Could machines be better than teachers?

COMPUTER-HUMAN INTERACTION IN LEARNING AND INSTRUCTION CHILI

École Polytechnique Fédérale de Lausanne

Farah Bouassida,*

Supervisor: Lukasz Kidzinski†

Supervisor: Pierre Dillenbourg‡

January 7, 2016

*Communication systems' Masters farah.bouassida@epfl.ch

†lukasz.kidzinski@epfl.ch

‡pierre.dillenbourg@epfl.ch

Abstract

The goal of this project is to implement a web platform that teaches a set of concepts through different kind of material and to gather different information about the learner. Each sequence of learning material is called a learning path and the different learning paths are a video on the first concept then a video on the second concept, or a video then a text or a text then a video or two consecutive texts on both taught concepts. The project's purpose is to demonstrate a possible correlation between the information we have on the learner and his learning gain, the improvement he made before and after he was taught on the concepts, in order to build afterwards, an adaptive learning platform. This platform would propose the best learning path based on the gathered learner information.

Contents

1 Introduction

Our goal is to establish a relationship between a person's learning behavioral characteristics and the type of teaching activities he takes and how it impacts the outcome of his learning experience. Later, this could be used to construct an adaptive learning platform based on a set of selected features of the learner that we could define as his learning profile. For this project we tested the difference between two equivalent explanations of the same concept, one presented with a text in an encyclopedia fashion and the other presented with a video selected from a MOOC. Previous studies classify learners based on common sense, such as high-aptitude versus low-aptitude, highly motivated versus poorly motivated. Each category of learner is then proposed a different learning strategy and psychological tests are available to identify the different styles of learners. Our goal would be to find automatically the best teaching material corresponding to each style and to widen the understanding of a learners' profile not only basing it on one common sense criteria. In fact, thanks to machine learning we can classify learners based on wider criteria as we collect information on different aspects of a person and we optimize the correlation of a combination of all of these aspects or just a subset of them to predict the learning gain a learner would achieve with a given combination of activities.

2 Learning platform

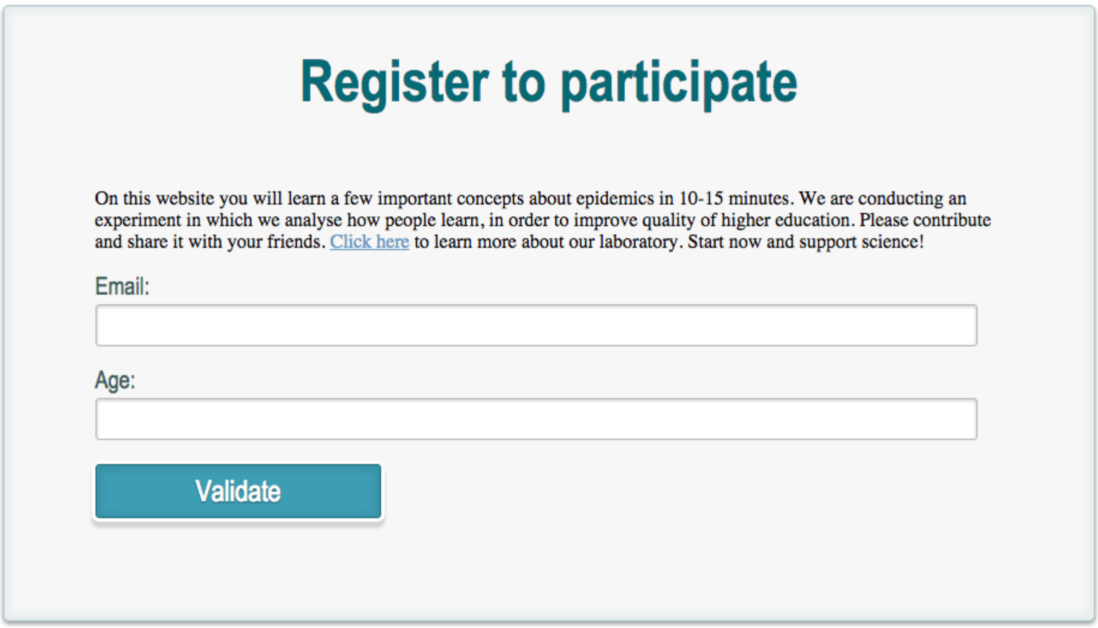


Figure 1: Learning interface

2.1 Personnality test interface

In order to gather the data for our analysis, we had to develop a learning web platform[1] coded in Django, inspired by a code on Github for a survey form [2] . In the following paragraph we call a "user" the person that participates to the experiment.

