Hochschule
Bonn-Rhein-Sieg
University of Applied Sciences

b-it Bonn-Aachen International Center for Information Technology

Analysis of Active Learning Mechanism Applied to Language Models for Computer Assisted Short Answer Grading

September 27, 2022

Elanton Fernandes

Advisors

Prof. Dr. Paul G. Plöger, M.Sc Tim Metzler

Table of Contents

1. Motivation
2. Problem Statement
3. State of the Art
4. Approach
5. Evaluation
6. Results
7. Summary
8. Future Work
9. Extra Slides

Motivation

In universities with an increase in number of student every semester, the number of tests conducted also increases. This means that:

- The professor spends more time in correcting student exams than preparing for lectures.
- If students are not assigned full scores for on a test, they expect a meaningful feedback from the professor.

Motivation

Consider the following dummy scenario:

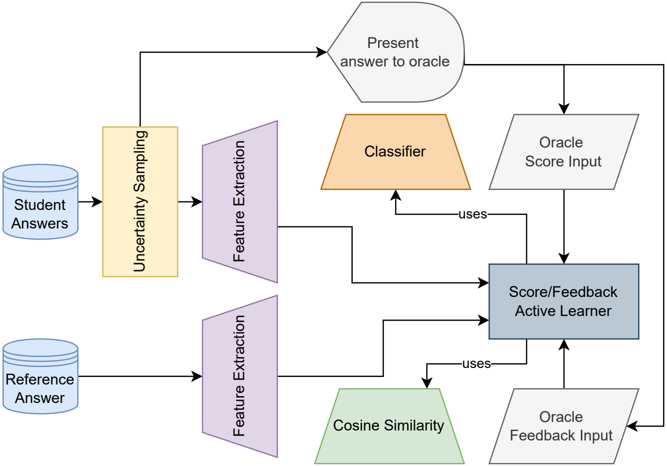
- 80 students enrolled in a class.
- Tests are conducted bi-weekly.
- Professor requires 15 minutes to evaluate one student test.
- Total time spent by the professor to evaluate all tests per week is 10 hours.

Problem Statement

- To automate the evaluation of student tests while still keeping the oracle/professor in the loop.
- Allow the assignment of meaningful feedback to student answers indicating their mistakes.

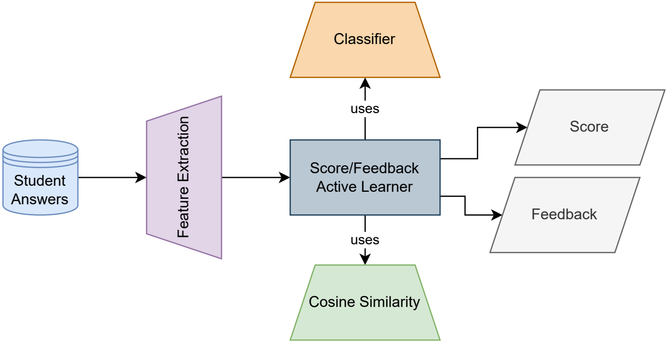
Approach

Training cycle



Approach

Prediction cycle



Approach

Uncertainty Sampling

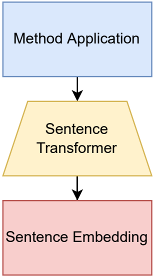
Uncertainty sampling is a query strategy that queries the instances about which it is least certain how to label. We use uncertainty sampling variant might query the instance whose prediction is the least confident:

$$x_{LC} = argmin_x P(\hat{y}|x;\theta) \tag{1}$$

Where x is the feature, y is the class label prediction, and $\hat{y} = argmax_y P(y|x;\theta)$ is the class label that has the largest posterior probability using model θ.

