

## Accepted Manuscript

An efficient instance selection algorithm for k nearest neighbor regression

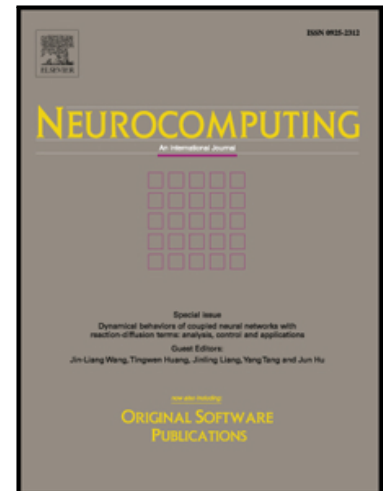
Yunsheng Song, Jiye Liang, Jing Lu, Xingwang Zhao

PII: S0925-2312(17)30688-4  
DOI: [10.1016/j.neucom.2017.04.018](https://doi.org/10.1016/j.neucom.2017.04.018)  
Reference: NEUCOM 18351

To appear in: *Neurocomputing*

Received date: 30 April 2016  
Revised date: 24 January 2017  
Accepted date: 3 April 2017

Please cite this article as: Yunsheng Song, Jiye Liang, Jing Lu, Xingwang Zhao, An efficient instance selection algorithm for k nearest neighbor regression, *Neurocomputing* (2017), doi: [10.1016/j.neucom.2017.04.018](https://doi.org/10.1016/j.neucom.2017.04.018)



This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

# An efficient instance selection algorithm for $k$ nearest neighbor regression

Yunsheng Song<sup>a</sup>, Jiye Liang<sup>\*,a</sup>, Jing Lu<sup>b</sup>, Xingwang Zhao<sup>a</sup>

<sup>a</sup> *Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, School of Computer and Information Technology, Shanxi University, Taiyuan, 030006, Shanxi, China*

<sup>b</sup> *Shanxi Meteorological Administration, Taiyuan, 030006, Shanxi, China*

---

## Abstract

The k-Nearest Neighbor algorithm(kNN) is an algorithm that is very simple to understand for classification or regression. It is also a lazy algorithm that does not use the training data points to do any generalization, in other words, it keeps all the training data during the testing phase. Thus, the population size becomes a major concern for kNN, since large population size may result in slow execution speed and large memory requirements. To solve this problem, many efforts have been devoted, but mainly focused on kNN classification. And now we propose an algorithm to decrease the size of the training set for kNN regression(DISKR). In this algorithm, we firstly remove the outlier instances that impact the performance of regressor, and then sorts the left instances by the difference on output among instances and their nearest neighbors. Finally, the left instances with little contribution measured by the training error are successively deleted following the rule. The proposed algorithm is compared with five state-of-the-art algorithms on 19 datasets, and experiment results show it could get the similar prediction ability but have the lowest instance storage ratio.

*Key words:* Instance selection, Nearest neighbor, Regression, Data reduction, Significant difference

---

\*Corresponding author

*Email addresses:* sys\_sd@126.com (Yunsheng Song), ljy@sxu.edu.cn (Jiye Liang), sxsqxj1ws@163.com (Jing Lu), zhaoxw84@163.com (Xingwang Zhao)

Download English Version:

<https://daneshyari.com/en/article/4947371>

Download Persian Version:

<https://daneshyari.com/article/4947371>

[Daneshyari.com](https://daneshyari.com)