# Text and Deep Learning 101

Brendan Herger, hergertarian.com
Slides: https://goo.gl/XVu3NY
Code: https://github.com/bjherger/spoilers_model

https://xkcd.com/1425/

Intro
Word Models
Letter Models
Case Study: Spoilers
Recap

# Intro

# Intro

**What is NLP?**

- Natural language processing is the area of ML / AI focused on human languages (e.g. English or Mandarin)

**Why DL + NLP?**

- DL can handle high-dimensionality data with complex, non-linear relationships

- DL can map words into real-valued vectors

**Why now?**

- The impossible is becoming possible (compute, techniques & data)

Word Models

# Tokenization / preprocessing

- **Tokenize:** Convert one long string into 'words'

- **Lemmatize:** Normalize words (e.g. running -> run, cats ->cat)

- **Pad:** Convert input into fixed length, by truncating or padding

# Tokenization / preprocessing

"Running can't be fun"

- **Tokenize:** [running, can, 't, be, fun]

- **Lemmatize:** [run, can, not, be, fun]

- **Pad:** [run, can, not, be, fun, ⍰, ⍰, ⍰, …]

# Architectures

- **Embedding:** Converts tokens to numerical vectors (Word2Vec)

  - Unseen words replaced w/ 'UNK'

- **Convolutional:** Similar to computer vision

- **RNN:** Able to 'read' document, one word at a time. Basic RNN, LSTM or GRU, generally bi-directional.

- **Output:** Whatever output layer(s) you want

# Frameworks

**Deep learning**

- torchtext: PyTorch's NLP data loaders

- keras: Treats text as 1D time series

**NLP**

- spacy: Common framework for lemmatization, part of speech extraction

- nltk, CoreNLP (java, python interace), OpenNLP (java)

# Word Models

Tokenization / preprocessing
Architectures
Frameworks

# Letter Models

**general discussion |** I just thought; Phasma must be obsessing over killing finn, not only because he betrayed the first order but hes also the only person (along with chewy) that knows about what she done on Starkiller base

(self.StarWars)

169

submitted 3 months ago by Regijack

39 comments    share    save    hide    give gold    report

# Letter Models

- **Preprocessing:** Longer padded sequences, fixed vocabulary (check your encoding!)

- **Architectures:** More / larger convolutions, due to larger sequence length. Otherwise the same

- **Frameworks:** Same

# Case Study

**HOT**    NEW    RISING    CONTROVERSIAL    TOP    GILDED    WIKI

**Join our Discord! | Question? Check our FAQ | Upcoming Discussion Threads**

84

**meta | Updated Rules (See Sidebar for Wiki Link)**
submitted 6 days ago by BlueSmoke95  **Mandalorian Enforcer**  - stickied
63 comments   share   save   hide   report
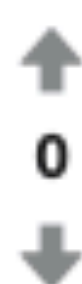
3972

**movies | Solo: A Star Wars Story trailer is here!**
submitted 7 hours ago by Jimmymgs  - stickied
1854 comments   share   save   hide   report

1

40.2k

**fun |** whoever manages the Star Wars twitter account deserves a raise.
submitted 10 hours ago by RealRipstar
847 comments   share   save   hide   report

(imgur.com)

r/StarWars · Posted by u/GregNextGen 19 hours ago

spoilers **Why did the ending of The Last Jedi feel so empty? (spoilers)**
spoiler

The final scene where the little boy is going "Luke Skywalker!" and it's trying to bring Luke back up to that legend status that Rian Johnson so thoroughly tore down. I don't know how or why word of Luke's last stand has spread, you'd think the First Order would want to cover that up. But when that scene came up, I just had a.... frown on my face. I did not smile at all. I loved Luke, he was my favorite character and just... this was supposed to be Rian Johnson's attempt to bring Luke back to being that legend, but why did it feel so empty?

💬 67 Comments    ➤ Share    🔖 Save    ✪ Give Gold    ⊘ Hide    ⚑ Report                48% Upvoted

**r/StarWars** · Posted by u/GregNextGen 19 hours ago

0

spoilers

spoiler

Why did the ending of The Last Jedi feel so empty? (spoilers)

The final scene where the little boy is going "Luke Skywalker!" and it's trying to bring Luke back up to that legend status that Rian Johnson so thoroughly tore down. I don't know how or why word of Luke's last stand has spread, you'd think the First Order would want to cover that up. But when that scene came up, I just had a.... frown on my face. I did not smile at all. I loved Luke, he was my favorite character and just... this was supposed to be Rian Johnson's attempt to bring Luke back to being that legend, but why did it feel so empty?

67 Comments     Share     Save     Give Gold     Hide     Report     48% Upvoted

**r/StarWars** · Posted by u/agentanthony 9 days ago

General Discussion **Prediction: Leia will not die in Episode IX**

Losing our beloved Princess Carrie Fisher obviously changed the outline of our Star Wars universe. But I don't think she will take the easy way out in Episode IX. Of course so many people are expecting a scene were maybe she's looking out a window before her ship explodes or her death will be mentioned in the opening crawl. Nope. I think she will have the ultimate tribute and live in the Star Wars universe forever. They already said filmed scenes are being used. Since the footage is limited, her part probably won't be big, but Episode IX is not going to be like Rogue One where every hero dies in the end. She will live.

💬 43 Comments    ➤ Share    🔖 Save    ⭐ Give Gold    ⊘ Hide    🚩 Report                    78% Upvoted

24

General Discussion **Prediction: Leia will not die in Episode IX**

Losing our beloved Princess Carrie Fisher obviously changed the outline of our Star Wars universe. But I don't think she will take the easy way out in Episode IX. Of course so many people are expecting a scene were maybe she's looking out a window before her ship explodes or her death will be mentioned in the opening crawl. Nope. I think she will have the ultimate tribute and live in the Star Wars universe forever. They already said filmed scenes are being used. Since the footage is limited, her part probably won't be big, but Episode IX is not going to be like Rogue One where every hero dies in the end. She will live.

43 Comments    Share    Save    Give Gold    Hide    Report                    78% Upvoted

# Letter Models

**Input**

• How old is Chewy?

**Converted text into lower case, removed non-standard characters**

• how old is chewy?

**Padded / truncated to 2000 characters**

• [h, o, w,  , o, l, d,  , i, s,  , c, h, e, w, y, ?, ▯, ▯, ▯, …]

# Recap

Word models
Letter models
Case Study: Spoilers

# Resources

- NLP Whitepaper, by Yoav Goldberg

- Deep Learning, book by Ian Goodfellow and Yoshua Bengio and Aaron Courville

- Introduction to LSTMs, by Christopher Olah

- LSTM / GRU intro & comparison

# Thanks!

Brendan Herger, hergertarian.com
Slides: https://goo.gl/XVu3NY
Code: https://github.com/bjherger/spoilers_model

# Thanks!

Brendan Herger, hergertarian.com
Slides: https://goo.gl/XVu3NY
Code: https://github.com/bjherger/spoilers_model

# Questions?

Brendan Herger, hergertarian.com
Slides: https://goo.gl/XVu3NY
Code: https://github.com/bjherger/spoilers_model