

UNIVERSAL COMPRESSOR- based statistical inference

ABSTRACT

New affordable convenient instruments often contribute to the development of science, music, etc.

My talk is on a new class of applications of a well-known instrument - UNIVERSAL COMPRESSORS (UC), such as zip, originated in the interface between the Information theory and Computer Science by the end of the last century.

In addition to their direct application, the lengths of compressed stationary ergodic strings give a good approximation to the likelihood of these strings which generally cannot be evaluated analytically. This fact is implied by the nearly minimal average length of compression which is achievable, iff the lengths of compressed strings approximate their likelihood.

This was formulated as the MDL principle by J. Rissanen in 1984 as a new principle of statistics since the Likelihood is the main tool of statistical inference.

My talk is on two celebrated implementations of the MDL-principle.

Literary texts are modeled as stationary ergodic sources after the groundbreaking Shannon work of 1949. I test homogeneity between literary texts thus contributing to their authorship attribution.

Another application is screening out active inputs of a general system modeled as a functional multivariate relationship disturbed by a 'colored noise'.

The latter topic is an alternative approach to the **Compressed Sparse Sensing** which became recently extremely popular.

Our theoretical results make feasible the following types of applications:

- i. 'Tagging Flags' to the users' accounts in a large computer network with abrupt change in users' profiles possibly caused by unauthorized intrusion into the system for their more detailed follow up study.
- ii. monitoring large corpora of audio/text strings, e.g. on-line forums or the phone call traffic in some areas, for 'flagging' matches to specific profiles of interest.