P4-17 Anomaly Detection of Sequence Data Based on T-Information

Ulrich Speidel† Stefan Jan Skudlarek†† Hirosuke Yamamoto††
† The University of Auckland †† The University of Tokyo

Compression-based distance of sequence data

Characteristics of T-information (Titchener, 1984): Suitable for short sequences, self-synchronization

Anomaly detection by distance matrix processing

Mapping to vector of medians

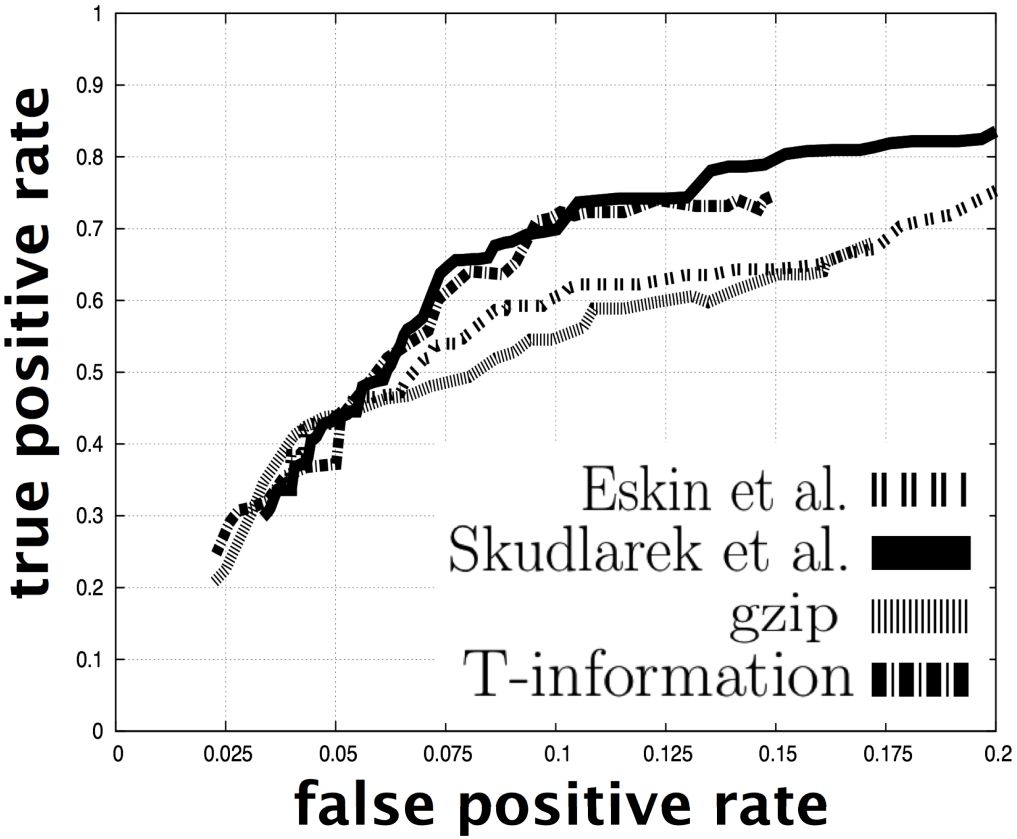
Final classification via modified k-means

Evaluation by non-numerical sequence data set (Schonlau, 2001)

Simulation of Masquerade attack

Short sequences

Large alphabet



Result: Superiority of T-information-based distance

References

[1] E. Eskin et al., A Geometric Framework for Unsupervised Anomaly Detection, Applications of Data Mining in Computer Security, Kluwer, 2002.

[2] S.J. Skudlarek, H. Yamamoto, Representative Sequence Selection in Unsupervised Anomaly Detection using Spectrum Kernel with Theoretical Parameter Setting, ICMLC2010, 2010.