Storylines

November 2, 2020

1 Introduction

Modern generative models of text are extremely flexible, with huge amounts of parameters trained on vast amounts of data [2]. However, black-box text generation models may not be satisfactory in all text-generation scenarios, especially when there is risk associated with the output. In such scenarios, a controllable model with explicit structure would allow risk mitigation via a human-in-the-loop process.

Our goal is to design structured, generative story models by modeling storylines. Prior work has noted that including structure in the generative model improves coherence in the generations [4, 10]. However, prior investigations into narrative structure use hand-crafted representations, such as entity coreference or keywords. We explore a different definition of what constitutes a storyline: We hypothesize that storylines are composed of segments that function similar to motifs in biological sequences. Motifs in biological sequences are recurring patterns that are believed to have biological signifiance. For example, they may determine how protein interactions take place. Analogously, we posit that storyline segments should determine how a reader interprets a story.

Modeling reader interpretation is very difficult, as eliciting such an interpretation from a human would be very expensive. Thankfully, biological sequence alignment algorithms offer a solution to this problem. Biological sequence alignment algorithms find motifs without directly modeling downstream effects, and only use similarity between elements of the sequences themselves.

Typically, these similarity measures are derived from manually-specified rules specific to biological applications. Our hypothesis, that storylines segments have a similar interpretive function, implies that the similarity measure for storylines should be semantic similarity. We therefore turn to large pre-trained sentence representations [8], as they have been shown to perform very well on related tasks such as natural language inference and entailment.

We perform an exploratory analysis of storyline induction by utilizing biological sequence alignment algorithms as well as large pretrained sentence representations. We show that alignment algorithms built upon measures of semantic similarity are able to extract common structure from stories.

2 Background: Pairwise Sequence Alignment

The goal of pairwise sequence alignment is to find a minimum cost alignment between sequences of sentences $\mathbf{x}_1$ and $\mathbf{x}_2 \in \Sigma^*$, where sentences are represented by a vector in $\mathbb{R}^n$. We additionally add a gap symbol '-' to Σ, such that $\Sigma = \mathbb{R}^n \cup \{-\}$. We refer to elements of Σ as tokens.

In order to align two stories, pairwise alignment extends edit distance with a distance measure $d : \Sigma \times \Sigma \to \mathbb{R}$ that operates on pairs of tokens, one from each story. The best alignment is given by the following:

$$\operatorname*{argmin}_{\pi} \sum_{(i,j)\in\pi} d(x_{1i}, x_{2j}), \tag{1}$$

where π is a joint path through $\mathbf{x}_1$ and $\mathbf{x}_2$. Valid paths are obtained by only inserting gaps into the original story while preserving order. Equation 1 can be solved exactly via dynamic programming. Let $\mathbf{x}_1^\pi, \mathbf{x}_2^\pi$ be the gap-augmented stories obtained by following path π, which are both of length L. We refer to $M = \begin{pmatrix} \mathbf{x}_1^\pi \\ \mathbf{x}_2^\pi \end{pmatrix}$ as the alignment.

When choosing a distance measure d, a key component is how the measure deals with gaps. There are two approaches: 1) gaps correspond to insertion or deletions and 2) gaps correspond to expansions or compressions. Classical algorithms from biology, such as Needleman-Wunsch, model insertions and deletions, as random mutations result in completely new elements that do not have a good interpretation as an expansion or compression. Other algorithms such as Dynamic Time Warping (DTW) model expansions and compressions, since warping may occur due to various reasons such as sampling at different frequencies. It is possible to model both of these phenomena in the distance measure. We assume that adding a sentence to a story, even if it builds on prior sentences, adds new information and is semantically different from preceding sentences, and therefore only model insertion and deletion.

3 Problem Setup: Multiple Sequence Alignment

In order to extract storylines, we would like to find patterns that are robust across multiple stories. Unfortunately, methods for solving pairwise alignment are not directly applicable, since they only consider pairs of stories. Instead, we turn to the problem of multiple sequence alignment, which generalizes alignment from pairs of stories to sets of stories.

