Team Skeletor at Touché 2021

Argument Retrieval and Visualization for Controversial Questions

Kevin Ros*, Carl Edwards*, Heng Ji, and ChengXiang Zhai

24 September 2021

*Equal contribution

Presentation Outline

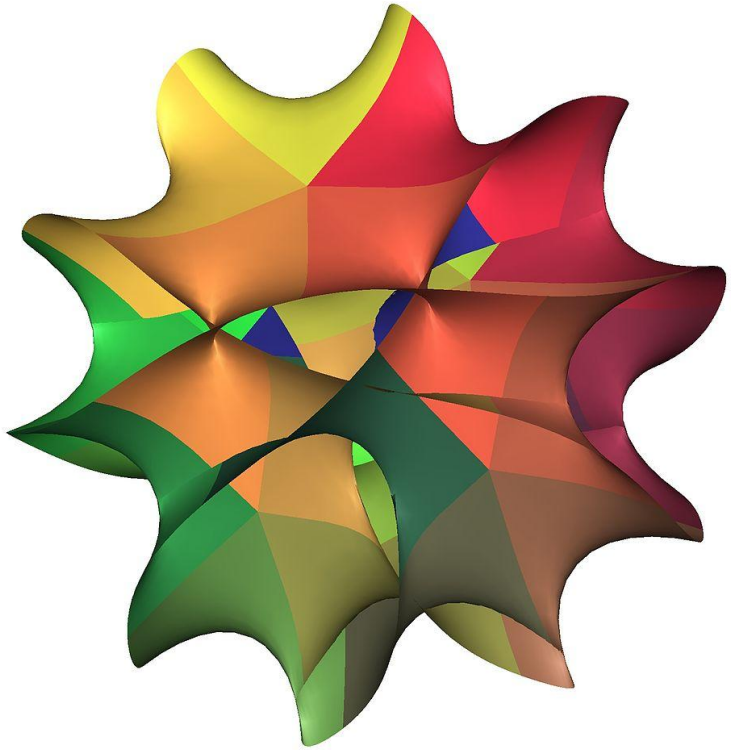
1. Argument retrieval methods
    a. Baseline and Manifold
2. Argument retrieval results
3. Argument retrieval visualization
    a. BM25 and Manifold
4. Future work

Methods: Baseline

- BM25
  - Run name: **bm25**
  - Standard implementation using Pyserini (with the default settings)
  - k1 and b fine-tuned on last year's topics and relevance scores
- Semantic
  - Run name: **semantic**
  - Question-answer similarity via msmarco-distilbert-base-v3 (BERT-based encoder)
  - Split each argument into similar-length passages (~200 words)
    - For a given topic, we retrieved the k=1000 most similar passages, and ranked the arguments using the maximum score of its passages
- Interpolated
  - Run name: **bm25-0.7semantic**
  - Interpolate the scores from the previous two methods
  - Coefficient fine-tuned on last year's topics and relevance scores

Methods: Manifold Approximation-based Reranking

- Argument importance estimated using manifold approximation
  - Manifold - topological space that locally resembles Euclidean space
- Algorithm based on Uniform Manifold Approximation and Projection (UMAP)
- Key idea: assume the top n initially-retrieved arguments are relevant, and perform manifold-based approximation on the nearest neighbors of those top n arguments
  - Gives us a score between an initially-retrieved argument and its neighbor(s)
  - We believe strong, complete, and relevant arguments will have a higher aggregated score over irrelevant arguments



Methods: Manifold Approximation-based Reranking

$$\rho_i = min\{d(x_i, x_{i_j}) | 1 \leq j \leq k, d(x_i, x_{i_j}) > 0\}, \tag{1}$$

$$\sum_{j=1}^{k} \exp(\frac{-max(0, d(x_i, x_{i_j}) - \rho_i)}{\sigma_i}) = \log_2(k), \tag{2}$$

$$w((x_i, x_j)) = \exp(\frac{-max(0, d(x_i, x_{i_j}) - \rho_i)}{\sigma_i}). \tag{3}$$

