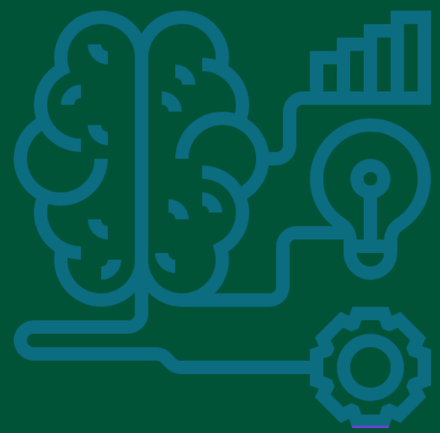
Automated Assessment using Machine Learning in blended learning : A review of Innovative Methodologies

Abbirah Ahmed
Arash Joorabchi
Martin J. Hayes



UNIVERSITY OF LIMERICK
OLLSCOIL LUIMNIGH

Outline of the Presentation

To provide an insight into the use of Machine Learning during the assessment process in teaching and learning.

Features and Limitations in relation to the automated assessment of responses.

Sample results for Short answer grading

Qualitative Analysis of Sentiment in Survey results

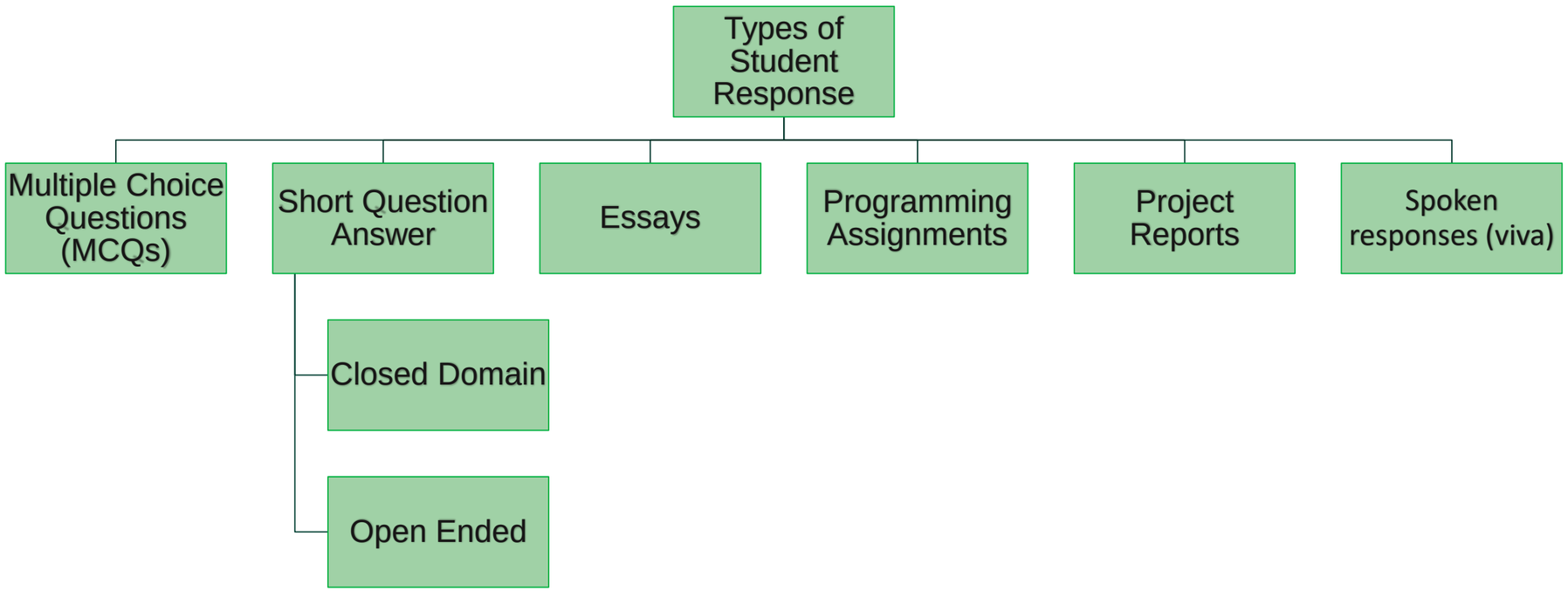
Automated Assessment

Definition:

