Milestone 6 Progress Evaluation

Academic Behavior Reccomendation System

Shreyas Ugemuge `sugemuge2014@my.fit.edu`

Yaqeen AlKathiri `yalkathiri2013@my.fit.edu`

Mohammed AlHabsi `malhabsi2013@my.fit.edu`

Shiru Hou `shou2015@my.fit.edu`

Faculty Sponsor: Dr. Phillip Chan `pkc@cs.fit.edu`

December 7, 2017

Contents

1 Progress of current Milestone

Task	Completed	Shreyas	Yaqeen	Shiru	Mohammed	Remarks
Finalize Feature Set	100%	100%	-	-	-	N/A
Finalize Model	100%	100%	-	-	-	N/A
Finish Recommender System	100%	100%	-	-	-	.
Data mining to find and select Readable Output models	100%	25%	25%	25%	25%	N/A
Split data for training and testing	100%	100%	-	-	-	N/A
Complete GUI	60%	-	-	50%	50	Back end to work with reccomender system and some design changes

1.1 Discussion of each task

1.1.1 Revise Behaviors

Some new behaviors were implemented including:

1. total number of unit discussions

2. number of fluency drill activities

3. number of crosswords attended (terminology)

4. number of game shows attended (terminology)

All behavior extraction was modified to produce only numeric values and behaviors were used in groups or one at a time to find the impact on the model.

1.1.2 Decide on data mining tools to use

The data mining tools finalized are as follows

- WEKA

- RapidMiner

- Python sklearn library

1.1.3 Modify merge program to provide output according to the DM tools' input requirement

The merge program from the previous milestone was scrapped and new one was written to achieve the following goals:

1. Provide more flexibility for behavior extraction. The new merge program now just uses the columns from the extracted behaviors as available from csv files

2. Provide a more logical workflow. Each person's behavior directory now contains one file for each weeks behavior extraction with a standardized naming convention

3. Only numeric data is output for all columns instead of above average. This eliminates any dashes or special characters

4. Update column headers so that file can be used with all the DM tools listed

5. Above average is now determined using the median value instead of the average which previously skewed the results to about 75% above average

1.1.4 Data mining to find and select Readable Output models

Different data mining algorithms were used to model the behaviors and the grades.
The behaviors were used as the X axis for the classifiers and whether or not the grade is above average was used as the Y value
The algorithms tested are:

1. Decision Tree

2. Naive Bayes

3. Random Forest Classifier

4. k nearest neighbour

Decision tree and naive bayes have a starting accuracy of 50% and hence had to be discarded. Random Forest Classifier has an accuracy of 86% at 11 weeks that drops to 75% gradually as weeks are reduced to 3.

1.1.5 Split data for training and testing

Simple java program was written to split the merged data into a training set that is 70% of the data and test set that is 30% of data. This in a way resembles having two different semesters one to create a model and the second one to test it on. The program preserves the ratio of above and below average for both the training and the testing set.

1.1.6 Complete GUI

GUI preliminary design finished. Design changes are to be made as per feedback and back-end needs to be written in order to work with the recommender system

1.2 Discussion of team member contribution

1.2.1 Shreyas

- implemented and worked with 3 correlation algorithms using sklearn python library

  1. Decision Tree
  2. Naive Bayes
  3. Random Forest Classifier

- Changed behavior extraction to provide only numerical values

- Changed the merge program

- Wrote program for train and test split

- Wrote program to use classifier models and find accuracy

- Compared accuracy of all 3 models over reduced weeks

1.2.2 Shiru

1. Update the behaviors file, merge all behaviors in one file that easier to use.

2. Solve the students 10 information lost problem.

3. Work on GUI

1.2.3 Yaqeen

Conceptualized, implemented and extracted 4 behaviors using Java as the language of implementation. Worked on and arranged the presentation for milestone 4. Explored the data mining tool: Weka, to find correlations between behaviors. Merged the behaviors in one file manually and converted it from .cvs to .arff file format to be used in Weka to learn correlations between some of the behaviors.

1.2.4 Mohammed

1. Created and designed GUI using Java GUI builder.

2. Decided on using WEKA for finding correlations.

3. Created a new format of the data set to be compatible with the data mining tool WEKA.

4. Fixed errors of the previous behaviors code.

5. contributed in the team presentation.

6. cleaned up behavior extractions to fit tools.

7. started looking for correlations with naive bayes and k-nearest neighbors algorithms.

2 Plan for the next milestone

Task	Shreyas	Yaqeen	Shiru	Mohammed
Use different algorithms for correlation	25% (Neural net, MLPClassifier)	25%	25%	25%
Find and finalize models to use for reccomender system	25	25	25	25
Finish GUI	-	50	-	50
Start implementing reccomender system with chosen model and data	25	25	25	25

3 Sponsor feedback on each task for current milestone

4 Sponsor Evaluation

Sponsor: Please detach this page and return to Dr. Shoaff

Score (0-10) for each member: circle a score (or circle two adjacent scores for .25 or write down a real number between 0 and 10)

Shreyas Ugemuge	0	1	2	3	4	5	6	6.5	7	7.5	8	8.5	9	9.5	10
Yaqeen AlKathiri	0	1	2	3	4	5	6	6.5	7	7.5	8	8.5	9	9.5	10
Mohammed AlHabsi	0	1	2	3	4	5	6	6.5	7	7.5	8	8.5	9	9.5	10
Shiru Hou	0	1	2	3	4	5	6	6.5	7	7.5	8	8.5	9	9.5	10

Faculty Sponsor

_____

Signature

_____

Date