Pattern Synthesis for Large-Scale Pattern Recognition

P. Viswanath  
*Indian Institute of Science, India*

M. Narasimha Murty  
*Indian Institute of Science, India*

Shalabh Bhatnagar  
*Indian Institute of Science, India*

**INTRODUCTION**

Two major problems in applying any pattern recognition technique for large and high-dimensional data are (a) high computational requirements and (b) curse of dimensionality (Duda, Hart, & Stork, 2000). Algorithmic improvements and approximate methods can solve the first problem, whereas feature selection (Guyon & Elisseeff, 2003), feature extraction (Terabe, Washio, Motoda, Katai, & Sawaragi, 2002), and bootstrapping techniques (Efron, 1979; Hamamoto, Uchimura, & Tomita, 1997) can tackle the second problem. We propose a novel and unified solution for these problems by deriving a *compact and generalized abstraction* of the data. By this term, we mean a compact representation of the given patterns from which one can retrieve not only the original patterns but also some artificial patterns. The compactness of the abstraction reduces the computational requirements, and its generalization reduces the curse of dimensionality effect. Pattern synthesis techniques accompanied with compact representations attempt to derive compact and generalized abstractions of the data. These techniques are applied with *nearest neighbor classifier (NNC)*, which is a popular nonparametric classifier used in many fields, including data mining, since its conception in the early 1950s (Dasarathy, 2002).

**BACKGROUND**

Pattern synthesis techniques, compact representations and its application with NNC are based on more established fields:

- **Pattern Recognition**: Statistical techniques, parametric and nonparametric methods, classifier design, nearest neighbor classification, curse of dimensionality, similarity measures, feature selection, feature extraction, prototype selection, and clustering techniques.
  - **Data Structures and Algorithms**: Computational requirements, compact storage structures, efficient nearest neighbor search techniques, approximate search methods, algorithmic paradigms, and divide-and-conquer approaches.
  - **Database Management**: Relational operators, projection, cartesian product, data structures, data management, queries, and indexing techniques.

**MAIN THRUST**

Pattern synthesis, compact representations followed by its application with NNC, are described in this section.

**Pattern Synthesis**

Generation of artificial new patterns by using the given set of patterns is called *pattern synthesis*. There are two broad ways of doing pattern synthesis: *model-based pattern synthesis* and *instance-based pattern synthesis*.

In model-based pattern synthesis, a model (such as the Hidden Markov model) or description (such as probability distribution) of the data is derived first and is then used to generate new patterns. This method can be used to generate as many patterns as needed, but it has two drawbacks. First, any model depends on the underlying assumptions; hence, the synthetic patterns generated can be erroneous. Second, deriving the model might be computationally expensive. Another argument against this method is that if pattern classification is the purpose, then the model itself can be used without generating any patterns at all!

Instance-based pattern synthesis, on the other hand, uses the given training patterns and some of the properties of the data. It can generate only a finite number of
3 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the product's webpage:
www.igi-global.com/chapter/pattern-synthesis-large-scale-pattern/10724?camid=4v1

www.igi-global.com/e-resources/library-recommendation/?id=1

Related Content

Business Data Warehouse: The Case of Wal-Mart
www.igi-global.com/chapter/business-data-warehouse/7798?camid=4v1a

Search Situations and Transitions
www.igi-global.com/chapter/search-situations-transitions/11052?camid=4v1a

Text Mining by Pseudo-Natural Language Understanding
www.igi-global.com/chapter/text-mining-pseudo-natural-language/11085?camid=4v1a

Recovery of Data Dependencies
Hee Beng Kuan Tan and Yuan Zhao (2005). Encyclopedia of Data Warehousing and Mining (pp. 947-949).
www.igi-global.com/chapter/recovery-data-dependencies/10732?camid=4v1a