Filtering and Prediction of Individual Sequences Corrupted By Noise Using the Lempel-Ziv Algorithm

Anelia Baruch and Neri Merhav
Dpt. EE, Technion, Haifa Israel
anelia@tx.technion.ac.il
merhav@ee.technion.ac.il

Abstract — **We address the problem of filtering and prediction of an individual binary sequence based on its *noisy* past, as an extension to [1]. The performance criterion investigated is the expected fraction of errors. We propose algorithms and compare their performance to that of the best finite state machine (FSM). We improve on previous results [1] by showing that optimum performance can be achieved by Lempel-Ziv-based estimation algorithms.**

I. Introduction

Let $\theta_1, \theta_2, \ldots$ be an arbitrary binary sequence corrupted by a Bernoulli noise process $\nu_1, \nu_2, \ldots$ with $Pr\{\nu_i = 1\} \triangleq p$. An observer accesses the noisy sequence $y_1, y_2, \ldots$, where $y_i = \theta_i \oplus \nu_i$, and $\oplus$ denotes addition modulo 2. The observer is interested in either estimating θ_i (filtering), or predicting θ_{i+1} (prediction), based on $y_1, y_2, \ldots, y_i$. We seek a universal estimator whose bit error probability is essentially as small as that of the best FSM, simultaneously for all θ. Previous work [3],[2] can be viewed as a special case of this filtering problem, where in [2] it was shown that there exists a sequential estimator whose asymptotic performance is as good as that of the best estimator that is implementable by a single-state machine. In [2], prediction without noise was considered and a sequential LZ-based predictor was shown to attain the finite state predictability of all infinite sequences. In this work, we improve on previous results of [1] where an asymptotically optimum sequential algorithm with growing memory was introduced. In this work we present a more practical, LZ-based algorithm that achieves the same goal.

A finite-state filter (FSF) with S states is a causal device that, upon receiving a sequence of observations $y_1, y_2, \ldots$, generates a sequence of estimates $\hat{\theta}_1, \hat{\theta}_2, \ldots$, while going through a sequence of states $s_1, s_2, \ldots$ that take on values in a finite set $\mathcal{S} = \{1, 2, \ldots, S\}$. The mechanism of the FSF is defined by a pair of deterministic functions f and g, where f is the *output function* that is given by $\hat{\theta}_i = f(s_i, y_i)$ for filtering and $\hat{\theta}_{i+1} = f(s_i)$ for prediction, and g is the *next-state function* that defines a recursive state update rule, according to $s_{i+1} = g(s_i, y_i)$. Let G_S be the set of all next-state functions of no more than S states. Henceforth, $x_i^j, i \leq j$, generically designates $(x_i, x_{i+1}, \ldots, x_j)$. Also, denote by g_k the k-th order Markovian next-state function whose state at time instant t is defined by $s_t = y_{t-k}^{t-1}$. For a given (f, g), let $e(\theta_1^n, \nu_1^n, (f,g)) \triangleq \frac{1}{n}\sum_{t=1}^{n} 1_{\{\hat{\theta}_t \neq \theta_t\}}$ be the fraction of errors attained when (f, g) is applied to y_1^n. Let $e_g(\theta_1^n) \triangleq \min_f E\{e(\theta_1^n, \nu_1^n, (f,g))\}$, and define the *FS filterability* of an infinite sequence θ by $e(\theta) = \lim_{S\to\infty} \lim_{n\to\infty} \min_{g\in G_S} e_g(\theta_1^n)$. The *aperiodic FS filterability*, $\bar{e}(\theta)$, is defined similarly, with the exception that the minimization is over the class of aperiodic machines, and the *Markovian filterability*, $\mu(\theta)$, is defined by $\lim_{k\to\infty} \lim_{n\to\infty} e_{g_k}(\theta_1^n)$.

[1]This work was supported by the ISF administered by the Israeli Academy of Sciences and Humanities.

