

Data Mining Hospital Readmissions and Mortality Rates using Multiagent Random Forest

Andrew Berry

December 1, 2016

Introduction

- ▶ Goal: Attempt to accurately predict location of hospital based on readmissions and death metrics
- ▶ Subgoal: Establish that there exists a correlation between such metrics and location
- ▶ Subgoal: Attempt to improve prediction accuracy by introducing multiagent communication

System Description

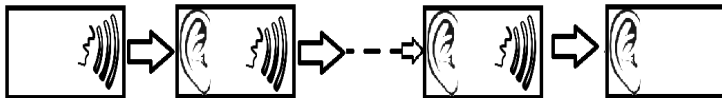
- ▶ Dataset: Hospital Readmission and Death Rates of US Hospitals
- ▶ Data analysis (dataframe manipulation): Pandas
- ▶ Machine Learning: Scikit-Learn
- ▶ Multiagent system: SPADE

System Description

- ▶ n agents
- ▶ Each agent has t trees in random forest classifier
- ▶ Experiments vary between using gini impurity and entropy as criterion (μ)
- ▶ Agents that “talk” communicate their predictions to “listening” agents, who decide between their own predictions and ones they have been told

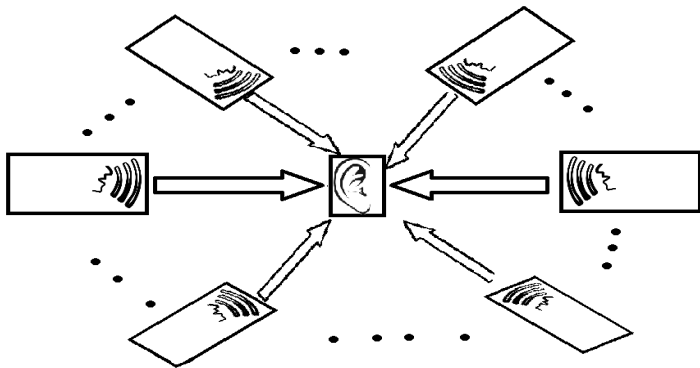
System Description

- ▶ Initial agent network:



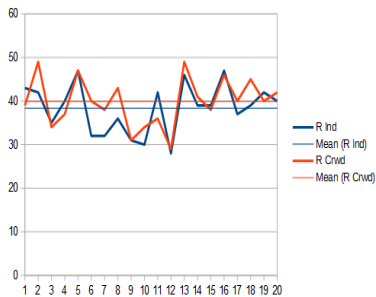
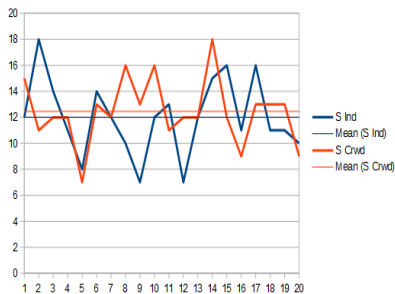
System Description

- Final agent network:

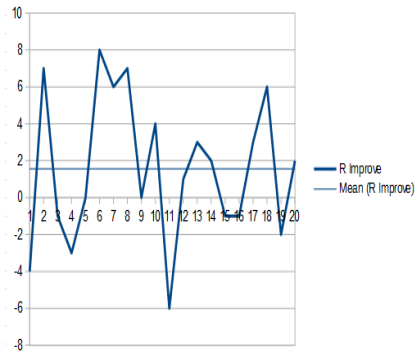
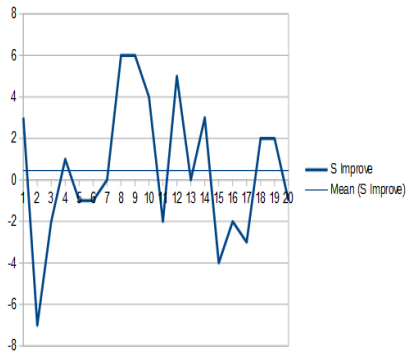


Experiment 1

- ▶ $n = 30$
- ▶ $t = 100$
- ▶ $\mu = gini$

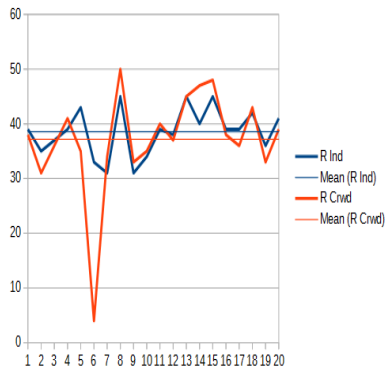
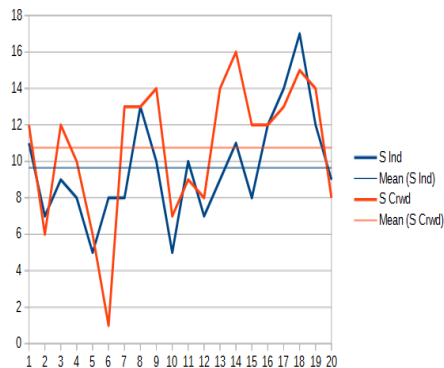


Experiment 1

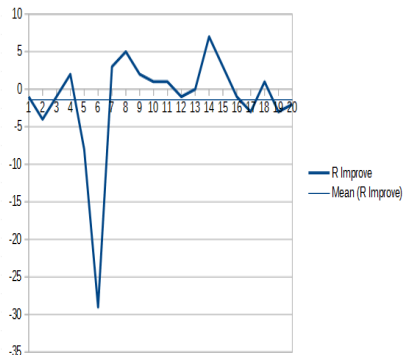
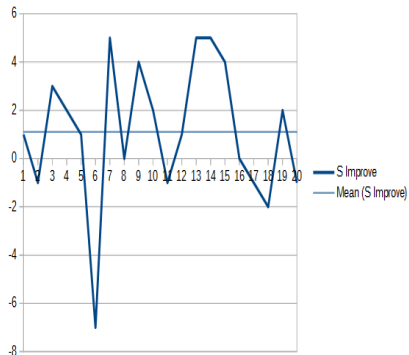


Experiment 2

- ▶ $n = 10$
- ▶ $t = 100$
- ▶ $\mu = \text{entropy}$



Experiment 2



Demo

Conclusions

- ▶ Better than random guessing for both states and regions
- ▶ Correlation is apparent
- ▶ Improvement from multiagent system fairly uncertain