Modeling Suspense in Reading as Uncertainty Reduction over Neural Representations

David Wilmot & Frank Keller (University of Edinburgh)
keller@inf.ed.ac.uk

Much of psycholinguistic research has focused on language processing at the sentence level, investigating factors such as ambiguity, expectation, and memory load. But most sentences are read in a textual context, and processing is also guided by text-level factors. Here, we focus on narrative texts, and on how reading stories is affected by suspense. Narratologists argue that suspense is important for keeping readers’ attention (Khrypko & Andreae, 2011), promoting their immersion (Hsu et al., 2014), and making stories enjoyable and interesting (Oliver, 1993). From a psycholinguistic perspective, we can regard suspense as text-level expectation: as we process a story incrementally, our expectations about how it will continue changes sentence by sentence.

We study suspense by generalizing concepts from human language processing (Hale, 2001, 2006) from the word level to the sentence level. Economists (Ely et al., 2015) have developed a general theory of suspense, arguing it is based on expectations over states. We compare two ways of modeling suspense: surprise, a backward-looking measure of how expected the current state is given the story so far; and uncertainty reduction, a forward-looking measure of how expected the continuation of the story is. Given a sentence at position $t$ with conditional probability $P(e_t)$, where $e_t$ is a sentence representation, surprise (Hale, 2001) is given by eq. (1). Uncertainty reduction (Hale, 2006) captures how much $e_t$ changes the entropy over possible next sentences $e_{t+1}$, eq. (2). Ely et al. (2015) model surprise and suspense as an expectation over states. Here, the states are the sentence representations $e_t$; eqs. (3) and (4) give the Ely definitions.

We use a hierarchical language model based on Generative Pre-Training (GPT; Radford et al., 2018) to encode sentences as neural representations; Figure 1 shows the architecture of our model. It builds a chain of representations (embeddings) that anticipates what will come next in a story, allowing us to infer measures of suspense. First, GPT turns each word in a sentence into a word embedding $w_i$. Then, an LSTM (Hochreiter & Schmidhuber, 1997) turns word embeddings into sentence embeddings $\gamma_i$. Based on these, another LSTM computes $e_t$, a contextualized representation of sentence $t$, for calculating surprise and uncertainty reduction. Training uses an autoregressive binary cross-entropy loss over state representations, using incorrect next sentences as negative targets. For inference, we obtain a distribution over continuations $P(e_{t+1})$ by either generating a set of continuations using GPT (Gen), or sampling them from a corpus (Cor).

Our model is trained on the WritingPrompts corpus of short stories (Fan et al., 2018); these were created as an exercise in creative writing and are interesting, natural, and of suitable length. To evaluate the model predictions, we selected 100 stories of 25–75 sentences in length. Crowd workers read a story and judge suspense for each sentence using a five-point scale. Five workers judged each story; the inter-annotator agreement was 0.57 (Krippendorff’s $\alpha$), which is substantial given the subjective nature of the task. To evaluate our model, we computed Spearman’s $\rho$ and Kendall’s $\tau$ between model predictions and suspense judgments, averaging over the five annotators and 100 stories. The human upper bound is the pairwise correlations between annotators.

Figure 2 shows surprise, uncertainty reduction, and suspense judgments for an example story; Table 1 gives the correlations between model predictions and judgments. The best performance was observed for $U_{Ely}$, uncertainty reduction over sentence representations. The definition of $U_{Ely}$ combines sentence probability ($S_{Hale}$) with the distance between sentence embeddings ($S_{Ely}$), which are both good predictors on their own. Hale uncertainty reduction, on the other hand, performs poorly. We conclude that our hierarchical neural architecture successfully models suspense judgments for short stories. This supports Ely et al. (2015) in that while uncertainty is a factor, it is consequential divergences of outcome states that are primary in narrative suspense.
Once upon a time

Figure 1: Architecture of our hierarchical neural model: word_enc computes word embeddings, sent_enc concatenates them to obtain sentence embeddings, and story_enc turns them into contextualized sentence representations $e_t$.

\begin{align}
S_t^{\text{Hale}} &= -\log P(e_t) \\
H_t &= -\sum_i P(e_{i+1}^i) \log P(e_{i+1}^i) \\
S_t^{\text{Ely}} &= (e_t - e_{t-1})^2 \\
U_t^{\text{Ely}} &= E[(e_t - e_{i+1}^i)^2] = \sum_i P(e_{i+1}^i)(e_t - e_{i+1}^i)^2
\end{align}

Table 1: Model results for the WritingPrompts corpus; confidence intervals in brackets.

<table>
<thead>
<tr>
<th>Prediction</th>
<th>$\tau$</th>
<th>$\rho$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>0.652</td>
<td>0.711</td>
</tr>
<tr>
<td>$S^{\text{Hale}}_{\text{Gen}}$</td>
<td>0.407</td>
<td>0.495</td>
</tr>
<tr>
<td>$S^{\text{Hale}}_{\text{Cor}}$</td>
<td>0.454</td>
<td>0.523</td>
</tr>
<tr>
<td>$U^{\text{Hale}}_{\text{Gen}}$</td>
<td>0.036</td>
<td>0.051</td>
</tr>
<tr>
<td>$U^{\text{Hale}}_{\text{Cor}}$</td>
<td>0.061</td>
<td>0.088</td>
</tr>
<tr>
<td>$S^{\text{Ely}}_{\text{Gen}}$</td>
<td>0.391</td>
<td>0.504</td>
</tr>
<tr>
<td>$S^{\text{Ely}}_{\text{Cor}}$</td>
<td>0.620</td>
<td>0.710</td>
</tr>
<tr>
<td>$U^{\text{Ely}}_{\text{Gen}}$</td>
<td>0.605</td>
<td>0.693</td>
</tr>
<tr>
<td>$U^{\text{Ely}}_{\text{Cor}}$</td>
<td>0.605</td>
<td>0.693</td>
</tr>
</tbody>
</table>

References