Given a set of K sequences, a multiple alignment is a matrix $M \in \Sigma^{K \times L}$, given by:

$$M = \begin{pmatrix} \mathbf{x}_1^\pi \\ \vdots \\ \mathbf{x}_K^\pi \end{pmatrix}, \tag{2}$$

where each $\mathbf{x}_i^\pi$ is obtained by inserting gaps into the corresponding $\mathbf{x}_i$ to ensure that each $\mathbf{x}_i^\pi$ is of length L, as in pairwise alignment.

There are two common measures of quality for a multiple alignment: 1) the sum of pairs (SP) score and 2) the consensus error or Steiner distance. All methods use the distance measure $d : \Sigma \times \Sigma \to \mathbb{R}$ from pairwise alignment. The SP score is given by

$$\mathrm{SP}(M) = \sum_{l=1}^{L} \sum_{i=1}^{K} \sum_{j=i+1}^{K} d(M_{il}, M_{jl}), \tag{3}$$

obtained by summing the pairwise distances between elements of M in the same column. The consensus error given a consensus sequence $\mathbf{z} \in \Sigma^L$ is defined as follows:

$$\mathrm{CE}(M, \mathbf{z}) = \sum_{l=1}^{L} \sum_{i=1}^{K} d(M_{il}, z_l), \tag{4}$$

the sum of the distances from each element in a column to z_l. The mean sequence $\mathbf{z} \in \Sigma^L$ that minimizes this error is known as the consensus sequence. Finally, the Steiner distance is closely related to the consensus error, and can be computed without explicitly constructing a multiple alignment. Let $D : \Sigma^* \times \Sigma^* \to \mathbb{R}$ be the distance measure obtained under global pairwise alignment using the token distance d, so that D operates on pairs of sequences. The Steiner distance is given by

$$\mathrm{SD}(\mathcal{X}, \mathbf{z}) = \sum_{i=1}^{K} D(\mathbf{x}_i, \mathbf{z}). \tag{5}$$

The optimal consensus error, $\min_{\mathbf{z}} \mathrm{CE}(M, \mathbf{z})$, is equivalent to the optimal Steiner distance, $\min_{\mathbf{z}} \mathrm{SD}(\mathcal{X}, \mathbf{z})$ [6]. Additionally, the optimal Steiner sequence, $\arg\min_{\mathbf{z}} \mathrm{SD}(\mathcal{X}, \mathbf{z})$, is equivalent to the optimal consensus sequence, $\arg\min_{\mathbf{z}} \mathrm{CE}(M, \mathbf{z})$, up to spaces. (Double check this last part carefully, maybe cut out consensus error definition.)

We utilize algorithms that optimize both the SP score and the Steiner distance.

4 Methods

As optimizing both the SP score and Steiner distance are NP-hard, we resort to heuristic optimization methods. We use a combination of three approaches: a progressive alignment method inspired by a classical algorithm from computation biology, an iterative averaging method similar to a popular algorithm from time-series analysis, and a greedy hill climbing algorithm that operates in the space of Steiner sequences.

4.1 Progressive Alignment

Inspired by the progressive alignment algorithm of Feng and Doolittle [5], we first take a naive approach to approximating a multiple sequence alignment. Progressive alignments aim to optimize the SP score.

Given a set of sequences $\mathcal{X}$ and an ordering σ, we progressively align the next sequence in the ordering to the already aligned sequences. Once a set of sequences are aligned, their columns of the alignment are frozen; elements of new sequences must align to a whole column from the existing alignment or a gap. This is referred to as the 'once a gap, always a gap' property [5].

In order to align a sequence to an existing alignment, we lift the definition of the token distance d to compare an element to a column of an alignment such that $d^+ : \Sigma^* \times \Sigma \to \mathbb{R}$, as follows:

$$d^+(\mathbf{y}, x) = \sum_{j=1}^{|\mathbf{y}|} d(y_j, x). \tag{6}$$

We can then extend pairwise global alignment with d^+, allowing us to align the columns of an alignment to a token. The full algorithm is given in Algorithm 1.

