FernUniversität in Hagen

When Argumentation Meets Cohesion: Enhancing Automatic Feedback in Student Writing

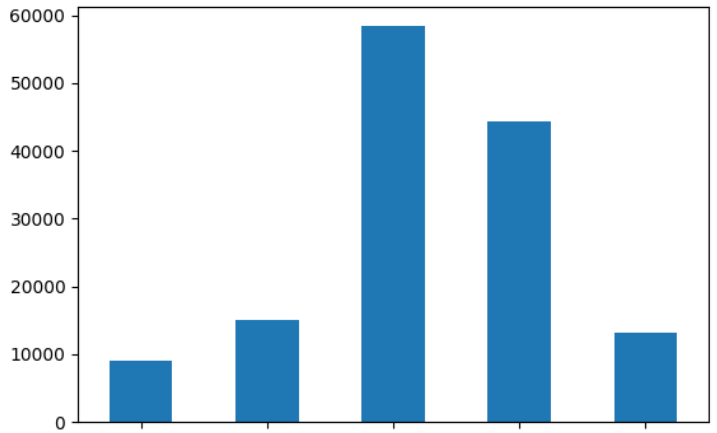



Yuning Ding[1], Omid Kashefi[2], Swapna Somasundaran[2], Andrea Horbach[1,3]
1. CATALPA, FernUniversität in Hagen, Germany
2. Educational Testing Service (ETS),Princeton, NJ, USA
3. Hildesheim University, Germany
{yuning.ding, andrea.horbach}@fernuni-hagen.de
{okashefi,ssomasundaran}@ets.org

Datasets

FB-Arguments

15,600 essays with 5 types of argumentative elements:
Lead, Position, Claim, Evidence, Conclusion

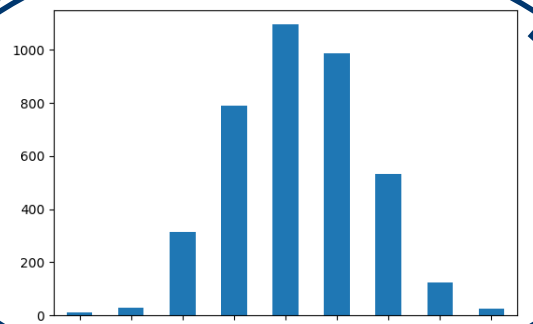


Overlap

452 Essays

FB-Score

3911 essays with cohesion scores ranging from 1.0 to 5.0 in increments of 0.5



Arguments in Student Writing

- Lead: an introduction to grab the reader's attention and point toward the thesis.
- Position: an opinion or conclusion on the main question.
- Claim: a claim that supports the position, refutes another claim or gives an opposing reason to the position.
- Evidence: ideas or examples that support claims, counterclaims, or rebuttals.
- Conclusion: a concluding statement that restates the claims.

Driverless cars have been a big topic lately.
I think that driverless cars shouldn't be allowed on public roads because they are not safe.
A driverless car could cause a major or even fatal crash. While most driverless cars require you to have hands on the wheel this does not mean you will be paying attention if something is about to happen.
I think that driverless cars are not safe and they should not be allowed on public roads.

Lead | Position | Claim | Evidence | Conclusion

Cohesion in Student Writing

- Definition of a High Cohesion Score:

  "The organization generally well controlled, a range of cohesive devices used appropriately such as reference and **transitional words** and phrases to connect ideas and generally **appropriate overlap of ideas** were found in this essay"

- Cohesion Features
  - Transitional Words: certain linguistic cues that enable the reader to establish connections between the ideas within the text (Crossley et al., 2016a).
  - Lexical Chains: chains of related words that contribute to the continuity of lexical meaning (Morris and Hirst, 1991)

