Fact-driven Storytelling with Large Language Models

Xinyan Guo; Jiaqi Mo; Ying Hua
Technical University of Munich, School of Computation Information and Technology, Research Group Social Computing

Motivation

1. How to generate argumentative messages in good quality?
- Use pyramid-structure for storytelling!

2. How to generate a pyramid in good quality?
- LLM helps to generate arguments
- Finetuning BERT-based models for quality control!

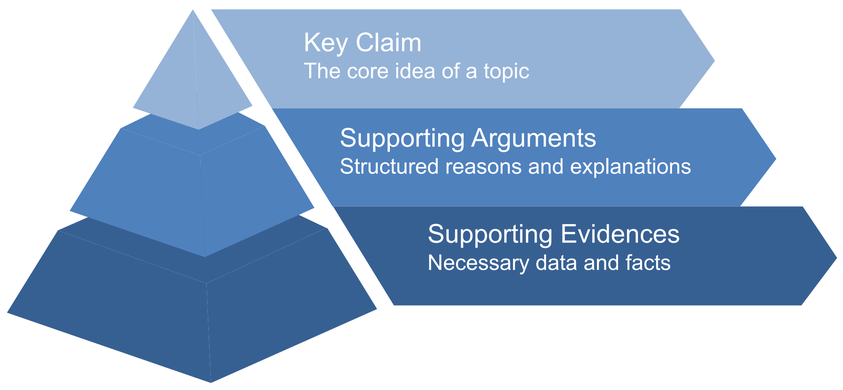

Figure 1. The Pyramid Principle for argumentative messages [2]

Evaluation

- Manually created pyramid-structured evaluation dataset including labels (e.g., “perfect” / “badclaim” / “badevidence”) [5]
- Total 152 pyramids
- 87% accuracy

	Relevance	Support	Coherence	Overall
Accuracy	0.78	0.87	0.80	0.87
Precision	0.88	0.96	0.71	0.92
Recall	0.83	0.84	0.89	0.75
F1-Score	0.85	0.89	0.79	0.83

Table 1. Evaluation metric scores for different models

Methods

- **Relevance:** how well the arguments are directly related to and support the main claim
- **Supportiveness:** how effectively the evidences back up each argument
- **Coherence:** the logical flow and overall cohesiveness of the entire argument structure
- **Completeness:** the thoroughness and comprehensiveness of the arguments

