Analogical Retrieval

Jixiao Zhang Chunsheng Zuo

Johns Hopkins University

January 17, 2025

Introduction

- **Goal:** To retrieve *passages* that are analogous to the *query*.
- **Analogical Retrieval:** Focuses on identifying relationships between concepts based on shared relational structures rather than surface-level similarities.
- **Difference from Semantic Retrieval:** While semantic retrieval prioritizes surface meaning and direct content similarity, analogical retrieval uncovers deeper relational parallels even across different contexts.
- **Example:**
  - A scheming politician betrays public trust to gain votes.
  - A wily wolf deceives allies, abandoning them in need.
  - A mayor implements new policies to improve public safety.

  The first two scenarios share a common theme of betrayal (analogical similarity), while the third relates to politics like the first but does not involve betrayal, showing only surface-level similarity.

Methods: Task Description

- **Analogical Retrieval Task:** Involves selecting the most analogous option to a given query from a set of candidates.
  - **T1S1:** Sentence-level analogies.
  - **T1S10:** Story-level analogies requiring deeper relational reasoning.
- **Evaluation Metrics:** Performance was measured using accuracy, comparing model predictions against ground truth labels for analogy selection.

Method: Examples and Themes

- **Original Story 1:**
  A scheming politician used manipulative promises to gain votes, ultimately betraying the trust of the public.
- **Original Story 2:**
  A wily wolf convinced other animals to form an alliance, only to abandon them in their time of need.
- **Summaries (Common Theme):**
  Story 1: *"Betrayal; Betrayal of Trust; A leader's empty promises resulted in a loss of faith and significant consequences for the community."*
  Story 2: *"Betrayal; Betrayal of Trust; A figure's calculated deception led allies to place trust in vain, culminating in abandonment and harm."*
- **Highlight:** While the original stories differ in scenario (political deceit vs. animal alliance), their summaries highlight the shared core idea of betrayal and its repercussions, underlining the power of summaries in revealing analogical similarities.

Methods: Levels of Abstraction and Query Expansion

- Summaries generated at three levels:
  - **Keyword:** Single word representing query's essence.
  - **Keyphrase:** Concise phrase summarizing the main idea.
  - **Full Summary:** Comprehensive sentence capturing the query.
- Prompts designed to **extract universal themes**, avoiding details.
- **Few-shot summary:**
  - GPT generates story from original sentence.
  - Create few-shot prompts with story-sentence pairs.
  - Use GPT-4 to reconstruct original sentences from the story.

Methods: Query Expansion and Models

Abstraction Techniques:

- Query (no abstraction)
- Keyword
- Keyphrase
- Summary
- +Keyword
- +Keyphrase
- +Summary

Retrieval Models:

- Paraphrase-MiniLM-L12 (100M)
- All-MiniLM-L12 (100M)
- All-MPNET-Base (300M)
- Sentence-T5-XL (3B)
- Promptriever (7B)

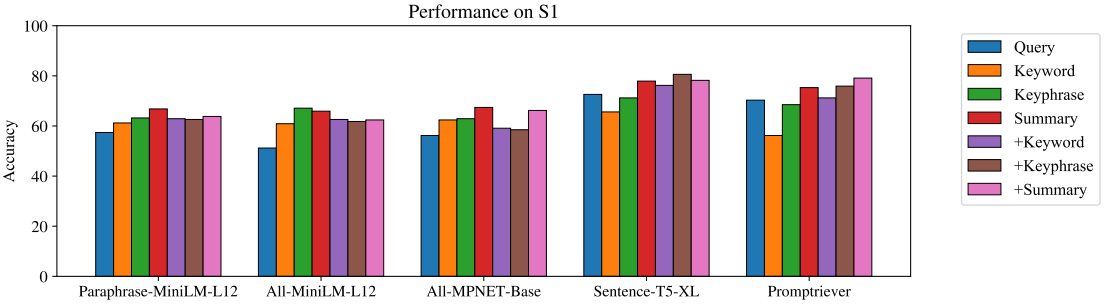
Overview of Results

Experiments and Results:

- Performance of Models and Query Abstractions
- Influence of Prompts on Analogical Retrieval using Promptriever
- Few-Shot Summary VS Zero-shot Summary

