Abstract

It is commonly asserted that the average amount of narrative time that elapses in a passage of fiction has decreased over the past several hundred years. But did this change in literary time really occur, and if so, how rapidly? Previous work on a small but carefully chosen set of hand-labeled passages finds that literary time becomes progressively shorter over the 19th and early 20th centuries. We explore computational methods for predicting literary time and apply these to a collection of more than 53,000 complete novels from the 1700s to the present day. We observe the same decrease in average narrative time up to the mid-20th century. After establishing this result we compare metrics of estimated literary time to patterns of language use.

1 Introduction

A goal of the study of the history of literature is to identify consistent changes in the writing of fiction. One of the more measurable characteristics is the narrative time elapsed in a passage. Table 1 shows example passages with long and short durations of elapsed narrative time. Narrative time by itself may not be particularly meaningful in isolation but if we can find consistent trends they may indicate larger changes in literary history.

We build on work that defines and measures elapsed narrative time in novels from the early 18th through late 20th centuries (Underwood, 2018). Underwood finds a consistent trend towards shorter elapsed time over the course of the 18th and 19th centuries from on average about one day down to about one hour, with an increase around the middle of the 20th century. The previous work suggests that the trend in elapsed narrative time is not subject to the discontinuities previously hypothesized by some literary theorists (Genette, 1980), but the work has a small, carefully chosen dataset and large variation.

We present a new extension of the existing labeled dataset, which we use to validate previous findings. We then use a simple baseline method, linear regression from unigrams, to predict the log of elapsed time per passage length. In addition to the simple baseline method, we try a more sophisticated ranking-based approach that improves predictions at the loss of temporal resolution.

After validating our predictive models on labeled data we extend these results to an unlabeled dataset that is four orders of magnitude larger by token count. We replicate the result from (Underwood, 2018) that there is a strong gradual trend towards shorter narrative time over the course of the 19th and 20th centuries that flattens around 1950.

Measuring trends in the narrative presentation of elapsed time is significant in that it can lend additional credence to historical claims previously based on near-anecdotal evidence. Such similar
gradual downward trends between the small labeled dataset and the large dataset serve to increase confidence in the gradual decrease in elapsed narrative time rather than the previously hypothesized discontinuity (Underwood, 2018). We interrogate the diachronic trends in narrative time using a learned topic model and preliminarily find that the trends are not due to a simple identification of century-specific style.

2 Related work

Much work has been done on computationally-assisted narrative interpretation, especially on such large-scale categorizations as genre, style, and nationality, as well as on quantifying literary-historical phenomena over time (Jockers, 2013; Piper, 2018). We focus on finer-grained textural information, as in work that quantifies intratextual qualities (Underwood, 2018; McGrath et al., 2018).

3 Datasets

We use three datasets of varying sizes: two small labeled datasets for learning to predict elapsed narrative time and to quantify the quality of that prediction. The third, much larger, dataset is for evaluating the extent to which the resulting literary historical claims about elapsed narrative time are justified on a larger corpus.

3.1 Existing labeled dataset

As a starting point, we begin with a dataset collected and analyzed by Underwood, which we refer to as the WLTMM dataset (Underwood, 2018). This dataset consists of 1,765 passages of about 250 words in length chosen from 108 canonical, best-selling, and randomly chosen novels.

3.2 Newly labeled dataset

In addition to the existing labeled passages, we created an additional set of labeled passages. These will be released publicly to the extent possible under copyright. In order to validate similarity within authors we label new passages from novels in the existing WLTMM dataset. To evaluate performance across authors, we also include a number of new, previously unlabeled novels.

We labeled elapsed narrative time for eight passages from each of seven novels for a total of 56 newly labeled passages. The novels were a combination of those already in the WLTMM dataset and new novels from across the time period covered by the WLTMM dataset: Robinson Crusoe (1719), Pride and Prejudice (1813), Frankenstein (1818), Jane Eyre (1847), Middlemarch (1871-2), Mrs Dalloway (1925), and The Recognitions (1955). Passages were 250-word segments chosen at random from the selected novels. No individual passages were repeated between this and the WLTMM dataset.

