

Learning Classification Trees for Personalized Risk Stratification

Anima Singh

Advisor: John V. Guttag

Massachusetts Institute of Technology



Problem Statement

Goal: Improving risk stratification of patients post heart attack, e.g., prediction of cardiovascular death and myocardial infarction within 90 days of the index event.

Common tool for cardiovascular risk stratification: TIMI Risk Score

For cardiovascular death,

Recall	Precision
44%	5%



Why is risk stratification hard?

1. Multiple physiological factors affect outcome, therefore many variables need to be considered.
2. Machine learning can be used to learn relationships between variables.
3. Learning risk stratification model is challenging due to high class imbalance in medical data sets.

For example, only ~2% of patients with non-ST elevation acute coronary syndrome suffer from cardiovascular death within 90 days of index event.



Our Method

Based on Learning Classification Trees from data

We use a novel non-symmetric entropy based method:

1. To generate personalized cut-points for numerical variables.
2. To determine the order of variables in the tree.



Some Results

When evaluated on a population of 4500 post ACS patients.

Mean Recall	Mean Precision
74%	6.3%

Improvement over state-of-the-art methods

Mean % Improvement		
	Recall	Precision
CI- SVM	3%	34%
Existing CI-CT method	4%	36%