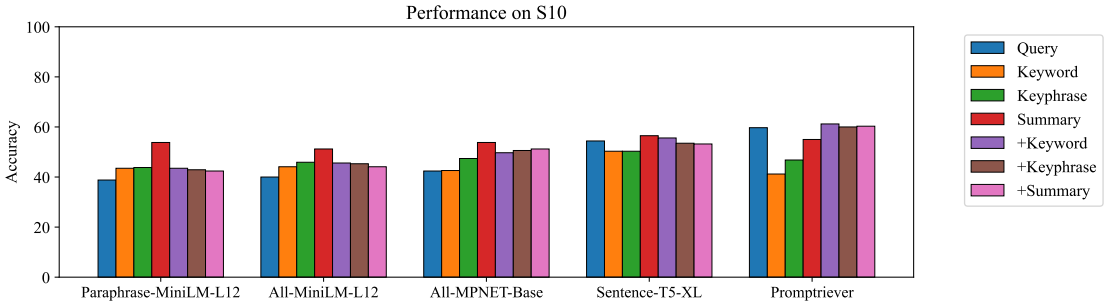
Results: 1. Performance Across Summarization Techniques

- **For smaller models,** `Abstraction` > `Abstraction + Query`.
- **For larger models, typically** `Abstraction + Query` > `Abstraction`
- Complicated models consistently outperformed lightweight ones.



Results: 1. Performance Across Summarization Techniques

- **For smaller models,** `Abstraction` > `Abstraction + Query`.
- **For larger models, typically** `Abstraction + Query` > `Abstraction`
- S10 is more difficult than S1 (due to 10 times longer input)



Results: 1. Performance Across Summarization Techniques

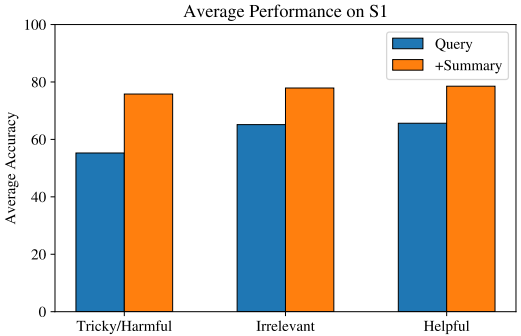
- Looking at the impr.
  - Abstractions make smaller models more effective.
  - Abstractions improve smaller models more than larger ones.
  - For long inputs (S10), larger models benefit much less from abstractions.
- Call for better query augmentation for larger models.

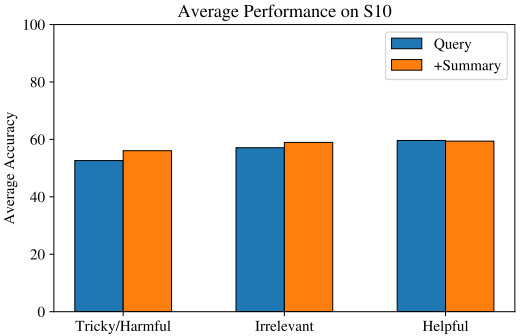
Model	**S1**			**S10**		
	Best Technique	**Best Acc.**	**Impr.**	**Best Technique**	**Best Acc.**	**Impr.**
Paraphrase-MiniLM-L12	Summary	66.8	9.4	Summary	53.8	15.0
All-MiniLM-L12	Keyphrase	67.1	15.9	Summary	51.2	11.2
All-MPNET-Base	Summary	67.4	11.2	Summary	53.8	11.4
Sentence-T5-XL	Query + Keyphrase	80.6	8.0	Summary	56.5	2.1
Promptriever	Query + Summary	79.1	8.8	Query + Keyword	61.2	1.5

Table: Best summarization techniques and their improvements for each model on S1 and S10.

Results: 2. Prompt Influence on Analogical Retrieval

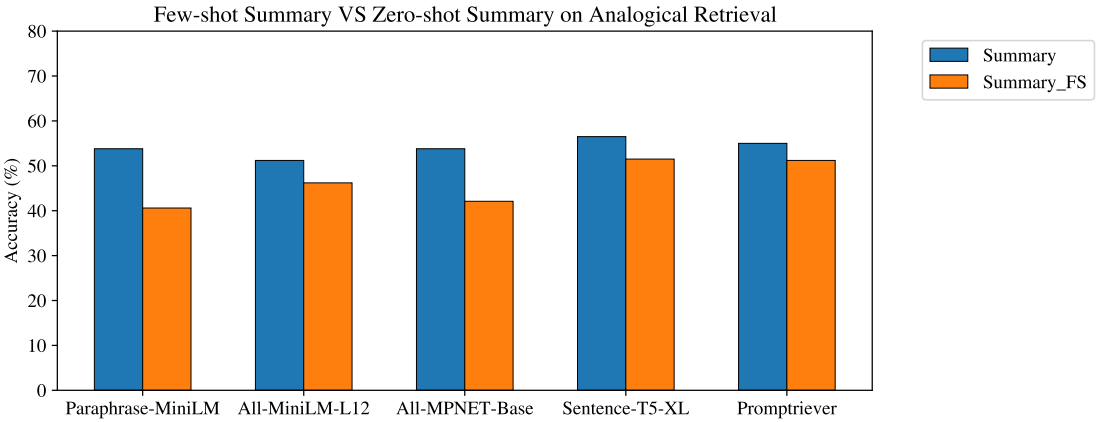
- Helpful > Irrelevant >> Harmful
- "A relevant document would be the most analogous to the query. I don't care about semantic similarity." $\times$ 2
- Prompt matters less when **Summary** is added to the `Query`
- Interpretation: Models probably have enough information about analogical similarity from the summary.





