Scikid-Learn: An AI tool for customized education

Andy Spezzatti, Mike Lawrence, Shine Shan, Niema El Bouri, Ada Tanyindawn, Aljaz Kosmerlj, James Hodson

May 29, 2019

Abstract

Education is at the hearth of societal, individual and collective development and is supposed to facilitate children empowerment. Yet, in many situations, it fails to do it well, whether because of economic problems or outdated approach. According to USI data, about 260 millions of children are out of school and a study conducted by Cindy Liu on US students concluded that 1 out of 5 students would consider suicide because of stress at school.

Education is currently not optimized, with the large level gap between students and individual differences not sufficiently considered. From these observations, we built SciKid-Learn: a mobile application for customized and adaptive learning that learns from children and recommends customized content as well as provide tests to evaluate progress. It was first developed thinking of application for students in developing countries, lacking an easy access to a good education. Yet, we also found it more generally relevant for every teaching institution, as a support to the traditional approach. It gives a way to easily learn from children, adapt and track progress more efficiently and during the whole life of each individual.

Clustering for personalization

Clustering of students We define a profile for each student, so we can know in which way the student is learning more efficiently. This profiling is done using learning style as in 2, personality as well as knowledge of academic subjects.

Clustering question according to cognitive abilities In order to test children on math questions and to customize the experience, we clustered the dataset according to three abilities: spatial representation, calculus and problem solving, in a semi-supervised way. We first hand labeled three hundred questions and then used a Support Vector Machine model with L2 regularization to cluster the remaining questions.
Clustering by content Questions were also clustered by their content, using NLP techniques for text preprocessing. Visualizing the feature space, we get an idea of the optimal number of cluster, confirmed using the elbow method. The latent space in Figure 3 (of dimension 500 here) is colored by cluster. It highlights the patterns behind the data. Figure 4 shows the distribution of words for a clustering focused on academic subjects.

Recommendations of learning contents The algorithm webscrapes several educational websites to find the most relevant articles and online classes. The first version was only based on wikifier API, a programming interface using ngrams to select relevant Wikipedia articles. It was improved using data from edX, academia.com and other online education platforms to provide a more powerful search tool.

The future of Scikid-Learn Neuroergonomics to optimize how we are using our brain It involves the collection of physiological data to gather important information such as attention level. It can help to tackle challenges like optimizing workload level in order to estimate how to ensure continued engagement with a high performance level.

The raise of more powerful NLP models A lot of learning material involves either written documents or vocal ones. Both of them could be analysed using NLP techniques. With the recent development of high performing NLP models (BERT, GPT2), more accurate and efficient learning tools can be developed.

Results

Interactive dashboard Students and teachers will interact through an interactive dashboard that track the students’ progress, display the recommendations and enable to visualize specific metrics, see Figure 5.

Partners

Aljaz Kosmerlj, AI for good foundation
James Hodson, CEO at AI for good foundation
Ikhlaq Sidhu, Sutardja Center, UC Berkeley
Alexander Fred-Ojala, Sutardja Center, UC Berkeley