Abstract

This thesis compares three feature selection methods: through Correlation Based Feature selection (CFS), Relief, and Wrapper methods. Three machine learning algorithms were used: J48 (a decision tree learner), naive Bayes (Bayesian Network), And Multilayer Perceptron (MLP) (Artificial Neural Networks). The purpose of comparison is to extract best set of features that leads enhance performance of classifiers. As the method is study_case_based SEER data is selected for this purpose.

The study showed that classification accuracy using the reduced feature set is equal and in some cases outperform the complete data set.

Moreover, as expected the performance of J48 decreases with the reduced data set.

CFS selected five features, WRAPPER returned eight features and RELIEF returned list of ranked features.

By comparing selected classifier methods Naïve Bayes is showed better results in this study. It produced a significant increase in accuracy with CFS, RELIEF, and WRAPPER methods.

المستخلص

يقدم هذا البحث دراسة مقارنة ثلاث خوارزميات Feature selection methods:

Correlation Based Feature selection (CFS), Relief and Wrapper methods.

تم إستخدام ثلاث خوارزميات من Machine Learning:

J48 (a decision tree learner), naive Bayes (Bayesian Network), And Multilayer Perceptron (MLP) (Artificial Neural Networks) .

الغرض من هذه المقارنة هو تقليل حجم مجموعة البيانات التي تحتوي على كمية كبيرة من البيانات وذلك بإختيار أفضل مجموعة من العدد الكلي لل Features والتي تؤدي الى تحسين أداء ال classifiers. وقد تم إختيار بيانات SEER لهذا الغرض.

قامت خوارزمية CFS بتحديد خمسة Features، أما ال WRAPPER قامت بإختيار ثمانية Features، وال RELIEF قامت بإرجاع قائمة ال Features مرتبة بعد تقييمها حسب الأفضلية.

أيضاً أثبتت التجارب أن خوارزمية Naïve Bayes تحصلت على أفضل النتائج في هذه الدراسة وحققت زيادة واضحة في الدقة مع الثلاث خوارزميات CFS, Relief and Wrapper methods.

Acknowledgements

First and foremost I would like to acknowledge my supervisor, Dr. Mohamed Elhafiz Mustafa Musa. Also I thank my best friends Neama Hussein Eman Abraheem thanks for their technical assistance and helpful comments. Special thanks must also go to my family. They have provided unconditional support and encouragement of my time in graduate university.

Table of Contents:

List of Figures

List of Tables