Abstract Representations of Plot Structure

Micha Elsner

Department of Linguistics
The Ohio State University

November 1, 2013

We have good models for short articles...
Less research on ***storytelling***.

A ***use of language*** we don't understand...
Lots of ***data*** our tools don't cover.

Challenges

Storytelling is **typical** language use...

- But still little formal understanding of what a story ***is***!

McKeown NAACL plenary: an emerging challenge for CL

- Linguistic: what are good representations?
- Applied: classic NLP tools for stories
- Digital humanities: corpus insight on literary problems
- Sociology: writing as cultural behavior
  - $>$500k Harry Potter fan stories on fanfiction.net alone!

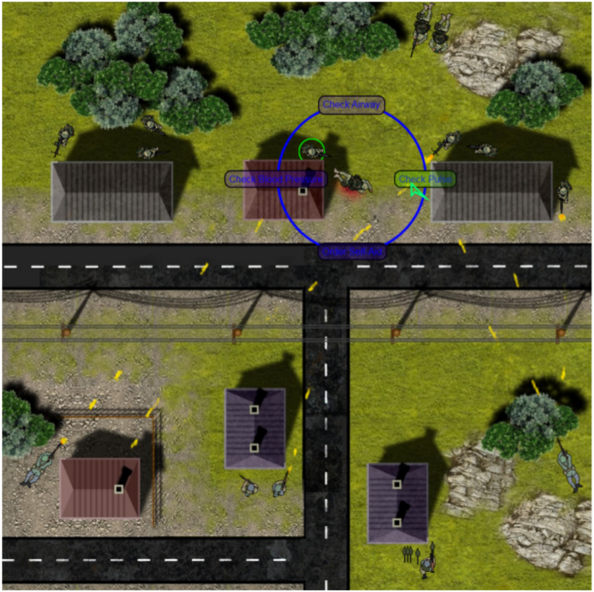
Applications

Applied reasons to study stories:

- Generating stories (games, education, etc.)
- Searching and recommendations
- Summarization

Story generation: games and training

Zook, Lee-Urban, Riedl et al 2012



When good summarizers go bad...

> ...follows the main character Elizabeth Bennet as she deals with issues of manners, upbringing, morality, education and marriage... (Wikipedia)

> The story turns on the marriage prospects of the five daughters of Mr. and Mrs. Bennet... (Amazon.com)

When good summarizers go bad...

...follows the main character Elizabeth Bennet as she deals with issues of manners, upbringing, morality, education and marriage... (Wikipedia)

The story turns on the marriage prospects of the five daughters of Mr. and Mrs. Bennet... (Amazon.com)

“Bingley.” Elizabeth felt Jane’s pleasure. “Miss Elizabeth Bennet.” Elizabeth looked surprised. “FITZWILLIAM DARCY” Elizabeth was delighted. Elizabeth read on: Elizabeth smiled. “If! “Dearest Jane! (Jason Huff: Microsoft Word ‘08)

What doesn’t work: sentences

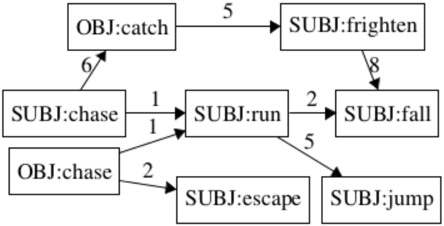
Why doesn’t the Microsoft summarizer work?

(cf Kazantseva+Szpakowicz)

- Extractive sentence selection
- Works ok in news: topic sentences cover the main idea
- Not so good for novels

What might work: event networks

- Schank scripts and (somewhat) Propp functions lead to:
- Chambers+Jurafsky event schemas
- Which lead to a variety of narrative representations
  - (Finlayson, recent Riedl, McIntyre+Lapata)



(McIntyre+Lapata)

Event networks are still pretty low-level

The emperor rules the kingdom: a fable

The emperor rules the kingdom. The kingdom holds on to the emperor. The emperor rides out of the kingdom. The kingdom speaks out against the emperor. The emperor lies.
–McIntyre and Lapata, ACL 10

- Does the story have a moral?
- Is there a hero? A villain?
- Is there a plot?
- What *is* a plot anyway?

