Toward Large-Scale Automated Scoring of Scientific Visual Models

Chee Wee (Ben) Leong[1] Lei Liu[1] Rutuja Ubale[1] Lei Chen[2]
[1]Educational Testing Service [2]Liulishuo.com

Motivation

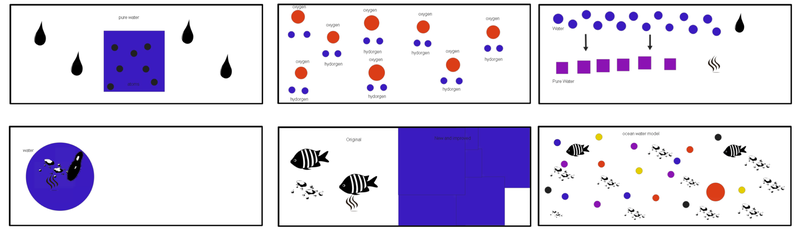
- The Next Generation Science Standards[1], developed for the U.S. K-12 educational system, explicitly identify modeling as one central and valued practice
- Visual modeling by students is a rich vehicle of information for educators. What do students know and what they can do?
- Learning progressions (LPs) to guide assessment development in order to develop useful diagnostic tools of knowledge acquired by students and the ability to inform instruction
- Combined visual modeling with LP in a computerized assessment for scientific concepts (e.g., Matter)
- Automated Scoring
  - fair and valid way to assess understanding of structure and behavior of Matter
  - disentangle the interaction between scientific modeling skills vs. artistic skills

Learning Progression (LP): A Standardized Evaluation Framework

	S				MI			B				D		
	1	2	3	4	1	2	3	1	2	3	4	1	2	3
LP-1	X	X			X	X		X				X		
LP-2		X	X			X		X	X			X		
LP-3			X	X		X	X		X	X			X	
LP 4				X		X	X			X	X		X	X
LP-5				X			X				X			X

- Scale (**S**)
  - composition of Matter beginning with the smallest units -- e.g., nanoscopic particles
- Material Identity (**MI**)
  - examines the anticipated number/identity of particles present
- Behavior (**B**)
  - examines if/how particle movement is represented
- Distribution (**D**)
  - examines positions of individual particles and space between them in a given Matter state

Visual Modeling Dataset



- Formative assessment prototype focused on core concept of Matter
- Pure water model vs. Ocean water model
- Drawing toolkit:
  - Micro-objects (circle, square, triangle, diamond)
  - Macro-objects (water drops, water steam, algae, salt, etc.)
  - Labels
  - Arrows
- 148 student-pair visual models of Matter
- Expert human annotators agreement
  - 100% for LP level; 97.1% for all sub-LP levels

Features and Modeling

Features	Brief Description	S	MI	B	D	LP
Counting-based						
mico-object types	# types of micro-objects	0.397	0.454	0.025	0.457	0.424
macro-object types	# types of macro-objects	-0.467	-0.449	-0.340	-0.414	-0.469
mico-object color types	# types of micro-object colors	0.252	0.291	0.014	0.315	0.265
EIC deviation	| EIC - # types of micro-object color-shape |	-0.376	-0.394	-0.132	-0.342	-0.373
arrows	# arrow instances	0.319	0.253	**0.772**	0.257	0.443
arrow strengths	# mean length of arrow instances	0.153	0.123	0.360	0.132	0.211
arrow randomness	variance of direction of arrow instances	0.182	0.158	0.597	0.173	0.324
Spatial-based						
k-NN-3	mean distance to 3 NN micro-objects	0.517	0.546	0.128	0.568	0.541
k-NN-10	mean distance to 10 NN micro-objects	0.512	0.550	0.128	0.573	0.544
dispersion	mean normalized spread between micro-objects	**-0.548**	**-0.589**	-0.128	**-0.614**	**-0.578**

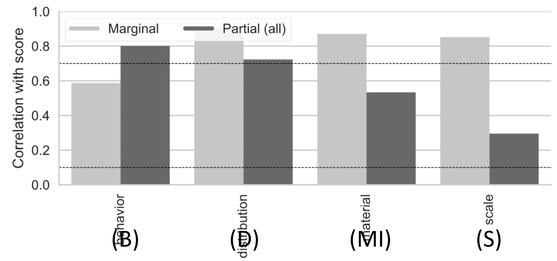
Scoring Model

- Supervised learning process: regression (rounding) vs. classification
- Matured statistical properties and explainable outputs
- 10 features extracted
- 10-fold cross validation using scikit-learn via SKLL
- QWK – penalize misclassification proportional to difference in agreement

Discussion

Learner	QWK	S.D.
Linear Regression	0.624	0.12
Decision Tree Regression	0.517	0.29
Random Forest Regression	0.623	0.17
Support Vector Regression (LibSVM)	0.551	0.17
Logistic Regression Classifier	0.581	0.10
Decision Tree Classifier	0.559	0.07
Random Forest Classifier	0.538	0.06
Support Vector Classifier (LibSVM)	0.443	0.29

Future Work



Improvements:

- Address skewed data distribution
- Informed feature engineering
- Multimodal-based scoring by using textual descriptions of students

References:

1. National Research Council and others. 2013. *Next generation science standards: For states, by states*.