International Journal of Innovative Computing, Information and Control
Volume **4**, Number **6**, June **2008**
 ISSN 1349-4198
pp. 1383–1392

A HIERARCHICAL SEARCH AND UPDATING DATABASE STRATEGY FOR LAZY LEARNING

Tianhong Pan[1,2] and Shaoyuan Li[2]

[1]School of Electrical and Information Engineering
Jiangsu University
301 Xuefu Road, Zhenjiang, Jiangsu 212013, P. R. China
thpan@ujs.edu.cn

[2]Institute of Automation
Shanghai Jiao Tong University
800 Dongchuan Road, Minhang, Shanghai 200240, P. R. China
syli@sjtu.edu.cn

Received February 2007; revised August 2007

Abstract. *Lazy learning is a memory-based learning technique that uses a query-based approach to estimate the best local model configuration by selecting neighbors of the query. The bottleneck of this algorithm is management of large datasets and searching for the relevant neighbors. This paper aims to deal with this issue and the k-Means cluster algorithm is cast into the lazy learning framework. By using this strategy, the nearest neighbors searching process can be converted into a hierarchical searching process with two levels and a lot of searching time for lazy learning can be saved. A novel criterion between two samples is proposed, and based on this criterion, the k-Vector Nearest Neighbors (k-VNN) is used to find the neighbors of the query, The database can quickly be updated without the need for any further computation. This updating strategy can save a lot of memory space and decrease the neighbor-searching time. These techniques have been successfully applied to estimate a nonlinear function.*
Keywords: Lazy learning, k-means cluster, k-vector nearest neighbors

1. **Introduction.** Lazy Learning is a local modeling technique that has attracted much attention in the past few years due to its simplicity and ease in use in identification of non-linear systems. It is also known as Just-in-Time Learning [1], Locally Weighted Learning [2,3] or Model-on-Demand [4] in the literature. Compared with the traditional modeling methods, lazy learning exhibits three characteristics [5]. Firstly, it defers processing of their inputs until it receives requests for information; it simply stores those inputs for future use. Next, it replies to information requests by combining the stored (e.g., training) data. Finally, it discards the constructed answer and any intermediate results. This algorithm has an advantage in avoiding the difficulty of finding an appropriate structure for a global model and is useful for modeling complex nonlinear systems. Another advantage is its inherently adaptive nature, which is achieved by storing the current measured data into the database [6]. Owing to those merits, lazy learning has been applied to a number of fields, such as: artificial intelligence [7], prediction of the stock market [8], robotic control [3], chaos system [9], chemical process [10,16-17], etc.

A key issue of this algorithm is how to select the neighbors of a query. Most existing lazy learning algorithms are based on k-Nearest Neighbors [6], k-Surrounding Neighbors [11] and k-Bipartite Neighbors [12]. Although these algorithms work well in many applications, they suffer from computational burden in the high dimensional samples or in a large database. Because the nearest neighbors of each query are obtained by using an exhaustive