92586 Computational Linguistics

Lesson 20. LSTM: characters and generation

Alberto Barrón-Cedeño

Alma Mater Studiorum-Università di Bologna
a.barron@unibo.it @_albarron_

07/05/2021

Previously

- Convolutional neural networks
- Recurrent neural networks
- Bidirectional Recurrent neural networks
- Long short-term memory networks

Table of Contents

Out of Vocabulary

The curse of OOV

Out-of-vocabularies cause big trouble

> The Mexico City Metro, operated by the Sistema de Transporte Colectivo, it is the second largest metro system in North America after the New York City Subway.

> The Mexico_City Metro, operated by the · de · ·, it is the second largest metro system in North America after the New_York City Subway.

Alternatives

- Replace the unknown with a random word, from the embedding space
- Replace the unknown word wit UNK, and produce a random vector
- **Turn into characters**

https://en.wikipedia.org/wiki/Mexico_City_Metro

Characters

Into Characters

Words are *just* a sequence of characters

By modeling the representations at the character level...

- We get rid of OOVs
- We can learn patterns at a lower level
- We reduce the variety of input vectors drastically

Let us see

Into Characters: outcome

- The training takes way longer (more than one hour)
- The training accuracy is great: ~ 92
- The validation accuracy is terrible: ~ 59
- **Overfitting**

Reasons/Solutions

- The model might be *memorising* the dataset
- Increase the dropout (try!)
- Add more labeled data (hard!)

A character-level model shines at its best at modeling/generating language

Text generation

Predicting the next word

- An LSTM can learn

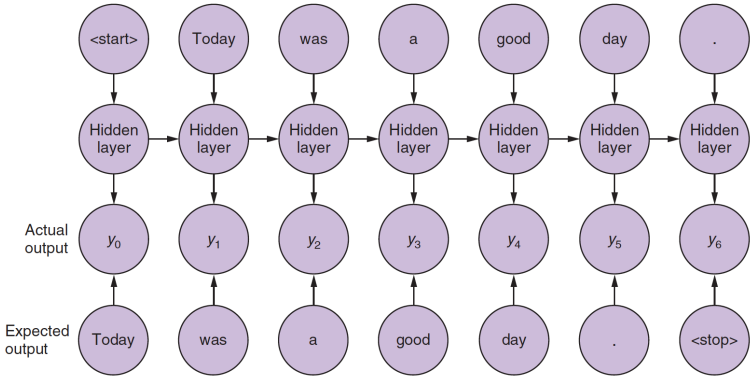
$$p(w_t \mid w_{t-1}, w_{t-2}, \ldots, w_{t-n}) \tag{1}$$

- It can do so **with a memory** (full context)
- It can do so at **character level**

From classification to generation

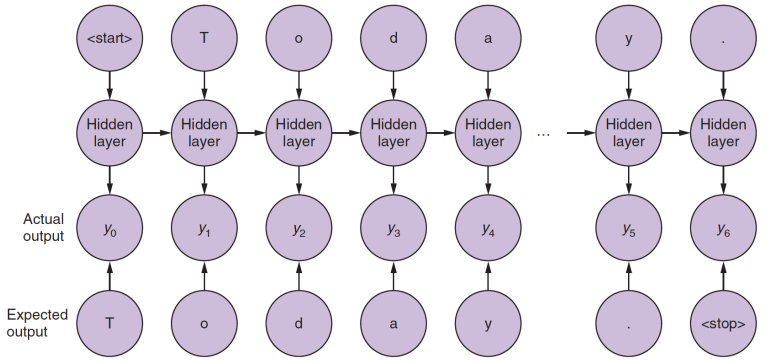
- No more classification layer at the end
- Now we want to predict the next word ($\sim$ word2vec?)

Unrolling the next-word prediction



(Lane et al., 2019, 299)

Unrolling the next-~~word~~character prediction



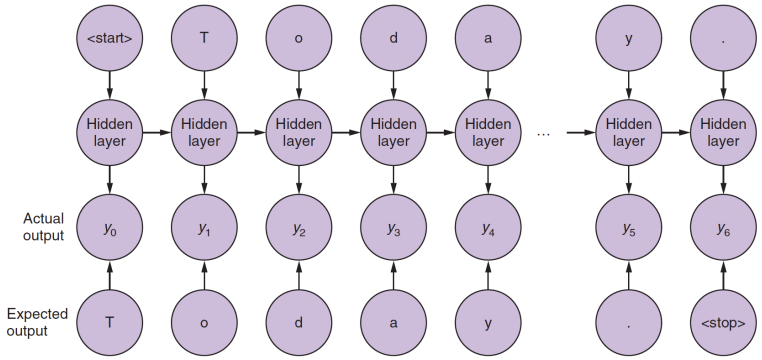
Expected output is the next token in the sample. Shown here on character level.

- Now the error is computed for every single output
- We still back-propagate until visiting a full instance

(Lane et al., 2019, 299)

New target labels

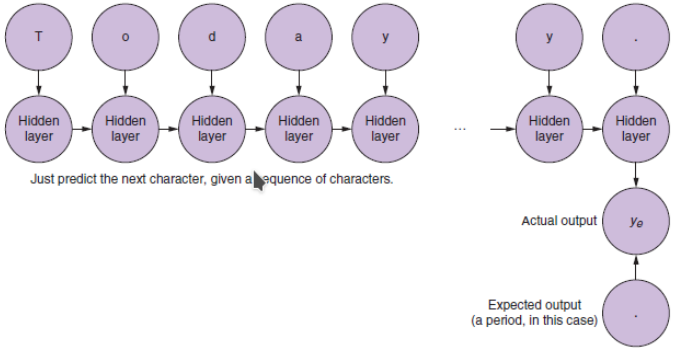
New output: a one-hot encoding (again) of the next character



Expected output is the next token in the sample. Shown here on character level.

(Lane et al., 2019, 299)

Predict after having looked at a sequence



Just predict the next character, given a sequence of characters.

(Lane et al., 2019, 300)

Generation example

Since we are interested in *style* and creating a consistent model, we wont use IMDB (multi-authored and small).

Let us try to mimic William Shakespeare

We made it up to Section 9.1.8 of Lane et al. (2019)

Recap

Recap: The path

1. Baby steps into computing
2. What is NLP? From rule-based to statistical
3. Pre-processing text: tokens, stemming, stopwording...
4. From words to vectors: the vector space model
5. A few supervised models
6. Training and evaluating in machine learning
7. From words to meaning: topic modeling
8. Using one neuron: perceptrons
9. Fully-connected neural networks
10. From words to semantics: word embeddings
11. Taking snapshots of text: CNNs
12. Texts as sequences: (Bi)RNNs
13. Using a better memory: LSTM
14. LSTM to produce text

Recap: The future path

That's 9 out of 13 chapters of Natural Language Processing in Action

1. Producing sequences: sequence-to-sequence models & attention
2. Named entity recognition
3. Question answering
4. Dialog systems
5. Multilingual models
6. Machine translation

You are ready to go on your own now and become a star

RAI News 24; 8 April, 2021

Now go and celebrate the end of the course

...and worry about your project from Monday!

- I'm available during the lesson times for 1-to-1 discussion on your project **upon request!**
- **Meanwhile, take care!**

References

Lane, H., C. Howard, and H. Hapkem
2019. *Natural Language Processing in Action*. Shelter Island, NY: Manning Publication Co.