

Application Development and Automating Scoring with Machine Learning



Wesley Smith, Alexis Fuller, Kyle Feuz

NIH Training in Gerontechnology, Computer Science

Introduction

One way psychologists and trained clinicians measure cognitive impairment is observing individuals as they perform a set of daily activities. While observing the participant, the

Methods

NOT Development

•In conjunction with the Psychology department at Washington State University we constructed the user interface and the determined the scoring requirements for the application.

Results

Using the inefficient, incomplete and inaccurate totals for each activity, we were able to correctly predict the correct activity score 99% of the time using a Naïve

clinician must record information about the performance such as start and stop times, and what types of errors were made. This is difficult to do in real-time and often requires reviewing taped sessions to verify data accuracy.

The purpose of this research was to develop an android application to give trained clinicians the ability to record efficiently and in real-time the performance of volunteers attempting to complete nightly tasks in a controlled environment. The Night Out Task or NOT app then uses the data entered to calculate a set of scores which the clinician can use in evaluating the level of cognitive impairment of an individual.

We hypothesize that many of the hard-coded formulas used in the NOT could be successfully replaced with machine learning algorithms and still maintain a high degree of accuracy. We therefore use the NOT data and Weka[1], a machine learning software application, to predict the same scores that are currently being calculate by the NOT app.

Background

•The NOT is designed to allow all buttons to be available to efficiently record the activities. Different interfaces are used for different sized devices.

Machine learning Approach

•We gathered data using a prototype of the NOT app at Washington State University. Two sets of data were gathered, one with older individuals, and the other with younger individuals.

•Using the data from the experiments, we calculated various attributes including number of inefficient, incomplete and inaccurate error totals for individual activities, and combined totals for each experiment. Other variables calculated include planning time and total time.

Bayes classifier.

Using the scores of each activity for each experiment the overall quality was predicted correctly 65% of the time.

With each activity score for each experiment we were able to predict the correct overall accuracy score 90% of the time. A linear regression classifier reproduced the exact formula we used to calculate the overall accuracy score in the NOT app.

Using the three error totals for each activity for every experiment we were able to predict the correct overall quality 65% of the time.

Using the three error totals for each activity for every experiment we were able to predict the overall accuracy score with a correlation coefficient of 0.766, and a mean absolute error of 2.05.

Similar tools for real-time annotation have been developed previously. The development of the Real-time annotation tool (RAT) [2], like the NOT app allows observers to annotate experiments in real-time. However, the RAT does not run on a mobile platform and is therefore less portable. The RAT also does not provide any aggregation or calculation of scores.

| 🖃 A A 🖻 | | | | | } 📢 🛜 43% 📕 3:: | |
|--|-----------------|--|-------------------------------------|--|-----------------|--|
| Recipe втор | INGREDIENTS | ADDITIONAL NTRA PERSEVE | RATION WANDERING REQUESTS HELP | Travel Bag | | |
| GATHERS ITEMS BEFORE READING RECIPE | | Movie CONTINUE | Snack CONTINUE | LOOKS IN MULTIPLE LOCATIONS | | |
| RETRIEVAL INEFFICIENT (>2 TRIPS EACH CUP.) | | LOOK IN MULTIPLE LOCATIONS LOOKS IN MULTIPLE LOCATIONS | | CARRIES ITEMS TO DOOR BY HAND | | |
| DOES NOT EFFICIENTLY LOCATE RECIPE (INDEX/TOC) | | RECORDS LEAVING BEFORE 6:25 | TAKES MORE THAN ONE SNACK | USES ALTERNATIVE CARRIER, NOT BAG | | |
| GATHERS CREAMY PB, NOT CHUNKY | | RECORDS >\$7 SENIOR, >\$11 ADULT | CHOOSES CANDY, NOT MILK CHOC. | DOES NOT PUT ITEMS IN BAG (EXCEPTION: TEA, \$, PHONE BY HAND) | | |
| GATHERS TABLE SALT, NOT KOSHER | | RETURNS TO SCHEDULE AFTER FINISHING TASK | CHANGES SNACK AFTER FINISHING TASK | DOES NOT FINISH TASK | | |
| GATHERS COFFEE, NOT ESPRESSO | | RECORDS COST FOR ONLY ONE PERSON | CHOOSES NON-SNACK ITEM | | | |
| GATHERS EXTRA ITEMS | | DOES NOT RECORD ANSWERS | CHOOSES DARK CHOC. | LOOKS IN MULTIPLE LOCATIONS | | |
| DOES NOT GATHER 1-2 NONESSENTIAL ITEMS | | RECORDS >\$3.50 SENIOR, >\$5.50 ADULT | DOES NOT BRING SNACK TO DOOR | WAITS FOR TEA, NOT MULTITASKING | STOP NOT | |
| CARRIES ALL ITEMS TO BAG, NOT BAG TO ITEMS | | RECORDS LEAVING AFTER 6:35 | DOES NOT FINISH TASK | GETS TEA, NOT DIRECTLY PUT IN THERMOS | s | |
| MAKES CHANGE AFTER FINISHING TASK | | DOES NOT FINISH TASK | | MAKES TEA MORE THAN ONCE | | |
| GATHERS GRANULATED SUGAR, NOT POWDERED | | | LOOKS IN MULTIPLE LOCATIONS | DOES NOT START TIMER | | |
| DOES NOT GATHER 1+ ESSENTIAL ITEMS | | LOOKS IN MULTIPLE LOCATIONS | GATHERS \$ BEFORE CHECKING SCHEDULE | DOES NOT WAIT ≥ 3 MINUTES | | |
| DOES NOT GATHER 3+ NONESSENTIAL ITEMS | | CALL IS NOT LAST TASK BEFORE EXIT | GATHERS MORE \$ THAN RECORDED | NO TEA IN THERMOS | | |
| LOCATES WRONG RECIPE AND GATHERS ITEMS | | CALLS BUT DOES NOT MENTION LEAVING | ADJUSTS \$ AFTER FINISHING TASK | NO WATER IN THERMOS | | |
| DOES NOT F | FINISH TASK | MAKES CALL MORE THAN ONCE | GATHERS LESS \$ THAN RECORDED | MAKES COFFEE | | |
| | E TRIPS TO DOOR | DOES NOT CALL | | D DOES NOT TAKE TEA TO DOOR | | |
| STOP NOT ONE OF LAST 2 TASKS | | | DOES NOT BRING \$ TO DOOR | DOES NOT FINISH TASK | | |

