



Transfer Learning for Style-Specific Text Generation

Leveraging pre-trained language models for creative uses of small datasets.

Katy Ilonka Gero, Giannis Karamanolakis, and Lydia Chilton

Motivation: Artistic applications of NLP

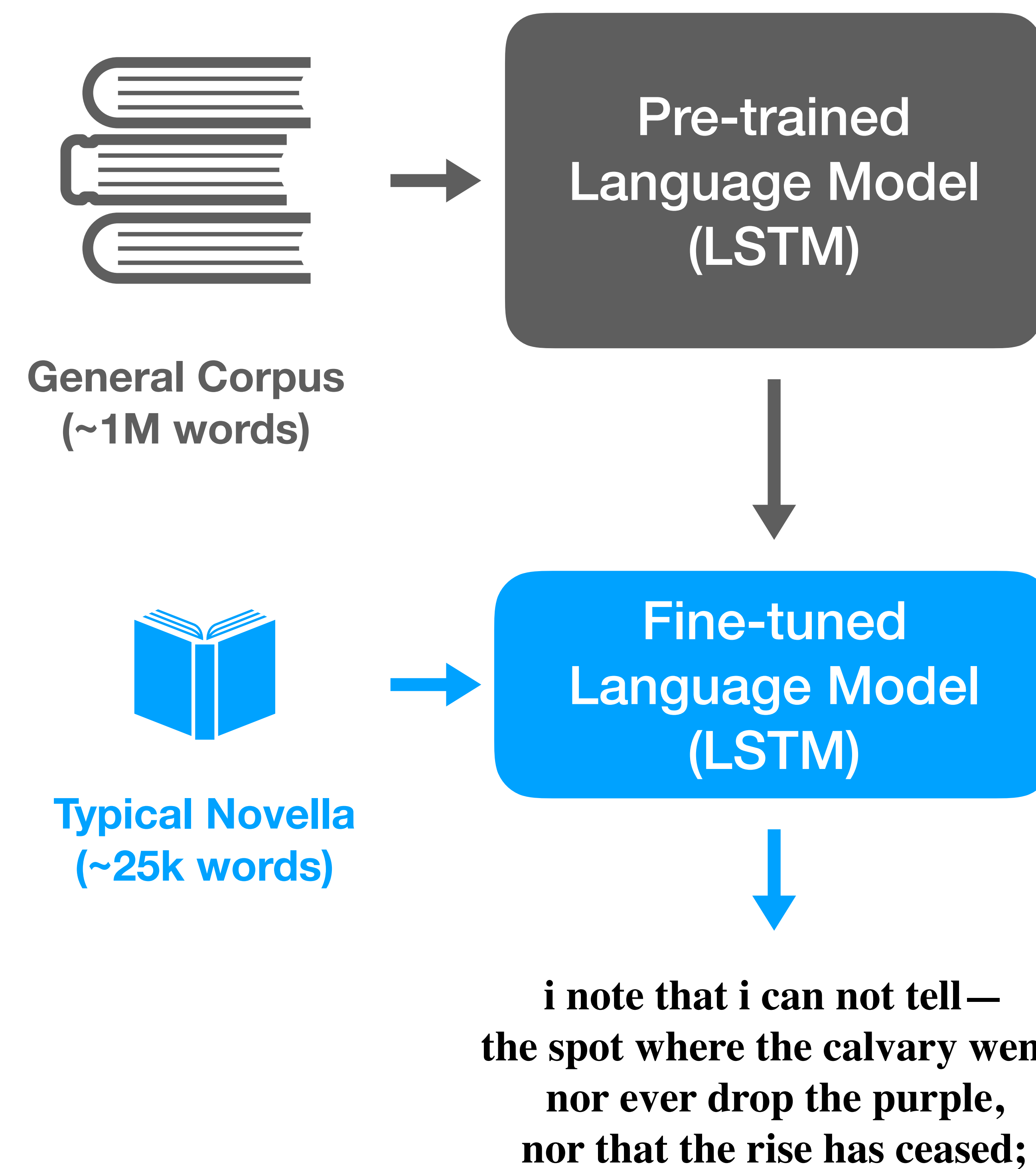
Neural networks have shown promising results for generating text for artistic purposes.

Challenge: Artistic datasets are small

Despite promising results, current methods require large data sets, often hundreds of millions of words. Yet most sources of literary text are relatively small.