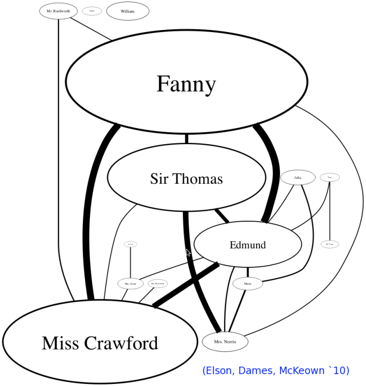
Literary theory

- Propp: characters have *functions*
  - Hero, villain, mentor...
- Crane and subsq.: plot creates moral and emotional involvement with the characters

> for some of the characters we wish good, for others ill, and depending on our inferences as to the events, we feel hope or fear, pity or satisfaction...
> —R. S. Crane

Plot is *high-level*...

Two basic insights:



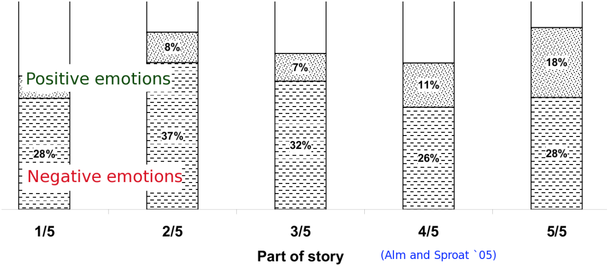
(Elson, Dames, McKeown `10)

Characters...
forming a *social network*

(Elson+al ‘10)

Plot is *high-level*...

Two basic insights:

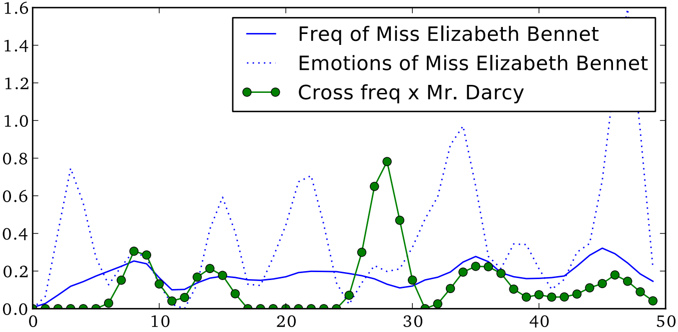


Story has an *emotional trajectory*

(Alm+Sproat ‘05)

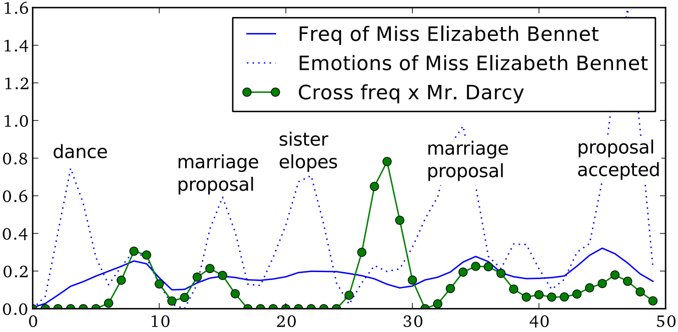
Combine the two:

- Compute a trajectory for each character
- Observe social relationships through time



Combine the two:

- Compute a trajectory for each character
- Observe social relationships through time



Preprocessing

- Chop the novel into paragraphs
- Parse everything and retrieve proper NPs
- Simple coreference on the NPs to find characters
- Emotion: “strong sentiment” cues from (Wilson+al ‘05)

Coreference

Similar to cross-document coreference:

- Shared name elements
- Presence in same documents
- List of gendered names and titles

Character (longest name)	gender	count
Mrs. George Osborne	F	662
Georgy Osborne	N	344
Capt. George Osborne	M	153
Mr. Osborne	M	146
Miss Jane Osborne	F	75
Master George	M	8
Mr. George	M	7
Lt. Osborne	M	7

Use this representation to measure similarity...

