

Interpreting Themes from Educational Stories

Yigeng Zhang¹, Fabio A. González², Thamar Solorio^{1,3}

¹University of Houston

²Universidad Nacional de Colombia

³MBZUAI

Theme interpretation

- Interpreting themes from literary compositions (stories) for educational purposes

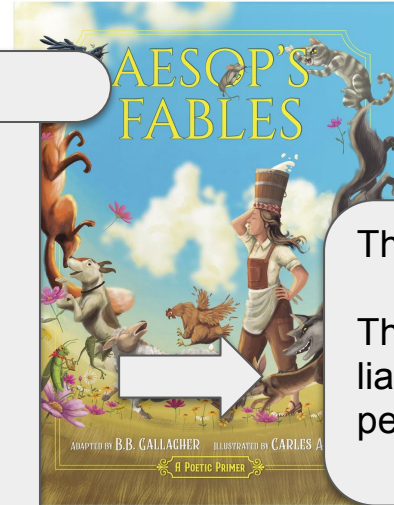
The Shepherd Boy & the Wolf

A Shepherd-boy, who watched a flock of sheep near a village, brought out the villagers three or four times by crying out, "Wolf! Wolf!" and when his neighbours came to help him, laughed at them for their pains.

The Wolf, however, did truly come at last.

The Shepherd-boy, now really alarmed, shouted in an agony of terror: "Pray, do come and help me; the Wolf is killing the sheep;" but no one paid any heed to his cries, nor rendered any assistance.

The Wolf, having no cause of fear, at his leisure lacerated or destroyed the whole flock.



Theme (moral/lesson):

There is no believing a liar, even when this person speaks the truth.

Theme interpretation

- **Interpreting the theme** is fundamental to understanding literal compositions written by natural language.
- Formulating and introducing a new benchmark to an underexplored topic
 - Higher-level of reading comprehension than literal comprehension
- Improving the quality of service on educational media content search/recommendation to children, parents, and educators.

EduStory: the dataset

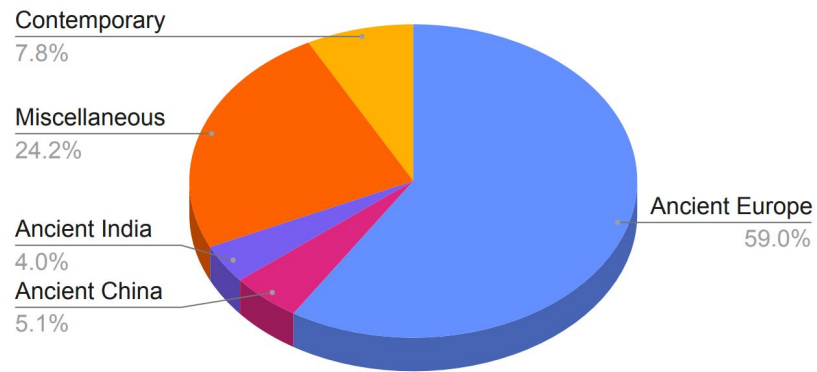
Educational stories with moral message in English

- Fable
 - Aesop's Fables (ancient Greece)
 - The Panchatantra (India)
- Folktale
 - Chinese Fables and Folk Stories (1908)
- Idiom story
- Success story
- Misc

