



# Fast AI Lesson 8: NLP



Language models  
And  
Much much more.



# About me: Jaco du Plessis

Mechatronic Engineer

Head of ML at [Notiv](#)

Focused building nlp and audio deep learning products

Hiring for NLP

[Twitter](#)

[linkedin.com/in/jaco-du-plessis-566b61105](https://www.linkedin.com/in/jaco-du-plessis-566b61105)

## Action Item Extraction

Suggested items

00:00 / 17:27

1x

would be good to, well, i was going to suggest that it would be good at least to review the previous retrospectives.

Actions

And i'll probably work on that... Later this afternoon or, tomorrow morning, get some, time slots in there for us to plan through.

00:56

And then, for the next day, then the week after we'll, kind of reassess and tickets that a week that we're going to set some objectives for the, cycle. We'll do that. We'll do that tomorrow. I think that was the planning in a row, just putting reassess that i think that's what we had discussed, right? Yeah.

02:04

We see it as a in maybe thinking through that lens might change your thoughts a little bit on the data that you pausing or,

06:36

set some objectives

- Context: cycle, tickets
- To do: reassess

# Notiv - summarization

## Creating a budget for the new financial year

Thu, Jun 17th 2021 6:09pm



Forecasting Tool, Creating A Budget, Financial Datum, Accounting System, Benchmarking Tool, Accountant, Interactive Dashboard

Speaker 1 - 00:20

All right. Welcome everyone to today's webinar with fathom creating a budget for the new financial year. Now I can still see there's quite a few of us joining in the session at the moment. I'm going to wait for about another minute or just to let everyone join in.

[unclear]

Hopefully that's given everyone the chance to join. And again, welcome to today's webinar today. We're going to be creating a budget for the new financial year using Fathom's new forecasting tool.

Now this is actually part two of a three-part webinar series. If you did attend the last one, welcome back just a few housekeeping things before we get started, the session is going to run for around 40 to 45 minutes. If you do need to jump off the session's going to be recorded and we'll make sure that recording gets to you at the very end,

Something else to note, we are using the zoom Q and a feature as well. If you do have any questions or comments throughout the session, feel free to pop those in. We'll leave about five to 10 minutes at the very end to go through some of the most frequently asked questions. If we don't get to yours, we'll make sure to follow up afterwards as well.

Now, just a quick introduction. My name's Jess, I'm a customer engagement specialist here at fathom. Basically that means I help our accountants and businesses get set up for success. Now, whether that sounds like you or not, you're welcome to take down my email at the bottom of the screen here. I'd love to hear from you if you have any questions or feedback at the end of the session as well.

Hey,

here's our agenda for today. I'll start out with a little bit of background on what fathom is for those of you who are brand new, who we are, and really what we do here. Then we'll jump into why businesses actually create a budget. Really what's the point of the session? Why are we really creating a budget for the new year?

And then we'll jump into the product itself and we'll show how fathoms forecasting tool can be used to build the budget and monitor our performance going forward.

## Summary of Creating a budget for the new financial year

Wed, Jun 23rd 2021 10:23am



Forecasting Tool, Accountant, Introduction, Creating A Budget, Business, Session, Interactive Dashboard

Speaker 1 - 00:00

Well, the session feel free to pop those in. We'll leave about five to 10 minutes at the very end to go through some of the most frequently asked questions. If we don't get to yours, we'll make sure to follow up afterwards as well.

Now, just a quick introduction. My name's Jess, I'm a customer engagement specialist here at fathom. Basically that means I help our accountants and businesses get set up for success. Now, whether that sounds like you are not, you're welcome to take down my email at the bottom of the screen here. I'd love to hear from you. If you have any questions or feedback at the end of the session as well.

Great.

Here's our agenda for today. I'll start out with a little bit of background on what fathom is for those of you who are brand new, who we are, and really what we do here. Then we'll jump into why businesses actually create a budget. Really what's the point of the session? Why are we really creating a budget for the new year?