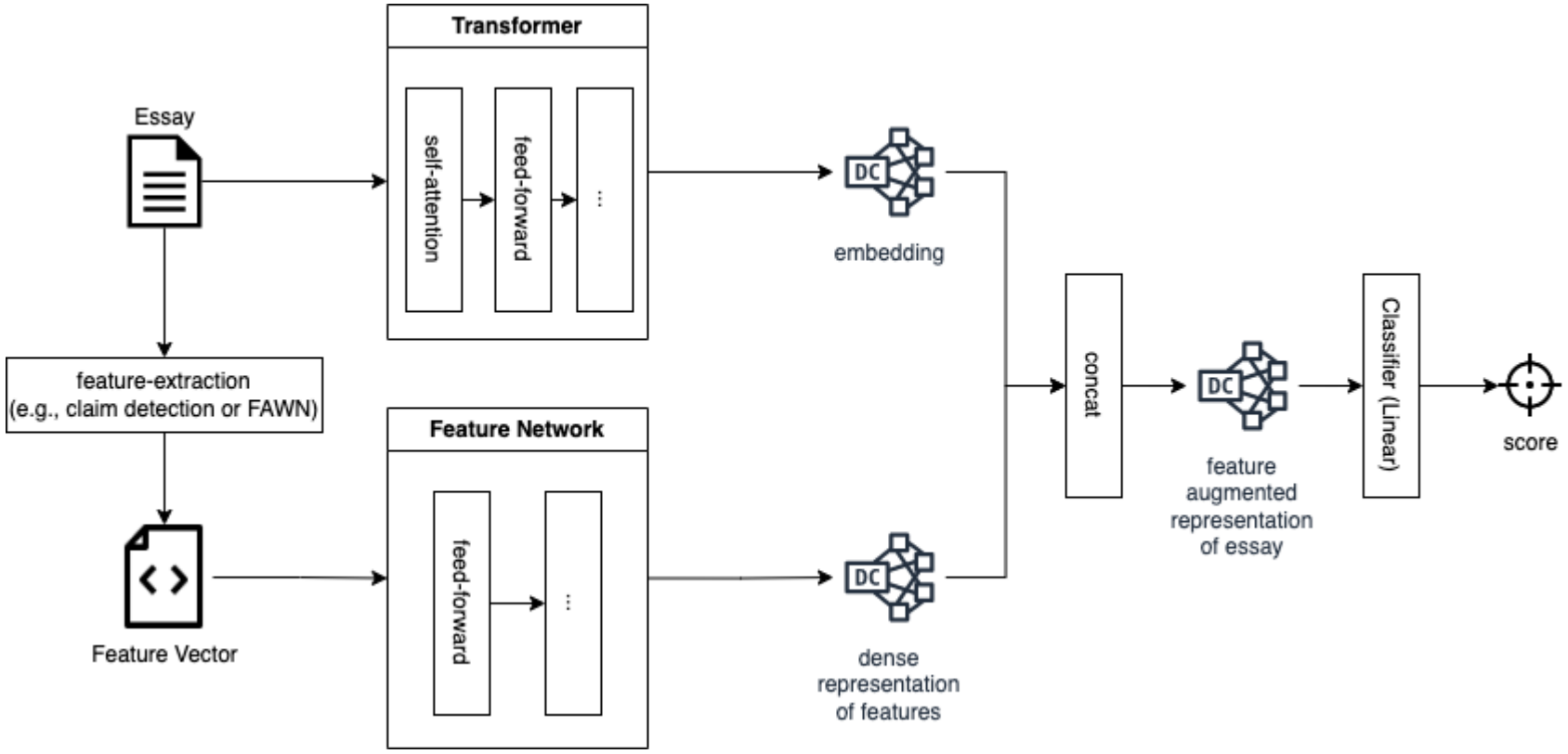
Feature Sets

- **ARG** (argument features) : the absolute and relative frequencies of the five types of argumentative elements (lead, position, claim, evidence, and conclusion) found in the essays.
- **TW** (Transitional Words) : absolute and relative frequencies of the connective words of different transitions, such as "so that" for the transition *belief*, "according to" for *consequence*, "because" for *evidence* … (Madnani et al., 2018)
- **LEX-1** : the counts, corresponding percentages, and normalized versions of different types of lexical chains, such as the total number of chains in the essay, the average chain size, and the number of large chains (lexical chains with more than four nodes).
- **LEX-2** : the interactions between discourse transitions using a discourse cue tagger. Each discourse cue was replaced with its tag, and the number of chains that (i) start after it, (ii) end before it, and (iii) continue over it (chains having nodes before and after the tag) was counted as features. (Somasundaran et al., 2014)
- **TW/LEX-1/LEX-2 in Claims** and **TW/LEX-1/LEX-2 in Evidence**

Research Questions

- To what extent does adding argument features enhance the performance of predicting cohesion scores in the context of automatic essay scoring?
- How does the inclusion of cohesion features, specifically transitional words and lexical chains, among argumentative elements impact the performance of automatic cohesion scoring when compared to the inclusion of these features alone?
- Does employing the automatic segmentation of argumentative elements as an auxiliary task result in better performance for predicting cohesion scores compared to the single-task approach?

