Does Structure Matter? Leveraging Data-to-Text Generation for Answering Complex Information Needs

Hanane Djeddal, Thomas Gerald, Laure Soulier. Karen Pinel-Sauvagnat, Lynda Tamine
Sorbonne Université, CNRS, Institut des Systèmes Intelligents et de Robotique (ISIR), in collaboration with Institut de Recherche en Informatique de Toulouse (IRIT)

Context, Motivation and Objectives

- Planning encoder-decoder: Encodes each document \(d \in D\) concatenated with the query \(q\) and decodes a plan \(p\).

\[ L_{\text{planning}}(q, p) = \sum_{k=1}^{K} \prod_{i=1}^{H_k} P(h_{k,i} | a_{k,i}, p, D_k) \]

- Content generation encoder-decoder: Encodes each heading \(h_{k,i}\) in the plan \(p\), concatenated with the embedding of the document list \(D_k\) and decodes an answer \(a_{k,i}\).

\[ L_{\text{content}}(q, h_{k,i}, \theta) = \sum_{k=1}^{K} \prod_{i=1}^{H_k} P(a_{k,i} | h_{k,i}, \theta, D_k) \]

- Final Training Loss:

\[ L = L_{\text{planning}}(q, p) + L_{\text{content}}(q, p, a) \]

Evaluation Setup & Results

- Data-Set

- Redefined the inputs and outputs of TREC CAR

- Plans only (Planning Module)

- Structured or plain answer (Content Generation Module)

- Results

- Planning models produce longer text (200 tokens), and do well in semantic evaluation metrics.

- On the plain answers setting, our models are the most effective, which confirms the importance of structure prior even if it’s not explicitly present in the final output.

- Our plans cover more facets, in correct order with a better relevant semantics.

Conclusion

- Contrary to simple information needs where a short and concise answer is sufficient; complex information needs require a detailed answer that tackles different aspects of the topic.

- Different solutions exist to present a search results:

  - Conversational search is more adapted for open-ended questions:
    - The user can express their needs in natural language.
    - Multi-turn dialogue.
    - Clarification questions.
    - Returning answers in natural language.

- Provide a complete and structured answer in natural language that covers the multiple facets of an open-ended query, using as input an initial ranked list of documents.

- Solve complex information needs with generative models, particularly from the perspective of data-to-text generation.