Better NLP 2.0:
one library rules them all!

29th June 2019, AI Labs,
DS for IoT meetup

Get the slide deck (just now) at

http://bit.ly/better-nlp-launch

(folder: presentations/29-Jun-2019)

About me

- Freelance Software, Data Engineer
- Interests: code quality, testing, performance, AI/ML, NN, etc...
- Strengthening teams and helping them _accelerate_
- Various developer communities: LJC, and others
- Attends and helps co-organise various events and conference

Disclaimer

- Sharing my ideas, but *YMMV*
- Possibly missed one or more things or made mistakes, *I ask for forgiveness*
- Not clear for some, *my apologies*
- Lots of info and resources
- Lots of takeaway, please go home and do some more research
- Please contribute and share, please tweet!

Thank you

- DS for IoT meetup
- Sponsor(s) for the event
- Guests and attendees
- Anyone else not named…

Agenda

- Background - previous presentation
- Inspiration
- NLP stuff
- Demo (show code)
- Resources
- Thanks
- Contact & feedback, Q &A

Background

- **AI Labs** initiative - *ask* **Ajit** *for more details*

- Learning DL from *Machine Learning is Fun!*

Background

See presentation from 9th March 2019

Inspiration

Top-down approach, OpenAI and fast.ai

NLP: libraries not yet covered…

- Many parts of SpaCy, Texacy and Rasa NLU, FastText
- Pytext
- Vowpal Wabbit
- StanfordNLP
- and many others

See extensive list of NLP resources

NLP: what is covered…

- Some aspects of Spacy and Texacy
- NLP concepts
    - Extract entities
    - Extract nouns (Proper, Common, etc..)
    - Gather facts from text
    - Obfuscate text
    - Parts-of-speech
    - Summarisations (multiple methods)

NLP: what is NOT yet covered...

Sentiment analysis | Relevance ranking | Taxomony | Bag-of-words | Text classification | Efficient Text Classification | Word representation learning | Enriching Word Vectors with Subword Information | Obtaining word vectors for out-of-vocabulary words | Sequence taggers | Joint intent-slot model LDA (Latent Dirichlet Allocation) or Topic Modeling | Dependency parser | Training domain-specific models | Text matcher...

NLP: what is NOT yet covered... (continued)

Learning word, sentence or document level embeddings | Metric/similarity learning | Content-based or Collaborative filtering-based Recommendation | Embedding graphs | Image classification, ranking or retrieval | Annotate and resolve coreference clusters | Contextual intent-slot models | Date matcher | Spell checking | Pretrained Model | Transfer learning | n-gram search | Word2Vec | WordNet | Vector space model | Clustering | SVM and many more...

Demo time!

Show and tell!

PyPi

Next: *publish* library on PyPi!

Jupyter notebooks

Now available!

Others are following suite...

Flair (by Zalando Research),

Pythia (by Facebook Research),

many others...

Vision and future!

Continue working on it...

Summary

- Little inspiration leading to bigger things
- Toy examples turning into notebooks
- Others have followed suite
- Checklist: so many features to add…
- Share and contribute on <u>github</u>: *create your first Pull Request on NLP, on AI, on ML*

Resources

- Source code of Better NLP
- Awesome AI, ML, DL resources
- Machine Learning is Fun site | Medium site

Thank you, again…

- DS for IoT meetup
- Sponsor(s) for the event
- Guests and attendees
- Anyone else not named…

Feedback

Please share your feedback,

@theNeomatrix369

to be applied to the slides for
everyone's benefit