(1) → the distance to x_i's closest neighbor
(2) → sigma is calculated to smooth and normalize the distances to the nearest neighbors
(3) → the weight of the edge from x_i to x_j (probability that this edge exists)

Methods: Manifold Approximation-based Reranking

- Initial retrieval using interpolated BM25 and semantic search.
- For each topic:
  - Assume the top 3 arguments are relevant
  - Find 50 nearest neighbors of the arguments
  - Compute manifold weights (previous slide) for each neighbor of each argument
  - Sum the weights across all top arguments
- Arguments are reranked by their incoming sum of weights
  - Two different approaches (by run name):
    - **manifold** → reranking all arguments according to their score
    - **manifold-c10** → only reranking arguments that appear in the top 10 from the initial run

Results

Table 1
Performance on Touché 2021 and 2020 Topics and Relevance Scores

Run Name	Relevance nDCG@5	Quality nDCG@5	2020 nDCG@5
bm25	0.661	0.822	0.6214
semantic	0.570	0.671	0.3475
bm25-0.7semantic	0.667	0.815	0.6347
manifold	0.666	0.827	0.5417
manifold-c10	0.666	0.818	0.5906

- Having a strong initial search method (e.g. BM25) is important.
- Interpolation with semantic search is helpful for relevance but not quality.
- Manifold improves on the quality of BM25-semantic even though semantic initially weakens the results.

Visualization

- Ranked lists of arguments may not be an optimal display method, especially during live debates
- Utilizing the args.me corpus, we explore various visualization techniques to help retrieve and summarize debates in real-time
  - Two techniques, each based on BM25 and manifold-based approximation, respectively
- Demonstrate the visualizations using the transcript of a debate and its relationships to the args.me corpus
  - Bill Nye Debates Ken Ham (about evolution and creationism), 110:53-114:04
  - Publically available on YouTube, and our paper contains the relevant transcript

Visualization: BM25

- YouTube provides the transcript in few-second windows, consisting of ~0-8 words per window
- As the debate progresses
  - Define a lookback size n=5
  - For the most recent lookback size windows
    - Use the text in the windows as a query, and search over the args.me corpus
    - Record the rankings of the top k=20 returned arguments
- Plot the ranking changes of the most frequent arguments as the debate progresses

Visualization: BM25

Table 2
Arguments from Figure 1

Argument ID	Representative Topic(s)
S4fde9bb-Aef3913d8	description of a bicycle incident
S9a8b0a09-A22358c86	understanding scriptures, the gospel, God
S690aacea-A986a10d7	creator of universe, infinite power, God
S23dda237-A69f9884f	having unlimited power, omnipotence of God
S5059e885-Abe1aa26d	the justness of God

