Abstract

Feature selection is the principal step in classification problems with attributes of high dimension. It may also be considered as a problem to determine the subset of terms in training corpus, which maximizes the classifier’s performance. Most of the machine learning algorithms have tainted performance in high dimensional feature space. In this paper, a novel feature selection method based on Ant Colony Optimization, a swarm intelligence algorithm is proposed. Ant Colony Optimization is a metaheuristic algorithm used to increase the ability of finding high quality solutions to NP-hard problems. The heuristic information required for the optimization process is obtained through a chi-square based statistical method, CHIR which results in fast convergence. Performance of the classifier with features selected by proposed method is compared to the feature selected by conventional chi-square and CHIR methods. It is found that the proposed algorithm identifies better feature set than the conventional methods.

Reference


Elias F. Combarro, Elena Montanes, Irene Diaz, Jose Ranilla, and Ricardo Mones. Introducing A Family Of Linear Measures For Feature Selection In Text Categorization, IEEE Transactions on Knowledge and Data Engineering, Vol., 17, No. 9, 2005, 1223-1232.


Xiao-Bing Xue and Zhi-Hua Zhou. Distributional Features for Text Categorization, IEEE Transactions on Knowledge and Data Engineering, Vol., 21, No. 3 2009, 428-442.


Thomas Stutzle and Holgar Hoos. Max-Min Ant System And Local Search For The Traveling Salesman Problem, IEEE Conference 1997


Index Terms

Computer Science

Pattern Recognition
Key words

Machine learning          feature selection

Ant colony optimization

chi-square method