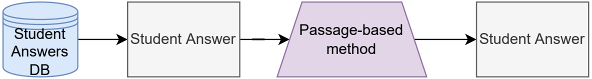
Approach

Feature Extraction: Overview



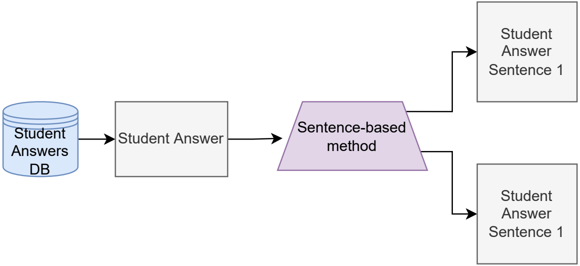
Approach

Feature Extraction: Passage-based method



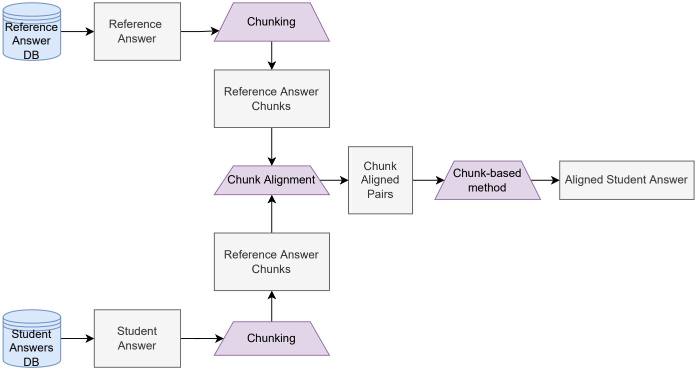
Approach

Feature Extraction: Sentence-based method



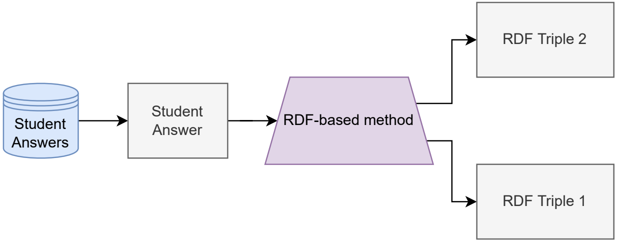
Approach

Feature Extraction: Chunk-based method



Approach

Feature Extraction: RDF-based method



Approach

Language Models

Language Model:	Base model	Number Training tuples
all-mpnet-base-v2[**?**]	microsoft/mpnet-base.	1.17B
all-distilroberta-v1[**?**]	distilroberta-base	1.12B
all-MiniLM-L12-v2[**?**]	microsoft/MiniLM-L12-H384-uncased	1.17B
multi-qa-distilbert-cos-v1[**?**]	distilbert-base	214M
all-MiniLM-L6-v2[**?**]	nreimers/MiniLM-L6-H384-uncased	1.17B

Table 1: Displays pre-trained language models with their base model used in training and number of training tuples used[**?**].

Evaluation

Score

- Pearsons Correlation

$$\rho(y,\hat{y}) = \frac{cov(y,\hat{y})}{\sigma_y \sigma_{\hat{y}}} \tag{2}$$

- RMSE Score

$$RMSE = \sqrt{\frac{1}{n}\Sigma_{i=1}^{n}(\hat{y_i} - y_i)^2} \tag{3}$$

Where y represents actual grade and $\hat{y}$ represents predicted grade with σ_y and $\sigma_{\hat{y}}$ computed as the standard deviation of y and $\hat{y}$

Evaluation

Feedback

Question	What is a variable?
Reference Answer	A location in memory that can store a value.
Student Answer	a value/word that can assume any of a set of values
Feedback A	correct
Feedback B	missing keywords: location in memory
Feedback C	A variable is a location in memory that stores a value

Table 2: Presented survey to participants.

$$Agreement\ Score = \frac{Model\ generated\ most\ rated\ feedback}{Total\ Number\ of\ Participants}$$

Results

Notations

Method	Notation
Passage-based Methods	M1
Sentence-based Method	M2
Chunk-based Method	M3
RDF-based Method	M4

Language Model	Notation
all-mpnet-base-v2	LM1
all-distilroberta-v1	LM2
all-MiniLM-L12-v2	LM3
multi-qa-distilbert-cos-v1	LM4
all-MiniLM-L6-v2	LM5

Results

Score: Pearson Correlation (Methods)

Dataset	M1	M2	M3	M4
Mohler []	**0.826**	**0.791**	**0.816**	0.782
NN Exam []	**0.941**	**0.828**	0.561	**0.846**
AMR Exam []	**0.658**	0.458	**0.640**	**0.428**

(a)

Dataset	M1	M2	M3	M4
Mohler []	0.689	0.627	0.687	**0.792**
NN Exam []	0.889	0.791	**0.638**	0.664
AMR Exam []	0.622	**0.474**	0.593	**0.428**

(b)

Table 3: Comparison of Pearson Correlation between Random Forest (a) and AdaBoost (b) classifiers. Where M1: Passage-based, M2: Sentence-based, M3:Chunk-based, and M4: RDF-based method.

