A SIMILARITY-BASED LEARNING ALGORITHM USING DISTANCE TRANSFORMATION

ABSTRACT

Numerous theories and algorithms have been developed to solve vectorial data learning problems by searching for the hypothesis that best fits the observed training sample. However, many real-world applications involve samples that are not described as feature vectors, but as (dis)similarity data. Converting vectorial data into (dis)similarity data is more easily performed than converting similarity data into vectorial data. This study proposes a stochastic iterative distance transformation model for similarity-based learning. The proposed model can be used to identify a clear class boundary in data by modifying the (dis)similarities between examples. The experimental results indicate that the performance of the proposed method is comparable with those of various vector-based and proximity-based learning algorithms.

EXISTING SYSTEM

In existing system vectorial data and pairwise proximity data used for machine learning. Vectorial data consist of a set of examples typically represented as data points in an n-dimensional euclidean space, in which each dimension corresponds to a feature.

Pairwise proximity data reveal pairwise (dis)similarities between data points, generally presented in the form of a two-dimensional (2D) matrix.
DRAWBACK OF EXISTING SYSTEM

- Hypotheses are restricted by prespecified features (i.e., dimensions), which might not always have been properly determined.
- Searching for the hypothesis that best fits the observed training data.

PROPOSED SYSTEM

The most natural method for (dis)similarity based learning is the k-nearest-neighbor (k-NN) algorithm, which relates to the exemplar model of human learning.

Several variants of the nearest-neighbor approach have been proposed to formulate the similarity-based classification problem into a standard vector-based learning problem by representing.

ADVANTAGE OF PROPOSED SYSTEM

- The nearest-centroid classifier, which adapts the prototype model of human learning transforming the pairwise (dis)similarity data directly to restore the applicability of the nearest-neighbor principle in classification.
- This study proposes a novel distance transformation method, Transform Distance, to clearly reveal a class boundary in data by modifying the proximities.
SYSTEM SPECIFICATION

Hardware Requirements

- System: Pentium IV 2.4 GHz
- Hard Disk: 40 GB
- Floppy Drive: 1.44 Mb
- Monitor: 15 VGA Colour
- Mouse: Logitech
- Ram: 512 Mb

Software Requirements

- Operating system: Windows XP
- Technology Used: Microsoft .NET
- Backend Used: SQL Server