And then we'll jump into the product itself and I'll show how fathoms,



# Agenda

- Recap of content
- Transformers aka sesame street
- How things go wrong in nlp
- Make your models better
- Code examples
- Paper review



## Recap

- Tokenizer
- numericalization
- Batching
- Embedding
- Model -> NN - with for loop
- Classification head



# Tokenization

```
first(spacy(['The U.S. dollar $1 is $1.00.']))
```

```
(#9) ['The', 'U.S.', 'dollar', '$', '1', 'is', '$', '1.00', '.']
```

```
coll_repr(tkn('&copy; Fast.ai www.fast.ai/INDEX'), 31)
```

```
"(#11) ['xxbos', '@', 'xxmaj', 'fast.ai', 'xxrep', '3', 'w', '.fast.ai', '/', 'xxup', 'index...']"
```



# Numericalization

```
nums = num(toks)[:20]; nums
```

```
tensor([ 2,  8, 21, 28, 11, 90, 18, 59,  0, 45,  9, 351, 499, 11, 72, 533, 584, 146, 29, 12])
```

This time, our tokens have been converted to a tensor of integers that our model can receive. We can check that they map back to the original text:

```
' '.join(num.vocab[o] for o in nums)
```

```
'xxbos xxmaj this movie , which i just xxunk at the video store , has apparently sit around for a'
```



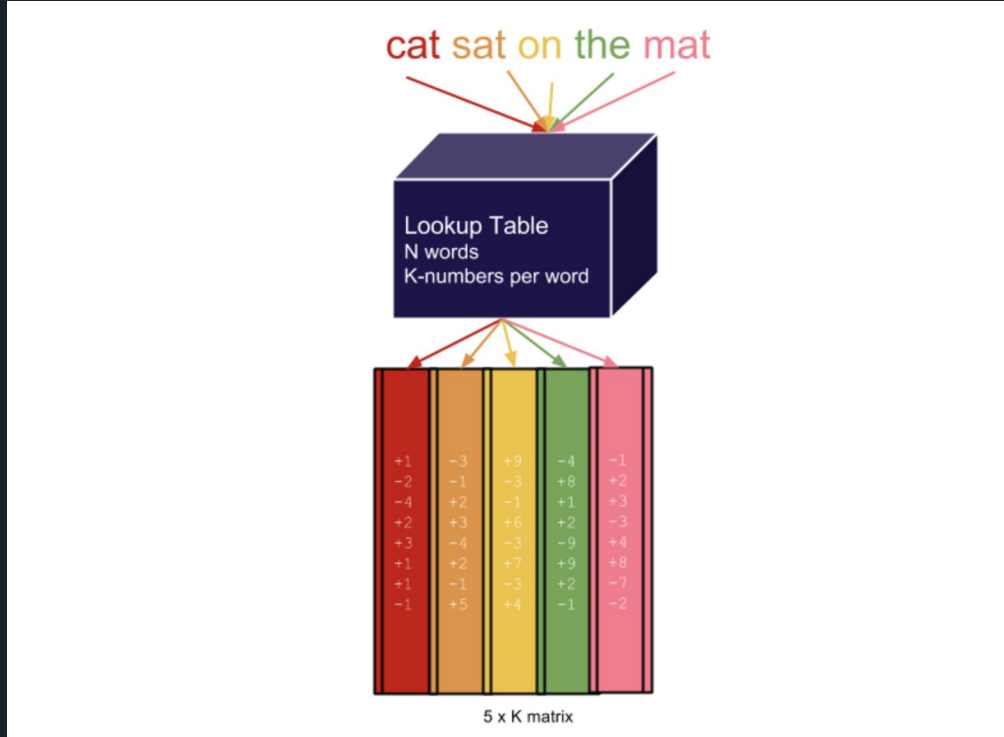
# Batching - Language model

xxbos	xxmaj	in	this	chapter	,	we	will	go	back	over	the	example	of	classifying
movie	reviews	we	studied	in	chapter	1	and	dig	deeper	under	the	surface	.	xxmaj
first	we	will	look	at	the	processing	steps	necessary	to	convert	text	into	numbers	and
how	to	customize	it	.	xxmaj	by	doing	this	,	we	'll	have	another	example
of	the	preprocessor	used	in	the	data	block	xxup	api	.	\n	xxmaj	then	we
will	study	how	we	build	a	language	model	and	train	it	for	a	while	.

