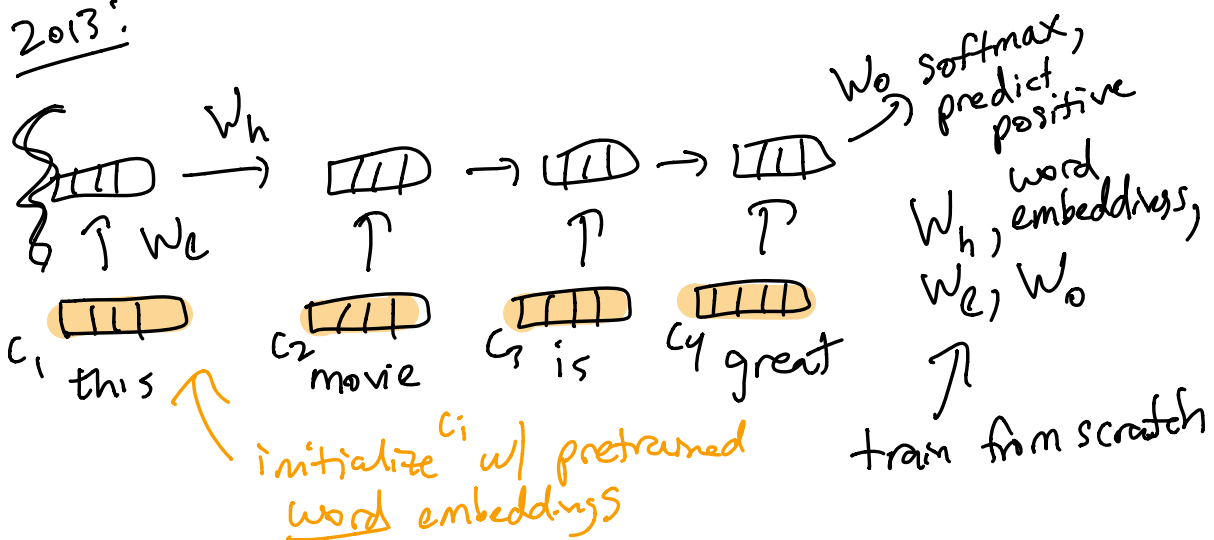


Using neural LMs for transfer learning

let's consider the task of sentiment analysis
2013:

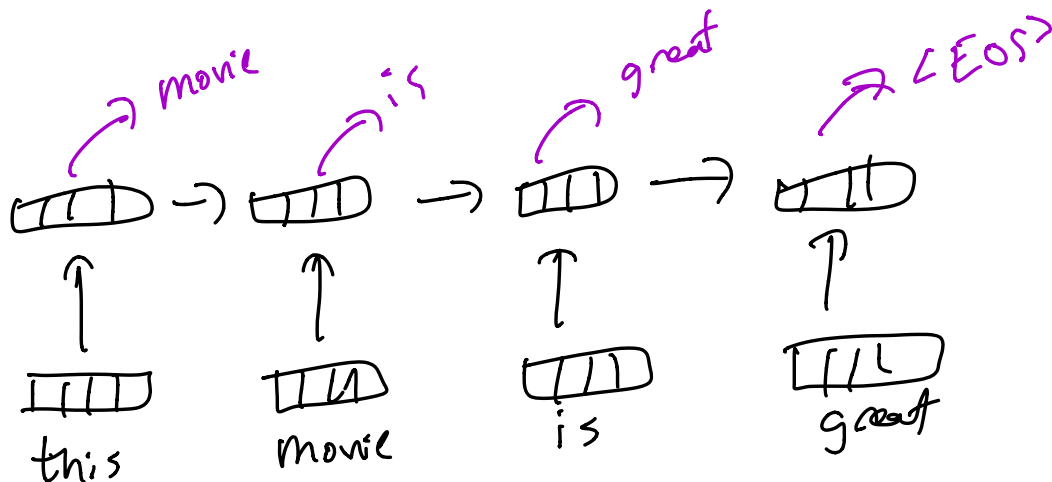


2018:

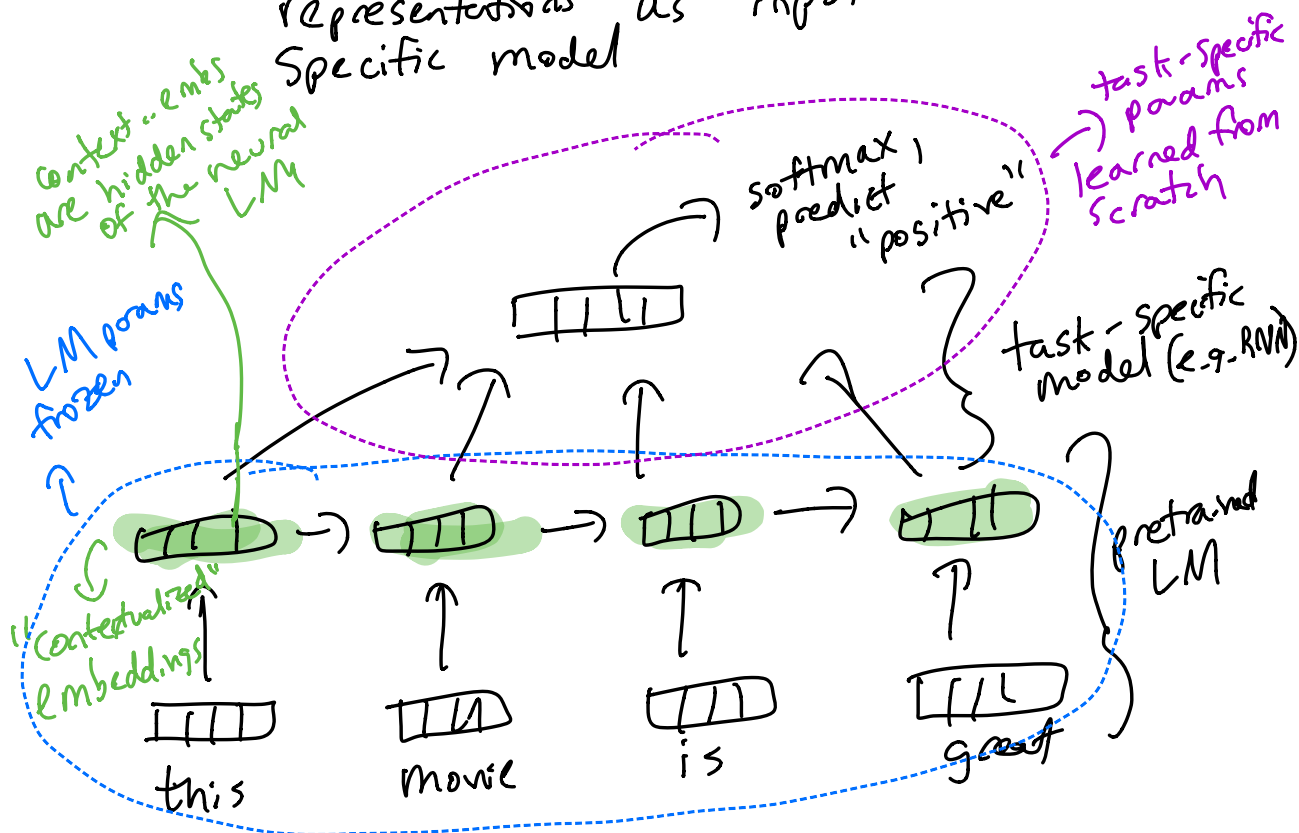
- issues w/ above: forcing the model to learn composition from a small labeled dataset is a bad idea
- what if instead of only transferring word embs, we repurpose a neural LM to solve the downstream task.