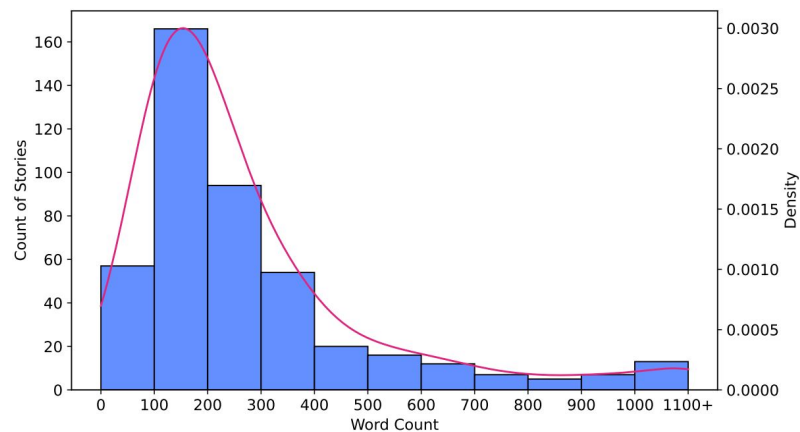


EduStory: the dataset

- Story cultural origins

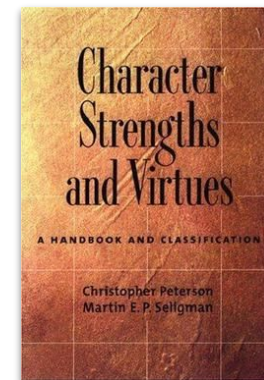


- Story length (word count)



EduStory: the dataset

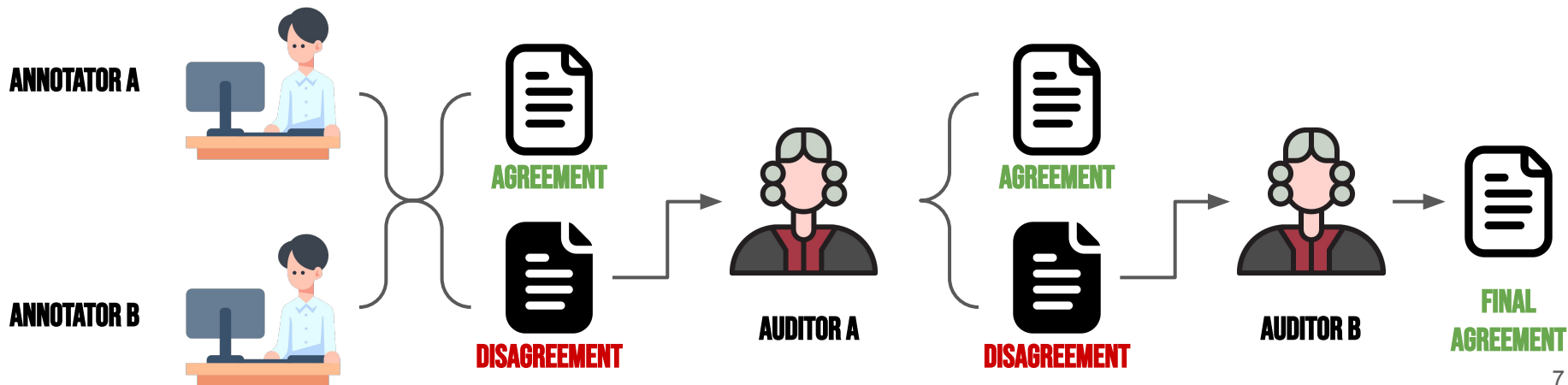
- Positive psychology: *Character Strengths and Virtues: A Handbook and Classification*



Wisdom and Knowledge 228					Humanity 59		
Creativity 3	Curiosity 0	Judgment 26	Learning 0	Perspective 199	Love 16	Kindness 31	Social Intelligence 12
Transcendence 25					Justice 22		
Appreciation 1	Gratitude 20	Hope 2	Humor 0	Spirituality 2	Citizenship 17	Fairness 4	Leadership 1
Courage 57				Temperance 60			
Bravery 5	Persistence 23	Integrity 29	Vitality 0	Forgiveness 2	Humility 25	Prudence 17	Self-Regulation 16

Data collection, annotation and audit

- Final dataset
 - 580 story-theme pairs
 - 451 unique stories
- 2 annotators, 2 auditors