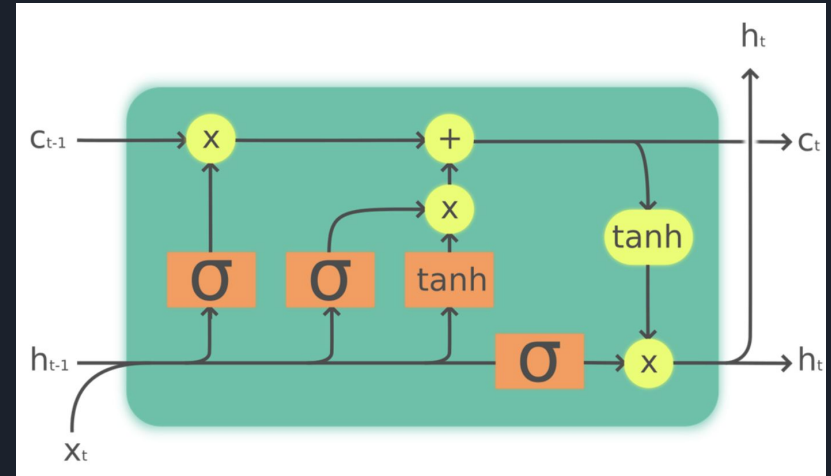
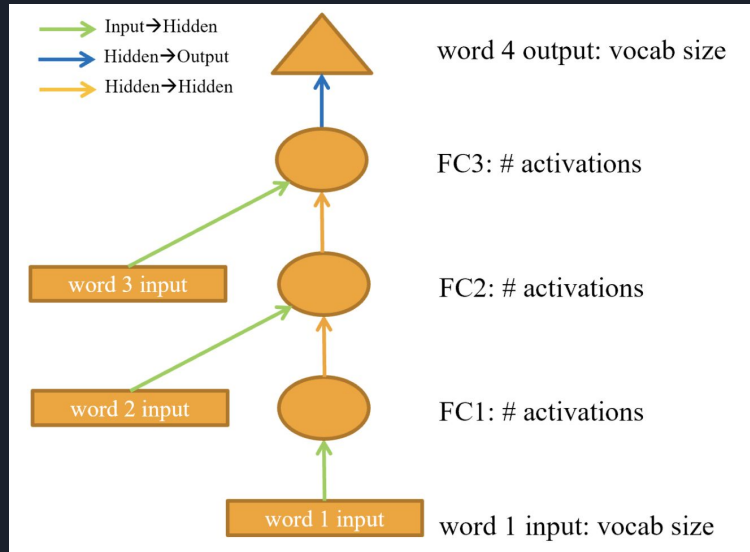
xxbos	xxmaj	in	this	chapter
movie	reviews	we	studied	in
first	we	will	look	at
how	to	customize	it	.
of	the	preprocessor	used	in
will	study	how	we	build

,	we	will	go	back
chapter	1	and	dig	deeper
the	processing	steps	necessary	to
xxmaj	by	doing	this	,
the	data	block	xxup	api
a	language	model	and	train

# Embedding



# Model



# Transformers

Attention is all you need.

BERT, ALBERT, GPT series


Large scale expensive  
to train

Not always needed



# Transformers

## Single Headed Attention RNN: Stop Thinking With Your Head

 **elvis** @omarsar0 · Jul 5  
A Primer on Pretrained Multilingual Language Models

This survey paper reviews the existing literature covering research around Multilingual Language Models (MLLMs) such as mBERT, XLM, XLM-R.

Great progress happening in NLP beyond English.

[arxiv.org/abs/2107.00676](https://arxiv.org/abs/2107.00676)

arXiv.org > cs > arXiv:2107.00676

Computer Science > Computation and Language

Submitted on 1 Jul 2021


### A Primer on Pretrained Multilingual Language Models

Sumanth Doddapaneni, Gowtham Ramesh, Anoop Kunchukuttan, Pratyush Kumar, Mitesh M. Khapra

Multilingual Language Models (MLLMs) such as mBERT, XLM, XLM-R, [xlm-roberta] have emerged as a viable option for bringing the power of pretraining to a large number of languages. Given their success in zero shot transfer learning, there has emerged a large body of work in (i) building bigger MLLMs covering a large number of languages (ii) creating exhaustive benchmarks covering a wider variety of tasks and languages for evaluating MLLMs (iii) analyzing the performance of MLLMs on monolingual, zero shot crosslingual and bilingual tasks (iv) understanding the universal language patterns of any learnt by MLLMs and (v) augmenting the (often) limited capacity of MLLMs to improve their performance on seen or even unseen languages. In this survey, we review the existing literature covering the above broad areas of research pertaining to MLLMs. Based on our survey, we recommend some promising directions of future research.

