

USING HYBRID FEATURE IDENTIFICATION  
& MACHINE LEARNING

---

**AUTOMATED ESSAY SCORING**

---

## ADVANTAGES OF COMPUTAERIZED SCORING

- ▶ Provide individual revision and feedback more quickly. The scores can be more descriptive than human raters.
- ▶ Increase the practicality in administering large-scale assessments of writing ability.
- ▶ Page (2003) states that "the automated ratings would surpass the accuracy of the usual two judges."

---

## AUTOMATED ESSAY SCORING (AES) SYSTEMS

- ▶ Project Essay Grader (PEG) : developed in 1966 by College Board, uses proxy measures (average word length, essay length, number of semicolons or commas) to predict the intrinsic quality of the essays.
- ▶ One of the strengths of PEG is that the predicted scores are comparable to those of human raters. Second, the system is computationally tractable.
- ▶ Methodology is straightforward

---

# AUTOMATED ESSAY SCORING (AES) SYSTEMS

- ▶ Intelligent Essay Assessor (IEA): uses a semantic text analysis method (Latent Semantic Analysis).
- ▶ Latent Semantic Analysis (LSA) is defined as “a statistical model of word usage that permits comparisons of the semantic similarity between pieces of textual information” (Foltz, 1996, p. 2). LSA first processes a corpus of machine-readable language and then represents the words that are included in a sentence, paragraph, or essay.

---

# E-RATER

- ▶ **E-rater: currently used by ETS for operational scoring of GMAT**
- ▶ Linear regression model with syntactic, rhetorical and topical features.
- ▶ **Syntactic Features:** parses sentences using the Microsoft Natural Language Processing. The identification of syntactic structure yields information about syntactic variety.
- ▶ **Rhetorical Structure Analysis:** identifies rhetorical relations, such as Parallelism and Contrast.
- ▶ **Topical Analysis:** evaluates the lexical and topical content of an essay by comparing the words it contains to the words found in manually graded training examples.

---

## DATASET/CORPUS

- ▶ The training and test data were acquired from a past competition from Kaggle.com sponsored by Hewitt- Packard.
- ▶ Each of the sets of essays was generated from a single prompt. Selected essays range from an average length of 150 to 550 words per response. Some of the essays are dependent upon source information and others are not. All responses were written by students ranging in grade levels from Grade 7 to Grade 10. All essays were hand graded and were double-scored. Each of the eight data sets has its own unique characteristics.

---

## POSSIBLE IMPROVEMENT OF E-RATER

- ▶ Syntactic Features: besides syntactic variety, also emphasize the importance of grammaticality and syntactic correctness.
- ▶ Topical Analysis: use machine learning algorithms to supplement the grading process. Linear regression, SVM, etc.
- ▶ When grading for different tests, different weights of these three modules.

---

## REFERENCES

- ▶ Jill Burstein, Karen Kukich, Susanne Wolff, Chi Lu, Martin Chodorow, Lisa Braden-Harder, Mary Dee Harris. 1998. Automated scoring using a hybrid feature identification technique. In the Proceedings of the 17th international conference on Computational linguistics - Volume 1.
- ▶ Page, E. B. (2003). Project Essay Grade: PEG. In M. D. Shermis & J. Burstein (Eds.). Automated essay scoring: A cross-disciplinary perspective. Mahwah, NJ: Lawrence Erlbaum Associates.
- ▶ Semire Dikli. 2006. Automated Essay Scoring. Turkish Online Journal of Distance Education - Volume 7, Issue 1.