Henning Wachsmuth, Khalid Al-Khatib, Benno Stein

Using Argument Mining to Assess the Argumentation Quality of Essays

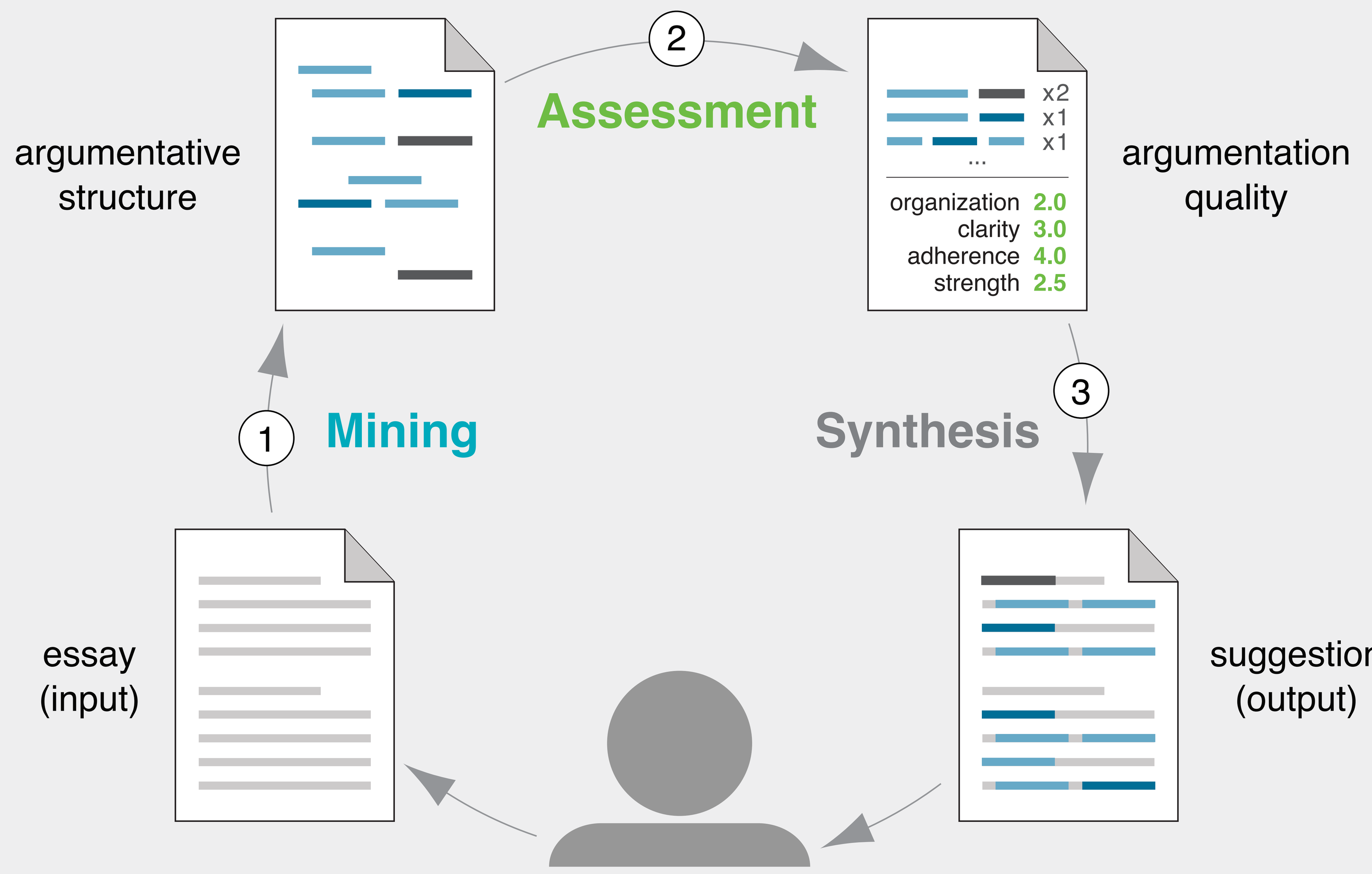
The first study of argument mining for argumentation quality assessment

Argument mining determines the argumentative structure of texts. The benefit of this structure has rarely been evaluated.

Argumentation quality assessment is needed for envisaged applications such as argumentative writing support.

Argumentative writing support for persuasive essays:

1. Mining of an essay's argumentative structure.
2. Assessment of argumentation quality dimensions.
3. Synthesis of suggestions for improvements (future work).