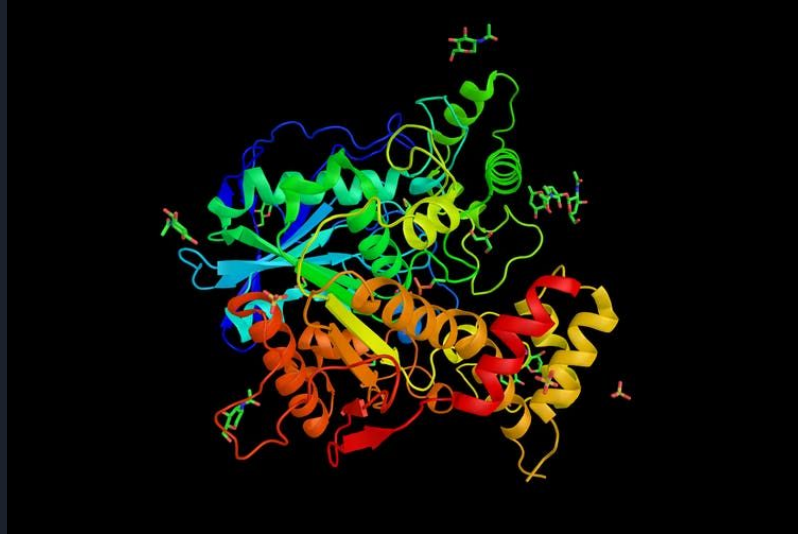
Subjects: Computation and Language (cs.CL)  
Cite as: arXiv:2107.00676 [cs.CL]  
See arXiv:2107.00676v1 [cs.CL] for this version

3 49 217 Tip

 **Jeremy Howard** @jeremyphoward · Jul 5  
They cite **multifit**, an lstm, but then suggest only transformers are used in this space. I still hear of a lot of folks in low resource languages getting better results with **multifit**/ulmfit style approaches

6 33 Tip

# Transformers - do learn about them



<https://jalammar.github.io/illustrated-transformer>

<http://nlp.seas.harvard.edu/2018/04/03/attention.html>

[The Transformer for language translation \(NLP video 18\)](#)

## How things go wrong - Noise

**Repetition** When is **Eas** ugh **Easter** this year?  
**Correction** When is **Lent** I meant **Easter** this year?  
**Restarts** **How much** no wait **when is** Easter this year?


Model	Train	Eval	HasAns-F1	NoAns-F1	Overall-F1
BERT-QA	ALL	SQUAD	83.87	70.55	77.46
		Heuristics	51.45 ↓ 32.42	74.49 ↑ 3.94	62.53 ↓ 14.93
		DISFL-QA	40.97 ↓ 42.90	75.97 ↑ 5.42	57.81 ↓ 19.65
	ANS	SQUAD	89.63	-	89.63
		Heuristics	80.52 ↓ 9.11	-	80.52 ↓ 9.11
		DISFL-QA	78.88 ↓ 10.75	-	78.88 ↓ 10.75

<https://arxiv.org/pdf/2106.04016.pdf>

# How things go wrong - Sentiment Classifier







# 1904.12848 : Unsupervised Data Augmentation for Consistency Training

<b>Datasets</b> (# Sup examples)		IMDb (25k)
Pre-BERT SOTA		4.32
BERT <sub>LARGE</sub>		4.51
S		
<b>Initialization</b>	<b>UDA</b>	IMDb (20)
Random	✗	43.27
	✓	25.23
BERT <sub>BASE</sub>	✗	18.40
	✓	5.45
BERT <sub>LARGE</sub>	✗	11.72
	✓	4.78
BERT <sub>FINETUNE</sub>	✗	6.50
	✓	<b>4.20</b>



# Unsupervised Data Augmentation for Consistency Training

Given the low budget and  
production limitations, this movie  
is very good.