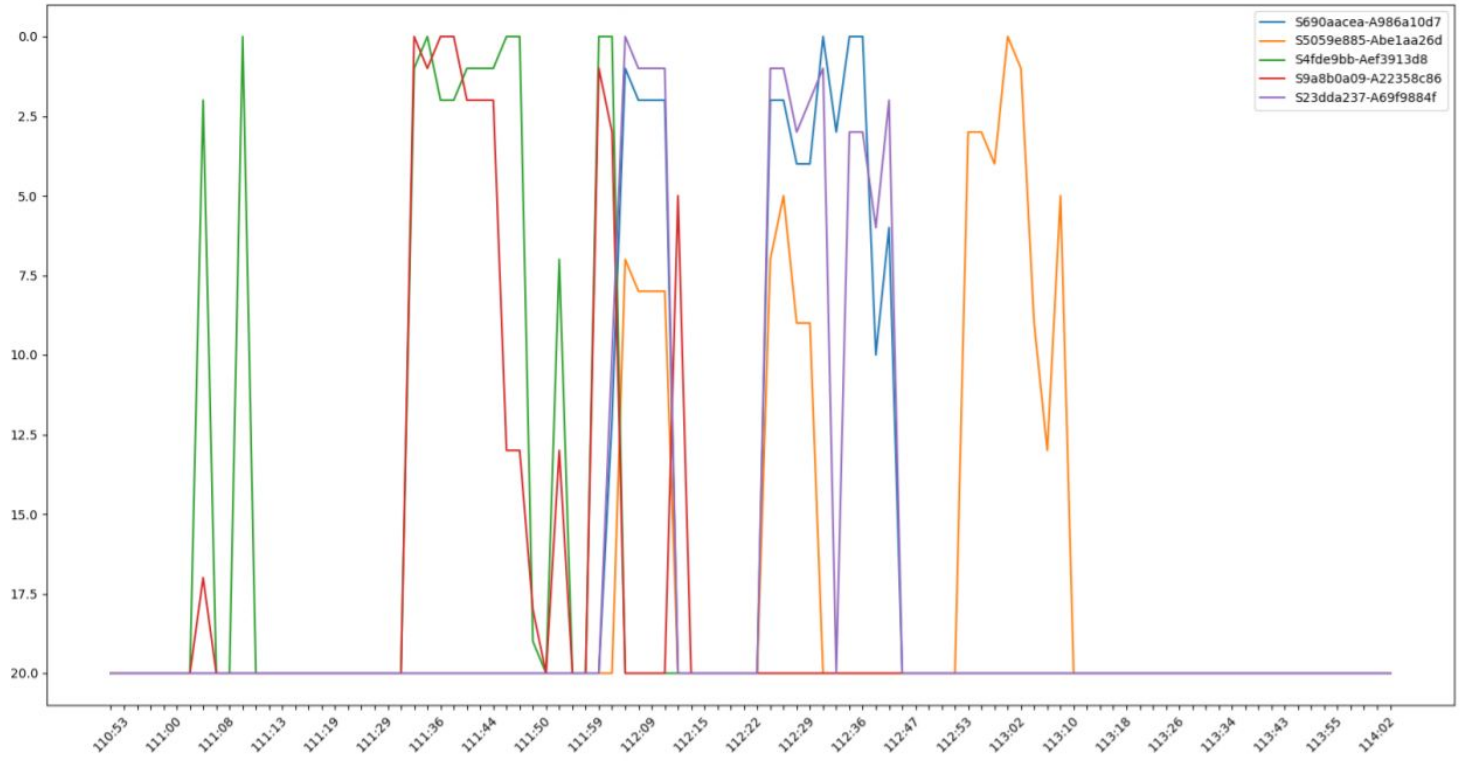
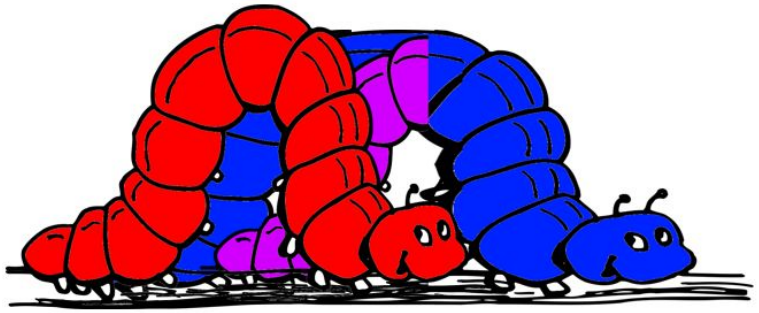


Figure 1: Rankings of the five most frequent arguments over the transcript window 110:53 - 114:04.

Visualization: Manifold

- Create “caterpillar embeddings” of the debate transcript.
  - Use a sliding window which expands and then contracts.



... we believe that in the beginning god created the heavens and the earth and I believe our uh creationist astronomers would say yeah you can observe the universe expanding uh why god is doing that in fact in the bible it even says he stretches out ...

- Split arguments as retrieved in BM25 into sentences.
- Encode sentences from argument and caterpillar embeddings using BERT-based sentence encoder.
- Project to two dimensions using UMAP.

