جامعة الموصل
كلية علوم الحاسوب والرياضيات

أساليب مقترحة لتحسين أداء تقنية آلة المتجه الداعم

محمد علي محمد علي الوزان

رسالة ماجستير

الرياضيات/الحاسوبية

بإشراف

الأستاذ المساعد

د. عمر صابر قاسم

١٤٤٠ هـ

٢٠١٨ م

الخلاصة

تمثلت هذه الرسالة في دراسة الية تطوير أداء التصنيف لتقنية الة المتجه الداعم Support Vector Machine (SVM) بالاعتماد على تقنية الخوارزمية الجينية Genetic Algorithm (GA) وذلك عن طريق جانبين أساسيين. الجانب الأول: إذ اختص هذا الجانب في دراسة مسألة اختيار الميزات Features Selection الضرورية في عملية تصنيف البيانات لاسيما عندما يكون حجم هذه البيانات كبيرا جدا مثل بيانات اللوكيميا Leukemia بنوعيه اللمفاوي الحاد ALL والحبيبي الحاد AML والتي تعتمد على القراءات الجينية. اذ تحتوي العينة على ٧٦ مشاهدة وفي كل مشاهدة توجد ٧١٢٩ ميزة. إذ تم استخدام طريقة مقترحة بين الخوارزمية الجينية وتقنية الة المتجه الداعم تقوم بتحويل البيانات من فضاء الأنماط العالي الابعاد High-D. Patterns Space إلى فضاء الخواص الواطئ Low-D. Feature Space لأجل تحديد الميزات الضرورية واللازمة لعملية التصنيف التي تتم عن طريق تقنية آلة المتجه الداعم إذ تم تقليل عدد الميزات في الطريقة المقترحة من ٧١٢٩ الى ١٠٠٠ ميزة فقط. أما الجانب الثاني: اما هذا الجانب فقد اشتمل على استخدام الخوارزمية الجينية كأداة لاختيار أفضل المعلمات σ و c في تقنية SVM. لقد تم بناء دالة اللياقة Fitness Function في الخوارزمية الجينية بالاعتماد على خطأ التصنيف وباستخدام المعلمات σ و c كمتغيرات معتمدة لهذه الدالة، وتم الحصول على قيم المعلمات σ و c في الطريقة المقترحة بشكل يناسب نوع التطبيق عن طريق الخوارزمية الجينية في حين تم اختيار هذه المعلمات بشكل عشوائي في تقنية SVM الاعتيادية. تبين من التطبيق العملي على بيانات اللوكيميا للطريقتين المقترحتين (الجانب الأول والثاني) بأن نسبة التصنيف بالنسبة للطريقة المقترحة تفوق نسبة التصنيف بالنسبة لتقنية SVM الاعتيادية، مما يدل على كفاءة الطريقة المقترحة مقارنة مع طريقة SVM الاعتيادية.

UNIVERSITY OF MOSUL
COLLEGE OF COMPUTER SCIENCES
AND MATHEMATICS

Suggested Methods to Improve the Performance of the Support Vector Machine Technology

Mohamed Ali Mohamed Ali Alwazan

M.Sc./Thesis

Mathematics/Computational

Supervised by

Assistant Professor

Dr. Omar Saber Qasim

2018 A.D. **1440 A.H.**

Abstract

This thesis was in the study of the mechanism of development of the performance of the classification of the technology of the support vector machine based on genetic algorithm technology through two main sides. The first side: This aspect is specific in the study of the necessary feature selection in the process of classification of data, especially when the size of this data is very large, such as leukemia data. A proposed method was used between the genetic algorithm and the support vector machine technology, which converts data from the high-dimensional space into the low-dimensional space in order to determine the necessary features required for the classification process performed by the support vector machine technology. The number of features in the suggested method was reduced from 7129 to 1000 feature only. The second side: This aspect included the use of the genetic algorithm as a tool to select the best parameters c and σ in the support vector machine technique. The fitness function was built on the genetic algorithm based on the classification error and using parameters c and σ as dependent variables for this function. The parameter values c and σ were obtained in the proposed method in a manner suitable for the type of application by the genetic algorithm. Normal support vector machine technology. The practical application of the leukemia data for the proposed methods (first and second side) showed that the rating ratio for the proposed method is higher than the rating for the conventional support vector machine technology, indicating the efficiency of the proposed method compared to the normal support vector machine method.