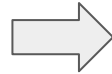
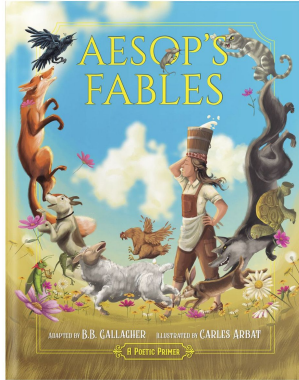


Theme interpretation task formulation

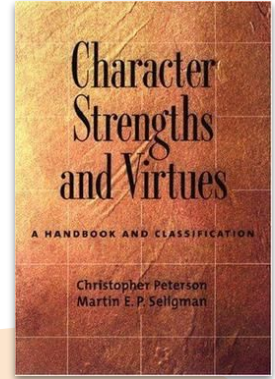
- Theme keyword identification
- Theme/story retrieval
- Theme interpretation as multiple-choice Q&A
- Theme interpretation as text generation

Task 1: Theme keyword identification

- Categorize theme from the perspective of **Positive Psychology**
 - 24 specific strengths under 6 broad virtues
- Identify the character strengths that the story conveys
 - Supervised classification



The shepherd boy and the wolf

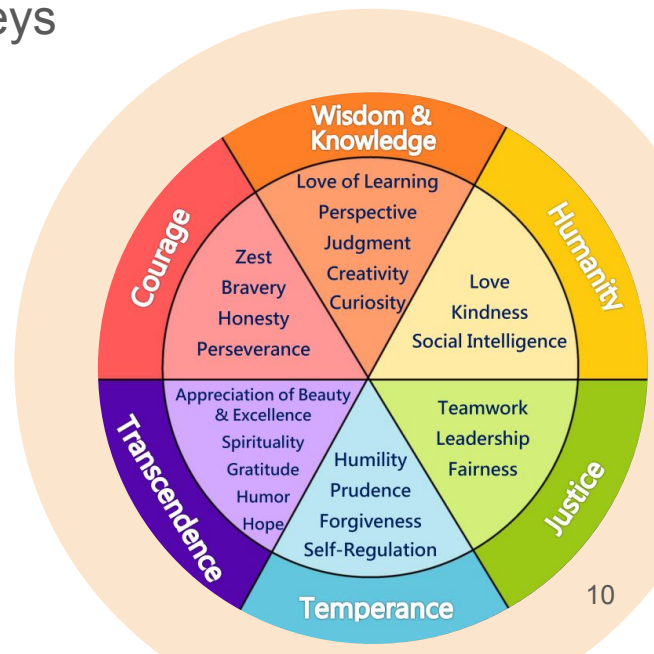


Task 1: Theme keyword identification

- Categorize theme from the perspective of *Positive Psychology*
 - 24 specific strengths under 6 broad virtues
- Identify the character strengths that the story conveys
 - Supervised classification

	Virtue		Strength	
	Dev	Test	Dev	Test
TF-IDF	15.4	11.2	4.8	6.5
BOWV	14.7	18.0	4.8	4.5
TextCNN	15.1	13.4	5.1	9.5
BERT	22.9	21.5	17.1	11.6
Flan-T5	19.9	14.5	13.8	5.9

Metric: macro F1 score



Task 2: Theme/story retrieval

- Text retrieval (document selection/ranking)
- Use a story (query) to retrieve themes (doc)
- Or vice versa

The shepherd boy and the wolf

There is no believing a liar, even when this person speaks the truth.



Never try to deceive a friend



Honesty is important



Greed will always lead to downfall.



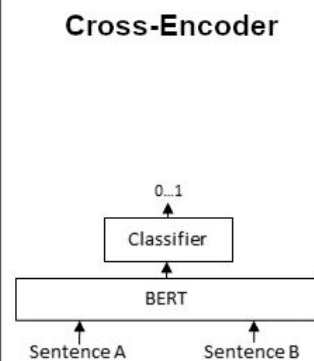
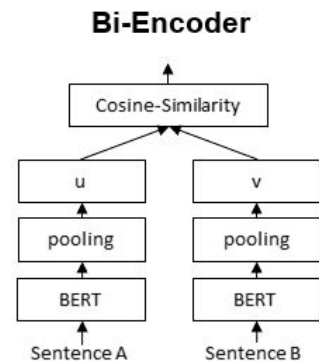
.....