Kernel function

$k(x, y)$: similarity between x and y

0: no similarity; $>$ 0: more similar

basic ML building block

Use this representation to measure similarity...

Kernel function

$k(x, y)$: similarity between x and y

0: no similarity; $>$ 0: more similar

basic ML building block

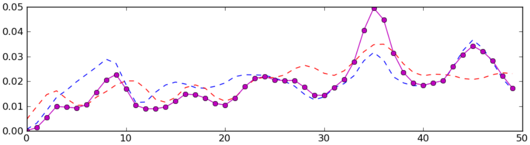
Use *convolution theorem* (Haussler '99) to build a complex kernel out of simpler ones:

$$k(x, y) = \sum_{ch_1 \in X} \sum_{ch_2 \in Y} \underbrace{c(ch_1, ch_2)}_{\text{kernel over characters}}$$

Similarity between characters

$e(ch_1, ch_2)$:

- Similarity for trajectory curves
- Normalized integral of the product
- Used for frequency and emotion



$d(ch_1, ch_2)$

- Nearby words

replied Elizabeth	17
Elizabeth felt	14
Elizabeth looked	10
Elizabeth's mind	7
...	

First-order character kernel

$$c_1(ch_1, ch_2) = d(ch_1, ch_2)e(ch_1, ch_2)$$

Adding social network features

Characters are more similar if:

- They each have close friends...
  - (Measured by co-occurrence frequency)
- ...who are also similar

Second-order character kernel

$$c_2(ch_1, ch_2) = c_1(ch_1, ch_2) \sum_{u' \in X} \sum_{v' \in Y} \underbrace{e(\widehat{u, u'}, \widehat{v, v'})}_{\text{relationship strength}} c_1(u', v')$$

Testing similarity

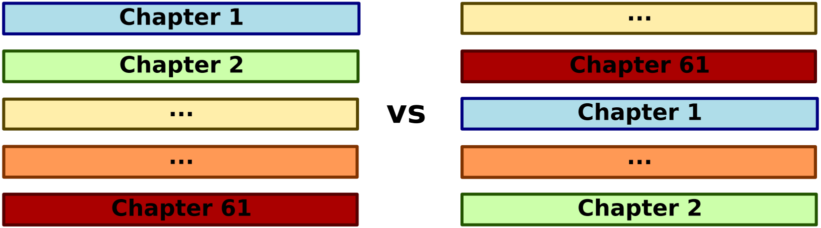
- First, simple proof of concept
- Independent of particular critical theory
- Difficult for very naive models

Testing similarity

- First, simple proof of concept
- Independent of particular critical theory
- Difficult for very naive models

Order discrimination

(Karamanis+al '04) (Barzilay+Lapata '05)



Results

	Random perm	Reversed

Results

	Random perm	Reversed
Whole-novel traj.	46	52

Results

	Random perm	Reversed
Whole-novel traj.	46	52
First-order k_1	60	51

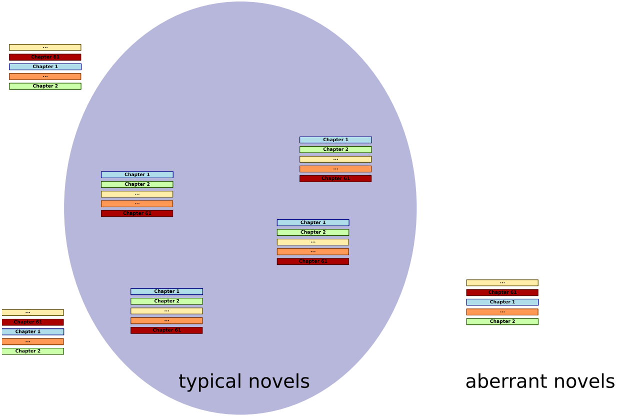
Results

	Random perm	Reversed
Whole-novel traj.	46	52
First-order k_1	60	51
Second-order k_2	**62**	**52**