Algorithm 1 Progressive Alignment

Given: A set of sequences $\mathcal{X} = \{\mathbf{x}_i\}_i$, ordering σ, and extended distance measure d^+
Initialize alignment M to $\mathbf{x}_{\sigma(1)}$
for all sequences $\mathbf{x}_{\sigma(i)}$ in order σ **do**
    Update $M = \textsc{PairwiseAlign}(M, \mathbf{x}_{\sigma(i)}, d^+)$
return M

4.2 An Iterative Averaging Algorithm

Next, we propose an iterative averaging (IA) algorithm which directly optimizes the Steiner distance. The IA algorithm is inspired by the DTW Barycenter Averaging (DBA) algorithm [7], which iteratively computes pairwise alignments between a mean sequence $\mathbf{z}$ and a set of sequences $\mathcal{X}$ then uses those alignments to recompute the mean sequence. As the DBA algorithm was designed to model expansion and compression, we adapt it for insertion and deletion.

The IA algorithm proceeds in a fashion similar to DBA. We first construct a multiple alignment from the pairwise alignments, then for each column of the multiple alignment compute the token that minimizes the distance to each element within that column. As we model insertions and deletions, there is ambiguity when mapping the pairwise alignments to a joint multiple alignment. Namely, there are multiple ways of aligning tokens from $\mathcal{X}$ that are all aligned to gaps in $\mathbf{z}$. We resolve this ambiguity heuristically by inserting gaps in the $\mathbf{x}_i^\pi$ obtained by pairwise alignment with $\mathbf{z}$ so that all tokens aligned to elements in $\mathbf{z}$ are aligned. The algorithm is detailed in Algorithm 2.

4.3 Hill Climbing Algorithm

As the previous iterative algorithm used a heuristic in the averaging step that was computationally cheap but not guaranteed to improve the Steiner distance, we also consider a greedy hill climbing (HC) algorithm that is more expensive but guaranteed to improve the objective.

The hill climbing algorithm proceeds as follows: Given an initial mean sequence $\mathbf{z}$, we first compute the pairwise alignments from $\mathbf{z}$ to each sequence in $\mathcal{X}$. We obtain an initial proposal sequence by averaging the token from each $\mathbf{x}_i$ aligned to each element of $\mathbf{z}$, rather than a gap. We then compute all one-step deviations from this proposal sequence, then find the proposal with the lowest Steiner distance to $\mathcal{X}$ for use as the mean sequence in the next iteration. One-step deviations are obtained by adding or deleting one element of $\mathbf{z}$. Candidates for addition are obtained by considering the elements of sequences in $\mathcal{X}$ that are aligned to gaps in-between elements of $\mathbf{z}$. The full algorithm is given below in Algorithm 3.

5 Experiments

We evaluate our MSA approaches on the WritingPrompts dataset [3], a dataset of 300K human-written short stories obatined from the WritingPrompts subreddit. Each story consists of a pair of a writing prompt and the story itself. In our experiments, we use only the story.

Algorithm 2 Iterative Averaging Alignment

Given: A set of sequences $\mathcal{X} = \{\mathbf{x}_i\}_i$ and distance measure d
Given: Initial mean string $\mathbf{z}$
function ITERATIVEAVERAGE($\mathbf{z}, \mathcal{X}, d$)
  for all iterations **do**
    for all sequences $\mathbf{x}_i \in \mathcal{X}$ **do**
      Compute pairwise alignments $M^i = \text{PAIRWISEALIGN}(\mathbf{x}_i, \mathbf{z}, d) \in \Sigma^{2\times L_i}$
    Compute multiple alignment $M = \text{STACK}(M^1, \ldots, M^K)$
    Update $\mathbf{z} = \text{AVERAGE}(M, d)$
  return M
function STACK($M^1, \ldots, M^K$)
  while Non-gap elements of $\mathbf{z}$ are not aligned in all M_2^i **do**
    Record column indices of M_2^i that have a non-gap element of $\mathbf{z}$
    Insert $(-,-)$ columns to the left of the columns with non-gap elements from $\mathbf{z}$ for each M^i
  return $\begin{pmatrix} M_1^1 \\ \ldots \\ M_1^K \end{pmatrix}$
function AVERAGE(M, d)
  Initialize $\mathbf{z} \in \Sigma^L$
  for all columns l in M **do**
    Set proposal z' to the average non-gap representation of column l
    Compute costs $c(z') = \sum_i d(M_{il}, z')$ and $c(-) = \sum_i d(M_{il}, -)$
    if $c(z) > c(-)$ **then**
      Set $z_l = -$
    else
      Set $z_l = z'$
  return $\mathbf{z}$

