GPA Prediction Based on Student's Lifestyle

4th Apirl 2017

TEAMMEMBERS

Minzhuo Jin, Jim Xu, Ivy Min

OVERVIEW

According to the general acknowledgement, each individual's lifestyle in college can have an influence on one's academic performance. We are curious about this phenomenon and want to explore further on how individual's lifestyle can influence his GPA. We plan to use decision tree algorithm with machine learning powered by Mlib and Spark to make reasonable GPA prediction for a specific student's lifestyle. Later on, we can visually present our result to user and advise user on how to getting into a better lifestyle in order to achieve academic success.

DATASET

We plan to use the dataset provided by the StudentLife Study in Dartmouth College. The dataset is approximately 12 GB in size, and well formatted in csv and db style. It covers a broad range of dataset about student college lifestyles, making it possible and reasonable to implement decision tree algorithm with machine learning.

- Data Documentation : http://studentlife.cs.dartmouth.edu/dataset.html
- Data Accessibility : http://studentlife.cs.dartmouth.edu/

PROBLEM CHARACTERIZATION

College Life is colorful. Students need to seek a balance between extracurricular activities and study. Many of us could not figure out what kind of lifestyle we live will have a potential influence on their academic performance, such as sleeping, working, entertaining, studying and mental status. How does lifestyle affect a student's academic performance? Can we extract useful information from students' smartphone sensor data and further predict their academic performance using the knowledge about their eating habits, sleep quality, activities, etc.? Our problem is derived from those

issues and tries to figure out how those potential factors contribute to individual's GPA. We decide to analysis the StudentLife Study's dataset in the manner of decision tree algorithm to build a reasonable relationship between one's lifestyle and his predicted GPA by using machine learning.

CURRENT RELEASE

The newest release from StudentLife Study is a simple model based on linear regression with lasso regularization that can accurately predict cumulative GPA. The predicted GPA strongly correlates with the ground truth from students' transcripts ($r = 0.81$ and $p < 0.001$) and predicts GPA within ±0.179 of the reported grades.

- *SmartGPA Study* : http://studentlife.cs.dartmouth.edu/smartgpa.pdf

ANALYTIC TASKS

1. Extract information from raw sensor data
2. Pair each student's lifestyle data, survey results, EMA results with their academic performance indicator, GPA in this case.
3. Split the dataset into 2 parts. One is for training the model. The other is for model evaluation.
4. Train a Decision Tree regression model that can estimate a student's GPA given lifestyle information.
5. Evaluate the model to see its accuracy, throughput, etc.

SOLUTION EVALUATION

The metrics we will use to assess the solution include throughputs, mean squared errors. The throughput will be evaluated for both the raw data extraction part and the training part.