A **student's response** to a question is automatically analyzed and either:

a) a **score** is generated in order to assess a **student's knowledge** and/or other **skills** and/or

b) some **actionable feedback** is provided



Challenges in Traditional Methods

Manual grading

Time consuming

Multiple correct answers

Ensuring fairness and consistency

Varying methods of grading

Example from Short Question Answer Dataset

Id	Question	Solution	Answer	Score-1	Score-2	Avg Score
1.1	Where are variables declared in a C++ program?	Variables can be declared anywhere in a program. They can be declared inside a function (local variables) or outside the functions (global variables)	inside the function scope and outside of the function scope in case of global variables	5	5	5
			variables can be declared in classes and methods.	3	4	3.5
6.5	What does the size of operator return?	The size in bytes of its operand.	the size of the list object.	4	2	3

Variability in expert assessments of a particular response poses a fundamental constraint on the use of Neural Networks to assess that response.
Practical solution: More training and/or more expert reviews. Increase the likelihood that the next response to be assessed is similar to text that has already been considered in the training of the network.

Existing Approaches

Classical text similarity Techniques

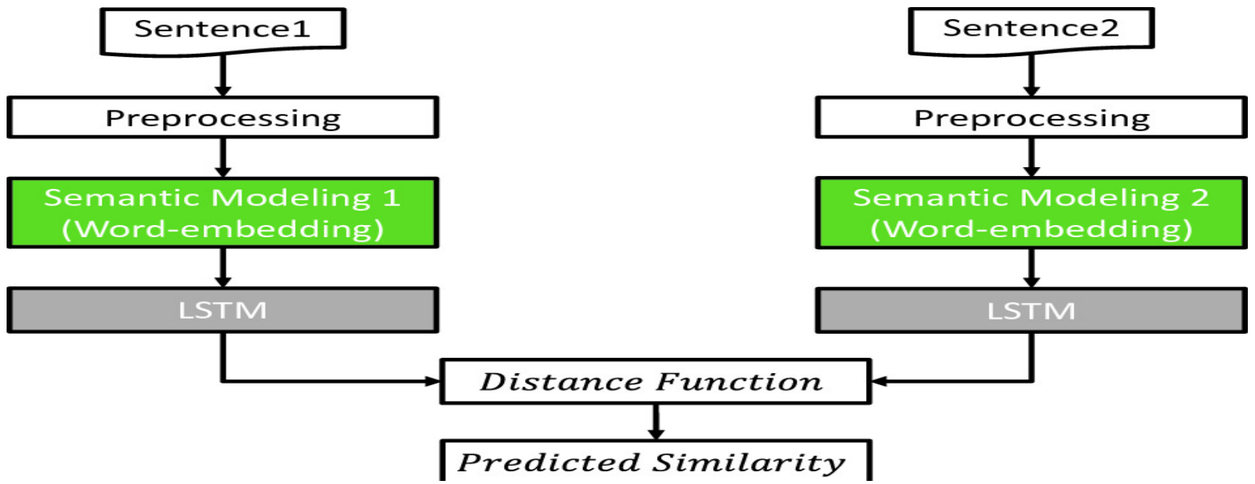
- *Time consuming*
- *Requires domain expertise*
- *Computationally expensive due to high dimensional vectors.*

Neural Network based Techniques

- *Eliminates the need of feature engineering*
- *Reduced human intervention*
- *Cost effective*

BiLSTM based ASAG System

BiLSTM: Bi-Directional Long Short Term Memory Networks. A Neural network approach to the analysis of content that assesses the likelihood of the use of other words ***before*** and ***after*** the word in question.



	id	solution	answer	score_me	score_other	score_avg	label	multiqa_predict	nli_distilbert_predict	distilroberta_predict
3	1.1	To simulate behaviour portions desired softwar...	Defined Specification phase stimulates behavio...	5.0	5.0	5.0	1.0	0.814893	0.906850	0.813368
4	1.1	To simulate behaviour portions desired softwar...	It used let users first idea completed allow c...	3.0	3.0	3.0	0.6	0.606470	0.667543	0.571624
5	1.1	To simulate behaviour portions desired softwar...	To find errors finalized	2.0	2.0	2.0	0.4	0.406924	0.436603	0.401964

Limitations in Deep Learning based Automatic Scoring Systems

- Unavailability of sufficient training data
- Training time of the model
- Generalizability of model
- Accuracy and reliability of the systems
- Limited existing integration with Virtual Learning Environments (VLEs)

Qualitative Analysis (Closed Domain vs Open Ended Questions)

Closed Domain	Open Ended
Are you satisfied?	How satisfied or dissatisfied are you with this process?
Do you think you would use this?	How would this fit into your work?
Does that work for you?	What do you think about that?

