Generating Effective Headlines for Social Media Posts

Raymond Hung, Jack Forlines, Dipika Kumar

DATASCI 266: Natural Language Processing with Deep Learning

UC Berkeley School of Information

Berkeley
UNIVERSITY OF CALIFORNIA

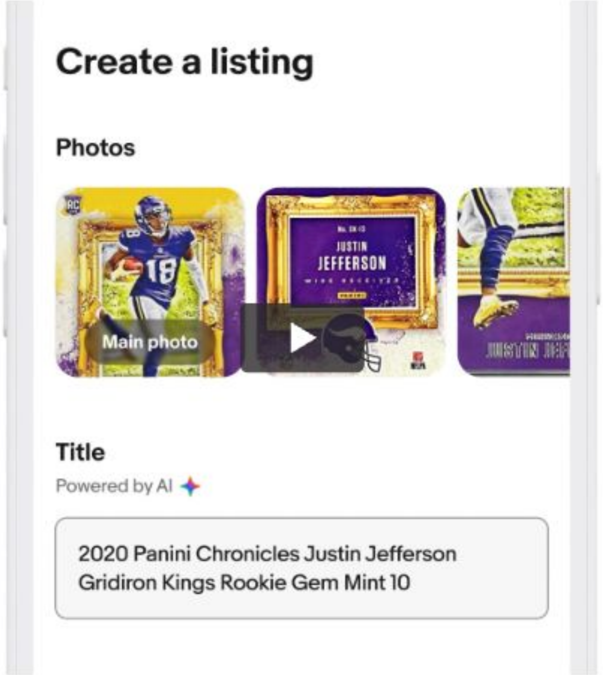
Introduction

What problem are we trying to solve?

- Advanced large language models (LLMs) are extensively utilized for summarizing news articles and documents
- Summarizing social media content poses unique challenges due to the variability in text length and need for captivating headlines

Practical Applications

- **Title Generations:** Automated suggestion of titles derived from the body text
- **Existing Implementations:**
  - Gmail incorporates suggested titles for email composition
  - eBay utilizes automated title generation for its listings



Source: eBay's Magical Listing Tool

Berkeley
UNIVERSITY OF CALIFORNIA

Data

- **Reddit datasets:** Sourced from Kaggle containing 500K posts from data science and machine learning related subreddits

- **Data cleansed**
  - NAN rows removed
  - Text normalization

- Added prefix, paired, shuffled, and produced 25K subset dataset
  - 20K training (80%)
  - 3.75K validation (15%)
  - 1.25K test (5%)

Model Experiments

- Models (same hyperparameters across models):
  - Pegasus (Baseline)
  - T5
  - BART
  - GPT/OPT350M
- Transfer learning on cleansed datasets
  - Fine-tuning with GPT/OPT350M models highlight complexity of the task
- T5 model hyperparameters fine-tuned further

Automatic Evaluation

Cand_Title-Ref-Title (N=1250)	rouge1	rouge2	rougeL	rougeLsum	bleu
PEGASUS-Xsum	0.157997	0.043143	0.14421	0.143903	0.019779
T5	**0.184472**	**0.051011**	**0.165052**	**0.165165**	**0.029681**
GPT-2	0.069966	0.023475	0.053999	0.053998	0.007463

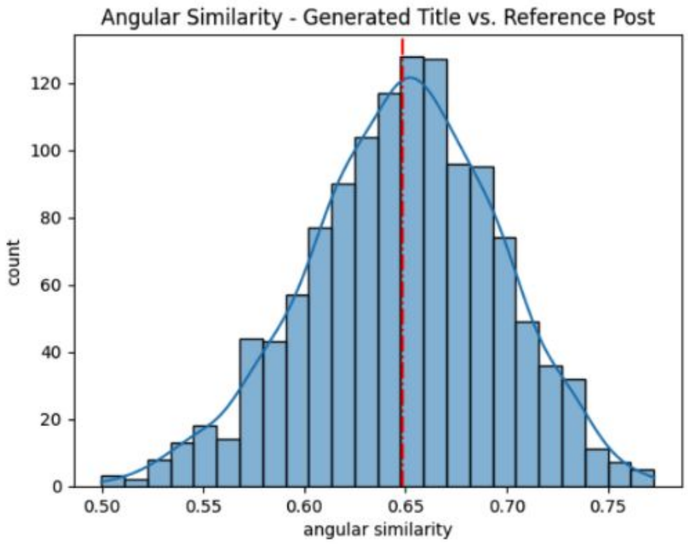
Cand_Title-Ref-Title (N=100)	rouge1	rouge2	rougeL	rougeLsum	bleu
T5 (max_length = 64)	0.239797	0.08335	0.219164	0.220178	0.059214

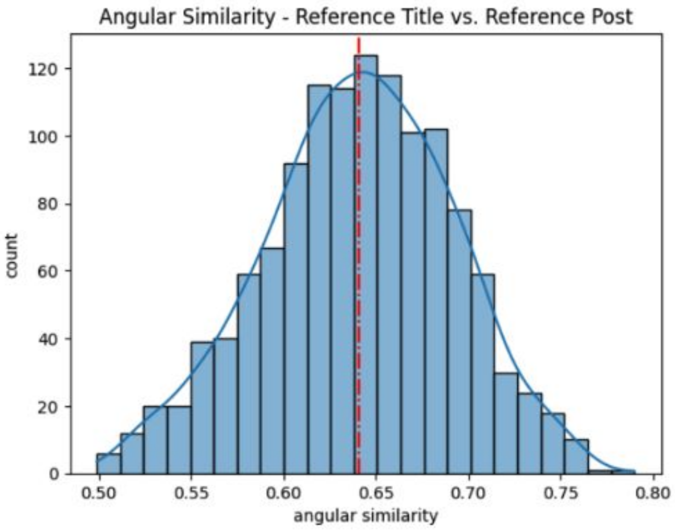
	AES		FD	
	Candidate Title vs Ref Post	Ref Title vs Ref Post	Candidate Title vs Ref Post	Ref Title vs Ref Post
PEGASUS-Xsum	0.63724	0.64045	1.07554	1.06696
T5	**0.64847**	0.64045	**1.04622**	1.06696
GPT-2	-	0.64045	0.01173	1.06696

Berkeley
UNIVERSITY OF CALIFORNIA

Automatic Evaluation

T5 Average Angular Similarity





Human Evaluation

	Reference Title	Candidate Title V1	Candidate Title V2 (Fine Tuned)	Candidate - Faithfulness	Candidate - Fluency	Candidate - Coherence
1	**is a pricey masters degree worth it**	what is the best way to get a job in data science	what is the best way to get a masters in ds	😩	😄	😩
2	how should I start	**how do i become a data analyst**	Data analyst job or not	😄	😄	😄
3	**ds masters subsequent phd studies**	MS in data science vs python	Should I do a phd in data science	😩	😄	😩
4	crime scene dataset photo and video database	**looking for datasets of crime scene images**	**looking for crime scene datasets**	😄	😄	😄

* **bolded** text indicates the best title based on the content in the original post

Conclusion

- T5 model produce the best ROUGE, BLEU, AES, and FD scores
    - Larger model
    - More adaptable to diverse tasks
- Further human evaluation needed to complement evaluations for meaningful conclusions
- Challenges from computing resources constraints