We perform MSA on story sets of size 5. As it would be intractable to run MSA on all story sets of size 5, we instead use the following procedure to obtain the sets $\mathcal{X}$:

1. For every story in the WRITINGPROMPTS dataset, we project each sentence using SBERT [8] into $\mathbb{R}^n$.
2. We compute the bigram bag-of-sentence (BoS) representations of stories by concatenating the SBERT representations of consecutive sentences and averaging over time.
3. We use each of the first 50k stories from the WRITINGPROMPTS dataset as centroids and find the 128 nearest neighbours for each centroid in bigram BoS space.
4. For each centroid, we then find the 4 closest stories from its 128 neighbours under the path-length normalized pairwise alignment distance. We avoid selecting duplicates by discarding stories with matching 10-grams.
5. Select 50 centroids (and their closest stories) based on the sum difference from the centroid to the closest stories.

For the token distance measure $d(x, y)$, we use

$$d(x,y) = \begin{cases} 0 & x = - \wedge y = - \\ \delta_x & x = - \wedge y \neq - \\ \delta_y & x \neq - \wedge y = - \\ \|x-y\|_2^2 & \text{otherwise,} \end{cases} \tag{7}$$

where δ_x, δ_y are gap penalties.

In choosing clusters, as well as our experiments with progressive alignment, we set $\delta_x = \delta_z \in \{125, 150\}$. For the Steiner distance, we set $\delta_x \in \{60, 75\}$ and $\delta_y \in \{180, 225, 300, 375\}$. We chose these gap penalties empirically based on preliminary analysis of the SBERT nearest neighbours of sentences as well as the resulting multiple alignments.

Algorithm 3 Hill Climbing Alignment

Given: A set of sequences $\mathcal{X} = \{\mathbf{x}_i\}_i$, and distance measure d
Initialize mean string $\mathbf{z}$
function HILLCLIMB($\mathcal{X}, \mathbf{z}, d$)
  for all iterations **do**
    for all sequences $\mathbf{x}_i \in \mathcal{X}$ **do**
      Compute pairwise alignments $M^i = \text{PAIRWISEALIGN}(\mathbf{x}_i, \mathbf{z}, d)$
    Compute proposal $\mathbf{z}' = \text{AVERAGEALIGNED}(M^1, \ldots, M^K)$
    Initialize list $Z = [\mathbf{z}']$
    for all indices $t \in [|\mathbf{z}|]$ **do**
      for all alignments M^i **do**
        Store tokens in $\mathbf{x}_i$ aligned to gaps between z_{t-1} and z_t in g_i^t
      for all elements $\mathbf{y}$ of the cartesian product of $g_1^t \times \cdots \times g_K^t$ **do**
        Set m to the average representation of $\mathbf{y}$
        Append addition proposal $[\mathbf{z}_{1:t}, m, \mathbf{z}_{t:|\mathbf{z}|}]$ to Z
      Append deletion proposal $[\mathbf{z}_{1:t}, \mathbf{z}_{t+1:|\mathbf{z}|}]$ to Z
    if no proposals improve the Steiner Distance **then return z**
    Set $\mathbf{z}$ to the proposal in Z with the lowest Steiner distance
  return z
function AVERAGEALIGNED($M^1, \ldots, M^K$)
  for all non-gap elements z_t in the alignments M_i **do**
    Set z_t to the average of all aligned tokens from each M_i
  return z

We compare multiple alignments based on the SP score and Steiner distance, and examine the output of each MSA algorithm by qualitatively evaluating the semantic closeness of alignments. For computing the MSA, we use the progressive alignment, iterative averaging, and hill climbing algorithm. For the IA algorithm, we initialize the mean sequence with the longest sequence $\mathbf{x}_i \in \mathcal{X}$. For the hill climbing algorithm, we initialize the mean sequence with two different configurations: either the mean sequence obtained from the progressive alignment or the IA algorithm.