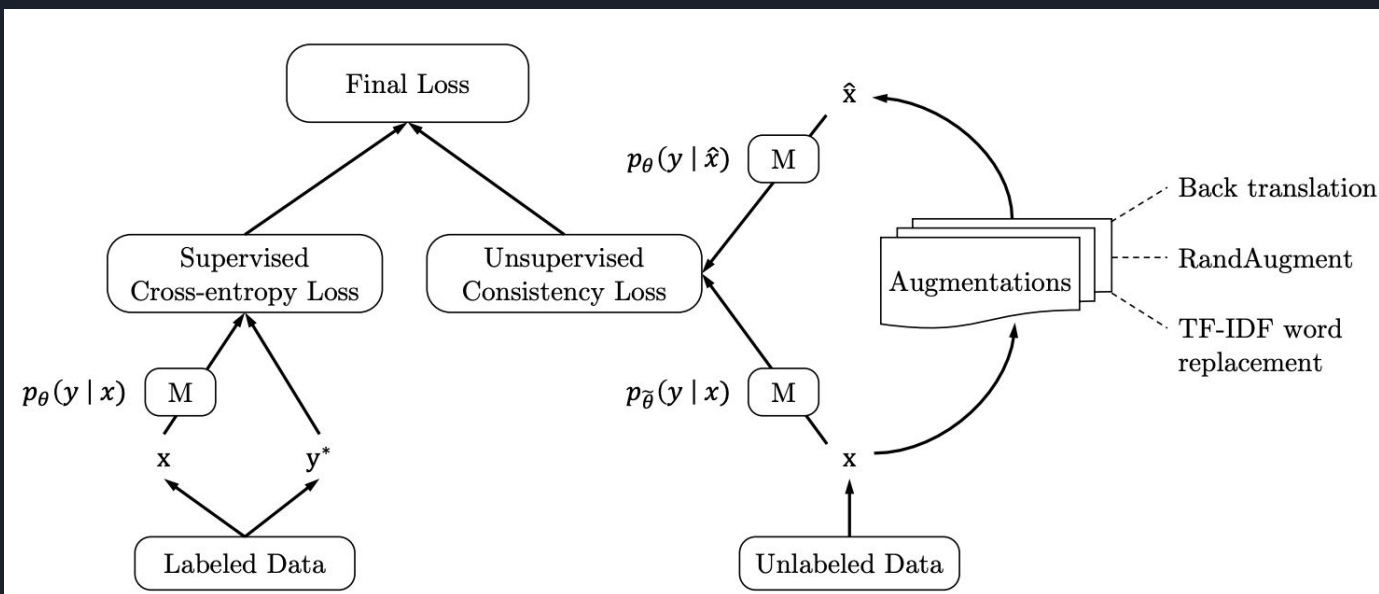
Back-translation

Since it was highly limited in terms of  
budget, and the production restrictions, the  
film was cheerful.

There are few budget items and production  
limitations to make this film a really good  
one.

Due to the small dollar amount and  
production limitations the ouest film is very  
beautiful.

# Unsupervised Data Augmentation for Consistency Training





# Unsupervised Data Augmentation for Consistency Training

**Confidence-based masking.** We find it to be helpful to mask out examples that the current model is not confident about. Specifically, in each minibatch, the consistency loss term is computed only on examples whose highest probability among classification categories is greater than a threshold  $\beta$ . We set the threshold  $\beta$  to a high value. Specifically,  $\beta$  is set to 0.8 for CIFAR-10 and SVHN and 0.5 for ImageNet.

**Domain-relevance Data Filtering.** Ideally, we would like to make use of out-of-domain unlabeled data since it is usually much easier to collect, but the class distributions of out-of-domain data are mismatched with those of in-domain data, which can result in performance loss if directly used [44]. To obtain data relevant to the domain for the task at hand, we adopt a common technique for detecting out-of-domain data. We use our baseline model trained on the in-domain data to infer the labels of data in a large out-of-domain dataset and pick out examples that the model is most confident about. Specifically, for each category, we sort all examples based on the classified probabilities of being in that category and select the examples with the highest probabilities.



# Unsupervised Data Augmentation for Consistency Training

$$\min_{\theta} \mathcal{J}(\theta) = \mathbb{E}_{x_1 \sim p_L(x)} [-\log p_{\theta}(f^*(x_1) | x_1)] + \lambda \mathbb{E}_{x_2 \sim p_U(x)} \mathbb{E}_{\hat{x} \sim q(\hat{x}|x_2)} [\text{CE}(p_{\tilde{\theta}}(y | x_2) || p_{\theta}(y | \hat{x}))]$$
(1)

$$\frac{1}{|B|} \sum_{x \in B} I(\max_{y'} p_{\tilde{\theta}}(y' | x) > \beta) \text{CE} \left( p_{\tilde{\theta}}^{(sharp)}(y | x) || p_{\theta}(y | \hat{x}) \right)$$

$$p_{\tilde{\theta}}^{(sharp)}(y | x) = \frac{\exp(z_y / \tau)}{\sum_{y'} \exp(z_{y'} / \tau)}$$