To meet the idea of having wider understanding of a learner's profile, we tried a set of different questions that could be related to different aspects of a person: psychological, demographic or IQ estimation. We list them below together with the corresponding rating system:

- Two questions derived from IQ tests that would be grouped together in a variable named "IQ". -1 was attributed to any wrong answer and +1 was attributed to the good answer.

- Ten personality questions from the International Personality Item Pool [3]. Five questions are related to the conscientiousness of a person and five questions related to the openness of a person. These two criteria have proven, in previous research[4], that they are correlated to a person's learning process. For each case, three items were positively marked and two negatively marked. For this part of the test, the answer varied from "I strongly disagree" to "I strongly agree" and the points ranged from 1 to 5 or -1 to -5 in the case we gave negative marks. The result for the consciousness was stored in the column named "conscious" whereas the openness' results was stored in a variable named "open".

- Five questions that are related to the logic of a person and his capacity for imagination. These five questions were taken from the paper folding test [5]. +1 was attributed to the right answer while -1 was attributed to a false one. The results were stored as a variable named "paperFolding".

  For the collection of demographic information, we asked the following questions:

- One question about the academic level of the test taker. The possibilities to pick from were: "Haven't graduated high school", "High school graduate", "Apprenticeship", "College student", "Bachelors", "Masters", "Doctorate". The answers in this order were assigned a number from 1 to 7. The results were saved as a variable named "level" and treated as a categorical variable in the data mining process.

- One question about the gender of the person. During the first test the user was asked his gender and the answer was stored in a variable named ?gender?.

[1] https://github.com/chili-epfl/learning-platform

[2] https://github.com/jessykate/django-survey

[3] http://ipip.ori.org The Big-Five Factor Structure

[4] Linking proactive personality and the Big Five to motivation to learn and development activity. Major, Debra A.; Turner, Jonathan E.; Fletcher, Thomas D. Journal of Applied Psychology, Vol 91(4), Jul 2006, 927-935. http://dx.doi.org/10.1037/0021-9010.91.4.927

[5] http://steinhardtapps.es.its.nyu.edu/create/assessment/vz2/start.html

- One question asking in which country the test taker is living.
- One question about the age of the person saved as the variable "age".

Other implied features used were the time a person takes to accomplish a test or Or one of the activities or the whole experiment.

The developed website guides you through different steps in a precise order explained later in section 2.2. The goal of this experiment is to teach a new concept through different paths assigned randomly and in a uniform way to each visitor of the web page.

At the end of the experiment, we also ask questions about how the test went and how people felt about it to have additional information that could infer relevant psychological features.

The taught concept is "Epidemics" and users are asked the same set of questions about this theme before and after they learn about it so that we could calculate the learning gain of a person as the difference between the scores he gets for the two tests.

2.2 Website Hierarchy

1. The registration form: It is the first step of the experiment where the user needs to register entering his email and his age. The email is an assertion that the user has a unique identifier. A user can only take the test once in order to measure his learning gain when he goes only once though the teaching activities. Note that the user is free to go through the activities as much as he wants during this one time experiment meaning that he could watch a video more than once while in the video page, pause it, move forward etc... Equally the user could read a text as much times as he wishes.
   Besides, doing the experiment only one time allow us to measure the time it takes a person to go through all the steps.

2. The psychological test: The questions of this test are enumerated in section 2.1 and the answer of the user to them constitues the features that will help us search for patterns among the subjects and to find the best path for them to improve their learning experience.

3. The pretest: It is the same as the post test and is composed of six questions with different degrees of difficulty about "Epidemics", the theme treated in the learning activities of this experiment. The user is not supposed to have previous knowledge on the subject. Thus, for each question, a user could always select the answer "I don't know" to avoid random guessing. The score of the test was the sum of the points collected for each question where a valid answer is marked +1, a false answer is marked -0.5 and the questions answered "I don't know" are set to 0.

4. The activities: There are two successive activities each one aims to teach one concept. The first concept is the epidemic's definition and it's causes and the

second one is the epidemic's reproductive number. Each concept could be explained with a video or a text. That way, we had four different teaching paths: text-text, text-video, video-text, text-text.

5. The post test: see pretest.

6. The assessment form: It is composed of some of the NASA-TLX [6] questions and gathering the opinion of the user about different aspects of the experiment taken. This questionnaire is known as an assessment tool that rates the perceived workload in order to assess a certain task.