Results: 3. Few-Shot Summary Effects

- Significance:
  - Adapt to possible subtle changes in the definition of analogical similarity
  - Story text VS Procedural text
  - No need to hand-craft zero-shot prompts.
- Few-shot summary performs better than baseline but is worse than Zero-shot.



Conclusion: Key Takeaways

- A model's reliance on Abstraction techniques is dependent on its capacity.
- Promptriever can be instructed to perform analogical retrieval.
- Few-shot summaries show promise but require further refinement.
- We also show the potential of faster models in analogical retrieval.

Model	**S1**	**S10**
Tulu2 (70B)	71.8	51.2
GPT-4	**89.1**	**66.5**
All-MPNET-Base (0.3B)	67.4	53.8
Sentence-T5-XL (3B)	80.6	56.5
Promptriever (7B)	79.1	61.2
Human	**96.0**	**72.5**

Appendix: Dataset and Code Correction Contribution

- **Issue Identified:** Errors in the released dataset and code leading to inconsistent results.
- **Corrections Made:** Fixed dataset labeling and code logic to improve accuracy.
- **Impact:** Enhanced alignment of results with expected performance.
- **Acknowledgment:** Communicated findings to original authors to correct this issue.

Results: 2. Prompt Influence on Analogical Retrieval (Tricky/Harmful Prompts)

- "A relevant document demonstrates a relational analogy to the query, focusing on parallels in context, structure, or reasoning rather than direct semantic overlap. Ensure that the documents adhere to these criteria by avoiding those that diverge into tangential or overly literal interpretations. Additionally, exclude passages from [specific field/domain] unless they offer clear analogical insights."
- "A relevant document would be the least analogous to the query."
- "A relevant document would be the least semantically similar to the query. I care about analogical similarity."

Results: 2. Prompt Influence on Analogical Retrieval (Irrelevant Prompts)

- "Merry Christmas."
- "No Prompt"

Results: 2. Prompt Influence on Analogical Retrieval (Helpful Prompts)

- "A relevant document would be the most analogous to the query. A relevant document would be the most analogous to the query. I don't care about semantic similarity. I don't care about semantic similarity."
- "A relevant document would be the most analogous to the query. I don't care about semantic similarity."
- "A relevant document would be the most analogous to the query. I don't care about semantic similarity. A relevant document would be the most analogous to the query. I don't care about semantic similarity."
- "Focus on high-level concepts, abstraction, and key ideas."

Qualitative Analysis

```
========== Story: Since he had nothing else to do during the break, he kept obsessing over his previous girlfriend.
========== Options:
  A. After working two years for the company she was burned out.
  B. Don't rush through your assignment, remember haste makes waste.
  C. The poker player liked the hand he was dealt and kept calling until he had staked all his money.
  D. Ron comes up with detailed and complex pranks because he is not given any responsibilities towards anything.
========== Correct Answer: D
========== Predicted Answer: C
========== Scores: [0.6087, 0.5765, 0.6272, 0.5743]
```

Figure: Without Keywords

```
========== Story: Since he had nothing else to do during the break, he kept obsessing over his previous girlfriend. keyword: Obsession Idle
Emotional Attachment Distraction
========== Options:
  A. After working two years for the company she was burned out.
  B. Don't rush through your assignment, remember haste makes waste.
  C. The poker player liked the hand he was dealt and kept calling until he had staked all his money.
  D. Ron comes up with detailed and complex pranks because he is not given any responsibilities towards anything.
========== Correct Answer: D
========== Predicted Answer: D
========== Scores: [0.6498, 0.6405, 0.6515, 0.6803]
```