ELMo: embeddings from language models

Step 1: pretrain an RNN LM on lots of data

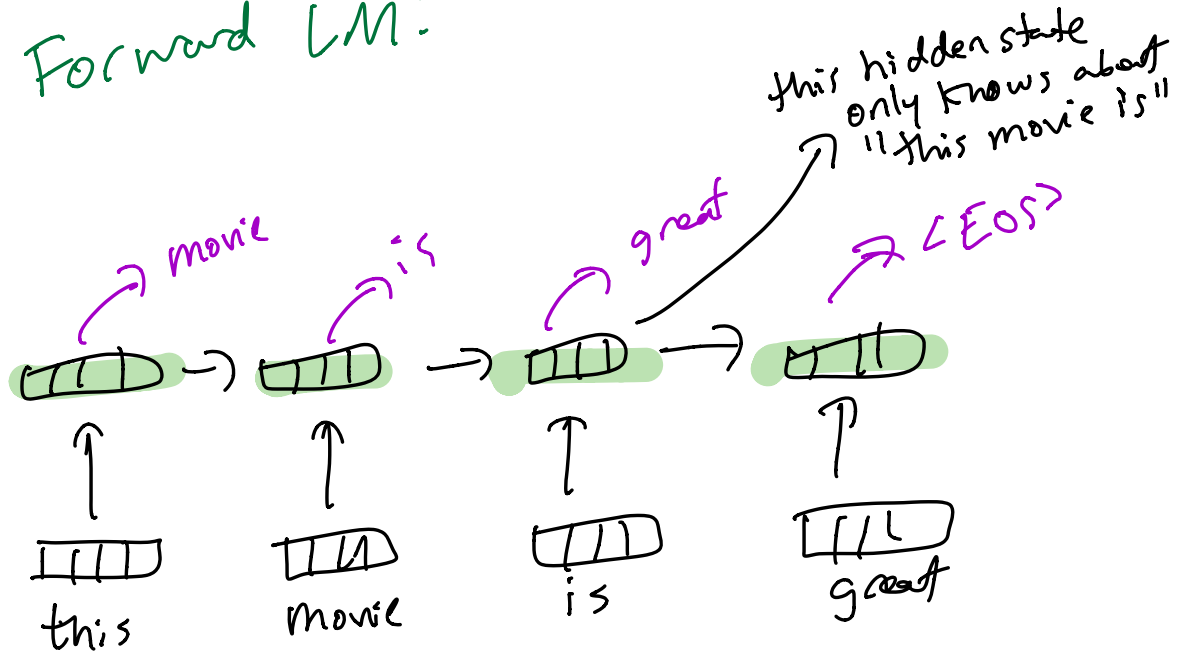


Step 2: freeze LM parameters, use its representations as input to a task-specific model

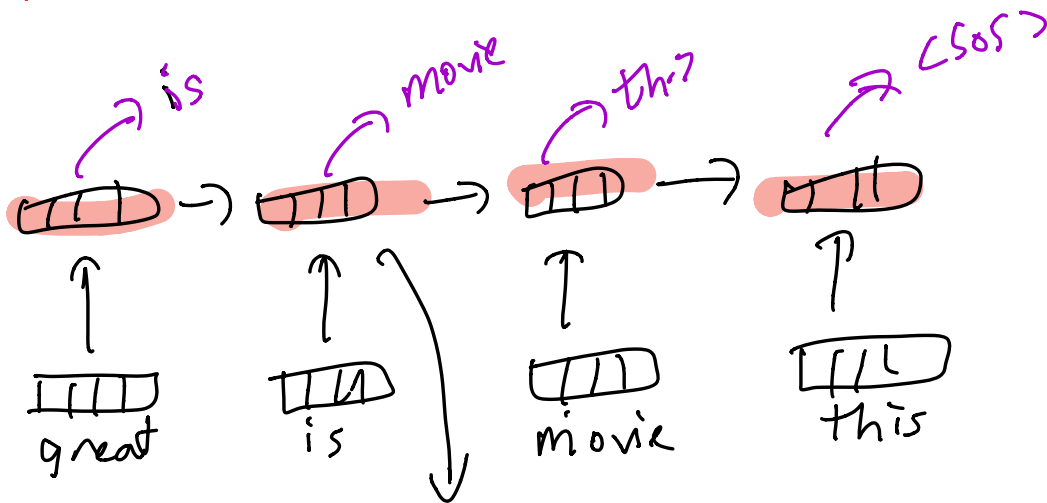


ELMo architecture:

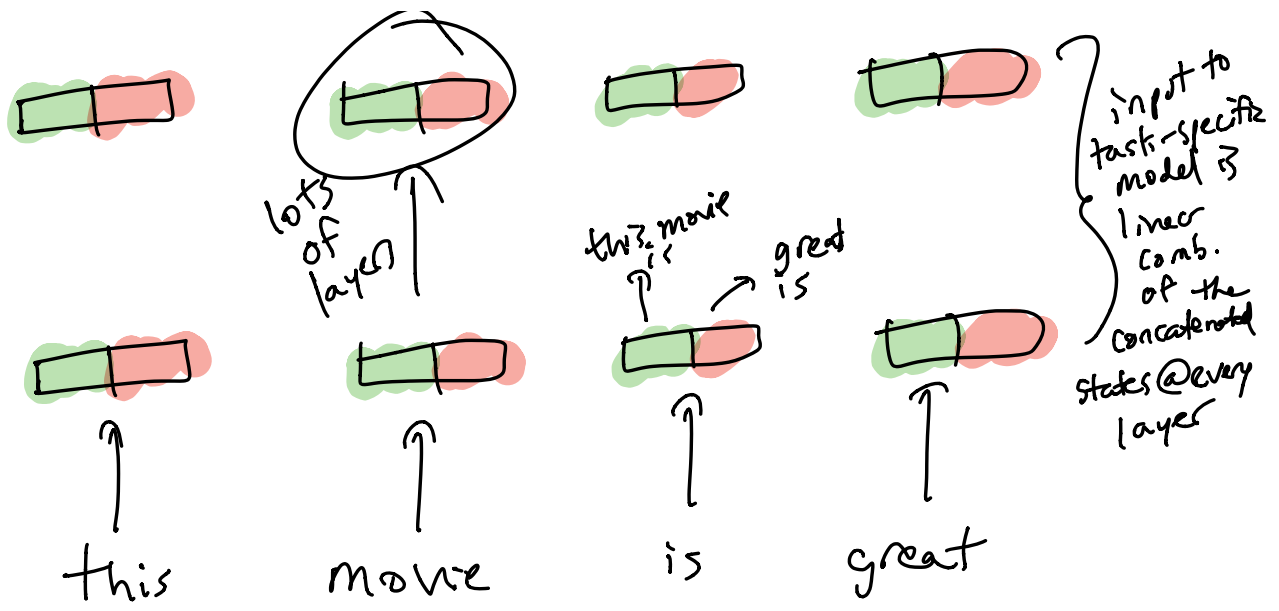
Forward LM:



BACKWARDS LM



this hidden state only knows about "great is"

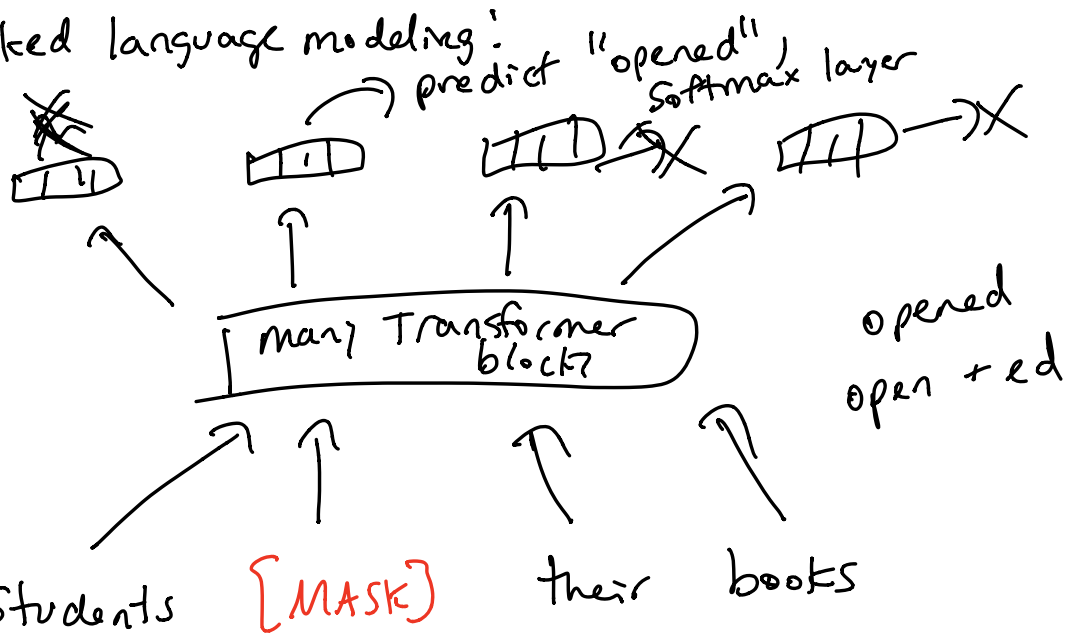


forward/backward LM is pretty clumsy
 can we replace these w/ a single model

ELMo \Rightarrow BERT
 2018 2019

- 2 unidirectional LMs \Rightarrow one masked LM
- recurrent NNs \Rightarrow Transformers
- fine-tune the entire model instead of keeping the LM frozen
- pretrain on way more data w/ a way bigger model

Masked language modeling:



- each representation at every time step is fully contextualized!
→ "aware" of both past + future words
- Same training obj. as basic LM