Survival of the Wittiest:

Evolving Satire with Language Models

Thomas Winters & Pieter Delobelle

PhD Students at KU Leuven, Belgium

@thomas_wint

thomaswinters.be

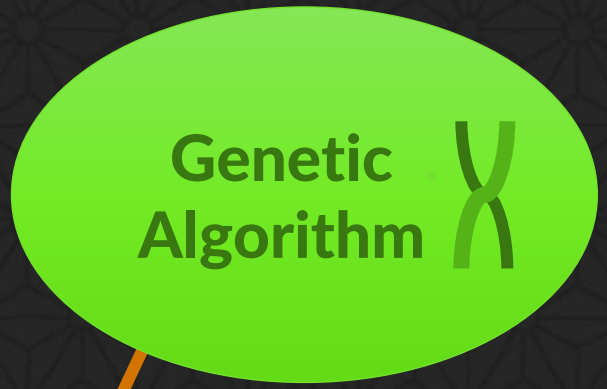
@pieterdelobelle

people.cs.kuleuven.be/~pieter.delobelle/

Paper overview



Fitness function
& mutation operator



Base GA components
+ pareto front

GALMET:
Genetic Algorithm using Language Models for Evolving Text

Apply to satire
as test / pilot

Satire GALMET
Evolves headlines into satire



Transformer models

Large **language models**, **pretrained** on large corpora

Outperform previous neural architectures
on **most language tasks**

OpenAI



GPT-2 & GPT-3

Completes any textual prompt left to right

→ Great for text generation

BERT

- Classifies text sequence / token
- Fills in token if left & right context given

Brown, Tom B., et al. (2020)."Language models are few-shot learners."
Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding

Masked Language Model

- BERT-like models trained using masked language model
  → Good at predicting tokens with left & right context given
- RoBERTa uses BPE tokenisation: subwords

Computational creativity is a subfield of <mask> science.

Tokenize

['<s>', 'Com', 'put', 'ational', 'Ġcreativity', 'Ġis', 'Ġa', 'Ġsub', 'field', 'Ġof', '<mask>', 'Ġscience', '.', '</s>']

Encode using vocabulary index

[0, 14721, 9179, 5033, 11140, 16, 10, 2849, 1399, 9, 50264, 2866, 4, 2]

Masked language model head (MLM)

Predictions:

0.721: computer
0.116: cognitive
...
0.010: social
0.002: rocket

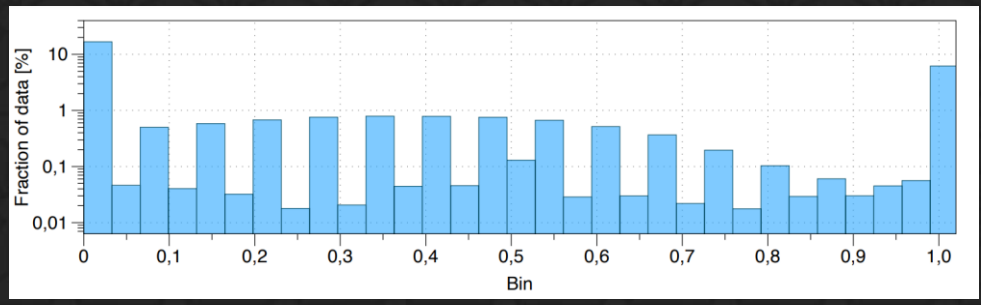
Datasets

Satire edited to be more like headline (+ funniness ratings)

- 2.7K from Unfun *(West and Horvitz 2019)*

Edited headlines & satire (+ funniness ratings)

- 15K from Humicroedit *(Hossain et al, 2019)*
- 8K from Funlines *(Hossain et al, 2020)*



Headlines vs Satire classification datasets

- 20K from Unfun *(West and Horvitz, 2019)*
- 26K Sarcasm Detection dataset *(Misra & Arora, 2010)*
- 22K from The Onion or Not *(https://www.kaggle.com/chrisfilo/onion-or-not)*

→ Also augmented data with all lowercase, titlecase and uppercase

Two RoBERTa models with different heads

Hey, I'll just alter the headline one word at a time!