Weighted nearest-neighbor

For training set T, is:

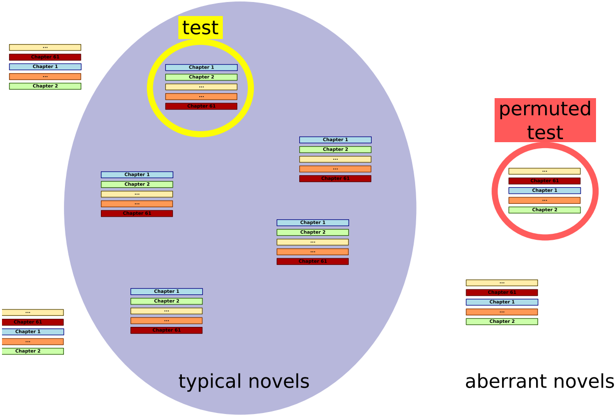
$$\sum_{t \in T} k(t, y) > \sum_{t \in T} k(t, y_{perm})?$$



Weighted nearest-neighbor

For training set T, is:

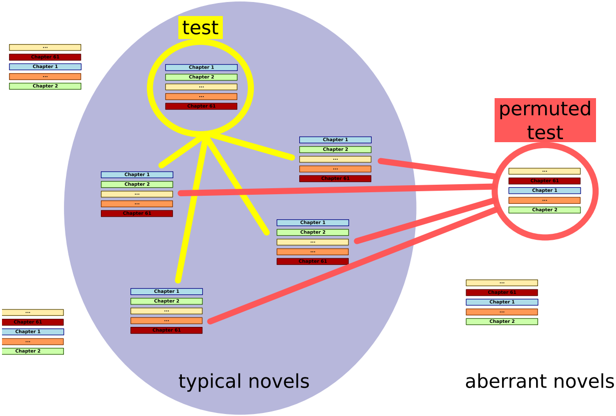
$$\sum_{t \in T} k(t, y) > \sum_{t \in T} k(t, y_{perm})?$$



Weighted nearest-neighbor

For training set T, is:

$$\sum_{t \in T} k(t, y) > \sum_{t \in T} k(t, y_{perm})?$$



Results

(30 19th.c novels from Project Gutenberg)

Binary classifications

Chance accuracy 50%
Significance via kernel-based non-parametric test (Gretton+al '07)

	Random perm	Reversed

Results

(30 19th.c novels from Project Gutenberg)

Binary classifications

Chance accuracy 50%

Significance via kernel-based non-parametric test (Gretton+al '07)

	Random perm	Reversed
Whole-novel traj.	50	53

Results

(30 19th.c novels from Project Gutenberg)

Binary classifications

Chance accuracy 50%
Significance via kernel-based non-parametric test (Gretton+al '07)

	Random perm	Reversed
Whole-novel traj.	50	53
First-order k_1	77	63

Results

(30 19th.c novels from Project Gutenberg)

Binary classifications

Chance accuracy 50%
Significance via kernel-based non-parametric test (Gretton+al '07)

	Random perm	Reversed
Whole-novel traj.	50	53
First-order k_1	77	63
Second-order k_2	**90**	**67**

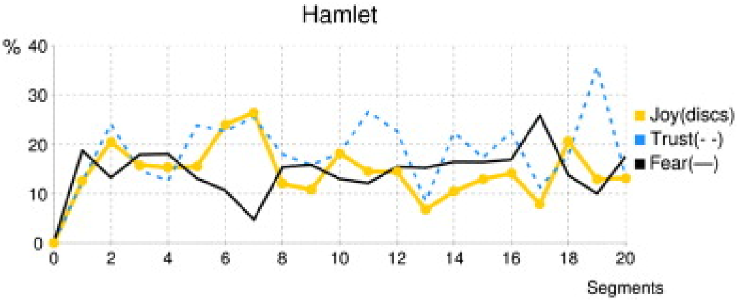
Improvements

Previous model was a little thrown-together...
In this section:

- More systematic investigation of features
- Some *attempts* at parameter tuning
- Results improve...
- ...but some frustrating issues remain

