Storylines

July 22, 2020

1 Introduction

Our goal is to improve the coherency of generative story models by modeling storylines. As an alternative to black-box models with little explicit structure, prior work has noted that including structure in the generative model improves coherence in the generations [1, 3].

However, prior investigations into narrative structure used hand-crafted representations, such as entity coreference or keywords. We hope to alleviate the need for hand-crafted structure via structured generative models.

2 Storyline Induction

As an initial exploratory step, we turn to sequence alignment algorithms to induce segmentations of stories.

Sequence alignment algorithms are used in biology to identify similar subsequences of proteins in an effort to determine whether a set of proteins share common ancestry. The algorithms start with a sequence of atomic representations, such as amino acids, then build up alignments using these atomic representations. Context is taken into account not in the representations, but through the dynamic program. We posit that these algorithms may also produce motifs that resemble storylines in short stories. We perform alignment with each independent sentence representation as an atom.

We will answer the following questions with our exploratory analysis:

1. At what level should we analyze narrative structure? Do storylines have a set or random length?
2. Is there a general structure, such as exposition, complication, falling action, and resolution? What are the narrative elements present in short stories?
3. How do we handle gaps that appear in this structure?
4. Can we leverage prompts in structure induction?
5. For assigning a given sentence to a narrative element, is the context (surrounding sentences) or sentence itself (ignoring context) more important?

2.1 Scale

3 Pairwise Alignment with SBERT and DTW

We first investigate whether we can recover structure that resembles a storyline using pretrained models. In particular, we use SBERT [2] to map sentences $\mathbf{x}_i \in \mathcal{X}^*$ to vector representations $\mathbf{y}_i \in \mathbb{R}^n$. As each story consists of several sentences, we obtain a sequence of representations corresponding to each sentence in a given story $\mathbf{Y} = \langle \mathbf{y}_0, \ldots, \mathbf{y}_T \rangle$. We then compute an alignment between the sentences of two stories using DTW.

We evaluate the efficacy of SBERT and DTW without fine-tuning by determining whether $DTW(Y_i, Y_j)$ correlates with an intuitive distance of stories. For a given story we obtain the top K closest stories under DTW in addition to K additional random stories, and compare the ranking under DTW with human annotation.

We additionally evaluate the alignments (TBD)

Analysis forthcoming.

If the metric is ill-behaved, is it possible that the alignments are still good? Would this imply that training through the metric is a bad idea?

3.1 Issues with DTW

1. Length: Fixing an alignment, if you copy one of the elements and insert it next to the original the DTW distance will increase. This biases DTW towards shorter sentences. This could possibly be fixed by jointly aligning and segmenting.

2. Not smooth: Alignment themselves are not smooth. Soft-DTW is a possible alternative (which has its own problems).

4 Multisequence Alignment

Although we found that the pairwise comparison using DTW correlated with intuitive distance, we aim to further extract robust storylines that are common to multiple stories. To accomplish this, we turn to multisequence alignment (MSA) which considers multiple stories and produces a global alignment across multiple stories.

We use the multisequence generalization of DTW [4]. We perform an evaluation of this measure: For a given collection of stories, we compare the K-subset with the lowest mDTW verus a human-annotated subset. We obtain a collection of stories by obtaining the $M > K$ closest stories to a given story, as measured by DTW.

Analysis forthcoming.

5 A Trainable Model

We wish to improve the accuracy of DTW and mDTW without human annotation. we optimize a nonparametric time-series modeling objective that extends builds on pairwise DTW:

$$\log p(Y_i) = \log \sum_j p(Y_i, Y_j)$$

Details TBD, but will use initial SBERT+DTW distance to approximate integral. if preliminary evaluation in Sections 3 and 4 are positive, then there is hope that the approximation grids do not have to be extremely large or updated as training progresses.

We also experiment with the multi-sequence extension.

References

[1] Angela Fan, Mike Lewis, and Yann N. Dauphin. Strategies for structuring story generation. *CoRR*, abs/1902.01109, 2019. URL `http://arxiv.org/abs/1902.01109`.

[2] Nils Reimers and Iryna Gurevych. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3982–3992, Hong Kong, China, November 2019. Association for Computational Linguistics. doi: 10.18653/v1/D19-1410. URL `https://www.aclweb.org/anthology/D19-1410`.

[3] Lili Yao, Nanyun Peng, Ralph M. Weischedel, Kevin Knight, Dongyan Zhao, and Rui Yan. Plan-and-write: Towards better automatic storytelling. *CoRR*, abs/1811.05701, 2018. URL `http://arxiv.org/abs/1811.05701`.

[4] F. Zhou and F. De la Torre. Generalized canonical time warping. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2):279–294, 2016.