Alright! I'll check if that edit actually makes it funnier!

Satire MLM

Finetuned RoBERTa MLM on all headlines and satire

Sentence with mask token → suitable token filled into mask

Dog <mask> bites man → Dog accidentally bites man

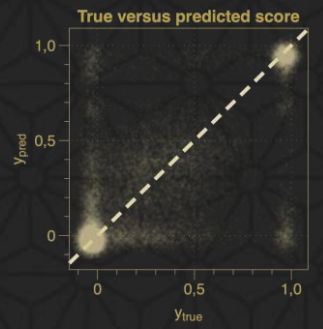
→Used for mutation operator: substituting & adding random tokens

Satire Regression

Finetuned RoBERTa regression head: predicts headline (0), satire (1) or on perceived funniness in between ([0,1])

"Most Americans Want Congress To Investigate Michael Flynn"
→ Predicted score: 0.01

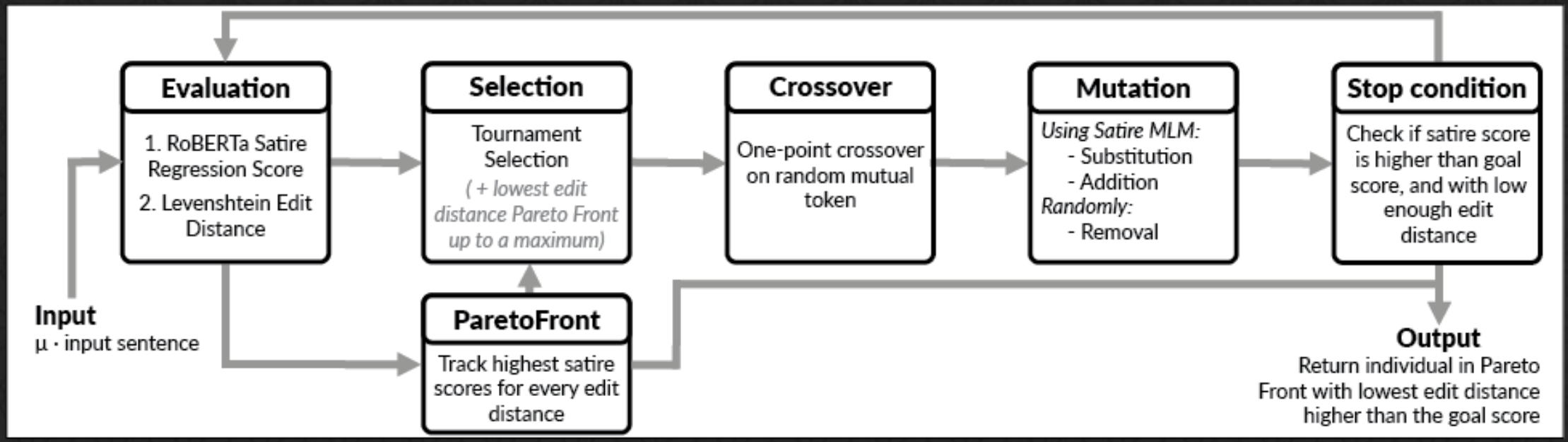
News: The Perils Of Climate Change: A Rise In Sea Level Of Just One Foot Would Be Devastating For The Man Tied To This Pier
→ Predicted score: 1.05



