عنوان فارسی مقاله:
یک الگوریتم تکاملی هیبریدی برای انتخاب ویژگی در داده کاوی

عنوان انگلیسی مقاله:
A hybrid evolutionary algorithm for attribute selection in data mining

توجه!
این فایل تنها قسمتی از ترجمه میباشد. برای تهیه مقاله ترجمه شده کامل با فرمت ورد (قابل ویرایش) همراه با نسخه انگلیسی مقاله، اینجا کلیک نمایید.
6. Conclusions

This paper has proposed a hybrid evolutionary algorithm for attribute selection in data mining. The GA-SVM hybrid incorporates the stochastic nature of genetic algorithms together with the vast capability of support vector machines in the search for an optimal set of attributes. The eradication of the redundant attributes using the GA-SVM hybrid improves the quality of the data sets and enables better classification of future unseen data. The proposed GA-SVM hybrid was validated upon 5 data sets obtained from UCI machine learning repository. Results collated have shown that the proposed hybrid is able to produce a high average classification accuracy that is comparable or better than some of the established classifiers in the data mining community. The simulations carried out also showcase the statistical consistency of the GA-SVM hybrid, which is evident from the histogram analysis and box plots. Secondary improvements to the hybrid included the utilization of a correlation measure to improve the average fitness of a chromosome population. The results obtained verify that the substitution of weaker chromosomes based on the correlation measure improved the hybrid's classification accuracy. This was observed from the higher classification accuracy attained upon testing on the same UCI data sets. The stability of the classifier was also enhanced as ascertained by the low variance of the results collected. The analysis hitherto has thus demonstrated the viability of the GA-SVM hybrid as a good classifier when the irrelevant attributes are removed.

In conclusion, the proposed GA-SVM hybrid was able to produce high average classification accuracy that is comparable or better than some of the established classifiers in the data mining community. The simulations carried out also showcase the statistical consistency of the GA-SVM hybrid, which is evident from the histogram analysis and box plots. Secondary improvements to the hybrid included the utilization of a correlation measure to improve the average fitness of a chromosome population. The results obtained verify that the substitution of weaker chromosomes based on the correlation measure improved the hybrid's classification accuracy. This was observed from the higher classification accuracy attained upon testing on the same UCI data sets. The stability of the classifier was also enhanced as ascertained by the low variance of the results collected. The analysis hitherto has thus demonstrated the viability of the GA-SVM hybrid as a good classifier when the irrelevant attributes are removed.