High Dimensional Anomaly Detection

Venkatesh Saligrama (NSF CPS: 0932114)

Anomalies with spatial and temporal stamps arise in a number of applications including communication networks, traffic monitoring and video analysis. In these applications anomalies are temporally or spatially localized but otherwise unknown. We propose a novel graph-based statistical notion that unifies the idea of temporal and spatial locality. This notion lends itself to an elegant characterization of optimal decision rules and in turn suggests corresponding empirical rules based on local K-nearest neighbor distances. We compute scores for each data sample based on these local distances and declare data samples as containing local anomalies based on this score. We show that such rules not only asymptotically guarantee desired false alarm control but are also asymptotically optimal. We show that our approach overcomes the deficiencies of some popular “estimation-followed-by-detection” methods that suffer from resolution issues in limited data scenarios. We verify our algorithms on synthetic and real data sets.