Sentiment analysis (or opinion mining) is a natural language processing (NLP) technique used to determine whether data is ***positive, negative or neutral.***

Name Entity Recognition automatically identifies named entities in a text and classifies them into predefined categories. Entities can be names of people, organizations, locations, times, quantities, monetary values, percentages, and more.

Qualitative Data based Analysis of ‘Sentiment’

Sentiment analysis is an ‘easier’ challenge than assessment in a Teaching and Learning context

Easier to get familiar with responses – less training

Research objectives are more straightforward – is a respondent positive or negative in response to a question

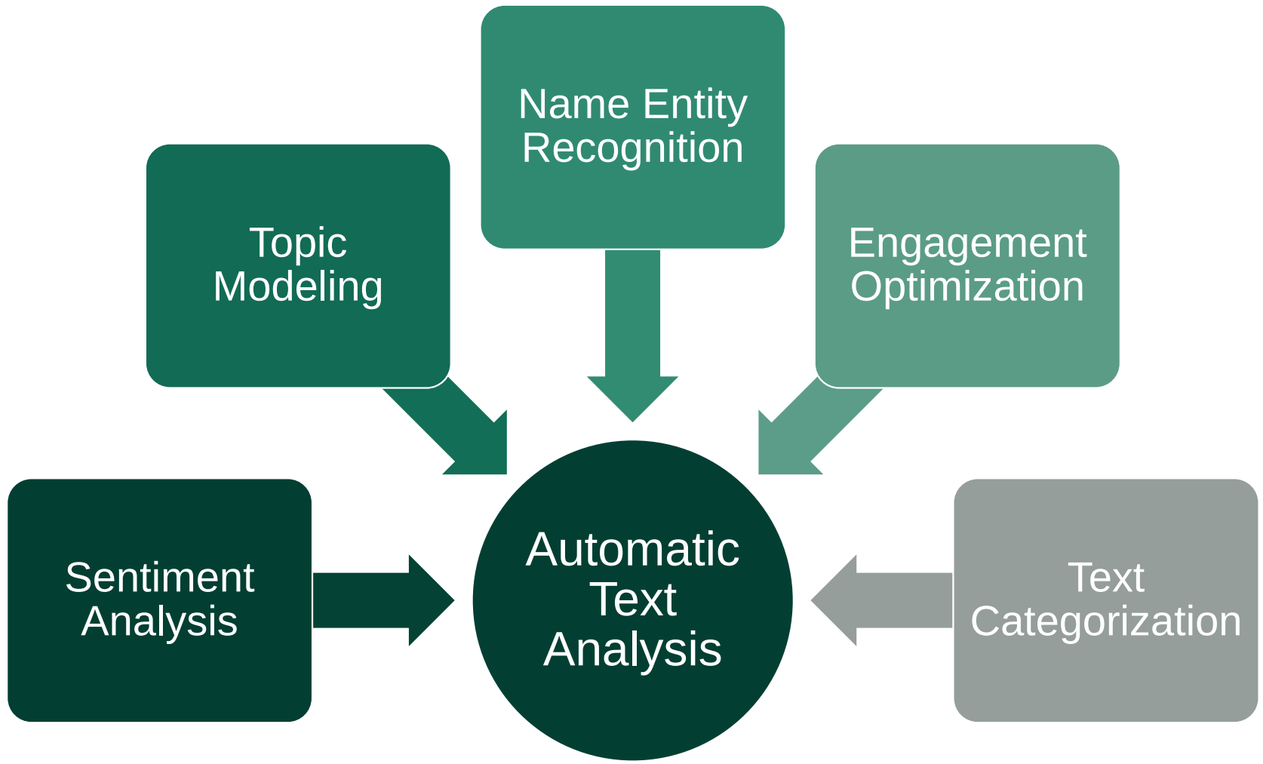
Response framework fewer responses patterns to match

Fewer connections between words and phrases to identify

Challenges

- Voluminous
- Unorganized
- Free human responses (vast vocabulary, sentence structure, sarcastic phrases, contradictory answers)
- No right/wrong answers. Grading more subjective?
- Thoughts, opinion, suggestions, etc.
- Still require two or more analysts to interpret the results for freeform responses.