•We predicted scores for each activity, an overall quality score, and an overall accuracy score using 10-fold cross validation.

•In order to get the best results for each data mining experiment we tried different classifiers including, naïve bayes, linear regression, and a decision tree.



Conclusions

The NOT app will continue to be used to collect data to assist other research projects. The machine learning is able to accurately predict individual activity scores and the overall accuracy score. However, it struggles to correctly predict the overall quality. With more data we hope to improve this prediction.

References

[1] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten (2009); The WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11, Issue 1

[2] K.D. Feuz and D. Cook. Real-time annotation tool (RAT). Proceedings of the AAAI Workshop on Activity Context-Aware System Architectures, 2013. pdf

| 🖃 A A 单 | | | | } ≮ } 🛜 43% 📕 3:29 AM | | | |
|--|------------------------|----------------------------|--------------------------|------------------------------|--|--|--|
| File Name: wes12 | EXPORT Summary of | of Results | RESTART | | | | |
| | Task Planning | g Time:4 | | | | | |
| | Total Execution | Total Execution Time: 30 | | | | | |
| | Overall Task (| Overall Task Quality: Poor | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| Inefficient: 9 Incomplete: 4 Inaccurate: 3 | | | | | | | |
| Movie Sche | edule Tea | Snack | Change | | | | |
| Completion S | core: 1 Completion Sco | re: 3 Completion Score | e: 1 Completion Score: 3 | | | | |
| Active Time: 2 | Active Time: 1 | Active Time: 6 | Active Time: 12 | | | | |
| w/multitask: 2 | 24 w/multitask: 1 | w/multitask: 6 | w/multitask: 12 | | | | |
| Sequencing: 2 | 2 - 5 Sequencing: 8 - | 1 Sequencing: 6 - 3 | Sequencing: 5 - 4 | | | | |
| Simultaneous | : 6 Simultaneous: 0 | Simultaneous: 2 | Simultaneous: 3 | | | | |



Inefficient errors: 0 Inefficient errors: 2 Inefficient errors: 0 Inefficient errors: 1 Incomplete errors: 1 Incomplete errors: 1 Incomplete errors: 0 Incomplete errors: 0 Inaccurate errors: 0 Inaccurate errors: 0 Inaccurate errors: 1 Inaccurate errors: 0 Recipe Travel Exit Phone Completion Score: 1 Completion Score: 3 Completion Score: 3 Completion Score: 3 Active Time: 17 Active Time: 28 Active Time: 3 Active Time: 24 w/multitask: 17 w/multitask: 28 w/multitask: 3 w/multitask: 24 Sequencing: 4 - 6 Sequencing: 1 - 7 Sequencing: 7 - 2 Sequencing: 3 - 8 Simultaneous: 7 Simultaneous: 5 Simultaneous: 1 Simultaneous: 7 Inefficient errors: 2 Inefficient errors: 3 Inefficient errors: 0 Inefficient errors: 1 Incomplete errors: 0 Incomplete errors: 0 Incomplete errors: 1 Incomplete errors: 1 Inaccurate errors: 0 Inaccurate errors: 0 Inaccurate errors: 0 Inaccurate errors: 2