3.3 HathiTrust

In order to extend our analysis to a large-scale data set including contemporary works we need access to passage-level word count information from volumes that are in copyright. The HathiTrust Digital Library provides, among other things, page-level word frequencies of millions of volumes of text. We sampled 116,434 volumes from HathiTrust that (a) had a genre metadata tag of fiction and (b) were published in English. We constructed bags of words for each volume. The vocabulary was limited for computational efficiency to the top 10,000 words in a sample of 1,000 HathiTrust volumes.

The year of publication field in the HathiTrust metadata is an insufficient proxy for year of first publication (Bamman et al., 2017). In order to estimate year of first publication, if a volume was published after the year of the author’s death, we estimated the year of first publication as the year of the author’s birth plus thirty years. We found this metadata-based strategy provided a reasonable estimation for our purposes, so we did not predict first publication from the text data as in (Bamman et al., 2017). Particular examples of failure were for novels first published posthumously or later editions published while an author was alive. Figure 1 shows the distribution of estimated years of first publication in our sample of the HathiTrust dataset.

![Figure 1: Estimated years of first publication in the sample of 53,539 HathiTrust volumes.](https://www.hathitrust.org/)
Table 2: Representative example passages with labeled elapsed time and predicted elapsed time. The top two passages show correct predictions for short duration and long duration. The bottom two passages show examples of extreme mismatches in prediction. Bright green: ten words with the greatest combined contribution in each passage indicative of long elapsed time. Dark green: next ten words in each passage indicative of long elapsed time. Bright purple: ten words with the greatest combined contribution in each passage indicative of short elapsed time. Dark purple: next ten words in each passage indicative of short elapsed time.

Table 3: Time durations converted to log of minutes per 250-word passages.

3.4 Volume deduplication

Individual works appear in HathiTrust many times over. Reducing the number of duplicates is important to avoid the distorting effects that such duplicates can have, especially on semantic models (Schofield et al., 2017).

We used random low-rank projection of each volume from a vector with 10,000 entries weighted by inverse document frequency to a vector with 100 entries, and, after normalization, compared cosine similarities between all volumes. Only one volume was kept from any set of volumes with pairwise cosine similarities above a certain threshold. The cutoff threshold was determined by running this process on a smaller labeled dataset of 700 volumes corresponding to 62 novels that were also in the WLTMM dataset. Figure 2 shows, via a histogram of pairwise cosine similarities, that the vast majority of such similarities were within the range of random while a small but noticeable fraction were very similar, as expected. We used the relatively conservative threshold of 0.75 that removed 87.5% ± 30.6% of each novel’s duplicates on average while only removing eight volumes that were not duplicates from two novels each.

4 Predicting elapsed narrative time

We learned to both directly predict elapsed narrative time using linear regression and to rank two passages’ elapsed times via logistic regression. Furthermore, we used a latent Dirichlet allocation topic model to interpret the output of the linear regression model (Blei et al., 2003).

4.1 Linear regression

Following the work of Underwood, we trained and tested a linear regression model on the WLTMM dataset. The input to the model was the unigram togram of pairwise cosine similarities, that the vast majority of such similarities were within the range of random while a small but noticeable fraction were very similar, as expected. We used the relatively conservative threshold of 0.75 that removed 87.5% ± 30.6% of each novel’s duplicates on average while only removing eight volumes that were not duplicates from two novels each.
Figure 3: Correlation between labeled and linear-regression-predicted elapsed narrative time for the newly labeled passages (blue) with the results for passages in the WLTMM dataset (orange) included for comparison. The diagonal lines indicate what would be perfect prediction. Points above the line correspond to passages with predictions that over-predict elapsed time, i.e. those with a predicted elapsed time greater than that which was labeled, and points under the line similarly correspond to under-predictions.

Table 4: Coefficients from the linear regression for (a) words with coefficients most associated with longer elapsed time and (b) words with coefficients most associated with shorter elapsed time.

