\right)_i = \sqrt{\frac{b}{a-1}\sum_{j=1}^{a}\left(\bar{Y}_{ij} - \hat{p}_i(r)\right)^2} \quad \text{and} \quad \hat{\sigma}_g = \max_i\left(\hat{\sigma}_g\right)_i. \tag{10}$$

The real parameters $a$ and $b$ both increase as $r$ increases. A convenient choice is $b(r) = O(\sqrt{r})$.

### 3.3. Algorithms for the stopping rule

The following algorithms are proposed to serve as the stopping rule for SDF. Under a nominal condition, the $\eta$–$\varepsilon$ relation function $f$ is computed offline via Algorithm 1. Having the function $f$ already computed, the relative error method [3] is used to compute the estimated state probability vector $\hat{\boldsymbol{p}}(r_{stop})$ online from a symbol block of length $r_{stop}$ by Algorithm 2. Finally, a Boolean variable FLAG is produced to indicate whether the deviation of the online estimation $\hat{\boldsymbol{p}}(r_{stop})$ of the stationary state probability vector from $\hat{\boldsymbol{p}}_{ref}$ (that is already computed offline in the learning phase) is within the absolute error bound $\varepsilon$.

## 4. Summary and conclusions

This short work presents a flexible stopping rule for online monitoring in the framework of symbolic dynamic filtering (SDF) [1] and Markov chain Monte Carlo (MCMC) [8] computations. The stopping rule uses the fixed-width methodology [4] to find the $\eta$–$\varepsilon$ relation offline, which serves as the information input to the relative error method [3]. This algorithm can be used for SDF-based anomaly detection [1] by checking the Boolean variable FLAG, where FLAG = 1 indicates that the (online estimated) state probability vector has deviated from its (offline computed) nominal value beyond a specified bound at a given confidence level.

**Algorithm 1** Algorithm of Offline Learning

**Input:** Symbol block $\boldsymbol{s}^L$ of length $L$, absolute error bound $\varepsilon$, number of states $n$, minimum block length $r_{min}$, and confidence level $(1-\alpha)$.
**Output:** Estimate of stationary probability vector $\hat{\boldsymbol{p}}_{ref}$ and $\eta$–$\varepsilon$ relation function $f$, i.e. $\eta = f(\varepsilon)$.
Select $\boldsymbol{y} = \{y_j\}$ for computing $\boldsymbol{x} = \{x_j\}$.
**for** each $y_j$ in $\boldsymbol{y}$ **do**
  Compute $r_j = \max\left(\text{ceil}\left(\frac{n}{y_j}\right),\ r_{min}\right)$.
  Obtain the symbol block $\boldsymbol{s}^{r_j}$.
  Use Eqs. (9) and (10) to estimate $\hat{\sigma}_g$ from $\boldsymbol{s}^{r_j}$.
  $x_j = \frac{\hat{\sigma}_g}{\sqrt{r_j}}\,\Phi^{-1}\left(1-\frac{\alpha}{2}\right)$ from Eq. (8).
**end for**
Regression analysis to obtain $f$ based on the pair $(\boldsymbol{x}, \boldsymbol{y})$.
Compute $\hat{\boldsymbol{p}}_{ref}$ from the whole symbol block using Eq. (1).

**Algorithm 2** Algorithm of Online Monitoring

**Input:** Offline estimated stationary probability vector $\hat{\boldsymbol{p}}_{ref}$, absolute error bound $\varepsilon_m$, minimum block length $r_{min}$, number of states $n$, and $\eta$–$\varepsilon$ relation function $f$.
**Output:** Stopping time $r_{stop}$, state vector $\hat{\boldsymbol{p}}(r_{stop})$ and FLAG.
Compute $\eta = f(\varepsilon_m)$.
Compute $r_{stop} = \max\left(\text{ceil}\left(\frac{n}{\eta}\right),\ r_{min}\right)$ by using Eq. (5).
Compute $\hat{\boldsymbol{p}}(r_{stop})$ from a symbol block $\boldsymbol{s}^{r_{stop}}$ by using Eq. (1).
**if** $\|\hat{\boldsymbol{p}}(r_{stop}) - \hat{\boldsymbol{p}}_{ref}\|_\infty \le \varepsilon_m$ **then**
  FLAG $= 0$.
**else**
  FLAG $= 1$.
**end if**

## Acknowledgement

The authors gratefully acknowledge the benefits of discussion with Professor Qiang Du.

## References

[1] A. Ray, Symbolic dynamic analysis of complex systems for anomaly detection, Signal Processing 84 (7) (2004) 1115–1130.
[2] X. Jin, K. Mukherjee, S. Gupta, A. Ray, S. Phoha, T. Damarla, Asynchronous data-driven classification of weapon systems, Measurement Science and Technology 20 (2009) 123001.
[3] A. Ray, Signed real measure of regular languages for discrete event supervisory control, International Journal of Control (12) (2005) 949–967.
[4] J.M. Flegal, M. Haran, Markov Chain Monte Carlo: can we trust the third significant figure, Statistical Science 23 (2) (2008) 250–260.
[5] R. Bapat, T. Raghavan, Nonnegative Matrices and Applications, Cambridge University Press, Cambridge, UK, 1997.
[6] G. Jones, M. Haran, B. Caffo, R. Neath, Fixed-width output analysis for Markov Chain Monte Carlo, Journal of the American Statistical Association 101 (2006) 1537–1547.
[7] S. Meyn, R. Tweedie, Markov Chains and Stochastic Stability, Springer-Verlag, 1993.
[8] B. Berg, Markov Chain Monte Carlo Simulations and their Statistical Analysis, World Scientific, Singapore, 2004.