

BONSAI: AN INTERACTIVE SYSTEM FOR VISUAL EXPLORATION OF MANY DATA MODELS

Tech ID: 18044 / UC Case 2009-045-0

ABSTRACT

Machine learning or statistical learning emphasizes the use of "black box" algorithms to model data and applies these models to make classification predictions when applied to new data. However, the growth in scale of applicable datasets and learning tasks has outstripped many tools for carefully supervising this modeling process. The result is that most real-world implementations involve humans preparing sets for training and testing, comparing baseline performance of a set of models, and optimizing parameters of a given modeling approach.

To address this problem, Researchers at UC Berkeley have developed a novel system called Bonsai, aimed at making transparent the inner workings of the "black box". Bonsai provides multiple visual lines of inquiry into the model development process and the interaction of the model with the data. This gives the user the ability to have a far deeper understanding of the data and specific modeling techniques and their strengths and weaknesses. It opens the door for development of alternative methods for modeling the data.

The system is especially valuable for classification problems arising from large and high dimensional data sets, where manual inspection or construction of classification models can be prohibitively time-consuming. In addition, the system encourages a machine learning "guided tour" through the data, improving the user's understanding of the data and participation in the modeling process. In contrast to much previous work, the emphasis is on considering the joint "space" of the data and multiple machine learning models, rather than providing either an interface for manual classification or for post-construction analysis of a single model.

APPLICATIONS

Data modeling tools and modeling suites, Visualization applications, Improvement of prediction models, Exploration of under-sampled or inadequately processed data, for purposes of improving data or challenging models

ADVANTAGES

- Improves user's understanding of data and modeling techniques
- Can help solve previously time-prohibitive classification problems

CONTACT

Curt A. Theisen (Deceased)
curt@berkeley.edu
tel: 510-643-7214.



OTHER INFORMATION

KEYWORDS

computer, copyright, copyrighted
content, software, internet

CATEGORIZED AS

- » [Computer](#)
- » [Software](#)

RELATED CASES

2009-045-0