6 Results

We find that the MSA algorithms perform better on the objective they optimize, as seen in Table 1. The progressive alignment outperforms the other approaches by getting a lower SP score, while the IA and hill climbing (initialized with the IA mean sequence) approaches perform well on the Steiner distance. Initializing the hill climbing approach with the mean sequence obtained from the progressive alignment results in an alignment that has a lower Steiner distance than progressive, but the SP score increases.

Algorithm	δ_x	δ_y	**SP Score**	**Steiner Distance**
Progressive	125	125	2,494,153.00	877,445.99
Hill Climbing (Progressive)	60	300	2,514,774.25	774,525.85
Iterative Averaging	60	300	2,643,249.25	716,350.33
Hill Climbing (IA)	60	300	2,647,854.25	709,076.44

Table 1: The SP scores and Steiner distances for each of the MSA algorithms. The progressive algorithm optimizes the SP score and achieves the lowest. The hill climbing algorithm initialized with the mean sequence obtained from IA obtains the lowest Steiner distance. We see the algorithms get stuck in local optima, as the hill climbing algorithm initialize with the progressive mean sequence obtains a much higher Steiner distance than if it had been initialized with IA.

We compare the method with the best SP score, the progressive alignment, with the best method in terms of Steiner distance, the hill climbing algorithm initialized with the output of iterative averaging (IA). The alignments obtained from the progressive alignments are longer and contain more gaps than the hill climbing (IA) alignments. This is shown in Figure 2, where the total number of columns is much larger for progressive than HC(IA), and the column densities are lower for progressive as well. We find

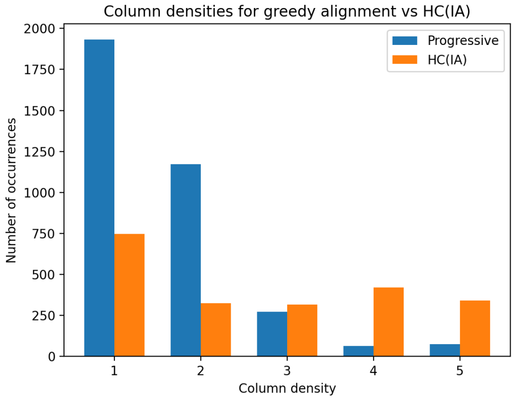

Figure 1: The progressive algorithm outputs alignments that are much sparser and longer than the best-performing method that optimizes the Steiner distance, hill climbing (IA). This is shown by a large number of columns from progressive alignments containing only a single non-gap element, and few columns containing more than 3 non-gap elements. These column density counts were obtained across all multiple alignments from the initial 50 clusters.

that the hill climbing algorithm aligns sentences that are not semantically similar, therefore we focus on analyzing the progressive alignments for evidence of storylines.

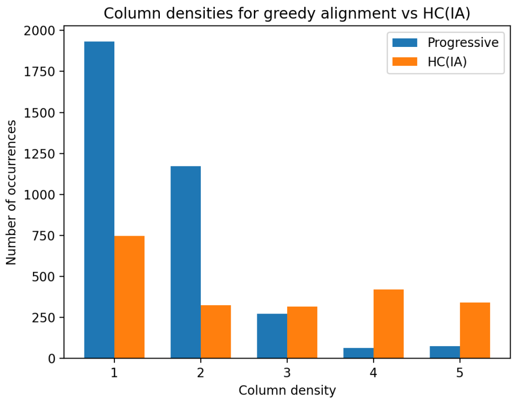

Figure 2: The progressive algorithm outputs alignments that are much sparser and longer than the best-performing method that optimizes the Steiner distance, hill climbing (IA). This is shown by a large number of columns from progressive alignments containing only a single non-gap element, and few columns containing more than 3 non-gap elements. These column density counts were obtained across all multiple alignments from the initial 50 clusters.

