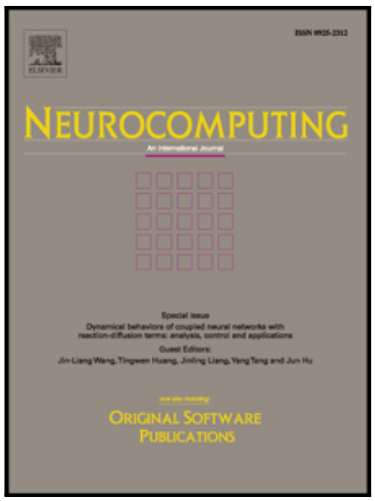

Probabilistic Group Nearest Neighbor Query Optimization Based on Classification Using ELM

Jiajia Li, Xiufeng Xia, Xiangyu Liu, Botao Wang, Dahai Zhou, Yunzhe An

PII: S0925-2312(17)31396-6
DOI: 10.1016/j.neucom.2017.05.095
Reference: NEUCOM 18786

To appear in: *Neurocomputing*

Received date: 30 September 2016
Revised date: 29 April 2017
Accepted date: 9 May 2017

Please cite this article as: Jiajia Li, Xiufeng Xia, Xiangyu Liu, Botao Wang, Dahai Zhou, Yunzhe An, Probabilistic Group Nearest Neighbor Query Optimization Based on Classification Using ELM, *Neurocomputing* (2017), doi: 10.1016/j.neucom.2017.05.095

Available online at www.sciencedirect.com

ScienceDirect

Neurocomputing

Neurocomputing 00 (2017) 1–14

Probabilistic Group Nearest Neighbor Query Optimization Based on Classification Using ELM

Jiajia Li[a], Xiufeng Xia[a], Xiangyu Liu[a], Botao Wang[b], Dahai Zhou[a], Yunzhe An[a]

[a]*College of Computer Science, Shenyang Aerospace University, China*
[b]*College of Computer Science, Northeastern University, China*

Abstract

The *probabilistic group nearest neighbor(PGNN)* query , which returns all the uncertain objects whose probabilities of being the group nearest neighbor (GNN) results exceed a user-specified threshold, is widely used in uncertain database. Most existing work for answering PGNN queries adopted a general framework which consist of three phases: *spatial pruning, probabilistic pruning, refinement*. In the probabilistic pruning phase, dividing the uncertain regions into many partitions to derive a tighter probabilities bounds is a common method. However, there is a tradeoff between the the computational cost of probabilistic pruning phase and refinement phase controlled by the granularity of the partitions. In this paper, we study the problem of setting the optimal granularity of the partitions for uncertain objects, and propose a new framework for PGNN queries based on granularity classification using ELM such that the overall cost is minimized. In addition, to improve the accuracy of classification and make the classifier applicable to the dynamic environment, a plurality voting method and a dynamic classification strategy are proposed respectively. Extensive experiments shows that compared with the default granularities of the partitions, the granularities chosen by ELM classifiers are more proper, which further improves the performance of PGNN query algorithm. In addition, ELM outperforms SVM with regard to both the response time and classification accuracy.

Keywords: Probabilistic group nearest neighbor query, Uncertain data, Classification, Extreme learning machine.

1. Introduction

Group nearest neighbor(GNN) query, which returns the object that minimizes the sum of distances to multiple query points, is one of the important variants of nearest neighbor(NN) query. This type of query has many applications in clustering [1], outlier detection [2] and facilities management [3], and has been studied extensively on certain data[3, 4, 5, 6].

However, uncertain data are inherent in numerous emerging applications due to limitations of measuring equipment, delayed data updates, or privacy protection. Probabilistic GNN(PGNN) query over uncertain data, which is to find all the uncertain objects whose probabilities of being the GNN results exceed a user-specified threshold, has been proposed by Lian et al. [7]. The algorithm bounds all the query points as an MBR to reduce the computation of distances, and pre-computes several (1-β)-hyperspheres for each object in a preprocessing step. The idea of this algorithm is that the object can be pruned with a probability of at least (1-β) if its corresponding sphere is pruned.

The algorithm above is not efficient if the objects have arbitrary shapes of their uncertain regions or the query points are distributed sparsely. To overcome these limitations, two pruning algorithms are proposed in our previous work

Email addresses: jiajia4487@163.com (Jiajia Li), xiaxiufeng@sau.edu.cn (Xiufeng Xia), neulxy@gmail.com (Xiangyu Liu), wangbotao@ise.neu.edu.cn (Botao Wang), zhoudahai.sau@189.cn (Dahai Zhou), anyunzhe@163.com (Yunzhe An)