What time is it? Temporal Analysis of Novels

Allen Kim, Charuta Pethe, Steven Skiena
Data Science Lab

What we are trying to do

Predict hour of day (0-23) throughout a book

Important for understanding story flow and how events in a book are segmented

Overview of Methods

1. Collect text with time references from dataset of books
2. Clean and label all time data extracted
   - Disambiguating AM vs PM - binary classifier
3. Train a model for time of day prediction for a window of text
   - 24-hour classifier
4. Use model to predict time over the entire book
5. Use dynamic programming to unify times throughout the book
   - Handling continuation of time

At nine o’clock, one morning late in July Gatsby’s gorgeous car lurched up the rocky drive to my door and gave out a burst of melody from its three noted horn...

Prediction: Hour 9 (not 21 due to mention of morning)

By six o’clock Michaelis was worn out and grateful for the sound of a car stopping outside. It was one of the watchers of the night before who had promised to come back so he cooked breakfast for three which he and the other man ate together.

Prediction: Hour 6 (not 18 due to mention of breakfast)

Core Details and Results

1. Collect time phrases using SUTime, a time parser from Stanford CoreNLP on two book datasets (Gutenberg, HathiTrust)
2. Implement three different models to predict times - Naive Bayes, LSTM, BERT
3. Models are trained for AM/PM classification as well as 24 hour classification
4. Times are predicted over n windows and we run DP to unify them into k segments

The graph on the left shows both human-annotated (blue) and our generated predicted times with DP (red) throughout The Great Gatsby. The errors in hours are shown below.

We had human annotators compare with our AM/PM predictions and show the agreement (above). We also show the error in hours (below) for our two datasets. BERT performs best overall.