

Katy Machine Learning Intro to NLP, Language Models, and Sentiment Analysis

K. Scott Ferguson, Ph.D.

<https://github.com/ksferguson/home/katyml>

8/11/2018

Natural Language - A Few Example Phenomena

► Movement

- (1)
- a. Mary saw the ball behind the tree.
 - b. What did Mary see t_{what} behind the tree?
 - c. Who does John think that Mary told Bill that she saw t_{who} at the reunion?
 - d. *What do you wonder whether Mary saw t_{what} behind the tree?

► Idioms

- (2)
- a. John kicked the bucket. (idiomatically 'John died')
 - b. John kicked the bucket over. (no idiomatic reading)
 - c. Jon re-kicked the bucket. (no idiomatic reading)

Natural Language - A Few More Examples

► Particles

- (3)
- a. John threw the ball up.
 - b. John threw up the ball.
 - c. The soldiers took up arms (to beat off the attack).
 - d. John re-threw the ball.
 - e. *John re-threw the ball up.

► Binding

- (4)
- a. Mary_i saw herself_i in the mirror.
 - b. Mary_i saw her_j in the mirror.
 - c. *She_i thinks highly of Mary_i.

Basic NLP

- ▶ Library: spaCy <https://spacy.io/>
- ▶ Example notebook:
<https://nbviewer.jupyter.org/github/ksferguson/home/blob/master/katyaml/spacy.ipynb>

Transfer Learning with Computer Vision

- ▶ Example notebook: <https://nbviewer.jupyter.org/github/ksferguson/home/blob/master/katyaml/fastai-MC-Dog-Breeds.ipynb>

Sequence Models

- ▶ Understanding LSTMs <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>
- ▶ Sequence Models: http://ttic.uchicago.edu/~shubhendu/Pages/Files/Lecture11_flat.pdf

Language Models

- ▶ Lang Models Blog - <https://thegradient.pub/nlp-imagenet/>
- ▶ ULMFit Blog - <http://nlp.fast.ai/classification/2018/05/15/introducing-ulmfit.html>
- ▶ OpenAI Transformer Paper - https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language_understanding_paper.pdf
- ▶ ELMo Paper - <https://arxiv.org/abs/1802.05365>

Sentiment Analysis

- ▶ Example notebook: https://nbviewer.jupyter.org/github/ksferguson/home/blob/master/katyaml/lang_model_sentiment.ipynb

Further Resources NLP

- ▶ NLP SOTA (state of the art)
`https://github.com/sebastianruder/NLP-progress/blob/master/sentiment_analysis.md`
- ▶ NLP News: `http://newsletter.ruder.io/`