Results

Score: Pearson Correlation (Language Models)

Dataset	LM1	LM2	LM3	LM4	LM5
Mohler []	**0.802**	**0.797**	**0.796**	**0.796**	**0.789**
NN Exam []	**0.732**	**0.670**	**0.705**	**0.755**	**0.760**
AMR Exam []	0.453	**0.518**	**0.525**	**0.523**	**0.503**

(a)

Dataset	LM1	LM2	LM3	LM4	LM5
Mohler []	0.659	0.673	0.211	0.544	0.499
NN Exam []	0.614	0.653	0.704	0.698	0.605
AMR Exam []	**0.502**	0.440	0.430	0.508	0.467

(b)

Table 4: Comparison of Pearson Correlation between Random Forest (a) and AdaBoost (b) classifiers with language models (LM).

Results

Score: Root Mean Square Error (Methods)

Dataset	M1	M2	M3	M4
Mohler []	**0.893**	**0.949**	**0.920**	0.942
NN Exam []	**0.296**	**0.520**	**0.433**	**0.522**
AMR Exam []	**0.596**	0.716	**0.596**	**0.736**

(a)

Dataset	M1	M2	M3	M4
Mohler []	1.218	1.226	1.169	**0.920**
NN Exam []	0.405	0.571	0.495	0.741
AMR Exam []	0.616	**0.707**	0.630	0.741

(b)

Table 5: Comparison of RMSE score between Random Forest (a) and AdaBoost (b) classifiers with methods (M).

Results

Score: Root Mean Square Error (Language Models)

Dataset	LM1	LM2	LM3	LM4	LM5
Mohler []	**0.931**	**0.941**	**0.941**	**0.941**	**0.956**
NN Exam []	**0.484**	0.591	**0.558**	**0.490**	**0.492**
AMR Exam []	0.735	**0.680**	**0.676**	0.684	**0.698**

(a)

Dataset	LM1	LM2	LM3	LM4	LM5
Mohler []	1.182	1.163	1.667	1.278	1.363
NN Exam []	0.632	**0.582**	0.587	0.587	0.650
AMR Exam []	**0.692**	0.748	0.718	**0.682**	0.736

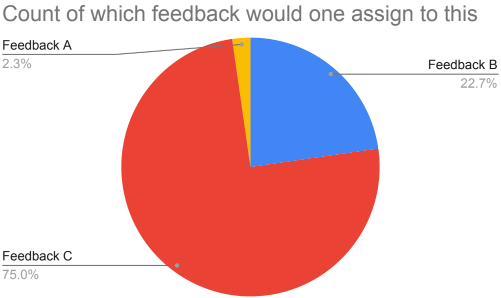
(b)

Table 6: Comparison of RMSE score between Random Forest (a) and AdaBoost (b) classifiers with language models (LM).

Results

Feedback: Survey Results

Question	What is a variable?
Reference Answer	A location in memory that can store a value.
Student Answer	a value/word that can assume any of a set of values
Feedback A	correct
Feedback B	missing keywords: location in memory
Feedback C	A variable is a location in memory that stores a value



Results

Feedback: Agreement Scores (Methods)

	M1	M2	M3	M4
Mean Agreement Score	**60.00**	22.73	31.82	35.91

(a)

	M1	M2	M3	M4
Mean Agreement Score	**60.00**	22.73	31.82	35.91

(b)

Table 7: Mean agreement scores for Random Forest (a) and AdaBoost Classifier (b) with methods.

Results

Feedback: Agreement Scores (Models)

Classifier	LM1	LM2	LM3	LM4	LM5
Random Forest	25.11	26.82	24.66	**37.05**	21.25
AdaBoost	25.11	26.82	24.66	**37.05**	21.25

Table 8: Mean agreement scores for Random Forest and AdaBoost Classifier with Language Models.

Results

Summary: Scores

Dataset	Method	Model	CL
Mohler	M1	LM1	RF
NN Exam	M1	LM5	RF
AMR Exam	M1	LM3	RF

Table 9: Pearson Correlation Performance Summary

Dataset	Method	Model	CL
Mohler	M3	LM1& LM4	RF
NN Exam	M1	LM2	RF
AMR Exam	M1& M3	LM3	RF

Table 10: RMSE Score Performance Summary

Results

Feedback

Dataset	Method	Model	Method-Model	Classifier
Mohler	M1	LM4	M1-LM4	RF

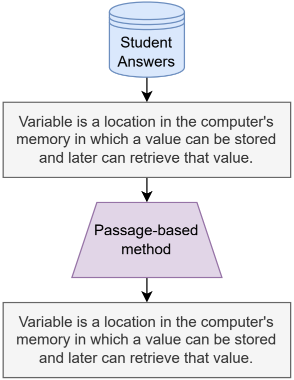
Table 11: Results of feedback evaluation

Summary

- Passage-based method and multi-qa-distilbert-cos-v1 model worked best for feedback assignment.

