Machine Learning, Natural Language Processing, and their Application in Educational Assessment

Jiangang Hao
2/24/2022
Educational Testing Service

Introduction

Digital learning and assessment generate a large amount of data in various forms. To deal with the complex data, psychometric researchers need to garner new techniques from data science, machine learning, and natural language processing (NLP). To facilitate the learning, an edited volume, Computational Psychometrics (Eds. von Davier, Mislevy, & Hao, 2022), was developed to introduce the new techniques in a way more accessible to researchers with psychometrics backgrounds. ML and NLP are two very extensive research disciplines themselves, and it is impossible to cover them fully in such a short workshop. As such, this workshop aims to provide the participants with a high-level overview of ML and NLP as well as some practical knowledge of using them in the context of educational assessment. We hope the workshop can help participants to get started with the acquisition of these important skills.

Prerequisites

Basic knowledge in statistics and python programming language.

Learning Goals

- Learn the basics of machine learning and some typical algorithms
- Learn how to use machine learning in Python and Orange
- Learn the basics of natural language processing (NLP)
- Learn how to use NLP in Python

References

Von Davier, A., Mislevy, R., & Hao, J. (Eds.), Computational Psychometrics: New Methodologies for a New Generation of Digital Learning and Assessment, Springer. https://link.springer.com/book/10.1007/978-3-030-74394-9

Bios of the Instructor

Dr. Jiangang Hao is a research director at the Research and Measurement Sciences Division of Educational Testing Service and is currently leading the Psychometric and Data Science Modeling group. His research centers on leveraging big data to measure people and their interactions, with a particular focus on behavioral data mining,

game-based assessment, learning analytics, AI/machine learning, natural language processing, automated scoring, learning, and assessment of 21st-century skills such as collaboration and communication. Dr. Hao got his Ph.D. in Physics and MA in Statistics from the University of Michigan. Before joining ETS, he worked on large-scale data in astrophysics and cosmology at Fermi National Accelerator Lab. His research spanned from developing CCD camera and astronomical digital image processing pipeline to developing algorithms for galaxy clustering and cosmological parameter estimation. Dr. Hao published extensively, and most of his publications can be accessed at the Google scholar site: https://goo.gl/pgggDK.

Schedule (subject to adjustments)

Description	Format	Duration
Overview of the data challenges from digital learning and assessment	Lecture	20 min
1.1 Introduction to machine learning basics • Supervised, unsupervised • SVM, Decision Tree, Random Forest, Gradient Boosting • Evaluation Metrics, ROC • Cluster analysis	Lecture	1 hour 30 min
1.2 Machine learning in Python: Scikit-learn	Demo & hands-on practice	30 min
1.3 Machine learning in GUI software: Orange	Demo & hands-on practice	30 min
Break		
2.1 Introduction to natural language processing basics • Text representation, N-gram, TF-IDF, word embedding • LSA/LDA • Automated Scoring • Deep learning-based language models	Lecture	1 hour 30 min
2.2 Natural language processing in Python: NLTK, spaCy	Demo and hands-on practice	1 hour
Wrap up/Q&A		30 min