

DETERMINING THE PREPROCESSING CLUSTERING ALGORITHM IN RADIAL BASIS FUNCTION NEURAL NETWORK

(Menentukan Alkhwarizmi Berkelompok Prapemprosesan dalam
Rangkaian Neural Fungsi Asas Jejari)

SAU LOONG ANG, HONG CHOON ONG & HENG CHIN LOW

ABSTRACT

Radial Basis Function Networks have been widely used to approximate and classify data. In the common model for radial basis function, the centres and spreads are fixed while the weights are adjusted until it manages to approximate the data. There exist some problems in finding the best centres for the hidden layer of Radial Basis Function. Even though some clustering methods like K-means or K-median are used in finding the centres, there are no consistent results that show which one is better. The main objective in this study is to determine the better method to be used to find the centres in the Radial Basis Functional Link Nets for data classification. Three types of method used in this study to find the centres include random selections, K-means clustering algorithm and also K-median clustering algorithm. The effects of K-means and K-median clustering algorithms on centres selection for Radial Basis Functional Link Nets in terms of accuracy and speed are shown in this study. To determine which clustering method is better, we calculate the preliminary Mardia's skewness. Therefore the skewness of the data is calculated to choose between the K-means or K-median clustering method in finding the centre of Radial Basis Function Network. Besides, the initial selection criterion using Mardia's skewness is able to show the improvement of efficiency in data classification. We use two sets of real data to demonstrate our result.

Keywords: Radial Basis Functional Link Nets; K-means; K-median

ABSTRAK

Rangkaian Fungsi Asas Jejari telah digunakan dengan meluas untuk menganggarkan dan mengelaskan data. Model biasa bagi Fungsi Asas Jejari menentukan pusat, serakan dan menyesuaikan pemberat sehingga ia dapat menganggarkan data. Terdapat beberapa masalah dalam mencari pusat terbaik bagi lapisan tersembunyi dalam Rangkaian Fungsi Asas Jejari. Walaupun beberapa kaedah berkelompok seperti min-K atau median-K digunakan untuk mencari pusat, namun tidak terdapat keputusan yang konsisten yang menunjukkan keputusan terbaik. Objektif utama dalam kajian ini adalah untuk menentukan kaedah yang lebih baik untuk mendapatkan pusat dalam Rangkaian Pautan Fungsi Asas Jejari untuk mengelaskan data. Tiga jenis kaedah digunakan dalam kajian ini untuk mencari pusat, antaranya pemilihan pusat secara rawak, algoritma berkelompok min-K dan juga Alkhwarizmi berkelompok median-K. Kesan Alkhwarizmi berkelompok min-K dan median-K dalam pemilihan pusat bagi Rangkaian Pautan Fungsi Asas Jejari, dari segi kejituan dan kelajuan ditunjukkan dalam kajian ini. Untuk menentukan kaedah kelompok yang lebih baik, kami mengira kepencongan Mardia awal. Oleh itu, kepencongan data dikira untuk memilih sama ada kaedah berkelompok min-K atau median-K dalam pencarian pusat bagi Rangkaian Pautan Fungsi Asas Jejari. Kriteria pemilihan awal yang menggunakan kepencongan Mardia dapat menunjukkan peningkatan keberkesanan dalam pengelasan data. Dua set data sebenar digunakan untuk memperlihatkan hasil yang diperolehi.

Kata kunci: Rangkaian Sambungan Fungsian Asas Jejari; min-K; median-K

References

- Ang S.L., Ong H.C. & Low H.C. 2006. Comparison between modified radial basis functional link nets and radial basis function. *Proceedings of the 2nd Indonesia-Malaysia-Thailand Growth Triangle (IMT-GT), Regional Conference on Mathematics, Statistics and Applications* 4: 120-125.
- Blake C. 1998. Repository of machine learning databases. University of California, Irvine. <http://www.ics.uci.edu/~mllearn/MLREpository.html> (27 January 2007).
- Buchtala O., Klimek M., & Sick B. 2005. Evolutionary optimization of radial basis function classifiers for data mining applications. *IEEE Transactions on Systems, Man, and Cybernetics* 35(5): 928-947.
- Ester M., Kriegel H.P., Sander J. & Xu X.W. 1996. A density-based algorithm for discovering clusters in large spatial database with noise. *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining* (KDD-96), 226-331.
- Haykin S. 1999. *Neural Networks. A Comprehensive Foundation*. Upper Saddle River, NJ: Prentice Hall.
- Khan S.S. & Amir A. 2004. Cluster center initialization algorithm for K-means clustering. *Pattern Recognition Letters* 25: 1293-1302.
- Khattree R. & Naik D.N. 1995. *Applied Multivariate Statistics with SAS Software*. Cary, NC: SAS Institute Incorporated.
- Lo J.T.H. 1998. Mutilayer perceptrons and radial basis function are universal robust approximators. *Neural Networks Proceedings, IEEE World Congress on Computational Intelligence* 2: 1311-1314.
- Looney C.G. 1997. *Pattern Recognition Using Neural Networks*. New York: Oxford University Press.
- Looney C.G. 2002. Radial basis functional link nets and fuzzy reasoning. *Neurocomputing* 48(1-4): 489-509.
- Macqueen J. 1967. Some methods for classification and analysis of multivariate data. *Proceedings of the 5th Berkeley Symposium on Probability and Statistics*. Berkeley: University of California Press.
- Mardia K.V. 1970. Measures of multivariate skewness and kurtosis with applications. *Biometrika* 57(3): 519-530.
- Mico L. & Oncina J. 2001. An approximate median search algorithm in non-metric spaces. *Pattern Recognition Letters* 22(10): 1145-1151.
- Pao Y.H., Park G.H. & Sobajic D.J. 1994. Learning and generalization characteristic of the random vector functional link net. *Neurocomputing* 6: 163-180.
- Schwenker F., Kestler H.A. & Palm G. 2001. Three learning phases for radial-basis-function networks. *Neural Networks* 14: 439-458.
- Sgarbi M., Cola V. & Reyneri L.M. 1998. A comparison between weighted radial basis functions and wavelet networks. *Proceedings of the 6th European Symposium on Artificial Neural Network* 10(4) :13-19.
- Sing J.K., Basu D.K. & Nasipuri M. 2003. Improved K-means algorithm in the design of RBF neural networks. *Conference on Convergent Technologies for Asia-Pacific Region, Tencon*. 2: 841-845.
- Tou J.T. 1974. Dynoc-a dynamic optimal cluster seeking technique. *Int J. Comput. Inf. Sci.* 8 (6): 541-547.
- Witten I.H. & MacDonald B.A. 1988. Using concept learning for knowledge acquisition. *International Journal of Man-Machine Studies* 27: 349-370.
- Zhang G.Z. & Huang D.S. 2004. Radial basis function neural network optimized by a genetic algorithm for soybean protein sequence residue spatial distance prediction. *IEEE Congress on Evolutionary Computational, CEC 2004* 1: 1015-1019.

School of Mathematical Sciences

Universiti Sains Malaysia

11800 USM

Penang, MALAYSIA

E-mail: dellphin1223@hotmail.com*, hcong@cs.usm.my, hclow@cs.usm.my

* Corresponding author