.txtLAB @ mcgill

UNDERSTANDING NARRATIVE

Computational approaches to detecting narrative frames

Eva Portelance and Andrew Piper, McGill University

PROJECT GOAL

➔ Understanding narrative structure at a large scale remains a challenging problem within the field of cultural analytics and computational linguistics. Our aim with this project is to develop novel methods to study the pacing of narrative scene changes and the overall distribution of different plotlines within novels.

➔ Being able to analyze such narrative features at large scale can give us insights into the way different genres, time periods, or cultures favor different modes of storytelling. In this project we formalize definitions of narrative scenes and implement new methods of detection and clustering using computational methods.

ABSTRACT

➔ This poster presents:
- Our formalisation of narrative events;
- A description of our approach to the segmentation of narratives by frames and results presenting its accuracy;
- Results from predictive clustering of frames into larger-scale “plotlines” using different clustering models.

WHAT IS A NARRATIVE ?

➔ “...narrative is essentially a mode of verbal presentation and involves the LINGUISTIC recounting or telling of EVENTS.” *-Dictionary of Narratology*, by Gerald Prince, 2003

➔ Here, we interpret EVENT using a simple semantic definition:

The relationship between an AGENT of an event (VERB) and its THEME (Object)

- For it to be have a truth value, it must have a domain of discourse (D), or world:

D = ENTITIES x TIME x SPACE

WHAT MAKES A FRAME ?

➔ Classic def.: a story within a story,e.g. 1001 Nights
- frame narratives are more of a GENRE.

➔ To include ALL narratives, we need a broader definition. A parallel with scenes in theater:

scene change = disruption in the sequencing of events > FRAME BOUNDARY

TESTING OUT MODELS

DATA SET: 12 000 word passages from 9 novels/novellas: *Arabian Nights*, *Frankenstein* (1818), *Wuthering Heights* (1850), *A Tale of Two Cities* (1859), *The Kreutzer Sonata* (1890), *Heart of Darkness* (1899), *Ethan Frome* (1911), *Catch-22* (1961), *A Dance with Dragons* (2011)

BASELINE: As a baseline, we will use Hearst’s Texttiling Algorithm (1994; 1997) used for text segmentation into subtopics; originally tested on journalistic texts.

HUMAN PERFORMANCE: As an attainable goal, we would hope to attain a performance level comparable to that of Human annotators given the same task.

GOLD STANDARD: 3 human annotators were asked to annotate the 9 passages by hand for characters, setting and/or time change, as well as narrator change and narrative digression. If at least ⅔ of the annotators agreed on a frame location within +/-100 word margin, the average location of those annotation was considered to be in the gold standard. Over all passages this represented 208 frame boundaries.

OPERATIONALIZING THIS THOUGHT PROCESS

HYPOTHESIS: a frame is a sequence of events within its own domain of discourse. Therefore, a frame boundary would consist in a change in the three components of D.

➔ **NARRATIVE SEGMENTATION ALGORITHM:**
- Uses parts of speech (POS) to determine in what category a content word belongs.
- Compares sets of words from adjacent passages of a narrative.
- Records the level of dissimilarity over a determined threshold as frame boundaries.
- Combine predicted results from Character changes (Proper Nouns), Setting and Time changes (Contentful nouns excluding proper nouns, and all words excluding stopwords).

➔ **PLOTLINE CLUSTERING ALGORITHM:**
- Cluster predicted frames from segmentation algorithm using bag-of-words and clustering model.

RESULTS

VARIABLES:
- **Window size**: size of chunks of texts to compare for similarity.
- **Step size**: size of incrementation along word count.

TEXTTILING : BASELINE

Using the window and step size reported in Hearst (1994): step(w)=20, window=120 (k=6, k*w=120). +/-40 (2*w) word margin.

Precision	Recall	F1 Accuracy
0.18232044	0.190751445	0.186440678

*Note this version of texttiling did not normalise to the closest paragraph boundary, since annotators were not asked to do so either.

HUMAN AGREEMENT

Given the Gold Standard as defined above and a +/-100 word margin, The average human performance over all passages is:

Precision	Recall	F1 Accuracy
0.81556332	0.862222222	0.825064711

The average Kappa over all frames and 3 annotators is 0.61, with a lot of variation (Catch 22 -0.07 ; Frankenstein 0.95)

NARRATIVE SEGMENTATION

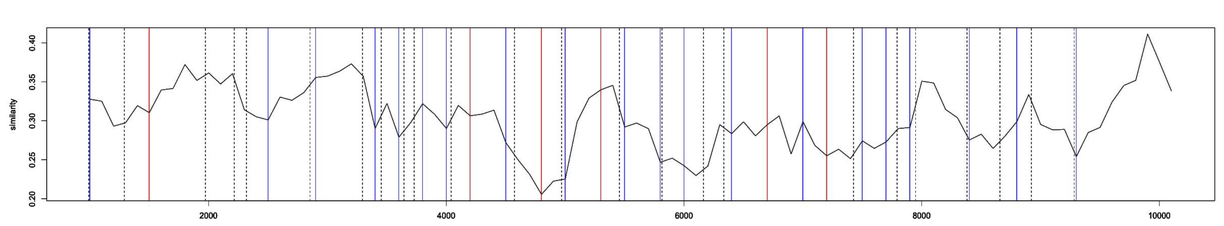
➔ Best performance:
- 1000 word window and a step of 100 word
- measure similarity using a cosine similarity test
- select local minimums under the 95 percentile confidence interval as frame boundaries
- COMBINATION of frame boundaries predicted from changes in the set of proper nouns (i.e. Entities) as well as those of all content words (exclude stop words), and all contentful nouns (excluding proper nouns) (i.e. Time and Setting). +/-200 (2*w) word margin.

Precision	Recall	F1 Accuracy
0.681372549	0.668269231	0.674757282

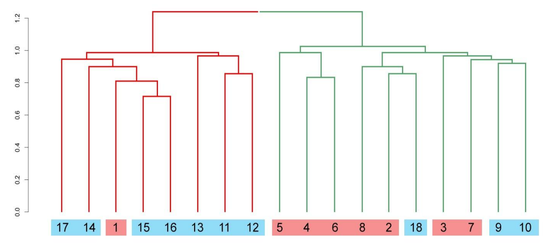
➔ A very simplified analysis of our initial thought process does in fact predict frame boundaries!
- We tried adding in pronoun resolution to increase proper noun counts, but the change in performance was negligible.

➔ **GRAPH:** The following represents the cosine similarity of the sets of contentful nouns within a 1000 word window at every step for words 1000 to 10000 of Conrad’s *Heart of Darkness* (1899). Additionally, it superimposes the Gold Standard frame boundary locations with those predicted by our algorithm.

BOUNDARIES
..... Gold
— Correct Prediction
— Over Prediction



PLOT DISTRIBUTION



➔ Clustering of predicted frames from R.R. Martin’s *A Dance with Dragons* (2011) into two distinct plotlines
- The coloured contours of frames indicate gold standard distribution of plotlines
- Uses Ward’s method of hierarchical clustering

➔ Average performance over all passages of matching every 100 word step to the correct subplot grouping using this model:

Precision	Recall	F1 Accuracy
0.916807001	0.648351648	0.75955665

➔ Best performance for Hierarchical clustering uses Ward’s method and bag-of-words.

NEXT STEP: Testing other clustering models.

THE END