CCIT Report No. 313

Universal Prediction of Random Binary Sequences in a Noisy Environment

Tsachy Weissman and Neri Merhav

ABSTRACT

Let $X=\{(X_t, Y_t)\}_{t\in Z}$, be a stationary time series where X_t is binary-valued and Y_t the *noisy observation* of X_t, is real-valued. Letting **P** denote the probability measure governing the joint process $\{(X_t, Y_t)\}$, we characterize $U(l,\mathbf{P})$, the optimal asymptotic average performance of a predictor allowed to base its prediction for X_t on $Y_1, \cdots, Y_{t-1}$, where performance is evaluated using the loss function l. It is shown that the stationarity and ergodicity of **P**, combined with an additional "conditional mixing" condition suffice to establish $U(l, \mathbf{P})$ as the fundamental limit for the almost sure asymptotic performance. As is made apparent, this finding can be regarded as an extension to the celebrated Shannon-McMillan-Breiman Theorem and $U(l, \mathbf{P})$ as a generalized notion of the Shannon entropy which can capture the sensitivity of the clean binary sequence to noise. For the case where $\mathbf{X}=\{X_t\}$ is governed by **P** and Y_t given by $Y_t = g(X_t, N_t)$ where g is any deterministic function and $\mathbf{N}=\{N_t\}$, the noise, is any i.i.d. process independent of **X** (namely, the case where the "clean" process **X** is passed through a fixed memoryless channel), it is shown that, analogously to the noiseless case, there exist *universal predictors* which do not depend on **P** yet attain $U(l, \mathbf{P})$. Furthermore, it is shown that in some special cases of interest (e.g., the Binary Symmetric Channel and the absolute loss function), there exist *twofold universal predictors* which do not depend on the noise distribution either. The existence of such universal predictors is established by means of an explicit construction which builds on recent advances in the theory of prediction of individual sequences in the presence of noise.