Automatic Scoring with Augmented Features



Picture from Writing AI Project @ ETS
https://confluence.ets.org/display/WAI/Feature+Augmented+Scoring

Feature Augmentation

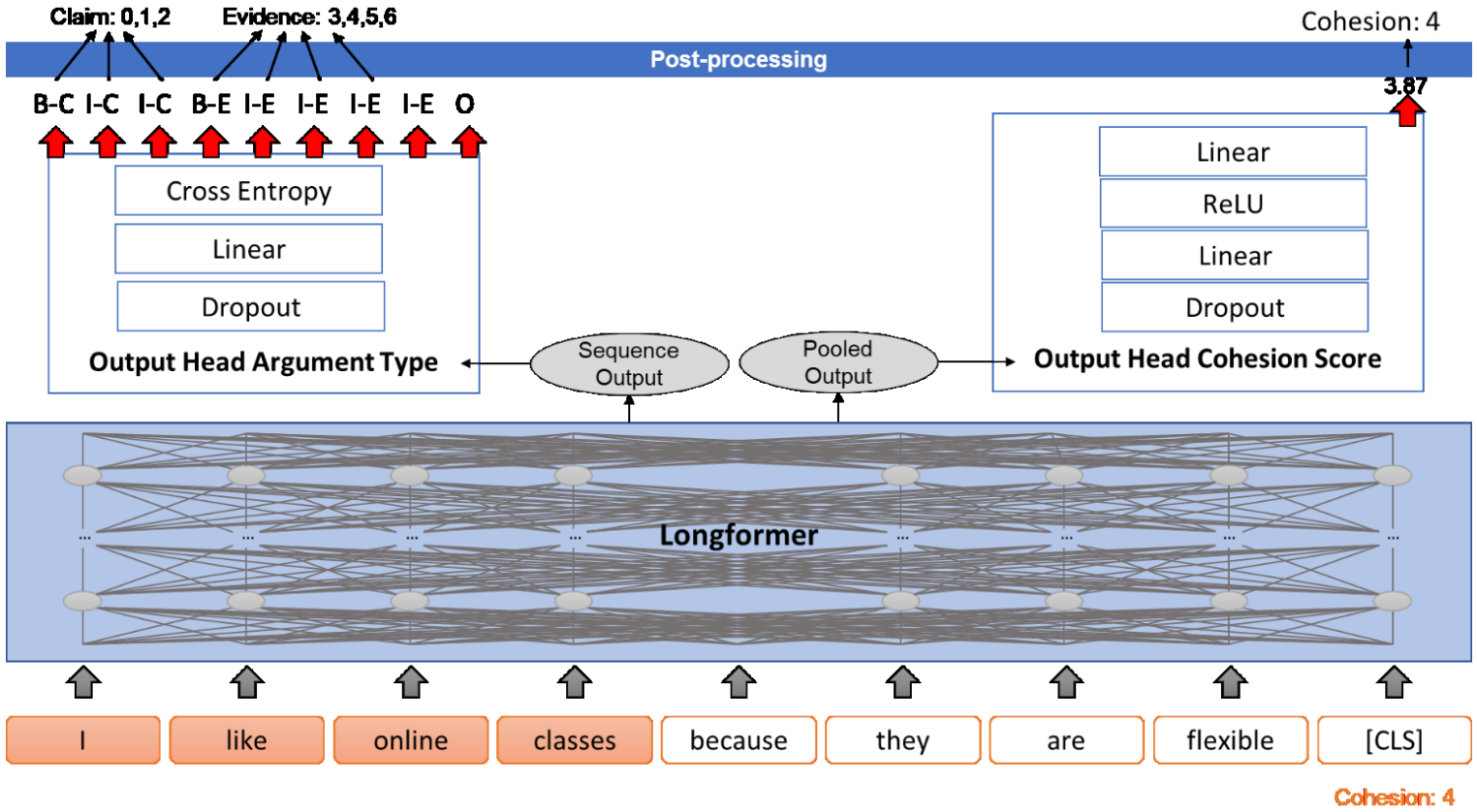
- Only argumentative features

- Only cohesion-related features

- Cohesion features × arguments

Setting	F1
Baseline	.21
+Arguments	-.03

Multi-Task Learning



Multi-Task Learning

- The cohesion score is learned more accurately in a multi-task learning process by adding the automatic segmentation of argumentative elements as an auxiliary task.

Setting	Span F1	Cohesion F1	Cohesion QWK
Arguments Span Only (with goldstandard)	.66	-	-
Arguments Span Only (with predicted span)	.63	-	-
Cohesion Only	-	.21	.54
Longformer+ARG	-	.18	.57
MTL_{linear} (with predicted span)	.63	**.30**	**.60**
$MTL_{dynamic}$ (with predicted span)	.62	.23	.59

Conclusion

- Augmenting the feature set with argumentative elements and lexical chains can improve the performance of transformer-based automatic cohesion scoring.
- Jointly learning the segmentation of argumentative elements as an auxiliary task can improve the performance of cohesion score prediction.
- In practical terms, our model's ability to assess cohesion and comprehend argumentative structures showcases its potential for providing feedback to learners.