

Comparison of Gradient Boosting Decision Tree Algorithms for CPU Performance

Haithm Haithm ALSHARI^{*1}, Abdulrazak Yahya SALEH², Alper ODABAŞ³

^{*1} Eskişehir Osmangazi Üniversitesi Matematik ve Bilgisayar Bilimleri Bölümü, ESKİŞEHİR
² FSKPM Faculty, Malaysia Sarawak University (UNIMAS), 94300 Kota Samarahan, SARAWAK
³ Eskişehir Osmangazi Üniversitesi Matematik ve Bilgisayar Bilimleri Bölümü, ESKİŞEHİR

(Alınış / Received: 14.02.2021, Kabul / Accepted: 15.04.2021, Online Yayınlanma / Published Online: 28.04.2021)

Keywords

Decision Tree,
Gradient Boosting,
XGBoost,
LightGBM,
CatBoost

Abstract: Gradient Boosting Decision Trees (GBDT) algorithms have been proven to be among the best algorithms in machine learning. XGBoost, the most popular GBDT algorithm, has won many competitions on websites like Kaggle. However, XGBoost is not the only GBDT algorithm with state-of-the-art performance. There are other GBDT algorithms that have more advantages than XGBoost and sometimes even more potent like LightGBM and CatBoost. This paper aims to compare the performance of CPU implementation of the top three gradient boosting algorithms. We start by explaining how the three algorithms work and the hyperparameters similarities between them. Then we use a variety of performance criteria to evaluate their performance. We divide the performance criteria into four: accuracy, speed, reliability, and ease of use. The performance of the three algorithms has been tested with five classification and regression problems. Our findings show that the LightGBM algorithm has the best performance of the three with a balanced combination of accuracy, speed, reliability, and ease of use, followed by XGBoost with the histogram method, and CatBoost came last with slow and inconsistent performance.

CPU Performansı için Gradyan Artırıcı Karar Ağacı Algoritmalarının Karşılaştırılması

Anahtar Kelimeler

Karar Ağacı,
Gradyan Artırıcı,
XGBoost,
LightGBM,
CatBoost

Öz: Gradyan Artırıcı Karar Ağacı (GBDT) algoritmalarının regresyon ve sınıflandırma problemlerinin çözümünde makine öğrenimindeki en iyi algoritmalar arasında olduğu kanıtlanmıştır. Kaggle gibi web sitelerinin düzenlediği birçok yarışmayı kazanması sebebiyle en popüler GBDT algoritması olan XGBoost son teknoloji performansa sahip tek GBDT algoritması değildir. LightGBM ve CatBoost gibi kimi zaman XGBoost'a göre daha fazla avantajları olan başka GBDT algoritmaları da vardır. Bu makale, en iyi üç gradyan artırıcı algoritmanın işlemci(CPU) performansını karşılaştırmayı amaçlamaktadır. Bunun için ilk olarak bu üç algoritmanın nasıl çalıştığını ve aralarındaki hiperparametre benzerliklerini açıklayacağız. Daha sonra performanslarını değerlendirmek için doğruluk, hız, güvenilirlik ve kullanım kolaylığı olarak dörde ayırdığımız performans kriterleri kullanacağız. Üç algoritmanın performansı beş sınıflandırma ve regresyon problemi ile test edilmiştir. Bulgularımız, LightGBM algoritmasının, dengeli bir doğruluk, hız, güvenilirlik ve kullanım kolaylığı kombinasyonu ile üçü arasında en iyi performansa sahip olduğunu, bunu histogram yöntemiyle XGBoost'un izlediğini ve CatBoost'un ise özellikle yavaş ve tutarsız performansla diğerlerinin gerisinde kaldığını göstermektedir.

*İlgili Yazar: aodabas@ogu.edu.tr