We observe that stories with very short sentences result in better qualitatively better alignments. This is corroborated by Figure 3, where the average distance from short sentences to their nearest neighbours is smaller than that of longer sentences. The shorter sentences may be better semantic matches, or there may be out-of-domain issues with the sentence representations from SBERT. The datasets SBERT was trained on were obtained from image captioning and other sources that may not transfer well to the particular setting of the subreddit where WritingPrompts was collected from; additionally, the average sentence lengths for the datasets SBERT were trained on were 14.1 and 22.3 [1, 9] whereas the average length of sentences in the WritingPrompts dataset is 28.4 [3]. We posit that SBERT has less exposure to longer, more complex (or ill-formed) sentences such as those found in WritingPrompts. The result of this bias is that shorter sentences are more likely to be matched, and stories that contain many short sentences (especially those with many trivial nearest neighbours) may be favored.

However, despite the sentence-level complexity of WritingPrompts, the stories themselves tend to have very simple structure. Due to their status as short stories, the traditional exposition, rising action, falling action, and conclusion segmentations do not fit very well. The short stories condense the structure

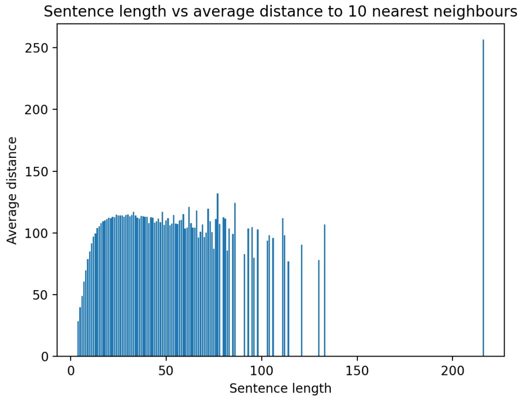

Figure 3: Longer sentences tend to be farther from their nearest neighbours. 2^{16} random sentences were sampled from WritingPrompts, and the plot shows their length versus the average distance to their 10 nearest neighbours. Distances were averaged across all sentences of the same length.

and combine or completely leave out parts. Additionally, as we truncate stories at 25 sentences, we may miss the conclusion as well as the falling action.

With the SBERT and data limitations in mind, we examine a particular set of 5 stories with nontrivial alignments as a case study. The prompt and first few sentences of each story are given in Figure 4. We find that the initial heuristics for finding sets of stories using the bigram BoS representations, and subsequently the pairwise alignment distances is effective at finding similar stories. As the stories in our case study cluster are all about outer space, we refer to them as the space story cluster. The multiple alignment for the cluster, in Figure 5, although sparse, shows some evidence of storylines. Indices 1-14 of the alignment appear to be the introduction of stories 1, 2, 3, and 5, with story 4 taking a less conventional structure and style. Due to their nature as short stories, much of the remainder of the stories fall under exposition: the stories unfortunately do not obey the rule "show don't tell" so it is difficult to distinguish exposition from action.

We present a good sub-alignment in Figure 6, which involves indices 11 and 12 in Figure 5. This particular alignment occurs early in the multiple alignment, and captures shared structure across four of the five stories and marks the end of the introduction section mentioned previously.

We also examine a poor sub-alignment in Figure 7, which involves indices 31-34 from the multiple alignment. The fragment from story 4 occurs towards the end that story's introduction, yet is aligned to sentences that occur after story 5's introduction. We hypothesize that story 4 is dissimilar to the other four stories, causing errors in the alignment.

The full multiple alignment for this cluster, as well as other clusters, can be viewed by following the instructions at the following site: github.com/dongruoping/story-clustering.