We score persuasive essays based on the output of mining for four argumentation-related quality dimensions:

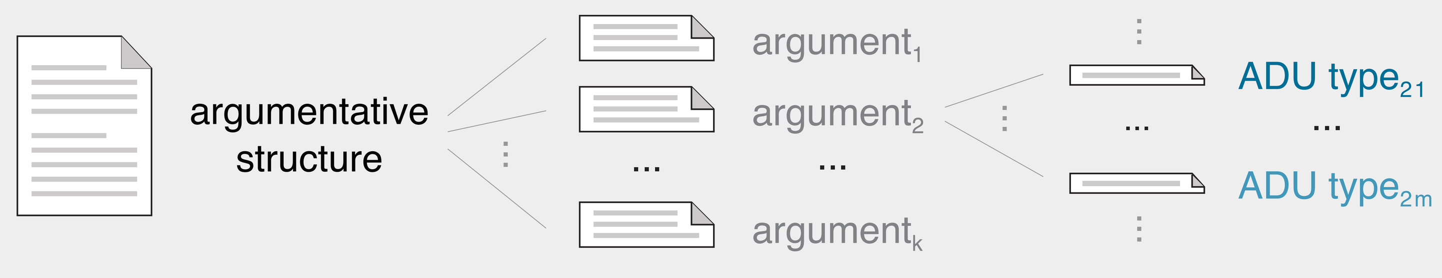
- Organization (Persing et al., EMNLP 2010)
- Thesis clarity (Persing and Ng, ACL 2013)
- Prompt adherence (Persing and Ng, ACL 2014)
- Argument strength (Persing and Ng, ACL 2015)

Main contributions of our work:

- The first study of the benefit of argument mining for argumentation quality assessment.
- Statistical insights into essay argumentation.
- The new state of the art for two quality dimensions.

Statistical insights into argumentation based on the output of mining

Modeling of an essay as a flow of paragraph-level arguments with sentence-level argumentative discourse units (ADUs).



Learning of mining four ADU types using standard features on the Argument Annotated Essays corpus (Stab and Gurevych, COLING 2014)

Argument mining approach	Accuracy	F_1-score
Majority baseline	0.525	0.361
State-of-the-art baseline (Stab and Gurevych, EMNLP 2014)	0.773	0.726
Our approach	0.745	0.745

Application of mining on all 6085 student essays from the International Corpus of Learner English (Granger et al., 2009).

prompt *Some people say that in our modern world, dominated by science and technology and industrialisation, there is no longer a place for dreaming and imagination. What is your opinion?*

essay *If we take a look back in time we are in a position to see man dreaming, philosophizing and using his imagination of whatever comes his way. We see man transcending his ego I a way and thus becoming a God - like figure. And by putting down these sacred words, what is taking shape in my mind is the fact that using his imagination Man is no longer this organic and material substance like his contemporary counterpart who is putting his trump card on science, technology and industrialization but Man is a way transcends himself through his imagination.*

none

conclusion

premise

For instance, if we take into account the Renaissance or Romantic periods of mankind and close our eyes we could see Shakespeare applying his imagination in the fancy world of his comedies: elf and nymphs circling the stage making it a dream that will lost forever in our minds. We could even hear their high-pitched weird chuckle piercing with a gentle touch our ears, but "open those eyes that must eclipse the day" and you'll wee the high-tech wiping out every trace of the human elevated spirit that have dominated over the previous centuries. What we see now is "deux aux machina" or the fake "God from the machine" who with the touch of a button could unleash Armageddon.

For poets and literate people of yore it was a common idea to transcend reality or to go beyond it by using their imagination not by using reason as we the homosapiens of our time do. For example, if we indulge in entertaining the idea of the film "The matrix" it has a lot to do with the period of Romanticism. But the difference is that a poet from that time could transcend reality, become one with Nature, and cruise wherever he wants using his imagination. Whereas now in the 21st century and in "The matrix" in particular the scientific type of Man thinks that at last he has succeeded in making travelling without boundaries via the virtual reality of his PC.