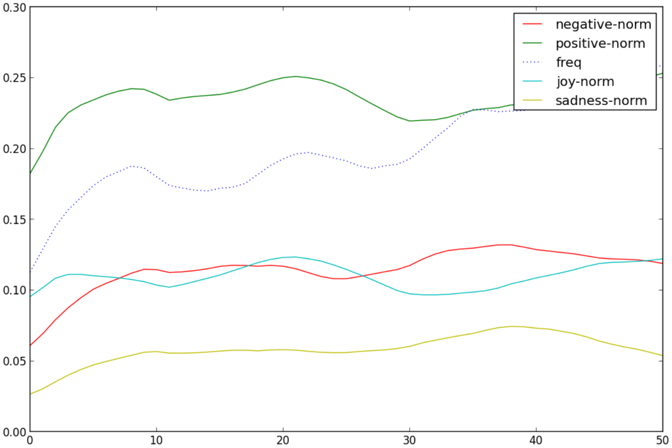
Better sentiment

- Earlier system used Wilson+al sentiment: *negative* vs *positive*
- Try Mohammad sentiment: 10 emotional classes:
- Anger, Anticipation, Disgust, Fear, Joy, Positive, Negative, Sadness, Surprise, Trust



(Mohammad '12)

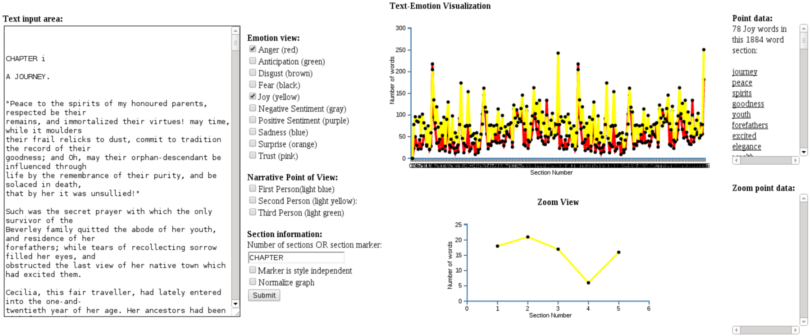
Character-specific version: Elizabeth Bennet again



A slight digression...

Investigated whether this would be useful as a writers' aid:

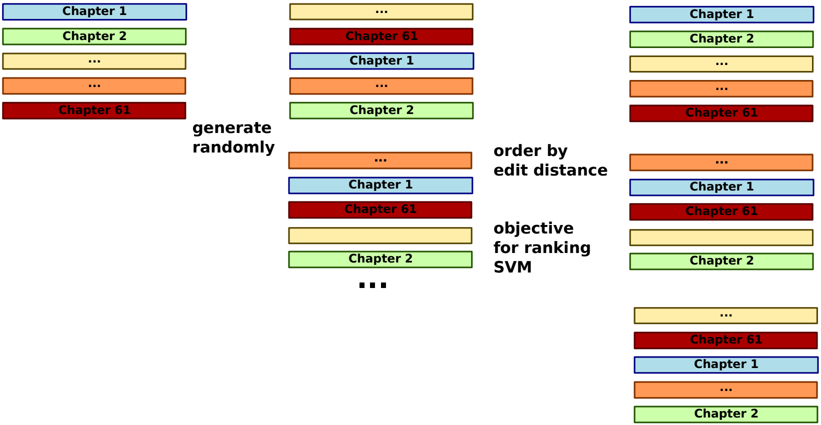
- MSc thesis: Robert Ang
- Cosponsored with U Edinburgh writer-in-residence Viccy Adams
- Writers found tool inspiring, but wanted sentence-level sentiment



Try it out! http://www.ling.ohio-state.edu/ melsner/sentivis/emotiongraph.html

Learning

Rank learning approach (Feng+Hirst '12)



Single-trajectory models

- Compute a single trajectory for the whole novel
- From above: this performed at chance!
- We'll see that this was due partly to feature issues