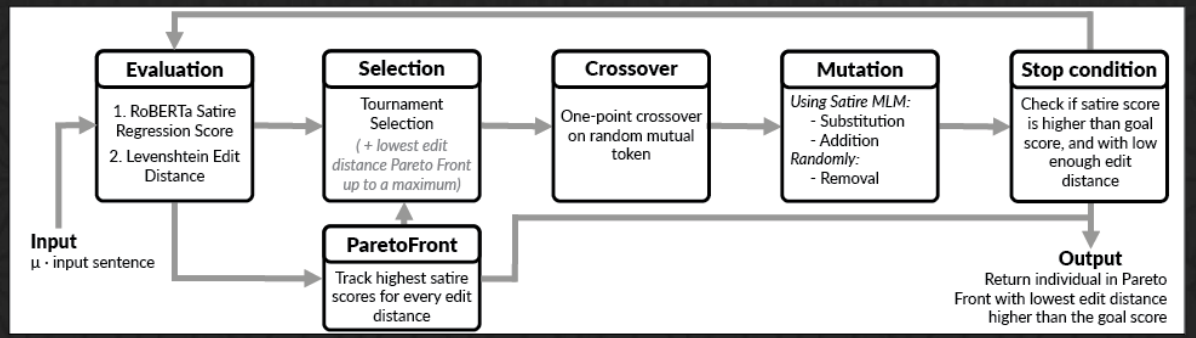
→ Used for fitness function

GALMET: Genetic Algorithm using Language Models for Evolving Text

General algorithm, which here evolves headlines into satirical headlines



Tiger bites zoo visitor

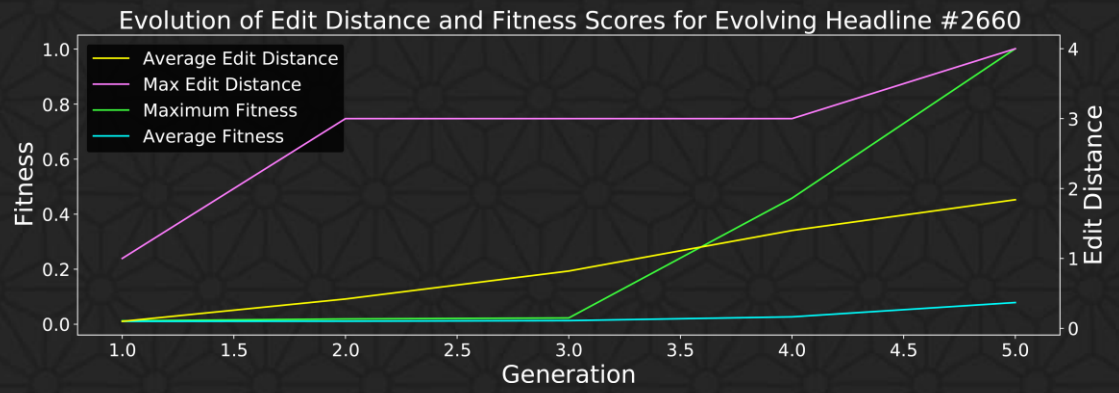


Survival of the Wittiest

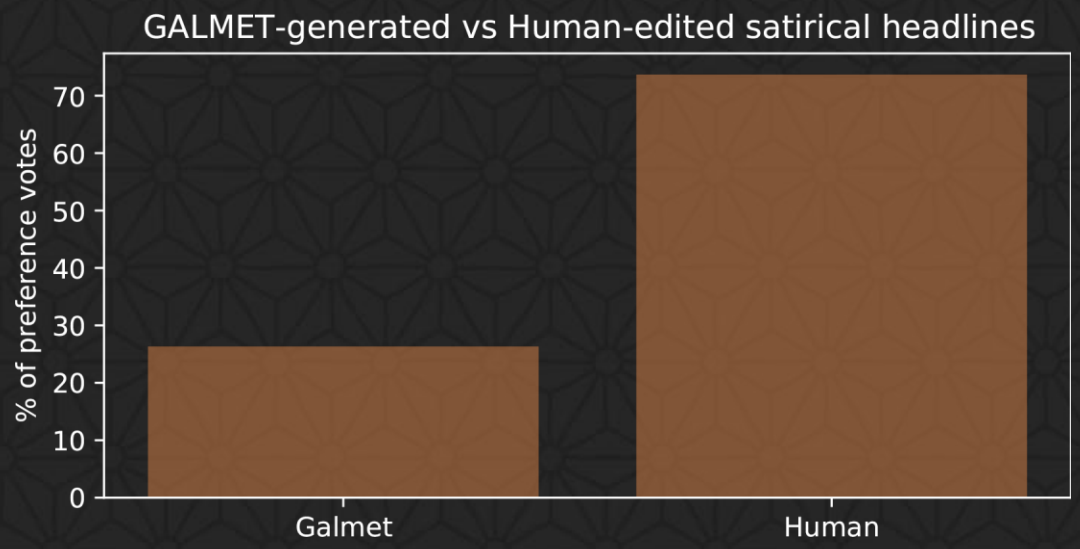
Most Americans Want Congress To Investigate Michael Flynn