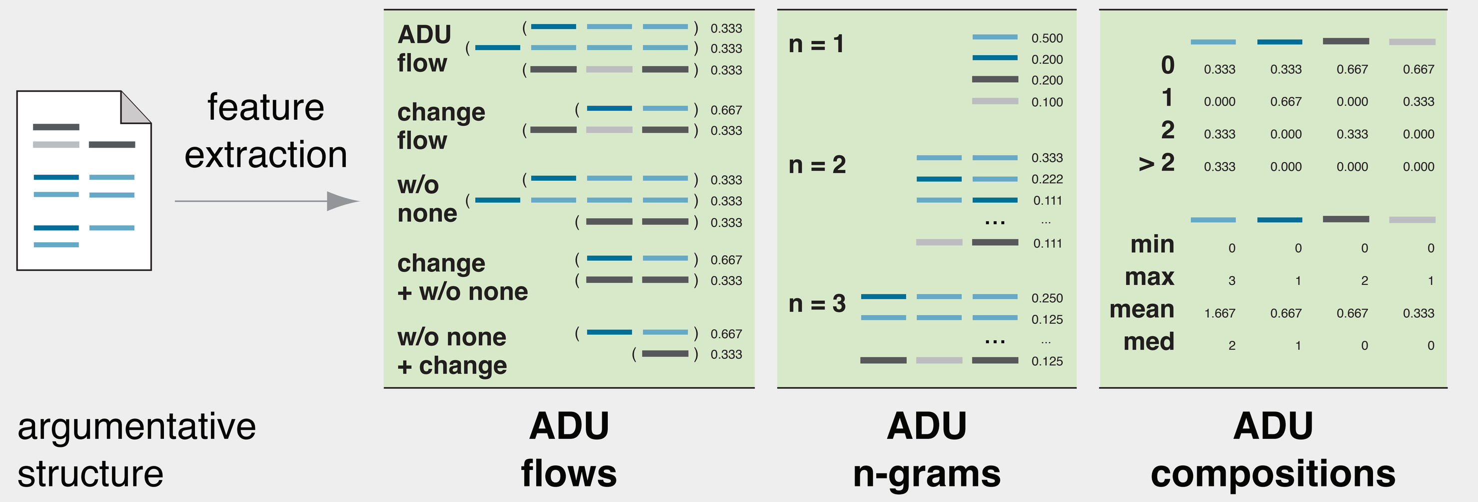
As a logical conclusion to my essay I would like to put only one thing. "Wouldn't it be better if imagination makes the world go round". If I was to answer this question, the answer would be positive, but given the aquisitive or consumer society conditions we live in let's make a match between imagination and science. It would be somewhat more realistic.

Analysis of common ADU change flows in all ICLE paragraphs.

		Paragraph of essay		
#	ADU change flow	average	first	last
1	(conclusion, premise)	25.1%	–	13.1%
2	(conclusion)	22.4%	0.9%	31.6%
3	(conclusion, premise, conclusion)	17.0%	–	27.2%
4	(none)	5.8%	42.7%	0.4%
5	(premise)	4.3%	–	1.4%
6	(none, thesis)	3.4%	25.9%	–
7	(premise, conclusion)	2.9%	–	2.7%

State-of-the-art assessment of essay organization and argument strength

Novel feature types for argumentation-related essay scoring based on the output of mining.



Evaluation on all 830–1003 ICLE essays that are labeled for each quality dimension with a score from [1, 4].

Demo

Experimental set-up exactly as in the papers of the (former) state-of-the-art approaches.

Essay scoring with several supervised approaches:

- Average score baseline
- State-of-the-art baseline (Persing et al. EMNLP 2010, Persing and Ng ACL 2013–2015)
- Content: Token n-grams, prompt similarities
- POS: Part-of-speech n-grams
- Flows: Sentiment flow patterns (Wachsmuth et al., COLING 2014, EMNLP 2015)
- Our approach: ADU flows, n-grams, and compositions

Mean squared errors in 5-fold cross-validation:

Essay scoring approach	Organization	Thesis clarity	Prompt adherence	Argument strength
Average score baseline	0.349	0.469	0.291	0.266
State-of-the-art baseline	0.175	0.369	0.197	0.244
Content	0.336	0.425	0.231	0.236
POS	0.326	0.461	0.231	0.233
Flows	0.228	0.481	0.257	0.259
Our approach	0.184	0.470	0.241	0.242
ADU flows	0.234	0.461	0.247	0.242
ADU n-grams	0.225	0.466	0.265	0.243
ADU compositions	0.194	0.457	0.239	0.239
Our approach + POS / Flows	0.164	0.496	0.232	0.246
ADU compositions + Content	0.178	0.435	0.216	0.226

(mean squared errors in green significantly improve the state of the art with a confidence of over 90%)