3 Data description

We gathered a total of 167 participants but only 77 of them accomplished both pre and post tests and went through all the experiment steps (from the registration to the assessment form). As our goal is related to the learning gain, we work with these 77 users. Among them, we have 25 female user and 52 male user. We have the following distribution for the possible learning paths:

- 27 users for text-text.
- 20 users for text-video.
- 12 users for video-text.
- 18 users for video-video.

4 Data visualization

As we have only have a few observations, we try to limit ourselves to a reasonable number of variables to avoid building bad models of regression due to the curse of dimensionality. The curse of dimensionality could be due to a large number of variables for a small number of observations. We choose 13 variables, among them 11 (open, conscious, IQ, PaperFolding, age, timePretest, timePosttest, timeAct1, timeAct2 and totalTime) are continuous and two (gender and education level) are categorical. We start by visualizing the box plot of the learning gain as a function of the learning path showed in figure 2. We observe that we cannot see a significant change in the learning gain that could support the hypothesis that one path is always the best for all types of learners.

We now try to see if the learning gain is correlated to some variable or some subset of variables given a learning path. For this experiment, we didn't try to use predefined categories of learners. Our approach was to collect as much data on the user's learning behavior as possible without assumed correlation between them. We then tried plotting each variable against the learning gain achieved by the user and

[6] http://humansystems.arc.nasa.gov/groups/tlx/

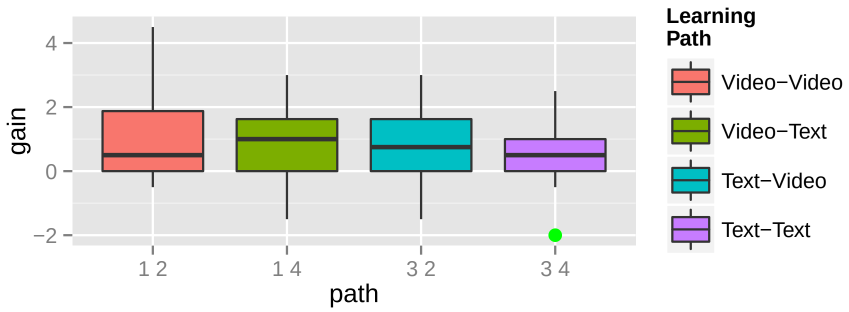


Figure 2: Learning gain according to the learning path

compare it for the four possible paths that a user could get. Figure 3 shows the box plot of the learning gain depending on the gender. The results cannot be fully trusted to make direct comparison between genders because of the high proportion of males compared to the number of female participants. But we can see that, within our observations, the video-video path and the text-video path worked better for females whereas video-text path was more successful for males.

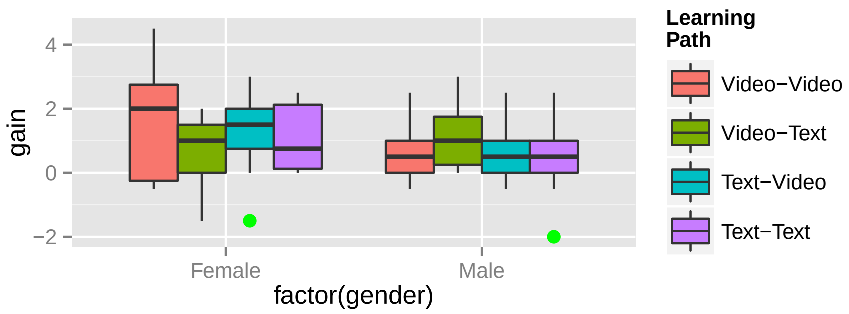
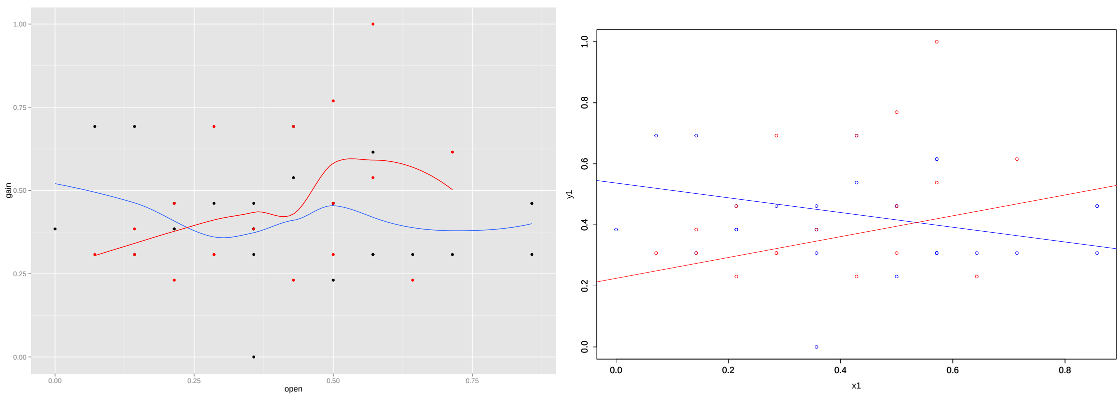


