

Machine Learning, Distant Reading, and American Literature

NeMLA 2016, Hartford, CT

James E. Dobson

Dartmouth College

Talk Outline

1. Presuppositions and Theory
2. Annotation
3. Text Mining
4. Sample Assignment



Presuppositions

- ▶ Interpretation is our key activity.
- ▶ Close reading is one of many ways of generating evidence.
- ▶ Computational approaches produce evidence.
- ▶ We cannot separate method (“data” collection) from interpretation.
- ▶ Interpretation in the humanities is a nihilistic hermeneutics.



Presuppositions

- ▶ Interpretation is our key activity.
- ▶ Close reading is one of many ways of generating evidence.
- ▶ Computational approaches produce evidence.
- ▶ We cannot separate method (“data” collection) from interpretation.
- ▶ Interpretation in the humanities is a nihilistic hermeneutics.



Presuppositions

- ▶ Interpretation is our key activity.
- ▶ Close reading is one of many ways of generating evidence.
- ▶ Computational approaches produce evidence.
- ▶ We cannot separate method (“data” collection) from interpretation.
- ▶ Interpretation in the humanities is a nihilistic hermeneutics.



Presuppositions

- ▶ Interpretation is our key activity.
- ▶ Close reading is one of many ways of generating evidence.
- ▶ Computational approaches produce evidence.
- ▶ We cannot separate method (“data” collection) from interpretation.
- ▶ Interpretation in the humanities is a nihilistic hermeneutics.



Presuppositions

- ▶ Interpretation is our key activity.
- ▶ Close reading is one of many ways of generating evidence.
- ▶ Computational approaches produce evidence.
- ▶ We cannot separate method (“data” collection) from interpretation.
- ▶ Interpretation in the humanities is a nihilistic hermeneutics.



Digital Annotation

- ▶ Standardize digitized texts
- ▶ Simultaneous readings
- ▶ Both public and private annotation
- ▶ Reading outside of and during class
- ▶ Promotes variety of reading methods
- ▶ High-quality data for analysis



Lacuna Stories

- ▶ Developed by Stanford
- ▶ Open Source
- ▶ Drupal-based
- ▶ Extensible authentication (LTI, CAS)
- ▶ Scaled to 2,500 students
- ▶ Supports tens of thousands of annotations



Learning from Annotation

in the grain, and impregnates the waters of the world, is caught by man and sinks into his soul. The moral influence of nature upon every individual is that
Who can estimate this? Who can guess how much firmness
fisherman? how much tranquillity has been reflected
unspotted deeps the winds forevermore drive flocks of stars
how much industry and providence and affection we have
What a searching preacher of self-command is the varying
Herein is especially apprehended the unity of Nature,
everywhere. All the endless variety of things make an idea
in his old age, that, look where he would, all things haste
the same entity in the tedious variety of forms. The fab
drop, a crystal, a moment of time is related to the whole, and partakes of the perfection of the whole.
Each particle is a microcosm, and faithfully renders the likeness of the world.

The world is reflected in small bits of nature. You can find truth in the world in some of nature's objects

Comment
Everyone
SOldfield

Missing text: Every particular in nature:

Information
Everyone
JCMorrison68



Text Mining: Sample Session (65 Minutes)

1. Quick exploration
2. Text Statistics
 - 2.1 Total words and lines
 - 2.2 Unique words
 - 2.3 Top Terms
 - 2.4 Top Phrases
3. KWIC (Keyword in Context)
4. Text Searching
5. Collocations
6. Word similarity
7. Topic Modeling (tf-idf)
8. Parts of Speech Tagging (locations, people)
9. Advanced Searching



- ▶ Python wrapper around two popular tools (nlk & sklearn)
- ▶ Open Source and Available: github.com/jeddobson/nlktool
- ▶ Extensible through functions
- ▶ Designed to support teaching and discovery

- ▶ Python wrapper around two popular tools (nlk & sklearn)
- ▶ Open Source and Available: github.com/jeddobson/nlktool
- ▶ Extensible through functions
- ▶ Designed to support teaching and discovery

- ▶ Python wrapper around two popular tools (nltk & sklearn)
- ▶ Open Source and Available: github.com/jeddobson/nltktool
- ▶ Extensible through functions
- ▶ Designed to support teaching and discovery



- ▶ Python wrapper around two popular tools (nltk & sklearn)
- ▶ Open Source and Available: github.com/jeddobson/nltktool
- ▶ Extensible through functions
- ▶ Designed to support teaching and discovery

Word Similarity

Sample Execution

```
[jed@punctum]$ nltktool -f Moby-Dick.txt similarity -t 'monomania'  
opening file: Moby-Dick.txt  
intention lakeman bearing duty all willingness escape mind spars  
commander knife ended comrades deliverance have stern bones hell ship  
crew
```

Sample Execution

```
[jed@punctum]$ nltktool -f Moby-Dick.txt collocations
```

```
opening file: Moby-Dick.txt
```

```
Sperm Whale; Moby Dick; White Whale; old man; Captain Ahab; Right  
Whale; sperm whale; Aye aye; Captain Peleg; Mr. Starbuck; New Bedford;  
Cape Horn; 'ye see; years ago; cried Ahab; lower jaw; never mind;  
chief mate; Mrs. Hussey; Father Mapple; white whale; ivory leg; cried  
Stubb; St. George; n't know
```

Moby-Dick Topic Model (ngram)

Topic #0: sperm whale,great sperm,sperm whale head,whale head,beneath surface,lower jaw,whale line,mortal man,greenland whale,whale fishery,feet length,white whale,looks like,guernsey man,general thing,good bye,good time,mast head,main mast,loose fish

Topic #1: white whale,old ahab,st thou,ha ha,aye aye,mr starbuck,cried ahab,ye ye,look ye,sperm whale,moby dick,guernsey man,young man,greenland whale,ivory leg,great sperm,good time,half way,like old,king post

Topic #2: moby dick,thee thou,lower jaw,snow white,thou st,cried ahab,feet length,captain ahab,white whale,great sperm,good time,guernsey man,ha ha,half way,ivory leg,good bye,greenland whale,life buoy,king post,like old

Topic #3: old man,cried starbuck,ere long,st thou,cruising ground,king post,life buoy,dost thou,board pequod,ahab did,aye sir,art thou,ivory leg,half way,ha ha,guernsey man,like old,look ye,look outs,great sperm

Topic #4: captain ahab,ahab did,ahab stood,young man,thou hast,dost thou,good bye,oh thou,st thou,stubb flask,starbuck stubb,board pequod,aye aye,mr starbuck,art thou,mr stubb,good time,captain bildad,loose fish,town ho

Topic #5: right whale,whale head,greenland whale,stubb flask,sperm whale head,starbuck stubb,feet length,good time,loose fish,town ho,fast fish,like old,life buoy,general thing,good bye,mast head,main mast,great sperm,lower jaw,looks like



Sample Assignment

Assignment Prompt:

- ▶ Use evidence obtained from a topic model, collocation, or similarity search to make an argument about a text that you have already read and annotated in Lacuna Stories.
- ▶ Produce your interpretive claim from this evidence.
- ▶ Reflect on the possible significance of this evidence.



Acknowledgements

Thank You

- ▶ Erin DeSilva (Dartmouth)
- ▶ Michael Goudzwaard (Dartmouth)
- ▶ Brian Johnsrud (Stanford)
- ▶ Michael Widner (Stanford)
- ▶ DartmouthX