International Journal of Cognitive Informatics and Natural Intelligence
Volume 10 • Issue 4 • October-December 2016

Natural Neighbor Reduction Algorithm for Instance-based Learning

Lijun Yang, Chongqing University, Chongqing, China
Qingsheng Zhu, Chongqing University, Chongqing, China
Jinlong Huang, Chongqing University, Chongqing, China
Dongdong Cheng, Chongqing University, Chongqing, China
Cheng Zhang, Chongqing University, Chongqing, China

ABSTRACT

Instance reduction is aimed at reducing prohibitive computational costs and the storage space for instance-based learning. The most frequently used methods include the condensation and edition approaches. Condensation method removes the patterns far from the decision boundary and do not contribute to better classification accuracy, while edition method removes noisy patterns to improve the classification accuracy. In this paper, a new hybrid algorithm called instance reduction algorithm based on natural neighbor and nearest enemy is presented. At first, an edition algorithm is proposed to filter noisy patterns and smooth the class boundaries by using natural neighbor. The main advantage of the algorithm is that it does not require any user-defined parameters. Then, using a new condensation method based on nearest enemy to reduce instances far from decision line. Through this algorithm, interior instances are discarded. Experiments show that the hybrid approach effectively reduces the number of instances while achieves higher classification accuracy along with competitive algorithms.

KEYWORDS

Instance-based Learning, Instance Reduction, Natural Neighbor, Nearest Enemy

1. INTRODUCTION

Instance reduction is a crucial task in instance-based cognitive machine learning algorithms. Instance-based learning algorithms use the whole prototypes from the training set to construct inference structures. The k-Nearest Neighbor Rule (KNN) (Cover & Hart, 1967) is a well-known example of instance-based learning algorithms. It requires a large memory space since the whole training set has to be stored which may be an excessive amount of storage for large dataset and leads to a large computation time in the classification stage.

To overcome the shortages, the Condensing Nearest Neighbor (CNN) (Hart, 1968) algorithm is proposed by Hart to reduce the number of training set by using the nearest neighbor decision. It removes any well classified instance using the nearest neighbor rule and obtains a consistent subset that does not affect the classification accuracy of the whole training set. The basic idea is that patterns located in decision boundary are crucial to the classification, but those far from the boundary have little effect. The Edited Nearest Neighbor (ENN) (Wilson, 1972) rule is another attempt of instance reduction, it aims to remove noisy patterns that are not correctly classified by their nearest neighbors. Moreover, hybrid methods are often used to compute a subset of the training set combining the characteristics of edition and condensation methods, and is widely used in the field of instance reduction problems.

DOI: 10.4018/IJCINI.2016100103

Usually, Edition method has been used in various condensation methods as a noisy pre-processing filter to eliminate the noisy patterns.

CNN is classic reduction algorithm, but it is very sensitive to noisy patterns and it may keep some patterns that far from the decision boundary. In recent years, some new reduction techniques are developed to overcome drawbacks and improve the performance. In 2007, Angiulli (Angiulli, 2007) introduced a novel algorithm, called Fast Nearest Neighbor Condensation (FCNN) rule which computing a training-set-consistent subset for the nearest neighbor decision rule and can deal with large data classification. Fayed and Atuya (Fayed & Atiya, 2009) presented a sample and effective prototype reduction algorithm, namely, the template reduction for KNN (TRKNN). The basic idea is based on defining the so-called chain which is a sequence of nearest neighbors and set a cutoff value for distances to obtain the patterns close to the classification boundary as the "condensed set". Nikolaidis (Nikolaidis, Goulermas, & Wu, 2011) proposed a class boundary preserving algorithm (CBP) which uses the reachable set to locate the border points, then prunes the training set. Later, Nikolaidis (Nikolaidis, Rodriguez-Martinez, Goulermas, & Wu, 2012) introduces spectral instance reduction (SIR) algorithm to partition the dataset into border and internal instances. In 2015, the binary nearest neighbor tree algorithm (BNNT) (Li & Wang, 2015) is presented to obtain the prototype set through selecting and generating prototypes from the binary nearest neighbor tree. Two strategies are used to select and generate prototype. Those selected patterns have different class labels in a tree. When the tree locates in a class interior, the centroid pattern is generated to replace these tree nodes.

The mentioned algorithms above are to keep the patterns have higher contribution for pattern classification and remove the vast number of inner patterns and all of outlier patterns. There is another strategy to lower the high computation cost and storage requirements when dealing with the large datasets, which utilizes some partitioning strategies to divide the whole space into several smaller spaces. The typical method is cluster-based prototype reduction algorithm (Lumini & Nanni, 2006; Mollineda, Ferri, & Vidal, 2002). An example is the Prototype Selection by Clustering (PSC) (Olvera-López, Carrasco-Ochoa, & Martínez-Trinidad, 2010) that divide the training set into many clusters and analyzes nonhomogeneous clusters to find the border instances. Instance Reduction Algorithm using Hyperrectangle Clustering (IRAHC) (Hamidzadeh, Monsefi, & Sadoghi Yazdi, 2015) is the second example. The different between PSC and IRAHC is different cluster methods and IRAHC uses the mean of the interior instance as its representative.

Moreover, some new techniques (Perlovsky & Kuvich, 2015) have been emerged for solving instance reduction problem. For instance, neural network learning algorithm (Kim, 2006; Wang, 2015), swarm intelligence optimization algorithm (Cano, Herrera, & Lozano, 2003; Miloud-Aouidate & Baba-Ali, 2012; Nanni & Lumini, 2009), fuzzy set method (Fan, 2015; Verbiest, Cornelis, & Herrera, 2013; Xiao-Meng, Xiao-Peng, Jun-Hai, & Meng-Yao, 2011) etc.

Instance reduction is a multi-objective optimization problem. The overall performance of reduction technique is not characterized only by the classification accuracy they exhibit, but also by the reduction ratio they achieve. However, most of the existing instance reduction methods do not obtain good results in terms of both accuracy and reduction percentage. In addition, besides, much noises in data set will greatly affect the performance of the algorithm. To solve these problems, a new hybrid algorithm called instance reduction algorithm based on Natural Neighbor and Nearest Enemy (3NE, for short) is introduced. Natural neighbor is a new form of neighbor just like *k*-nearest neighbor. The original purpose of proposed natural neighbor is to solve the choice of parameter *k* of *k*-nearest neighbor. In this paper, we use it as a pre-processing to eliminate noisy patterns, and present an adaptive edition method called Edited Natural Neighbor rule (ENaN). Subsequently, in order to reduce redundant instances, we propose a new condensation method based on nearest enemy to select instances close to decision line. Through the condensation method, we effectively divide the patterns into border set and internal set. Patterns nearby the class boundary are reserved and the interior patterns are removed.

This work is organized as follows. Section 2 describes the details of the proposed algorithm. Section 3 shows the details of experiments. Finally, Section 4 is the conclusions.

Symmetry vs. Duality in Logic: An Interpretation of Bi-Logic to Model Cognitive Processes Beyond Inference

Giulia Battilotti (2014). *International Journal of Cognitive Informatics and Natural Intelligence (pp. 83-97).*

www.igi-global.com/article/symmetry-vs-duality-in-logic/133297?camid=4v1a

Analysis of Cognitive Machines in Organizations

Farley Simon Nobre, Andrew M. Tobias and David S. Walker (2009). *Organizational and Technological Implications of Cognitive Machines: Designing Future Information Management Systems (pp. 99-110).*

www.igi-global.com/chapter/analysis-cognitive-machines-organizations/27875?camid=4v1a