Figure 3: Learning gain according to the learning path

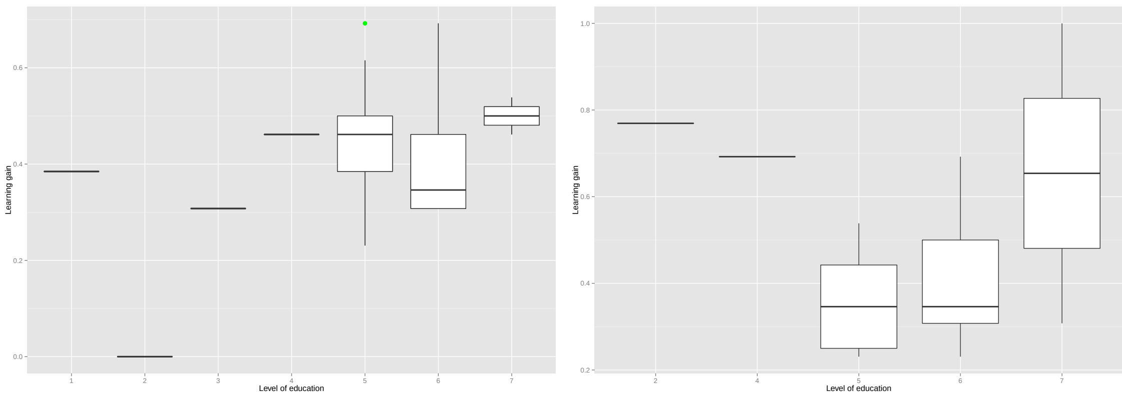
Hereafter, we decided to only focus on the video-video path and the text-text path to make the first comparisons and find the variables' correlations with the learning gain. In some cases we can observe a potential relation between the variable and the obtained learning gain. For instance, we can see that in figure 4a representing the case of the text only path, the learning gain doesn't seem to be impacted by the openness of the person whereas for the video path showed in figure 4b we get better results if the user is more open.

We can also see in figures 5a and 5b that the highest the educational level of the learner is the better his learning gain is, and this is more obvious in the case of the video only learning material. In fact, people with a doctorate perform much better than the other categories. Again, for the users that are assigned video only paths, we have a correlation of 0.4986 between the time they spend watching the first video and their learning gain. This result is quite logical as people that don't watch the whole video won't achieve good results in the post test. In contrast, other



(a) Local Polynomial Regression Fitting **(b)** Linear Regression

Figure 4: Learning gain depending on openness



(a) Users with text only material **(b)** Users with video only material

Figure 5: Learning gain depending on the achieved level of studies.

variables as the consciousness of the user doesn't seem to affect the learning gain for the observed cases. Thus, we could think that some variables could indicate a difference in the learning experience of the learner depending on the activities he undertakes to learn the new concepts.Demographic factors such as the gender, the level of studies of the learner or his his psychological features could also influence his learning experience.

5 Building a regression model

In this step of the project, our goal is to optimize the prediction of the learning gain of a person based on some variables. Thus, we would like to have a predictive model that could show the correlation between the learner's personality and the teaching outcome. Here, machine learning could help us find the most relevant personality features for a good prediction. The first approach is to have a model by learning path so that we could compare the different results we would expect from a learner depending on his activities.

