

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



Lacuna Stories

For nonconformity the world whips you with its displeasure. And therefore a man must know how to estimate a sour face. The by-standers look askance on him in the public street or in the friend's parlour. If this aversion had its origin in contempt and resistance like his own, he might well go home with a sad countenance; but the sour faces of the multitude, like their sweet faces, have no deep cause, but are put on and off as the wind blows and a newspaper directs. Yet is the discontent of the multitude more formidable to a man who knows the world than to a young man in the college. It is easy enough for a firm man to stand up to the rage of the masses. Their rage is decorous and prudent, for they are timid as mice. But when to their feminine rage is added the indignation of the people, it is a terrible force that no man can withstand. The habit of magnanimity and reliance on the future is a reverence for our past act or word, because the eyes of others are loath to disappoint them.

But why should you keep your memory, lest you contradict yourself; why should you contradict yourself; why should you stand alone, scarcely even in acts of piety, and live ever in the eyes of the Deity; yet when the devout should clothe God with shape and harlot, and flee.

A foolish consistency is the hobgoblin of little minds, and the great ones do not die, but are buried. With consistency of shadow on the wall. Speak what you think now in hard words, and to-morrow think in hard words again, though it contradict every thing you said to-day. —'Ah, so you shall be sure to be misunderstood.' — Is it so bad, then, to be misunderstood? Pythagoras was misunderstood, and Socrates, and Jesus, and Luther, and Copernicus, and Galileo, and Newton, and every pure and wise spirit that ever took flesh. To be great is to be misunderstood.

To me, this states that people can change and that that is normal.

Response
Everyone
Ernest24

don't be complacent...don't be lulled into the familiar for his secure consistency that is the demon

Comment
Everyone
JDuo

Here is another passage, like the opening paragraph, where Emerson appeals to past "genius" minds in order to further his argument that every mind is capable of genius. His writing is rife with references like these.

Comment
Everyone
Nicole_Saraviera

alissa

7



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 g
Who can estimate this? Who can guess how much firm
fisherman? how much tranquillity has been reflected
unspotted deeps the winds forevermore drive flocks of st
how much industry and providence and affection we ha
What a searching preacher of self-command is the varyin

Herein is especially apprehended the unity of Nature,
everywhere. All the endless variety of things make an ide
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 (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



- ▶ 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

