The "what", "why", and "how" of Mining Noisy Data

Raghu Krishnapuram
IBM Research - India
Bangalore, India

Abstract:

Uncertain data is expected to comprise almost 80% of all available data in the near future. The problem of uncertain or noisy data is particularly severe in many of the newer sources of data such as Internet blogs, social media sites, mobile devices, and sensors. There is a strong need to devise new approaches in order to extract useful insights and intelligence from the abundant and rich information contained in such sources. This talk will present various types of uncertainty that occur in such data an outline a few possible approaches to realize the benefits. The talk will particularly focus on sources of noisy text data and highlight certain applications such as customer relationship management.