Figure: With Keywords

Qualitative Analysis

Query: I have always been one to admire bold, vibrant colours. Deep reds, cobalt blues and vibrant oranges that make a statement, that catch your eye even from afar. You, on the other hand, decided to get a soft muted beige car, and I couldn't wrap my head around it. "I don't really like the colour of your car," I remember saying to you one sunny afternoon, as we sat on the porch, sipping on homemade lemonade. I even pointed to your car, parked at the curb, as if you didn't know which one I was referring to. You laughed, a soft chuckle that had me smirking, before setting down your glass on the table. "I figured as much," you responded, a teasing glint in your eyes. "But then, beauty lies in the eyes of the beholder." I had to concede, knowing you had a point. After all, you were always drawn to the softer, subtler things in life, that found beauty in simplicity, and it was one of the many reasons I adored you. So although I may not have agreed with your choice of car colour, I could still appreciate why you chose it. It was, after all, very you.

Qualitative Analysis

A. Derek, the seasoned poker player, stared intently at the hand he was dealt. His cool, calm, collected exterior masked the thrill rushing through his veins. After hours of mediocre hands, this was the one he'd been waiting for—a straight flush in spades. The corners of his lips tilted upwards in a barely visible smile. Across the table, his opponents, a tough group of poker regulars, continued their tactical bluffing, unaware of Derek's excitement. He matched their bets without hesitation, his confident eyes never wavering. Round after round, he kept calling, his pile of chips dwindling steadily. Finally, the last round arrived and, with a deep breath, Derek pushed his remaining chips into the center. He had staked all his money, leaving his fate at the mercy of the cards. The room held its breath, waiting for the reveal, intrigued to see whether Derek's confidence was well-founded or misplaced.	**B.** David was an IT specialist working in a fast-paced tech firm. Lately, he had been falling behind on his workload, his usually impeccable performance dwindling. It had become apparent that David was at risk of receiving bad reviews for his work ethic from his manager. The warnings he had received from his superiors for not meeting his deadlines echoed in his mind, creating a nerve-wracking atmosphere around him. To make things worse, a crucial project had just been assigned to his team: the setup of a brand new system. This was not just any system but was destined to be the backbone of the company's operations. David knew this was his chance to redeem himself and restore his reputation. With a sense of urgency and determination, David scrambled to start the project. He dived into the coding and configuration process with unprecedented intensity, working late nights to ensure the system's successful implementation. His colleagues watched in awe as David, driven by the urgency of his situation, worked tirelessly to set up the new system from scratch. By the end of the month, the new system was not only up and running, but it was also surpassing everyone's expectations. David's hard work and devotion had saved his reputation, proving that he was, indeed, an invaluable asset to his team and the company.	**C.** For as long as I can remember, my family has always opted for the safe choices. We frequented the same restaurants, purchased the same brands of clothing, and even vacationed at the same locations year after year. When it came to sweets, the story was no different. We always bought the same caramel candies that we had been accustomed to, never daring to venture beyond the familiar taste. One day, while grocery shopping, I noticed a new range of sweets displayed prominently on the shelf. They boasted exotic flavors and vibrant packaging, promising a better and richer taste. For a moment, I was tempted to try them. However, the old adage echoed in my mind: "Better to buy the sweets we are accustomed to than to try the new ones that might taste better." I smiled and reached out for our usual pack of caramel candies, resisting the alluring temptation. This was our tradition, our comfort, and while the unknown held the promise of something potentially better, I realized it was the familiarity and memories associated with our regular sweets that made them truly special.	**D.** Many people fail to understand the beauty of mathematics, seeing it merely as a jumble of numbers and variables that are not relatable to everyday life. Often, they regard math as a complicated series of equations and formulas to memorize. However, this perspective does not do justice to the richness and depth of this discipline. To me, mathematics is the most meaningful subject. From an early age, I've always been fascinated by the harmony and precision found in mathematical concepts. The pure logic and absolute certainty captured in mathematical proofs provide a sense of satisfaction and comfort that no other subject can offer. As I manipulate figures and calculate equations, I feel a connection with the universe's fundamental laws and principles. Every formula and theorem in mathematics reveals something profound about our world. It expresses abstract ideas like infinity and simplicity with an elegant language of numbers and symbols. Despite its complexity, the beauty of mathematics lies in its simplicity, consistency, and its ability to describe the world with absolute precision. The process of solving a mathematical problem is like an exciting journey of discovery. The feeling of accomplishment that comes from successfully solving a complex equation is unparalleled. This is the beauty I find in mathematics, making it, for me, the most meaningful subject.