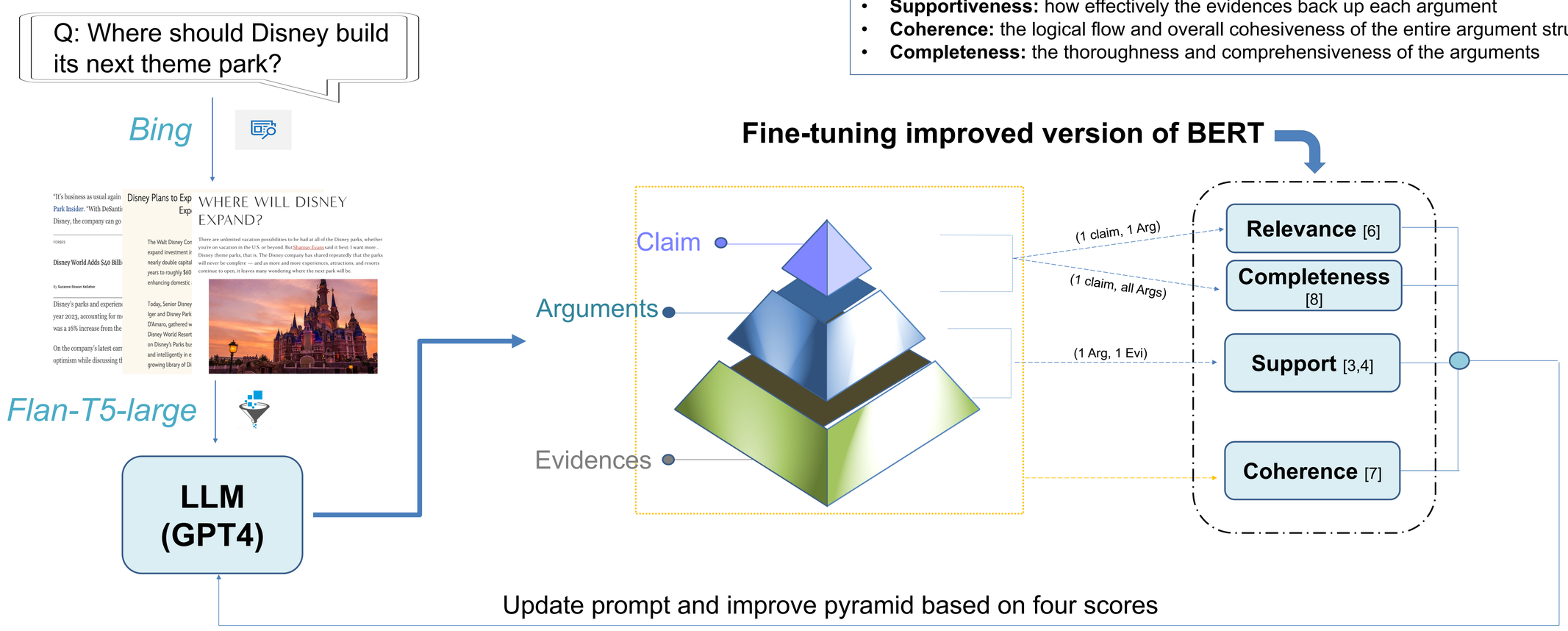

Figure 2. Pipeline of argument generation based on LLM

Results

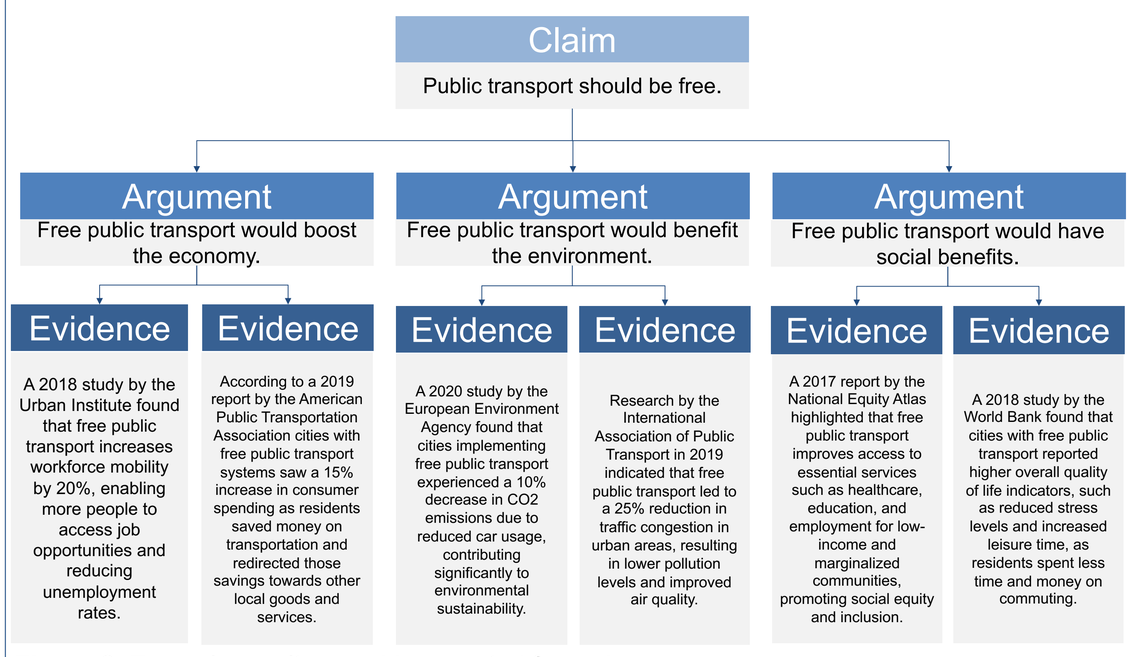

Figure 3. Example result pyramid generated from above process

Conclusions

- We proposed a machine learning approach to optimize the quality of a pyramid-structured argumentative message based on LLM.
- We developed a quality measure system by finetuning BERT-based models and calculating 4 different scores for each input pyramid.
- We tested the performance of our models using a manually created pyramid-style dataset and achieved an overall accuracy of 87%.

Check our code and data on GitHub!

Contact

Xinyan Guo, Jiaqi Mo, Ying Hua
Technical University of Munich
Email: xinyan.guo@tum.de, ge74toz@mytum.de, ying.hua@tum.de

References

1. Chen, G., Cheng, L., Tuan, L. A., & Bing, L. (2023). Exploring the potential of large language models in computational argumentation. *arXiv preprint arXiv:2311.09022*.
2. Minto, B. (2009). *The pyramid principle: logic in writing and thinking*. Pearson Education.
3. Gleize, M., Shnarch, E., Choshen, L., Dankin, L., Moshkowich, G., Aharonov, R., & Slonim, N. (2019). Are you convinced? choosing the more convincing evidence with a Siamese network. *arXiv preprint arXiv:1907.08971*.
4. Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020, November). A simple framework for contrastive learning of visual representations. In *International conference on machine learning* (pp. 1597-1607). PMLR.
5. https://www.kialo.com/
6. Gretz, S., Friedman, R., Cohen-Karlik, E., Toledo, A., Lahav, D., Aharonov, R., & Slonim, N. (2020, April). A large-scale dataset for argument quality ranking: Construction and analysis. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 34, No. 05, pp. 7805-7813).
7. Zhu, D., Wu, W., Song, Y., Zhu, F., Cao, Z., & Li, S. (2024). CoUDA: Coherence Evaluation via Unified Data Augmentation. *arXiv preprint arXiv:2404.00681*.
8. Stab, C., & Gurevych, I. (2017, April). Recognizing insufficiently supported arguments in argumentative essays. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers* (pp. 980-990).