II. Main Results

Our main result is a derivation of a scheme that asymptotically achieves $e(\theta)$. This scheme is based on the incremental parsing (IP) procedure of the LZ '78 algorithm, and can be viewed as a Markovian machine of increasing order. The transition between states is identical to that of the equivalent scheme in [2], apart of the fact that it is the noisy sequence $\{y_t\}$ which determines the states sequence rather than the clean one. The state at time instant t is the string of bits observed since the last phrase has terminated.

The estimation is as follows: denote by $N_t^y(s,x) = \sum_{i=1}^{t-1} 1_{\{s_i=s, y_{i+1}=x\}}$ the joint count of state s and the value of the *next noisy* bit being x, and let $\tilde{N}_t^y(s,x)$ be a randomized version of it, i.e., $\tilde{N}_t^y(s,x) = N_t^y(s,x) + z_x(t)\,(N_t(s))^{1/2}$ where $\{z_x(t)\}_{t=1}^n, x \in \{0,1\}$, are independent r.v.'s uniformly distributed over the interval $[0,1]$ and $N_t(s) = \sum_{i=1}^{t} 1_{\{s_i=s\}}$.

Let $N_t^\theta(s,x) = \sum_{i=1}^{t-1} 1_{\{s_i=s, \theta_i=x\}}$ be the joint count of state s and the value of the *current clean* bit being x. Now, an estimation of $N_t^\theta(s,x)$ is performed: $\hat{N}_t^\theta(s,x) = \frac{1}{1-2p}\sum_{i=1}^{t-1} 1_{\{s_i=s, y_i=x\}} - \frac{2p}{1-2p}N_t(s)$, and some auxiliary randomization is introduced which results in $\tilde{N}_t^\theta(s,x) = \hat{N}_t^\theta(s,x) + z_x(t)\,(N_t(s))^{1/2}$.

For prediction, the decision rule is $\hat{\theta}_{t+1} = x$ if $\tilde{N}_t^y(s_t, x) > \tilde{N}_t^y(s_t, 1-x)$. For filtering, the decision rule is $\hat{\theta}_t = x$ if $\tilde{N}_t^\theta(s_t, x) > \frac{p}{1-p}\tilde{N}_t^\theta(s_t, 1-x)$, and otherwise $\hat{\theta}_t = y_t$, where ties are broken arbitrarily. Denote by $e^{IP}(\theta_1^n)$ the expected fraction of errors made by this scheme.

Our first result is that, when prediction is concerned $e_{g_k}(\theta_1^n) - \min_{g\in G_S} e_g(\theta_1^n) \leq \frac{1}{1-2p}\sqrt{\frac{\ln S}{2(k+1)}}$ and therefore $\mu(\theta) = e(\theta)$. When both filtering and prediction are concerned we show that $e_{g_k}(\theta_1^n) - \min_{g\in G_s, g\ aperiodic} e_g(\theta_1^n) \leq O(\alpha(S,p))^k + \frac{k}{n}$ where $|\alpha(S,p)| < 1$ which implies that $\mu(\theta) \leq \bar{e}(\theta)$ for filtering. We further show that $e^{IP}(\theta_1^n) - e_{g_k}(\theta_1^n) \leq O\left(\frac{1}{\sqrt{\log n}}\right) + O\left(\frac{k}{\log n}\right)$. Combining these two observations it follows that the above described scheme achieves the FS filterability.

References

[1] A. Baruch, N. Merhav "Universal Filtering of Individual Sequences Corrupted by Noise", *Proc. ISIT 1998* p. 331 .

[2] M. Feder, N. Merhav and M. Gutman, "Universal Prediction of Individual Sequences," *IEEE Trans. Inform. Theory*, vol. 38, pp. 1258–1270, July 1992.

[3] J. V. Ryzin, "Asymptotic solutions of the sequential compound decision problem," *Ann.Math. Statist.*, vol. 37, pp. 954-975, 1966.

0-7803-5857-0/00/$10.00 ©2000 IEEE.

Authorized licensed use limited to: Technion Israel School of Technology. Downloaded on January 20, 2009 at 07:50 from IEEE Xplore. Restrictions apply.