Results are equivocal: single-traj model

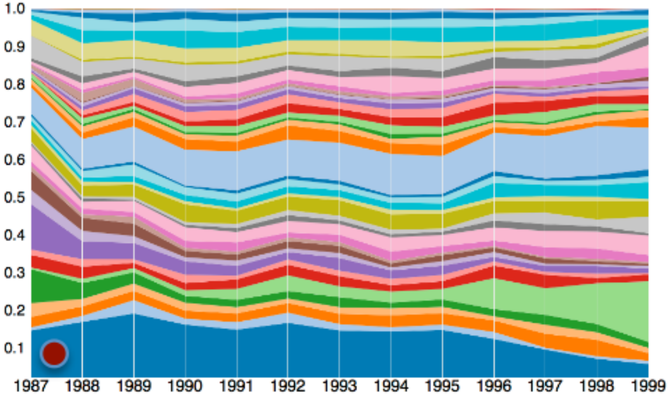
	Order	Rev
Anger	64	62
Anticipation	55	55
Disgust	62	53
Fear	71	65
Joy	51	51
Negative	65	60
Positive	60	65
Sadness	61	64
Surprise	60	55
Trust	60	53
Sentiment	57	65

Suggests dataset is very variable

Sentiment is useful— but dev parameters fit test poorly

Alternative approach: LDA topics

- Standard digital humanities way to do this...
- Also a variety of fancy models used on scientific journals



(Kim+Sudderth ‘11)

Our approach: vanilla LDA

- Fast, probably good enough as proof-of-concept
- 10 topics for comparison with Mohammad sentiment

0	man	reply	lady	gentleman	boy	head
1	love	life	heart	world	thought	soul
2	mother	father	woman	make	year	day
3	return	person	time	receive	give	place
4	character	world	men	feeling	opinion	mind
5	hand	eye	face	voice	speak	word
6	form	light	woman	air	pass	beauty
7	thing	make	letter	time	write	read
8	room	house	door	night	time	day
9	men	foot	man	round	horse	time

Results

	Order	Rev
Sentiment single-traj	57	65
LDA single-traj	67	89

Order and reverse are very different

- Order benefits from character features
- Reverse works better with single trajectory!
- Beginning and end vs overall coherence

Results

	Order	Rev
Sentiment single-traj	57	65
LDA single-traj	67	89
Sentiment characters	53	50
LDA characters	57	59

Order and reverse are very different

- Order benefits from character features
- Reverse works better with single trajectory!
- Beginning and end vs overall coherence

Results

	Order	Rev
Sentiment single-traj	57	65
LDA single-traj	67	89
Sentiment characters	53	50
LDA characters	57	59
Sentiment 2nd-ord	59	51
LDA 2nd-ord	60	58

Order and reverse are very different

- Order benefits from character features
- Reverse works better with single trajectory!
- Beginning and end vs overall coherence

Symmetrization by matching

- Kernel involves an *all-to-all* alignment
- In machine translation, often helps to have a symmetric alignment
  - IBM-1: If we learn to translate *Hund* as *dog*...
  - ...shouldn't we also learn *dog* is *Hund*?
- Various ways to enforce this
- (Matusov+al '04) compute a bipartite matching
- We'll do the same...
- Compute character-to-character scores
  - *Pride and Prejudice: Elizabeth* vs *Vanity Fair: Amelia*
- Use the best matching as the overall score

Results with matching

	Order	Rev
Sentiment single-traj	57	65
LDA single-traj	67	89
Sentiment characters	53	50
LDA characters	57	59
Sentiment 2nd-ord	59	51
LDA 2nd-ord	60	58

Results with matching

	Order	Rev
Sentiment single-traj	57	65
LDA single-traj	67	89
Sentiment characters	53	50
LDA characters	57	59
Sentiment 2nd-ord	59	51
LDA 2nd-ord	60	58
Sentiment matching	**81**	54
LDA matching	**72**	56
Sentiment (2-ord), matching	77	54
LDA (2-ord), matching	69	56

Matching helps!

- But second-order is no longer helpful...
- Maybe we need to apply the matching principle to relationships too?

