Data Shapley: Equitable Valuation of Data for Machine Learning

He Wang

As data is the fuel that drives the rapid growth of artificial intelligence, an underlying challenge is how to quantify the value of (particular) data. To address such a challenge, this paper proposes data Shapley value, leveraging powerful results from game theory, to evaluate the contribution of individual data points to a specific learning task using a particular learning algorithm. The authors first formulate the necessary properties of equitable data evaluation in machine learning and show that data Shapley satisfies such properties uniquely. Nevertheless, computing the exact data Shapley requires calculating all the possible marginal contributions which is exponentially large in the data size, and has to learn a performance evaluation model for the particular learning algorithm. Therefore, the authors apply two methods to approximate the data Shapley value and the performance metric. First, the authors employ the Monte-Carlo method and truncate the calculation of marginal contributions in a sampled permutation to approximate the exact data Shapley value, which reduces the computational costs to a large extent without introducing significant estimation bias. On the other hand, they employ the gradient method to train the performance evaluation model for only one epoch, and group the data by similarity to quantify the value of each group instead of each data point. The experiment results demonstrate several benefits and implications of data Shapley: Lower Shapley value data effectively capture outliers and corruptions, while higher Shapley value data imply what type of new data needed to improve the learning performance.