New Paradigm for Knowledge Discovery

The Need for better approaches to knowledge discovery is paramount. A person can read these words and understand the message in real-time (one second or less) via neuron-based processing that has a cycle time of about 10 milliseconds, corresponding no more than 100 neural hops per second. In sharp contrast, modern high-performance computers run at trillions of operations per second, yet cannot perform real-time processing of images, text, or audio data. Consequently, this new paradigm is motivated by the human brain's far superior speed and depth of insight.

The Solution under development by ORNL's Cyber Space and Information Intelligence Research Group involves application of patented technology to a novel paradigm for knowledge discovery. This new paradigm is described in Jeff Hawkins' 2004 book, "On Intelligence," which elucidates four unique features of the human brain, but does not provide any approach for computer implementation. One feature is an *irreducible representation* for each item that the brain processes. A second feature is *auto-associativity* among items (e.g., recall of one line of a song that enables remembering the remainder). A third feature is *hierarchical* processing (e.g., relate the simplest spoken sounds [phonemes] to words, then to phrases, sentences, concepts, then prediction in similar circumstances as a measure of "understanding" or "knowledge"). The fourth feature involves *feed-forward* links in the processing hierarchy to make the appropriate connections among words, phrases, sentences, and concepts in the context of previous knowledge. The fourth feature also includes *feed-back* from higher-to-lower levels in the hierarchy for self-consistent extraction of the knowledge in terms of known words (rather than nonsense words), proper syntax, correct grammar, etc. Likewise, image processing extracts (for example) points, lines, polygons, object identification, scene familiarity, and scene changes. Indeed, the same neocortical processing paradigm extracts a hierarchical sequence of patterns for all sensory observations (time-serial data).

The Capability can be computationally implemented by application of patented ORNL technology. This ORNL technology converts time-serial data into a discrete state (or item) in an irreducible representation, $y(k)$, as a unique identifier for the k-th state. State-to-state transitions, $y(k) \rightarrow y(m)$, are captured by an irreducible representation for the connectivity between these connected states, thus providing auto-associativity among items. The discrete states can be viewed as nodes, and the linkages can be viewed as edges in a network (a graph in the mathematical) of relationships among items (ontology). Clusters of relationships in this complex network form the hierarchical structure of data processing to information and then to knowledge for decision support. Recall of auto-associative sequences enables feed-forward, $y(k) \rightarrow y(m)$, for higher-level inferences and feed-back, $y(m) \rightarrow y(k)$, for self-consistent knowledge extraction. The occurrence frequency of these discrete states (items) is called a "feature vector," from which knowledge discovery inferences are extracted via processing by a "vector machine." Moreover, links among the discrete states are nodes that form a directed graph, or "feature graph," which is processed by a "graph machine" for knowledge discovery.

Lee Hively, PhD
Cyberspace Sciences & Information Intelligence Research (CSIIR) Group
Oak Ridge National Laboratory P.O. Box 2008, Oak Ridge, TN 37831-6418
www.ioc.ornl.gov