WEBSITE CHANGES AND USER BEHAVIOR
USING PANJIVA DATA TO EXAMINE CODE CHANGES

JOHN WANG
14.27 FINAL PAPER

Abstract. Premised on information theoretic results, I create an algorithm that attempts to maximize the entropy of a created ExamSet. Students are differentiated by weighted scores of the questions they answer correctly, where the weights are determined by the difficulty of the question. An ExamSet's entropy is defined by the probability of a student obtaining a given score.

First, this paper shows how I estimate the difficulty of a question (the probability r_j of answering question j correctly) using an iterative hill-climbing algorithm. Then, the paper shows how entropies of questions can be computed in order to approximate the entropy of an ExamSet. I show theoretical runtime results for a randomized algorithm to select questions based on entropy levels. Finally, I show how the results from this randomized algorithm are used as inputs into a genetic algorithm.

Contents

Date: November 27, 2012.

1. Problem Formulation