الخلاصة

دعت الحاجة لفهم المجاميع الكبيرة ، المعقدة ،الغنية بالمعلومات الشائعة بكل مجالات العمل ،العلم ،الهندسة ، وفي عالم الاعمال ، والاعتراف ببيانات الزبون والشركة وبفائدتها الاستتراتيجية الى استكشاف طرق اكثر فعالية في التعامل مع هذه البيانات. في هذه الاطروحة حاولنا ان نستخرج اكثر المعلومات الممكنة من البيانات المتوفرة وذلك باقتراح تقنية تدعى (العنقدة من اجل التصنيف). ومن ثم تطبيق هذه التقنية في بيئتي قواعد البيانات الموزعة المعروفة بنظام قواعد البيانات الموزعة المتجانسة ونظام قواعد البيانات الموزعة الغير متجانسة ، في هذه الاطروحة تم تقديم خوارزميتين تصف وتقارن تطبيق التقنية المقترحة في نظامي قواعد البيانات الموزعة.

الخوارزمية المقترحة الاولى هي:

"خوارزمية العنقدة الموزعة المتجانسة للتصنيف" : هذه الخوارزمية تهدف الى تدريب نموذج تصنيف من مجموعة من البيانات الغير معنونة الموزعة على الشبكة ، وهذا ببناء نموذج عنقدة محلي على مجاميع من البيانات الموزعة على ثلاث مواقع في الشبكة ومن ثم بناء نموذج تصنيف محلي بالاعتماد على البيانات المعنونة الناتجة من نموذج العنقدة ، ثم يتم بناء نموذج تصنيف عام في الحاسبة المركزية ومن ثم استخدام هذا النموذج في التنبؤ المستقبلي .

الخوارزمية المقترحة الثانية هي:

"خوارزمية العنقدة الموزعة الغير متجانسة للتصنيف" : والتي تهدف لبناء نموذج تصنيف على مجاميع من البيانات الموزعة بشكل غير متجانس على مواقع الشبكة، في هذه الخوارزمية تتجمع هذه المجاميع من البيانات في الحاسبة المركزية ومن ثم يتم بناء نموذج عنقدة وبعده نموذج تصنيف.

في هذه الاطروحة سوف يتم تقديم مقارنة لهذه الخوارزميتين وتبيان الفرق بعدة مقاييس، مثل الدقة للمصنف الناتج ،الوقت المصروف بالتنفيذ ، الكلفة الخزنية المطلوبة.

PDF created with pdfFactory trial version www.pdffactory.com

Abstract

The need to understand large, complex, information-rich data sets is common to virtually all fields of business, science, and engineering and in the business world invite to discovering more activate approach to deal with this data, corporate and customer data are becoming recognized as a strategic asset. In this thesis tries to extract the most information can be from the data available, this is by suggest technique called clustering for classification and formalize this technique in the two environments of distributed database system known as homogeneous and heterogonous distributed database systems. In this thesis two algorithms are introduced to describe and compare the applying of the proposed technique to the two types of distributed database systems

The First Proposed Algorithm is: **HOM**ogeneous **D**istributed **C**lustering **F**or **C**lassification (HOMDCFC) Algorithm; try to learning a classification model from unlabeled datasets distributed homogenously over network, this by building a local clustering model on the datasets distributed over three sites in the network and then build a local classification model based on labeled data that produce from clustering model, in the one computer considered as control computer a global classification model is built and then use this model in the future predictive.

PDF created with pdfFactory trial version www.pdffactory.com

Second Proposed Algorithm: **HE**Trogeneous **D**istributed **C**lustering **F**or Classification (HETDCFC) Algorithm; that try to build a classification model over unlabeled datasets distributed heterogeneously over sites of network, the datasets in this algorithm collected in one central computer and then build the clustering model and then classification model.

In this work a comparison of these two proposed algorithm is introduced and show the different in the criteria, accuracy of the produced classifier, time spent in exe cution, cost require for central storage.

PDF created with pdfFactory trial version www.pdffactory.com