Abstract

The Internet has been in a state of explosive expansion over the last decade and a half. The addition of numerous web pages to the World Wide Web by a vast array of authors on a plethora of topics leaves behind the problem of organizing these web pages in order to improve search results leading to more relevant information. In this paper, a modified attribute weighted dynamic k-Nearest Neighbor classification algorithm, using k-Means clustering, is proposed. This presents a solution to the automatic classification of Web Pages on the WWW, supported by the adaptive dynamic nature of the algorithm. Web pages are classified based on the class distribution of the pages in their neighborhood. Attribute weighting is used primarily to improve classification accuracy in cases of imbalanced class distribution. Empirical results observed show good classification accuracy, while at the same time, improving on other shortcomings of the traditional k-NN classification model.

References

- Xu Y. , and Wang H. , A new Feature Selection method based on support vector
Dynamic k-NN with Attribute Weighting for Automatic Web Page Classification (Dk-NNwAW)

- Eui-Hong Han, Karypis G., and Kumar V., Text Categorization Using Weight Adjusted k-Nearest Neighbor Classification, 1999.
- Yong Z., Youwen L., and Shixiong X., An Improved KNN Text Classification Algorithm

Index Terms

Computer Science
Information Sciences

Keywords