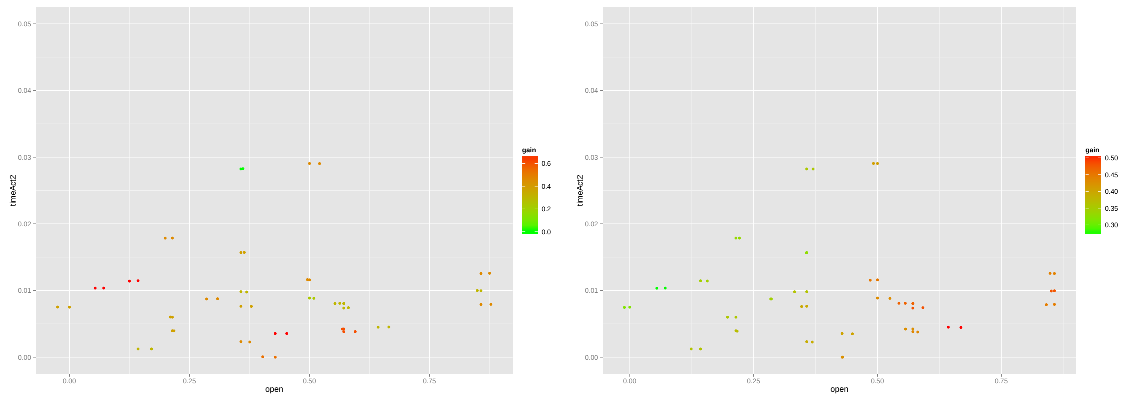
5.1 Model implementation

Despite of the number of complete observations (77 complete experiment results), we try to build a regression model with machine learning algorithms to predict the learning gain of a user.

Before applying machine learning algorithms, we scale the data with the minimum and maximum values to obtain, in the case of the continuous variables, values between 0 and 1. The level and gender variables are considered as factors in R. We choose the Support Vector Machines (SVM) regression technique with the package e1071 in R. SVM provides the usage of kernels what allow us to have a non-linear regression. The radial kernel had performed better for our model.

5.2 Results

After the tuning of the parameters, we obtain a Root Mean Squared Error (RMSE) of 0.14 for the text only path and of 0.2 for the video only path. We used a radial kernel and performed a 5-fold Cross Validation (CV) with 80% of the observations used for the train set and 20% of the observations for the test set. With dummy encoding of the categorical variable for SVM we don't achieve better results. We then try to perform simple feature selection, taking only variables that have correlation with the learning gain above 0.1 taking their absolute values. We again don't see an improvement in the predictions. We can think that the poor prediction results are due to the lack of data as we have only 27 observations for text-only path and 18 observations for video only path. Moreover, the fact that we base our prediction on human variables could also introduce some unpredictability to our regression. Nonetheless, we use the model built for video only path to predict the learning gain of the users that had text only path to estimate their learning gain if they had a different path and compare the results. We can observe a change for certain users. In fact, when we plot the learning gain for a pair of variables as the variable "open" that indicates the openness of a person and the variable "timeAct2" for the time a person took for the second activity, certain points have different learning gains. In fact, figure 6b in comparison with figure 6a shows that people with high openness would have higher learning gain with video path while people with very low openness perform better with text material. This could confirm what we saw as correlation between the openness of the user and his learning gain and could suggest that we can adapt the activities proposed to the learner depending on this psychological feature.



(a) Users with text only material

(b) Users with video only material

Figure 6: Comparison of the gain of the users that had text material with their estimated gain with video material.

6 Work limitations

Due to time limitation, the activities chosen differ simply in how they are presented and not in the content. Changing the content and the teaching style(visual vs aural) introduce two variables at a time. But our current paths could lead to very small changes on the learning gain results and makes it harder to correlate the features to the results. Paths following the inductive or deductive approach for example could be more differentiating for learners' profile. We could also have opted for MOOC only material but some videos would have more graphs and photos than others. Besides, completely different features (time, demographics, psychological traits...) could lead to a kitchen sink effect, which doesn't help to find the right regression [7]. This approach is still good for a first exploration as we didn't need to restrain ourselves with predefined learning styles and this work wanted to find new groups of relevant features that defines learners' styles. In future work, we could target more the wanted features to study. Most importantly, we have to gather more data to improve the construction of regression models. Some rules of thumb indicates that we have to gather 10 observations for one predictor. The goal is to have a statistically sufficient number of observations by category, for instance a good number of leaners in both male and female categories or in all four paths.

For this experiment, the recruitment of users was done via social media advertisement. This is a good way to gather different styles of learners than if we limited our users to one environment like EPFL. Some users feedback was in respect to the website's saying that they needed more explanation about the experiment's steps by explaining more the experiment or by showing the progress of the test (A progress bar or the page number out of the total number of steps to accomplish). In addition, the fact that the test about the concepts is taken twice is not highlighted enough during the experiment. Warning the user about to could have encouraged more him

[7]https://www.kellogg.northwestern.edu/faculty/dranove/htm/dranove/coursepages/Mgmt%20469/choosing%20va

to pay attention to the presented content and extract the useful information. Our concerns were more focused on doing an easy to follow not very time demanding learning platform. Some people also asked for feedback which was not given after the end because in our perspective, scores to the test weren't important but the improvement between the two test scores was. Even if emailing scores doesn't hep the experiment results, it could encourage people to share the test with friends.