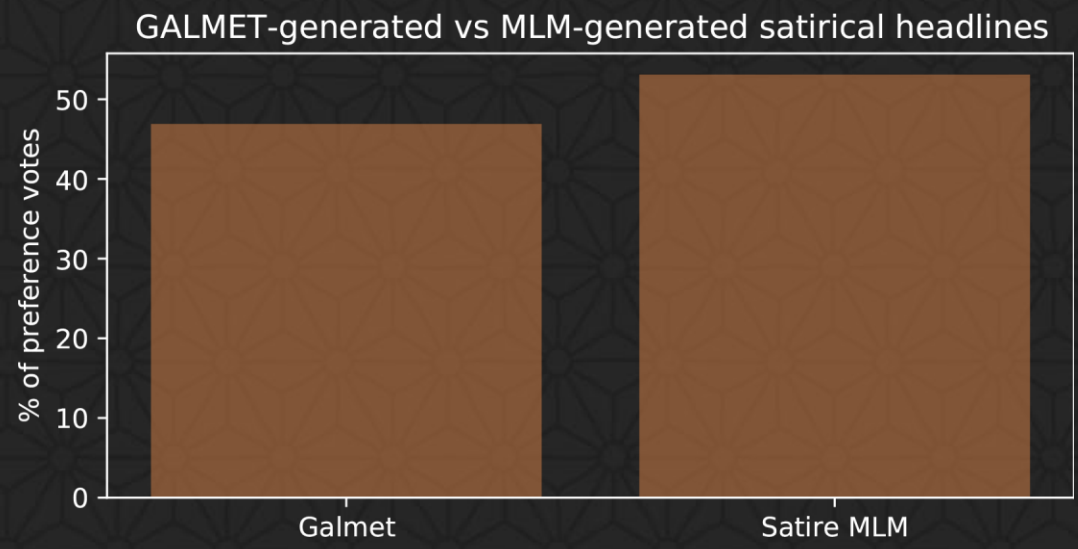
→ 224 Americans Asked To Investigate Michael Jordan

d_{edit}	Score	Phenotype
0	0.010341	Most Americans Want Congress To Investigate Michael Flynn
1	0.019262	Why Americans Want Congress To Investigate Michael Flynn
2	0.070894	Most Americans Want To Investigate Michael Jordan
3	0.458300	224 Americans Asked To Investigate Michael Flynn
4	1.001400	224 Americans Asked To Investigate Michael Jordan



Evaluation





- Human edited headlines from Humicroedit test set *(1 word changed)*
- GALMET was allowed up to 7 edits away *(otherwise genetic algorithm would be just search)*
- Satire MLM just 7 random masking substitutions

→ GALMET not better than its MLM component for evolving satire in this setting

Why? Possible explanations:

Imbalanced dataset

- Much more 0s and 1s than in-betweens: harder to smoothly evolve

Language-breaking mutator operators

- inventing words and breaking grammar due random adding & removing locations
  - *"Vomoted"* from adding token to *"Vote"*
  - *"Haveared"* by removing token from *"Have Cleared"*
- Fitness function not trained to deal with *"broken sentenes"*
  → out of domain

Not very effective Crossover?

- If just one-point cross-over on random words not enough → ~ Parallel hill climbing

Missing information

- True satire needs some context about current state of world: here always same
  (just following regression fitness)

Conclusion

GALMET is an interesting framework for deploying BERT models to **evolve text into other style**/type of text

- *Possibly better for other text generation, such as poetry or adversarial example generation*

→ Co-creative applications where program **suggests improvements** by evolving parts of texts?



Thanks!

Survival of the Wittiest:
Evolving Satire with Language Models

Thomas Winters & Pieter Delobelle

@thomas_wint

thomaswinters.be

thomas.winters@kuleuven.be

@pieterdelobelle

people.cs.kuleuven.be/~pieter.delobelle

pieter.delobelle@kuleuven.be

Some images (based on the works) of dooder on freepik.com