

Nom et Prénom : ZETTAM MANAL

Date de soutenance : 12/11/2020

Directeur de Thèse : J. LAASSIRI

Sujet de Thèse :

Parallel predictive methods for big data analysis in healthcare field

Résumé :

L'analyse de données est devenue un problème extrêmement important et difficile dans des disciplines telles que l'informatique, la biologie, la médecine, la finance et la sécurité intérieure. Comme des quantités massives de données sont disponibles pour l'analyse, les techniques d'intégration évolutives deviennent importantes. Dans le même temps, de nouveaux problèmes de confidentialité se posent, où des informations sensibles peuvent facilement être déduites de la grande quantité de données. Cette thèse pourrait être discernée en deux parties. Dans la première partie, nous commençons par présenter un état de l'art des big data et des outils d'apprentissage automatique. Nous avons ensuite introduit les méthodes d'apprentissage statistique et d'analyse des données de masse. Après cela, nous nous sommes concentrés sur la régression multiple. Nous avons identifié les méthodes permettant de résoudre le problème des moindres carrés et présenté une méthode paralysante algébrique adjointe. Nous avons comparé les performances des méthodes permettant de résoudre le problème des moindres carrés à l'aide d'un modèle d'estimation. Le modèle introduit a été testé sur plusieurs ensembles de données de santé. Dans la deuxième partie de cette thèse, nous avons présenté un modèle de prédiction basé sur la méthode de prédiction TLBO-FLANN pour prédire la valeur du statut de la maladie angiographique. En effet, les modèles de réseaux de neurones artificiels (RNA) sont une alternative aux différentes méthodes conventionnelles basées sur des statistiques. Les ANN sont capables de générer une correspondance complexe entre les espaces d'entrée et de sortie, créant ainsi des limites non linéaires arbitrairement complexes pour la prise de décision.

MOTS-CLES:

MapReduce; hadoop; régression multiple; méthode adjointe; réseau de neurones; TLBO-FLANN prédiction.

Abstract :

Data analytics has become an extremely important and challenging problem in disciplines like computer science, biology, medicine, finance, and homeland security. As massive amounts of data are available for analysis scalable integration techniques become important. At the same time, new privacy issues arise where one's sensitive information can easily be inferred from the large amounts of data. This thesis could be discerned in two parts. In the first part we begin by presenting a state of arts of big data and machine learning tools. Then we introduced the Statistical learning and big data analytics methods. After that we focused on the multiple regression. We identify the methods solving the least square problem and presented a parallelized algebraic adjoint method. We compared the performance of the methods solving the least square problem on the basis of an estimation model. The introduced model was tested on several health data sets. In the second part of this thesis we presented a prediction mapreduce-based TLBO-FLANN model for predicting angiographic disease status value. Indeed, artificial neural network (ANN) models are alternatives to various conventional methods based on statistics. ANNs are capable of generating complex mapping between input and output space, thus creating arbitrarily complex nonlinear boundaries for decision making.

KEY WORDS:

MapReduce; hadoop; multiple regression; adjoint method; artificial neural network; mapreduce-based TLBO-FLANN model; prediction.