Relevance

Task 2: Theme/story retrieval

- Text retrieval (document selection/ranking)
- Use a story (query) to retrieve themes (doc)
- Or vice versa



	Story-theme		Theme-story	
	Dev	Test	Dev	Test
BM25	0.35	0.28	0.22	0.14
DPR	0.49	0.31	0.52	0.33
BERT	0.46	0.24	0.46	0.28
MPnet	0.65	0.40	0.60	0.43
MiniLM _{CE}	0.42	0.27	-	-

Metric: Mean Reciprocal Rank (MRR)

Task 3: Theme interpretation with multiple-choice Q&A

- Typical machine reading comprehension

The shepherd boy and the wolf

What is the main idea of this story?

Never judge anyone by the way they look **A**

There is no believing a liar, even when this person speaks the truth. **B** ✓

Greed will always lead to downfall. **C**

Task 3: Theme interpretation with multiple-choice Q&A

- Method:
 - BERT/MPnet: Use special token concat context and choices, modeling the $P(a_1, a_2, \dots, a_n | s)$
 - Flan T5: Prompt design and choice letter generation
- Distractor creating strategy
 - Random, sampling from different virtue, sampling from same virtue

	Diff Virtue		Random		Same Virtue	
	Dev	Test	Dev	Test	Dev	Test
BERT	0.67	0.58	0.74	0.68	0.72	0.64
MPnet	0.37	0.26	0.52	0.29	0.37	0.36
Flan-T5	0.56	0.46	0.50	0.41	0.56	0.46

Task 4: Theme interpretation as text generation

- Text-to-text generation for question answering

The shepherd boy and the wolf



What is the main idea of this story?



There is no believing a liar, even when this person speaks the truth.



Task 4: Theme interpretation as text generation

- Method
 - Flan T5 finetune on the training dataset
 - OPT-175B zero-shot (comparable to GPT-3)
 - ChatGPT (Feb 2023) zero-shot
- Human evaluation
 - Quality of individually generated themes
 - Best-choice generated theme for the story

Source	Model generated theme			Model generated and original theme			
	Flan-T5	OPT-175B	ChatGPT	Flan-T5	OPT-175B	ChatGPT	Original
Human eval score	11	23	50	13	20	46	36
Best interpretation	7%	17%	77%	0%	10%	50%	40%

Task 4: Theme interpretation as text generation

- Method
 - Flan T5 finetune on the training dataset
 - OPT-175B zero-shot (comparable to GPT-3)
 - ChatGPT (Feb 2023) zero-shot
- Human evaluation
 - Original theme with model-generated themes under one-sentence restriction

Source	Model generated and original theme			
	Flan-T5	OPT-175B	ChatGPT	Original
Human eval score	12	18	34	44
Best interpretation	3%	13%	30%	53%

Findings

- In all task formulations, this research problem has proven challenging, even for pre-trained LMs, highlighting the need for further attention from the research community.
- In the generation task, although a strong proprietary LLM, ChatGPT, demonstrates impressive quality, it still falls short of the original human-written theme sentences.

Contribution

- Pioneer an underexplored research area: interpretive comprehension
 - Propose NLP solutions for each task
 - Benchmarking NLP models with extensive experiments
- EduStory: the dataset
 - Story and theme pairs from various sources
 - Theme keyword annotations
- Empower next-level natural language understanding with open-ended formulations

Data & Code

- <https://github.com/RiTUAL-UH/EduStory>
- Contact: Yigeng Zhang, yzhang168@uh.edu