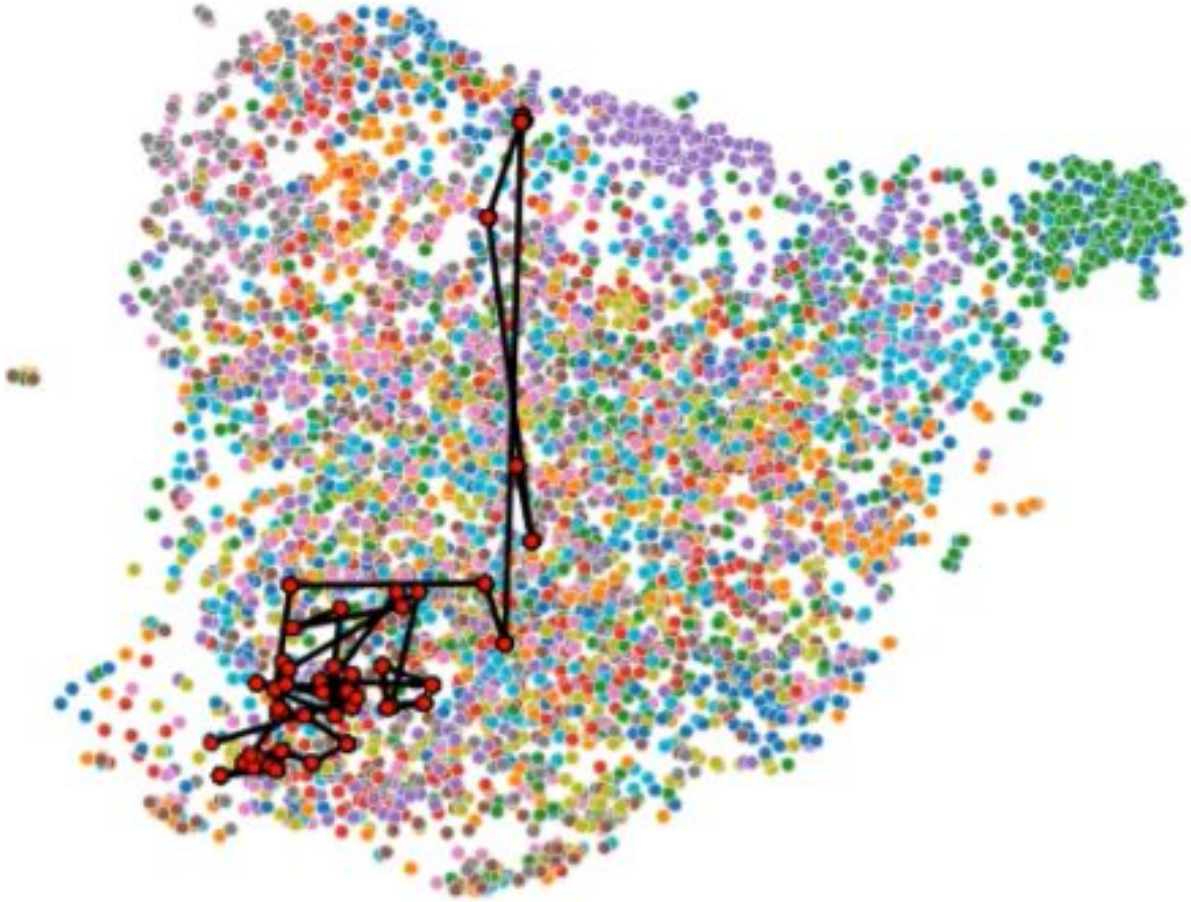
Visualization: Manifold

The argument begins in the middle.