Machine Learning in Qualitative Research

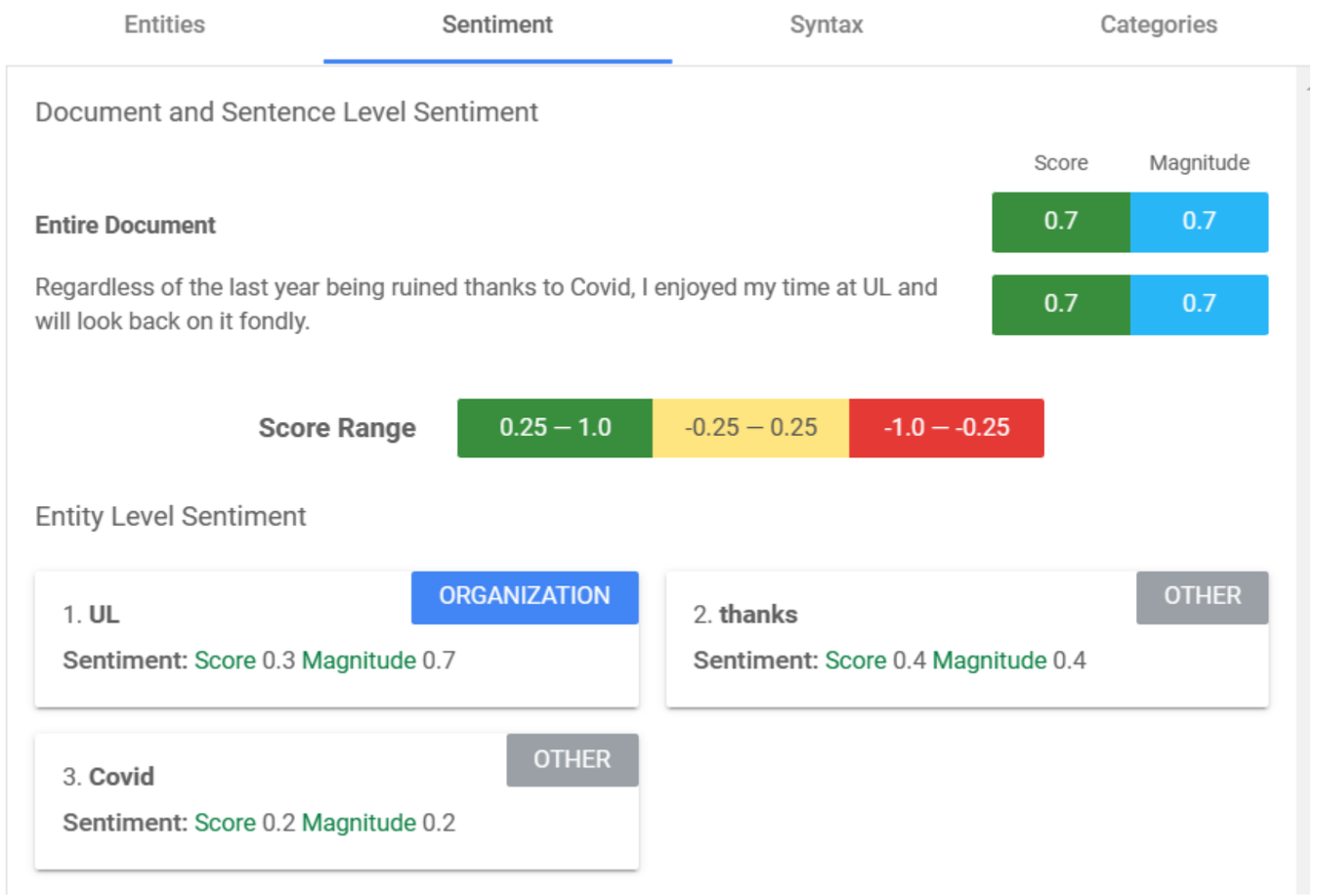


Sentiment Analysis Examples from Student Exit Survey

```
text      :Opportunity to partake in the Erasmus program.
score     :0.5
magnitude :0.5
Sentiment:  Positive

text      :The transition to online classes wasn't too bad.
score     :0.20000000298023224
magnitude :0.20000000298023224
Sentiment:  Positive

text      :Space out assessment pls
score     :-0.30000001192092896
magnitude :0.30000001192092896
Sentiment:  Negative
```