<table>
<thead>
<tr>
<th>Word</th>
<th>Coefficient</th>
<th>Word</th>
<th>Coefficient</th>
<th>Total contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>years</td>
<td>0.7331</td>
<td>woman</td>
<td>-0.3986</td>
<td>1.52</td>
</tr>
<tr>
<td>life</td>
<td>0.6059</td>
<td>black</td>
<td>-0.4176</td>
<td>1.25</td>
</tr>
<tr>
<td>return</td>
<td>0.5924</td>
<td>best</td>
<td>-0.4193</td>
<td>0.68</td>
</tr>
<tr>
<td>family</td>
<td>0.5065</td>
<td>pocket</td>
<td>-0.4215</td>
<td>0.48</td>
</tr>
<tr>
<td>failure</td>
<td>0.5043</td>
<td>room</td>
<td>-0.4243</td>
<td>0.46</td>
</tr>
<tr>
<td>history</td>
<td>0.4867</td>
<td>thou</td>
<td>-0.4315</td>
<td></td>
</tr>
<tr>
<td>wrote</td>
<td>0.4772</td>
<td>struck</td>
<td>-0.4330</td>
<td></td>
</tr>
<tr>
<td>during</td>
<td>0.4694</td>
<td>wish</td>
<td>-0.4536</td>
<td></td>
</tr>
<tr>
<td>nor</td>
<td>0.4337</td>
<td>moment</td>
<td>-0.4935</td>
<td></td>
</tr>
<tr>
<td>aunt</td>
<td>0.4309</td>
<td>replied</td>
<td>-0.6549</td>
<td></td>
</tr>
</tbody>
</table>

Table 5: Words that contribute most positively (top) and most negatively (bottom) to the prediction of elapsed time for the passage from *The Recognitions* shown at the top of Table 2.

<table>
<thead>
<tr>
<th>Word</th>
<th>Frequency</th>
<th>Coefficient</th>
<th>Total contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>you</td>
<td>13</td>
<td>0.12</td>
<td>1.52</td>
</tr>
<tr>
<td>ve</td>
<td>4</td>
<td>0.31</td>
<td>1.25</td>
</tr>
<tr>
<td>to</td>
<td>9</td>
<td>0.08</td>
<td>0.68</td>
</tr>
<tr>
<td>and</td>
<td>6</td>
<td>0.08</td>
<td>0.48</td>
</tr>
<tr>
<td>eyes</td>
<td>2</td>
<td>0.23</td>
<td>0.46</td>
</tr>
<tr>
<td>her</td>
<td>4</td>
<td>-0.12</td>
<td>-0.50</td>
</tr>
<tr>
<td>have</td>
<td>2</td>
<td>-0.30</td>
<td>-0.60</td>
</tr>
<tr>
<td>him</td>
<td>4</td>
<td>-0.18</td>
<td>-0.71</td>
</tr>
<tr>
<td>but</td>
<td>4</td>
<td>-0.24</td>
<td>-0.95</td>
</tr>
<tr>
<td>thing</td>
<td>2</td>
<td>-0.49</td>
<td>-0.99</td>
</tr>
</tbody>
</table>

We further validated the trained model on our newly labeled dataset. Table 2 shows representative examples of passages for which the classifier performed well and examples for which it performed poorly. Figure 3 shows regression predictions for each passage in our newly labeled dataset with comparisons to predictions on passages from the same novels in the WLTMM dataset, where available. The Pearson correlation coefficient between labeled and predicted elapsed time for all passages in our newly labeled dataset was $R=0.3474$ ($p=0.0087$). The regression performed best on passages from *Frankenstein, Pride and Prejudice, and The Recognitions*, despite no passages from the latter two appearing in the training set.

Table 4 shows the words with regression coefficients most indicative of long and short elapsed narrative time. Some words, examples of which include *you, to, and and* in Table 5, had coefficients
Figure 4: Moving means of average predicted elapsed narrative time per passage in the labeled training dataset (orange) and the larger HathiTrust dataset (blue). Shaded regions represent 95% confidence intervals, which do not account for prediction uncertainty.