7 Discussion

We present a comparison of objectives and algorithms for performing multiple alignment, with the goal of inducing storylines. We find that progressive alignment results in the best alignment qualitatively. Analysis of the output of the progressive revealed some structure in stories, but the results were also obscured with noise from three possible sources: 1) suboptimal choices of story sets, 2) suboptimal alignments from the algorithm, and 3) behaviour of the sentence representations obtained from SBERT on out-of-domain data. For our next step, we will improve the accuracy and scalability of our alignment algorithm in order to apply it to even larger subsets of stories, addressing problems 1 and 2. We hypothesize that by increasing the number of stories aligned, we will be able to improve the quality of the alignments by increasing the probability that there is a story with a similar storyline.

References

[1] Samuel R. Bowman, Gabor Angeli, Christopher Potts, and Christopher D. Manning. A large annotated corpus for learning natural language inference. *CoRR*, abs/1508.05326, 2015. URL `http://arxiv.org/abs/1508.05326`.

[2] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners, 2020.

[3] Angela Fan, Mike Lewis, and Yann N. Dauphin. Hierarchical neural story generation. *CoRR*, abs/1805.04833, 2018. URL `http://arxiv.org/abs/1805.04833`.

[4] 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`.

[5] DF Feng and RF Doolittle. Progressive sequence alignment as a prerequisite to correct phylogenetic trees. 1987.

[6] Dan Gusfield. *Multiple String Comparison – The Holy Grail*, page 332–369. Cambridge University Press, 1997. doi: 10.1017/CBO9780511574931.017.

[7] François Petitjean, Alain Ketterlin, and Pierre Gançarski. A global averaging method for dynamic time warping, with applications to clustering. *Pattern Recognition*, 44(3):678–693, 2011.

[8] 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`.

[9] Adina Williams, Nikita Nangia, and Samuel R. Bowman. A broad-coverage challenge corpus for sentence understanding through inference. *CoRR*, abs/1704.05426, 2017. URL `http://arxiv.org/abs/1704.05426`.

[10] 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`.

Story 1
After hundreds of years of sending messages into the sky , humanity receives its first message from intelligent life . Decoded it simply says , " Be quiet before they find you . " " Zebin exclaimed as he received yet one more channel of communication from the Earth . Twenty years ago , the ambivalence over whether KIC 8462852 was in actuality an " alien mega structure " had finally come to an end after nearly 200 years of joint scientific endeavour by the leading lieges of the Earth . Since then , humanity had been trying with fervor to try and communicate with the star classified as a Dyson Sphere around 1480 light years away hoping that the far advanced civilisation might be generous enough to show the earthlings a way to solve their own energy crisis .
Story 2
Sunday Free Write : Leave A Story , Leave A Comment - New CSS Edition ! In 2056 NASA intercepted a frequency that was not of Earth . With its point of origin unknown they began to study it in an attempt to discover from whence it came . As it was studied it became known as the whoa signal , mockingly after the famous " wow !
Story 3
By 2345 humanity has colonized most of the solar system and have started sending probes beyond sol , when one day , a frantic interstellar message is received saying " do not leave the sol system " Listen well childen , for I have a Story of Old Earth to tell . Long ago , in ages past , the Men of ages past had but one True world , and had not yet learned to truely swim across the stars . In the nuturing embrace of Sol , humanity sowed life in the wake of the places Man gone .
Story 4
As the universe resonates for its final time , heat death approaching , the last energy of the universe is used to send a message . We knew something was going on , we knew something at any time would happen as time went on . Millennia after millennia we saw the lights fade out year by year pondering if new ones will be born to re-convey balance in the universe , but no the night sky kept getting darker star by star . We did n't care at first but the scientists all around our universe that we spread around knew something was wrong but were too busy with inorganically creating life on a possibly inhabitable planet around Zylon-B .
Story 5
Astronauts discover an abandoned space station orbiting a small moon . It is the far future . " Alright , we ' re here , " Captain Schiff said as the blocky craft approached the small disc-like object orbiting around Io , one of the five Galilean moons of Jupiter . An astronomer working for Global Colony Services , GSC , a young but highly profitable organization with fledgling colonies on the Moon , Mars , and in the asteroid belt , had spotted an irregular object in lunar-synchronous orbit that wasn ' t there last year . The GSC , or rather the league of nations sponsoring it , was concerned about a potential threat to the admittedly massive investments it had made in the asteroid belt , so it had sent the great ship MecaBubo 1 out to investigate .