Approach:Extra Slides

Feature Extraction: Passage-based method



Student Answers

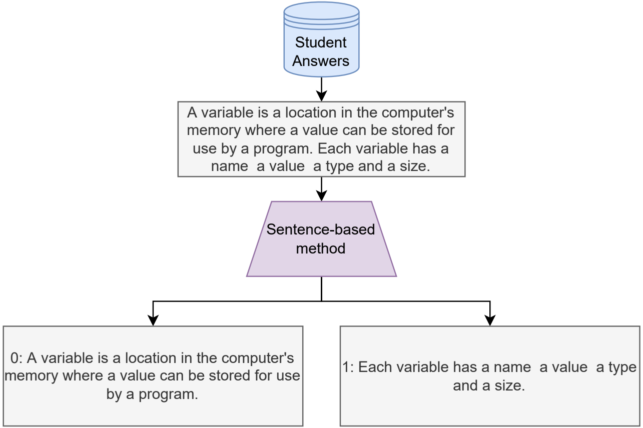
Variable is a location in the computer's memory in which a value can be stored and later can retrieve that value.

Passage-based method

Variable is a location in the computer's memory in which a value can be stored and later can retrieve that value.

Approach

Feature Extraction: Sentence-based method



Student Answers

A variable is a location in the computer's memory where a value can be stored for use by a program. Each variable has a name a value a type and a size.

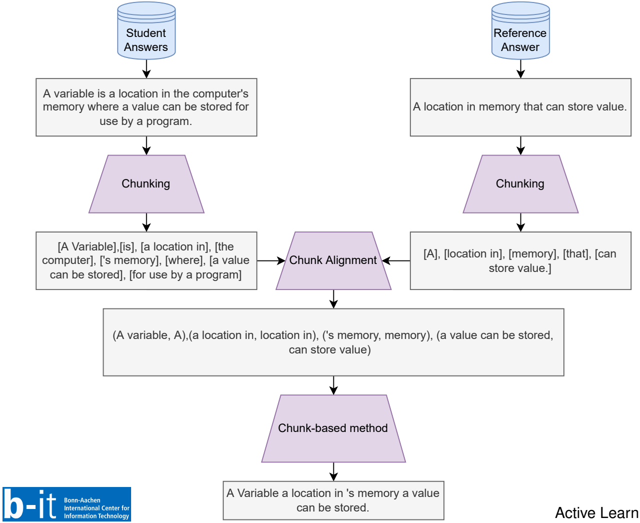
Sentence-based method

0: A variable is a location in the computer's memory where a value can be stored for use by a program.

1: Each variable has a name a value a type and a size.

Approach

Feature Extraction: Chunk-based method



Student Answers

A variable is a location in the computer's memory where a value can be stored for use by a program.

Chunking

[A Variable],[is], [a location in], [the computer], ['s memory], [where], [a value can be stored], [for use by a program]

Reference Answer

A location in memory that can store value.

Chunking

[A], [location in], [memory], [that], [can store value.]

Chunk Alignment

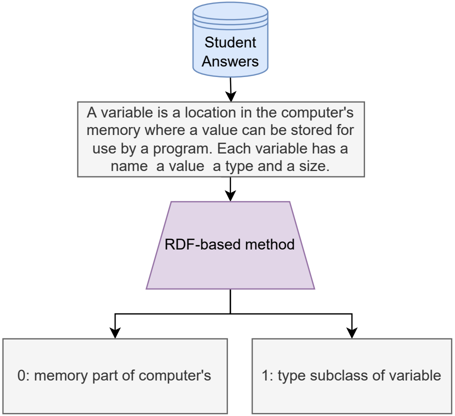
(A variable, A),(a location in, location in), ('s memory, memory), (a value can be stored, can store value)

Chunk-based method

A Variable a location in 's memory a value can be stored.

Approach

Feature Extraction: RDF-based method



Student Answers
A variable is a location in the computer's memory where a value can be stored for use by a program. Each variable has a name a value a type and a size.
RDF-based method
0: memory part of computer's
1: type subclass of variable