(a) Initial frame. Time: 111:10

Visualization: Manifold

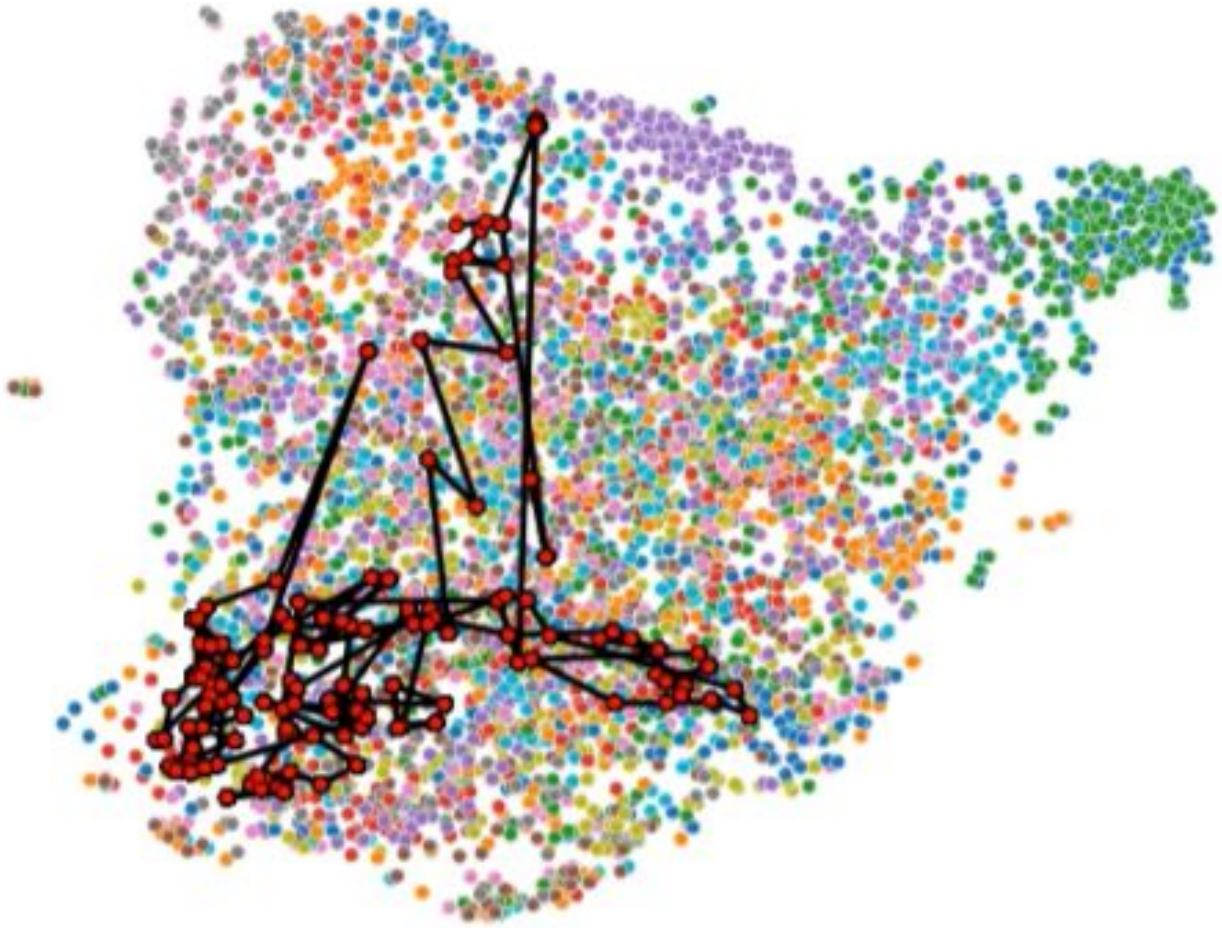
The argument quickly moves to the lower left quadrant, which we find to signify the creation of the universe and heavens, particularly in relation to God