4.2 Ranking is an easier task than regression

In order to address some of the issues of the linear regression we trained a ranking classifier by performing logistic regression on the difference between two passages’ bags of words. The goal was to see if determining which of the two input passages had a longer elapsed narrative time could be predicted with greater accuracy than directly predicting the log of the elapsed time. Specifically, we present the classifier with two passages and code a “correct” prediction as 0 if the first passage has a shorter duration and as 1 if the second passage has a shorter duration. The training data consisted of 100,000 pairwise comparisons for random pairs of passages in the WLTMM dataset, and test data was pairwise comparisons between all pairs of passages in our newly labeled dataset.

Figure 5 shows receiver operating characteristic (ROC) and precision-recall curves for the test set; the classifier had an area under the ROC curve (AUC) of 0.7448, precision of 0.72, and recall of 0.67. Logistic regression outperforms the baseline of ranking passages by the linear regression’s predicted elapsed time, which achieved an AUC of 0.6517, with precision and recall of 0.69 and 0.64, respectively. The words that are most associated with longer and shorter elapsed time remain mostly unchanged from those in linear regression. There do not seem to be specific novels on which the classifier performs better or more poorly.

4.3 \(\ell_1\) regularization for word importance

We find that although there is room for improvement, a number of different approaches for predicting elapsed time computationally have similar em-
empirical performance. We next consider the limits of the set of input features. How few word types can we use without seriously impacting performance?

We use an $\ell_1$-regularized ranking model with varying penalty parameter to progressively drive feature weights to zero. Surprisingly, a minimal set of eight words is sufficient to obtain 0.74 AUC on the training and test sets without any degradation in test performance. These words are of, in, to, had to indicate longer duration and you, said, it, he to indicate shorter duration.

While it is difficult to extrapolate from such words, all are short words that often appear in varieties of phrases. Said is additionally very indicative of dialogue, as—to a lesser extent—is you indicative of second-person dialogue as well.

5 Learned topics correlate with elapsed time

It is interesting to be able to predict human labels for elapsed time, but this by itself is only the beginning of the process of making empirical literary historical claims. We hope to use the larger corpus to reason about other trends that might be associated with such prediction. One type of trend is whether or not shorter and longer elapsed narrative time is grounded in specific language in the form of word co-occurrence patterns.

To this end, a topic model with 50 topics was trained on the HathiTrust volumes, with words that were found to be most predictive of elapsed narrative time excluded. Documents’ topic proportions were used in a linear regression to predict a document’s average elapsed narrative time.

To avoid simply learning from exactly the same data, the documents on which the topic model was trained had the eight words most indicative of elapsed time removed. The target value for the regression was the elapsed time predicted from the linear regression model on just the eight words excluded with words that were most predictive of elapsed narrative time. Some appear to be getting at other features of narrative, such as dialogue (replied, asked, cried), or content, such as government (people, government) and Rome (city, great, rome). Just as said is one of the words most associated with shorter elapsed time when ranking with $\ell_1$-regularization, so, too, is a topic consisting of words that mark dialogue associated with shorter elapsed time here.
Though the distribution of topics does not preclude the possibility that the models are also learning other correlates of publication date, they do show that there are word co-occurrence patterns that distinctly correlate with longer and shorter elapsed time while not simply reproducing the gradual downward trend in narrative duration over time seen in Figure 4.

6 Discussion

Additional work can be done with our existing formulation of narrative time to detect narrative time with greater accuracy, to investigate whether our literary historical claims extend to even larger corpora, and to investigate whether other aspects of literary texture contribute to elapsed narrative time. The concept of literary time could be formulated otherwise; for instance, other formulations might detect phrases that signal the rapid passage of time, such as “The next morning . . . ”. We have presented models that learn to predict elapsed narrative time and have furthermore used them to evaluate empirical historical claims about the shortening of literary time, finding that such claims remain valid even with the increased evidence afforded by narrative time prediction.

References


