RSMTool: collection of tools building and evaluating automated scoring models

Nitin Madnani[1] and Anastassia Loukina[1]

1 Educational Testing Service

Summary

RSMTool is a collection of tools for researchers working on the development of automated scoring engines for written and spoken responses in an educational context. The main purpose of the tool is to simplify the integration of educational measurement recommendations into the model building process and to allow the researchers to rapidly and automatically generate comprehensive evaluations of the scoring model.

RSMTool takes as input a feature file with numeric, non-sparse features extracted from the responses and a human score and lets you try several different machine learning algorithms to try and predict the human score from the features. RSMTool allows the use of simple OLS regression as well as several more sophisticated regressors including Ridge, SVR, AdaBoost, and Random Forests, available through integration with the SKLL toolkit (Blanchard et al. 2014). The tool also includes several regressors which ensure that all coefficients in the final model are positive to meet the requirement that all feature contributions are additive (Loukina et al. 2015). The primary novel contribution of RSMTool is a comprehensive, customizable HTML statistical report that contains feature descriptives, subgroup analyses, model statistics, as well as several different evaluation measures illustrating model efficacy. The various numbers and figures in the report are highlighted based on whether they exceed or fall short of the recommendations laid out by (Williamson, Xi, and Breyer 2012). The structure of RSMTool makes it easy for researchers to add new analyses without making any changes to the core code structure thus allowing for a wide range of psychometric evaluations.

The tool is written in Python and works on all platforms (Windows/Linux/Mac OS X). The source code is available from Github (Madnani and Loukina 2016)

References

Blanchard, Dan, Nitin Madnani, Michael Heilman, Nils Murrugarra Llerena, Diane M. Napolitano, Aoife Cahill, Keelan Evanini, and Chee Wee Leong. 2014. "SciKit-Learn Laboratory (Skll) 1.0.0." doi:10.5281/zenodo.12825.

Loukina, Anastassia, Klaus Zechner, Lei Chen, and Michael Heilman. 2015. "Feature selection for automated speech scoring." In *Proceedings of the Tenth Workshop on Innovative Use of Nlp for Building Educational Applications*, 12–19. Denver, Colorado.

Madnani, Nitin, and Anastassia Loukina. 2016. "RSMTool." https://github.com/EducationalTestingService/rsmtool.

Williamson, David M., Xiaoming Xi, and F. Jay Breyer. 2012. "A framework for evaluation and use of automated scoring." *Educational Measurement: Issues and Practice* 31 (1): 2–13. doi:10.1111/j.1745-3992.2011.00223.x.