WNN results:

	Order	Rev
Previous result (2-ord)	90	67
LDA single-traj	80	93
Sentiment matching	93	70

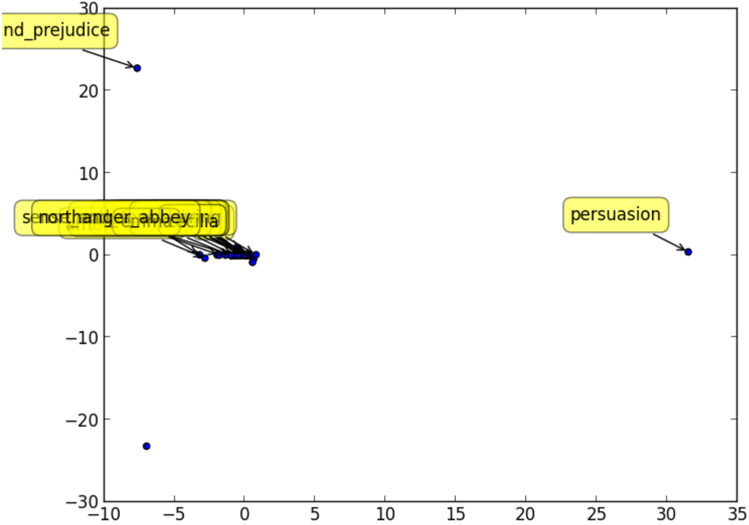
Textual similarity

What sorts of similarities is the system pulling out?

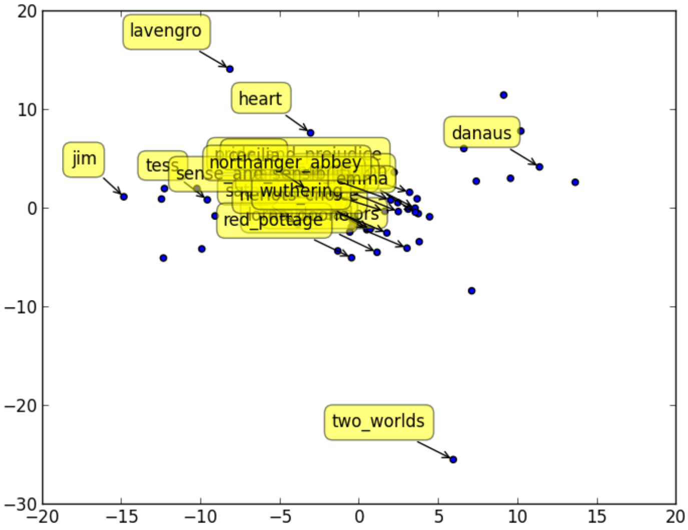
Nearest-neighbor clustering

- Convert distances to ranks
- Nearest-neighbor graph
- Spectral clustering (von Luxburg '06)

Single-traj (sentiment)

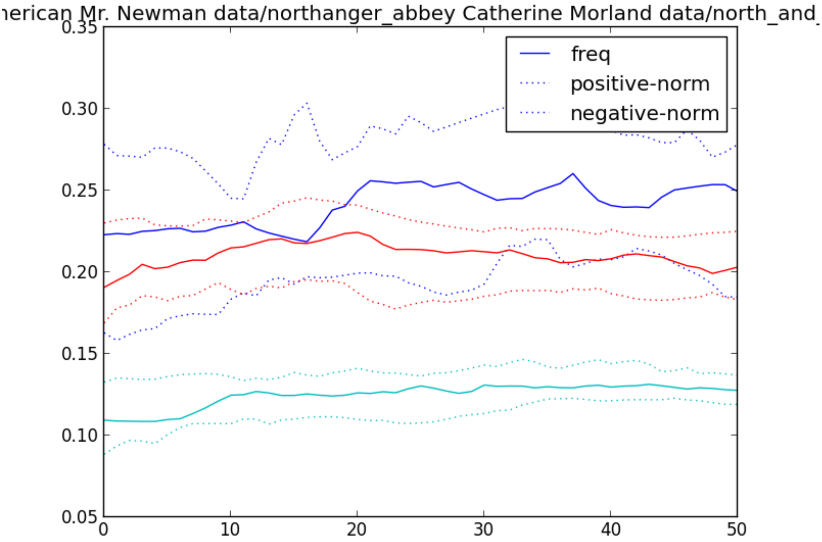


Characters (sentiment matching)



