Competence region estimation for black-box surrogate models

Tapan Shah
GE Global Research
San Ramon

Abstract

Machine Learning based black-box models are increasingly used as surrogate models for complex system under test (SUT). Building high fidelity surrogate models involves sophisticated modeling (deep neural networks, random forests, support vector machines) and feature engineering. Additionally, high quality data is collected from SUT to train the surrogate models. Due to the effort spent in developing these models, organizations re-use/leverage models developed by other organizations with suitable attribution. Each machine learning model has an associated model competence, i.e. the data region over which the model performs as desired. The model competence pre-dominantly depends on the amount and distribution of training data as well as the model type/features. This information is typically confidential, and developers would not like to share this information to the users. However, it is imperative for the users to estimate the ”competence” region to suitably deploy/fine-tune the model.

The main contributions are:

1. Define model competence as a function of fidelity.
2. Assuming a simulator/test system which can provide ”infinitely” many true responses, describe a simple algorithm to estimate the model competence. The algorithm draws inspiration from trust-region based optimization.
3. Typically, it is very expensive to obtain true responses from the SUT (or a physics-based simulator). Often, it is not possible to obtain a single response. We discuss potential methods to determine model competence under following two regimes a) Constraint on number of responses required from SUT and b) No response available from SUT, but the model outputs prediction uncertainty in addition to point prediction.

Copyright © 2021by the authors. All rights reserved.