(b) Time: 111:53

Visualization: Manifold

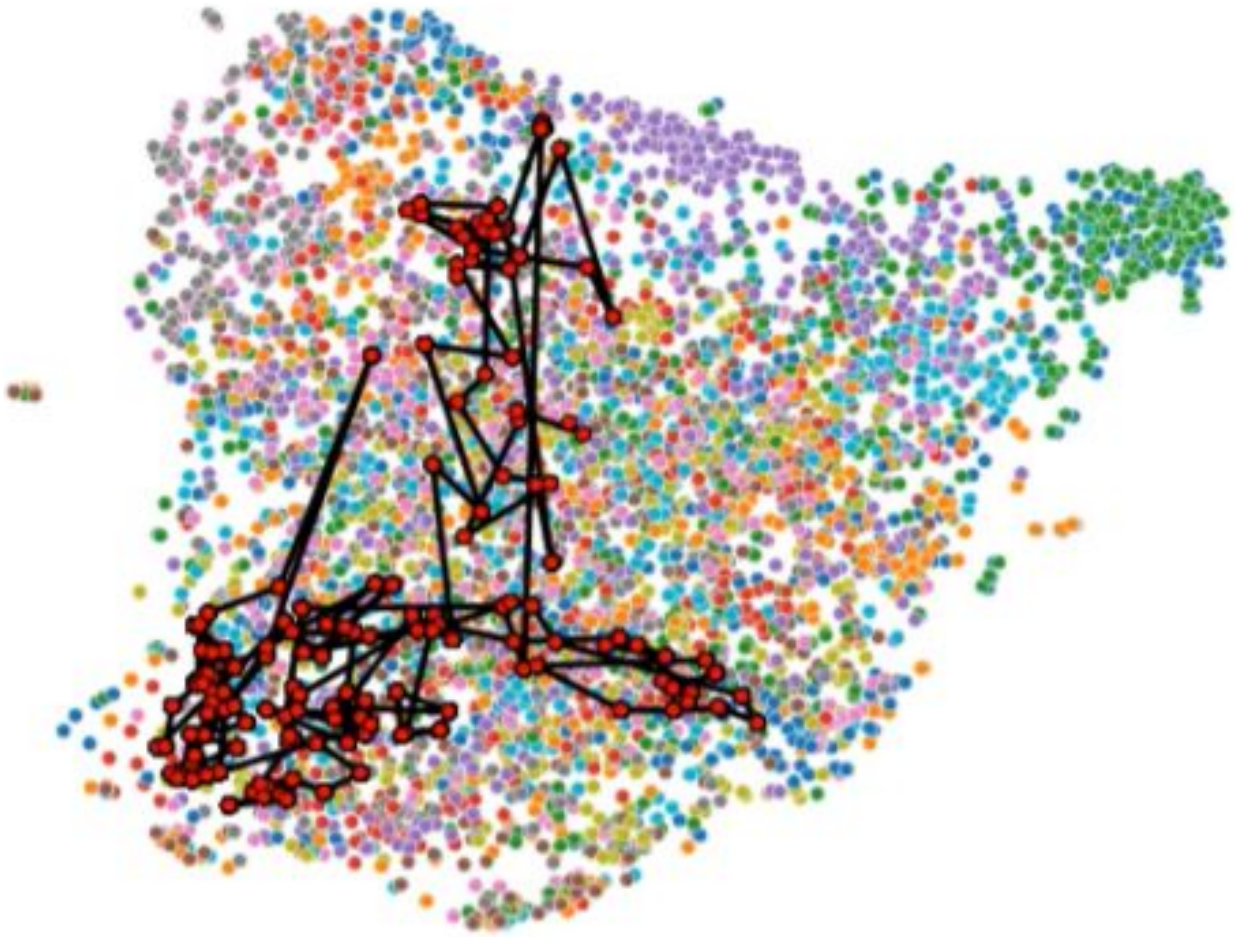
The path briefly moves to the right, when the debate focuses more on the omnipotence and omniscience of God.



(c) Time: 113:27

Visualization: Manifold

The debate moves upward when the discussion changes to physics, life science, and astronomy.



(d) Final frame. Time: 114:04

Future Work

- Further explore the effects of manifold-based reranking on arguments
  - Let more arguments contribute to the manifold weights
  - Positive and negative examples, statistical significance
  - Different corpora
- Expand on visualization techniques
  - Allow users to define topics
  - Better match spoken debate domain with written debate domain

Thank You!
If you have any questions, please email
Kevin Ros: kjros2@illinois.edu
Carl Edwards: cne2@illinois.edu