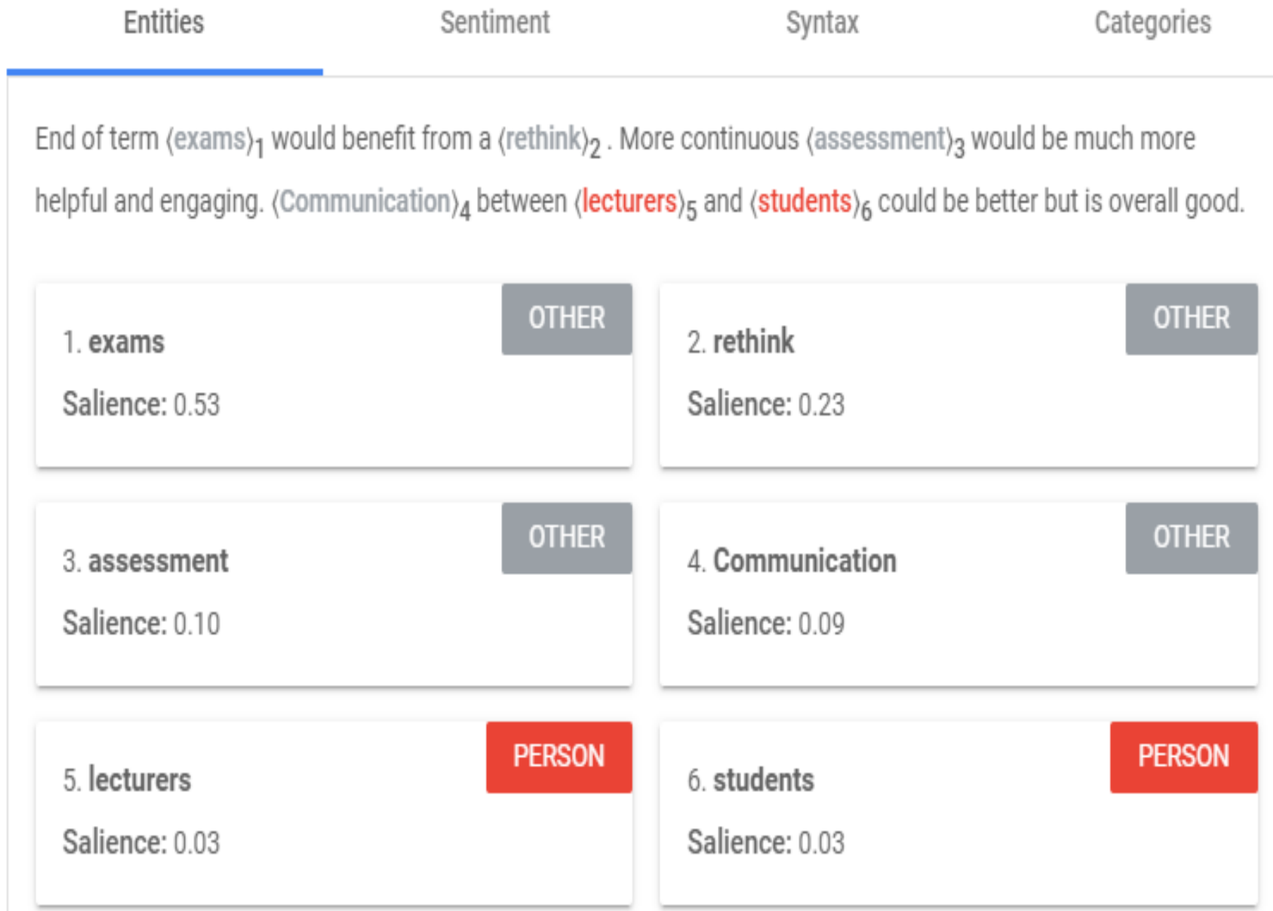


NER Examples from Student Exit Survey

```
Library not accessible during pandemic,
although other university libraryâ€™s were operating
Entity Name: Library
Entity type: OTHER
Salience score: 0.596

Entity Name: pandemic
Entity type: EVENT
Salience score: 0.208

Entity Name: university
Entity type: ORGANIZATION
Salience score: 0.195
```



Thank you

Looking forward to collaborate with you ☺

You can reach us at:

Martin.J.Hayes@ul.ie
Arash.Joorabchi@ul.ie
Abbirah.Ahmed@ul.ie

UNIVERSITY OF LIMERICK
OLLSCOIL LUIMNIGH

University of Limerick,
Limerick, V94 T9PX,
Ireland.

Ollscoil Luimnigh,
Luimneach,
V94 T9PX, Éire.

+353 (0) 61 202020

ul.ie