به نام خدا

امیرمحمد کمیجانی ۴۰۴۱۲۸۳۴

ارائه مقالات

عنوان مقاله

- A Survey on LLM-Generated Text Detection: Necessity, Methods, and Future Directions

چرا میخواهیم Text detection انجام دهیم

- قوانین
- میل کاربران
- آموزش LLM
- Science

مواردی که بررسی شدند

- Datasets and Benchmarks
- Detectors
- Evaluations
- Challenges and limitation

Datasets

- دیتاست ها در دامنه‌های مختلف اما محدودی مورد بررسی قرار گرفتند که بعدها موضوعات متنوع دیگری در حوزه NLP معرفی شدند.
- اکثرا از زبان انگلیسی استفاده شده ولی زبان‌های دیگر هم پیشنهاد شدند.
- مدل‌های زبانی بزرگ گوناگونی نیز پیشنهاد شد که البته تعداد آنها محدود بود.

Benchmarks

- بنچمارک ها باید از مدل‌های زبانی گوناگونی پشتیبانی و استفاده نمایند تا بتوانیم یک مدل جدید را با همتایان بسیاری بسنجیم.
- همچنین باید تنوع و گوناگونی زیادی را در موضوعات مختلف داشته باشیم تا مدل براساس موضوعات مختلف سنجیده شود.

Detectors

- Watermarking
- Statistics : همانند perplexity, token probability, Entropy
- Neural Based Models
- Human Assisted

Issues with Detectors

- Out of Distribution
- Real world Data => Noisy texts
- LLMs strength

Evaluation Metrics

- Precision, Recall, F1-score
- FPR,TPR, AUC-ROC score
- No Evaluation Framework, we must check all metrics and analyze them.

Conclusion and Future Directions

- Robust detectors
- Improve zero-shot detectors
- Low-resource
- Handling Data Ambiguity
- Developing Realistic Evaluation Frameworks

عنوان مقاله

- AI Generated Text Detection Using Instruction Fine-tuned Large Language and Transformer-Based Models

مقدمه

- Task-A : AI or Human
- Task-B : Which AI?

Dataset

- gemma-2-9b, GPT 4.0, llama-8b, mistral-7b, qwen2-72b and Yi-large
- Train,Validation,Test
- Train = Validation

Data Analysis

- Topics,
- Length,
- Lexical Diversity: Human > AI

Proposed Approach

- Ai or Human => Binary Classification
- Which AI => Multi-Class Classification

- GPT-4o-mini, LLaMa, Bert
- Instruction Tuning => Prompts on what exactly to do

Results and Future Works

- Task-A : Very good both GPT and Bert
- Task-B : Significant drop in Performance

- Larger Models,Longer context, Better Prompting
- => Task-B

عنوان مقاله

- Human vs Machine Generated Text Detection in Persian

Dataset

- Human: COPER(Covid19-articles) + Digikala Comments + news
- AI : Gpt-3.5-turbo

Models

- LSTM
- Dense Neural Networks
- CNN + LSTM + Dense Neural Networks

- Loss:BCE, Optimizer:Adam

Results

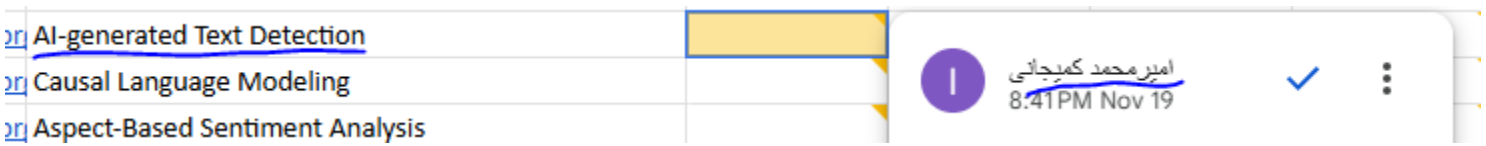
Model	Dataset		Accuracy	Precision	Recall
Dense NN + One-Hot	COPER		0.55	0.58	0.53
Dense NN + Sentence Transformers	COPER		0.61	0.62	0.59
LSTM + Sentence Transformers	COPER		0.68	0.69	0.67
CNN + LSTM + Dense	COPER		0.70	0.72	0.66
	pn_summary	Local	0.68	0.70	0.61
		International	0.73	0.75	0.71
		Economy	0.77	0.79	0.75
		Society	0.69	0.71	0.61
	Digikala		0.56	0.59	0.50

Table 1: A comparison between different human-machine text classifiers.

Future Works and Conclusion

- Machine learning can effectively detect AI-generated Persian text
- Contextual embeddings and pretrained models perform best

- Exploring additional linguistic features
- Expanding to other languages
- Studying how newer AI models affect detection performance


AI-generated Text Detection
Causal Language Modeling
Aspect-Based Sentiment Analysis
امیرمحمد کمیجانی
8:41PM Nov 19