Data and Data Analysis in the New IT World

By
Norden E. Huang
Research Center for Adaptive Data Analysis
National Central University

With the dawn of the new IT world, we will find data everywhere. In fact, we are practically drowning in data. Data are the only connect we have with the reality; therefore, data analysis is the only way we can find out the truth. Consequently, with the increasing volume of data, 'Data analysis' should be very different from the traditional 'Data Processing', for 'Analysis' implies taking the data apart and examination the components critically; while as 'Processing' means only going through a given set of algorithms. Examples of analysis are given first by textural categorization, which literally took the plays by William Shakespeare and Christopher Marlowe apart to the individual level and study the interrelationship among the components. Such examinations helped to settle the authorship of famous plays by Shakespeare and Marlowe. Examples of this technique applied to bio-medical data in disease diagnosis and DNA differentiation will also be given.

Categorization is useful, but it does not provide a quantification of the complexity within the data. As biological systems have to adapt to their complex environments, this evolution process makes all biological system extremely complex. It should be noted that complexity is not identical to randomness, which can be measured by entropy. To provide a measure for complexity, the Multi-Scale Entropy (MSE) method is introduced, which gives the means of separating healthy from the unhealthy quantitatively. Additional examples will be given on the heart failure and the balance measures.

Powerful as the MSE method is, stationarity is an implied requirement for the method to work properly. To overcome this hurdle, the HHT (Empirical Mode Decomposition, EMD, and Hilbert Spectral Analysis) is introduced to remove any trend in the data. HHT is an adaptive and versatile method for data analysis. Examples of using HHT are given for studying climate trends and cycles. Other applications of adaptive data analysis covering speech analysis, engineering and sciences problems will be discussed briefly during the talk.

As the natural processes are mostly nonlinear and nonstationary, adaptive data analysis methods are the only logical way to analyze our data and to further our understanding the underlying physical processes. With the variety of data from different processes, being adaptive is critical. As human, we evolved and survive because we are adaptive. So, we should learn to be adaptive in our treatment of data in this new IT World.