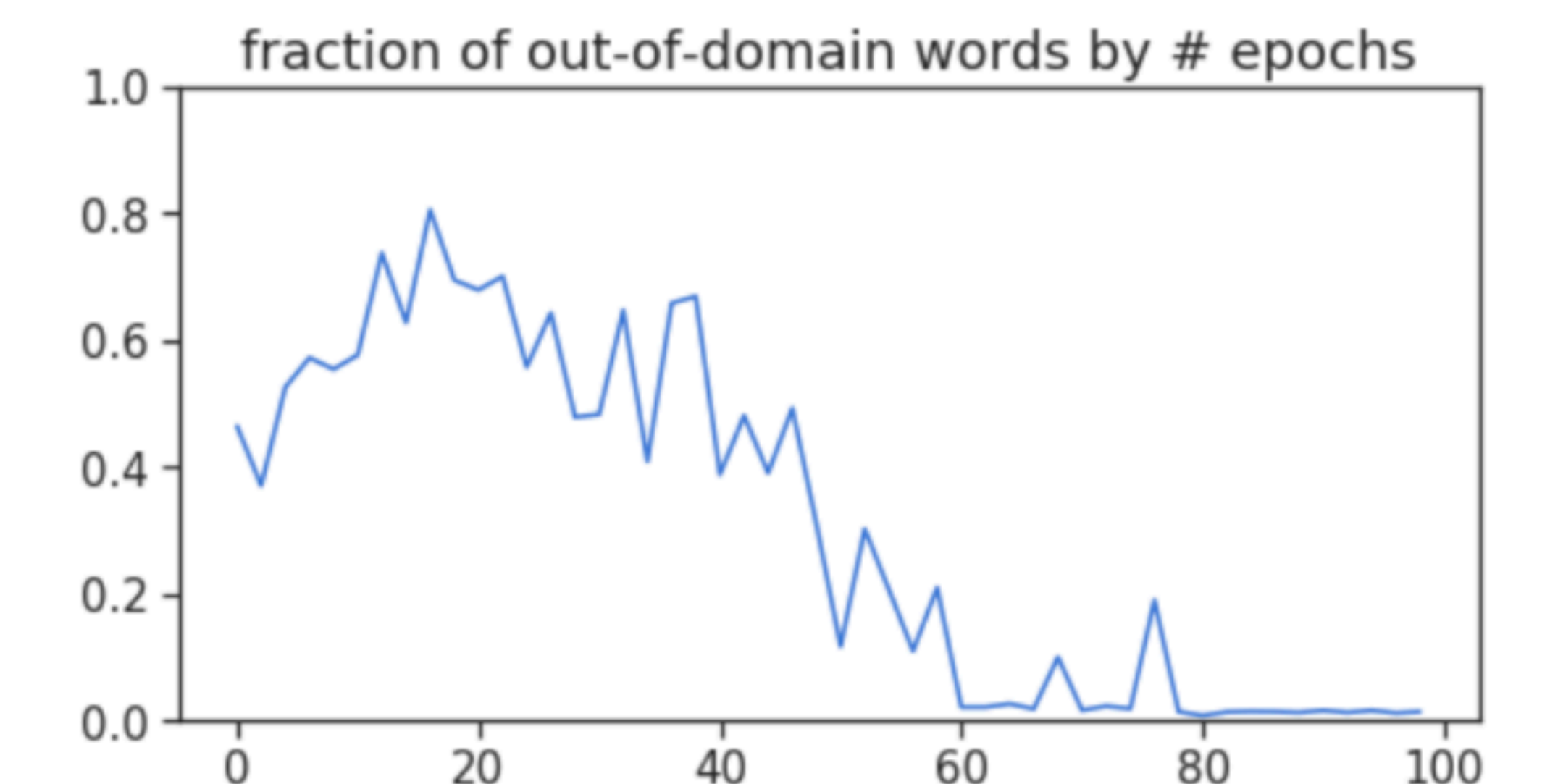
Approach: Use pre-trained models

We fine-tune a pre-trained language model using small data sets and qualitatively show that pre-training results in coherent generation for three literary styles.



Next steps: Mix vocabulary and style

How can we train a model that uses vocab of a different corpus? To test this, we allow words from the pre-trained model and generate text at each epoch of fine-tuning. If we fine-tune for too long, the model does not use out-of-domain words.



Next steps: Real-time web interface

Fine-tuning is fast compared to typical neural applications—fast enough to implement a web interface where anyone can upload a style corpus and start generating text within minutes.

IMAGINATIVE

Alice in Wonderland (25k words)

had down, there same be said burn gryphon the, , be lessons said 'who, would well it to rabbit began of over going child the be' done, not deal my quite time fun speak,

all that happens, you know,' said the duchess, 'as a poison! it must be kept in the safe.'

'do you know what you're saying!' said alice. caterpillar lobsters!

without pre-training

with pre-training

POETIC

Emily Dickinson (25k words)

wondered
t gain ; , ,
us mine troth morn warm. me the tenderer;
- are amid his me , perceived play . star

after god, surpasses god have
no nearer than heaven;
but heaven had not daffodils,
nor had it bound me.

HIGHBROW

Consider the Lobster (6k words)

the component rationales their to issue's imagine least known honors in boiling is three palates out and than gustatory on corrupt least brittlely their) get most pain torture are.

it's hard to deny in any meaningful way that this one knows what the pain is. even if you cover the 'm , it's hard to deny that there is any moral distinction between pain and pain.

