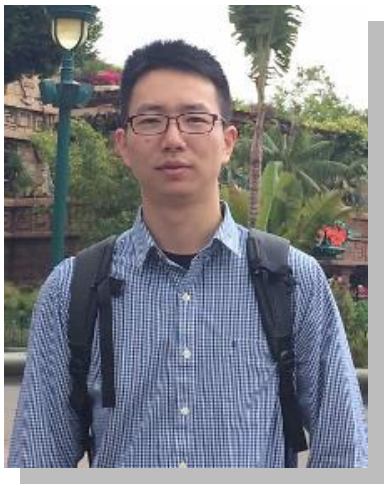
CoCo Seminar Series
Spring 2017

An Ensemble Machine Learning Approach for Robust Cancer Diagnosis



Haifeng Wang
Ph.D. candidate, Industrial and Systems Engineering
Binghamton University

Wednesday April 5th, 2017 8:30-9:30am
Engineering Building H-9 (Knoll-MacDonald Commons / Watson Commons)

Illness diagnosis plays a critical role in designating treatment strategies, which are highly related to patient safety. Nowadays, numerous classification models in data mining domains are adapted to cancer prediction and diagnosis based on patients' historical medical records. However, two challenges are gradually recognized by many data mining researchers for cancer diagnosis tasks: 1) enormous algorithms make it exhausting to select the appropriate one; 2) the performance of a good diagnosis algorithm not only depends on its high accuracy but also its stability (low diagnosis variance). To tackle the challenges, this seminar presents a support vector machine (SVM)-based ensemble learning algorithm to reduce the diagnosis variance and further increase diagnosis accuracy.

In this talk, a Weighted Area Under the Receiver Operating Characteristic Curve Ensemble (WAUCE) approach is discussed. The general performance of the model is investigated based on standard breast cancer datasets. The experiment results show that the WAUCE model achieves a higher accuracy with a significantly lower variance for breast cancer diagnosis compared to five other ensemble mechanisms and two common ensemble models, i.e., adaptive boosting and bagging classification tree. The proposed WAUCE model especially reduces the variance by around 70%, and it still increases accuracy by 0.9%, compared to the best single SVM model. In practice, the proposed methodology can be further applied to other cancer diagnoses, which offers an alternative to a safer, more reliable, and more robust illness diagnosis process.

Haifeng Wang is a Ph.D. candidate in the Industrial and Systems Engineering program at Binghamton University. His research interests include using system optimization, data mining and deep learning techniques to improve the performance of human illness diagnosis and pharmacy automation.

For more information, contact Sangwon Yoon (yoons@binghamton.edu).
http://coco.binghamton.edu/