Figure 4: The first few sentences from each of the 5 stories in our space story cluster analysis. The prompt is for each story is given in italics. All stories have a space-related theme.

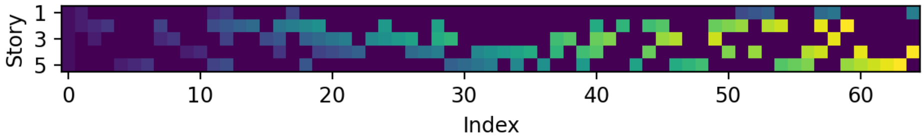

Figure 5: The multiple story alignment of the space story cluster, obtained via progressive alignment. We refer to the story in the top-most row as story 1, down through the bottom-most row as story 5. Each colored square corresponds to a sentence from the corresponding story, while a dark square corresponds to a gap. Sentences within a row proceed in order of their appearance in the story; only gaps are inserted within each row. An identical dummy beginning-of-story sentence token was inserted at the start of each story, yielding the full column of matches at index 0.

Story 1	**Story 2**	**Story 3**	**Story 5**
Twenty years ago , the ambivalence over whether KIC 8462852 was in actuality an “ alien mega structure ” had finally come to an end after nearly 200 years of joint scientific endeavour by the leading lieges of the Earth .	NASA discovered the signal was encrypted like nothing they had ever dreamed of ; the discovery of the encryption itself set technology hundreds of years ahead of where it once was .		An astronomer working for Global Colony Services , GSC , a young but highly profitable organization with fledgling colonies on the Moon , Mars , and in the asteroid belt , had spotted an irregular object in lunar-synchronous orbit that wasn ’ t there last year .
Since then , humanity had been trying with fervor to try and communicate with the star classified as a Dyson Sphere around 1480 light years away hoping that the far advanced civilisation might be generous enough to show the earthlings a way to solve their own energy crisis .	It sparked the golden age of exploration in our solar system ; Ceres , Vesta , Hektor , Thisbe , Diotina , Fortuna were among many asteroids in the asteroid belt that were to be mined and inhabited ; the once failed colonization of Mars was reattempted and achieved , Europa of Jupiter , Titan of Saturn and Triton of Neptune all were to be colonized and inhabited ; Man had even reached as far as the Oort Cloud in the outer reaches of our solar system as early as 2096 .	From the Illiuminious research Stations of Mercury , to the Ruins of Old Chicago , Humanity had taken to the space around Sol with great bravado .	The GSC , or rather the league of nations sponsoring it , was concerned about a potential threat to the admittedly massive investments it had made in the asteroid belt , so it had sent the great ship MecaBubo 1 out to investigate .

Figure 6: An example of a good alignment between 4 story segments from the multiple alignment. The first sentences from stories 1, 2, and 5 detail an event which sparked the following sentences. The next sentences from all four stories mention a form of exploration, investigation, or expansion.

Story 4	**Story 5**
We tried making stars but they all died out faster then the stars above but it seemed somehow to make our time last just a little bit longer .	Not even great technicians at that, since the ship could handle damn near everything itself .
We had time to formulate whats going on and a joint meeting of scientists from all around the corners of the universe .	We were there as human eyes and ears , witnesses to send back any special details , in case they were necessary , and mechanics to fix the ship in case one of those " one in a million " scenarios cropped up (which seemed to be happening more and more often since children of Adam and Eve began to colonize their small corner of the cosmos .
Some full robotic now , half robotic and humanoid , one was full humanoid from Earth 1B and some that looked so foreign from living on a planet with no sunlight for 3 years at a time with the vegetation growing above 40 feet in the air .	
Enough about the scientists though .	Schiff was our leader , and a good enough man , if a bit formal .

Figure 7: An example of a poor alignment between two story segments from the multiple alignment. The sentences are weak semantic matches, although both segments serve as exposition in their respective stories.