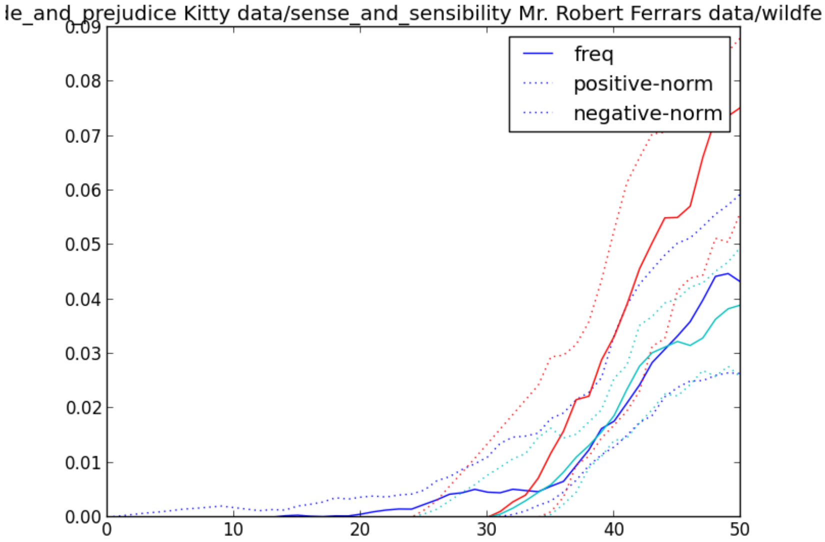
Character similarity

Main characters



Character similarity

Characters important to the climax of the novel



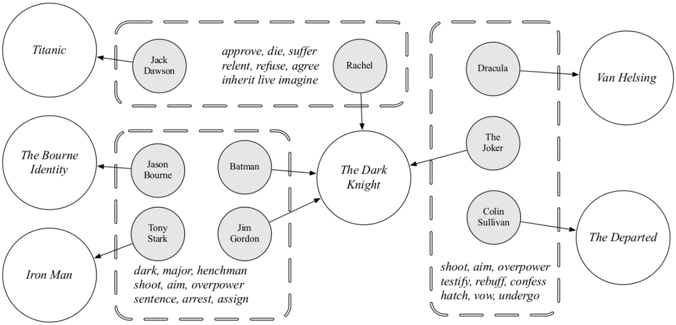
Related work

Recent paper:

Learning latent personas of film characters

Bamman, O'Connor and Smith

- Using summaries and IMDB metadata
- Bayesian model of character roles



Future work: from similarity to summarization

Insight from language and vision: (Mason+Charniak ‘13, Kuznetsova+al ‘12)

Original: Go all-out glam in the shimmering Dyeables Roxie sandals. Metallic faux leather upper in a dress thong sandal style with a round open toe. ...

Extracted: Shimmering snake-embossed leather upper in a slingback evening dress sandal style with a round open toe .

System: Shimmering upper in a slingback evening dress sandal style with a round open toe .

(Mason+Charniak ‘13)

Resource projection approach:

- Assumes resources (captions/summaries) for training
- Nearest-neighbor search for similar images in training set
- Extract captions from neighbors
- Use image features to edit caption

Conclusions

- Plot structure is complicated!
- Characters, emotions and events over time
  - Transfer from dev to test is complicated
- Simple ordering test as proof of concept
  - Reversals behave differently than randomized
- Analysis of clustered "roles"

Thanks: Robert Ang, Sharon Goldwater, Mirella Lapata, Victoria Adams, Rebecca Mason, Kira Mourão, Jon Oberlander

- Paper EACL 2012, followup in submission