June 2021

Machine-learning Based Automatic Formulation of Query Sequences to Improve Search

Massimiliano Ciaramita

Michelle Chen Huebscher

Benjamin Boerschinger

Yannic Kilcher

Machine-learning Based Automatic Formulation of Query Sequences to Improve Search

ABSTRACT

People use search engines to look up information on the Internet, using search queries related to their information needs. This disclosure describes the use of machine learning techniques, including supervised learning and reinforcement learning to train a search agent to search deeper for better, more accurate, better supported answers by interacting with the search engine. The interaction mimics strategies utilized by human experts to carry out accurate web search. The search agent can be modular, and to provide answers to a user query, performs operations such as formulation of new queries in a sequence, analysis of intermediate results, and selection of results based on a chosen success metric that can take into account factors such as accuracy, diversity, presence of justification, etc.

KEYWORDS

- Web search
- Search engine
- Query reformulation
- Search session
- Sensemaking
- Reinforcement learning
- Monte Carlo tree search
- Context free grammar
- Information retrieval
- Supervised learning

BACKGROUND

People use search engines to look up information on the Internet, using search queries related to their information needs. One of the most common use cases is searching for information on the web in response to a query or question. The response to the query may include the information provided in the form of short textual answers or documents, possibly including other media content (images, video etc.). Academic work (such as [1]) defines a simple action space for a reinforcement learning agent, e.g., based on user searches for a particular term from the query within the search results.

DESCRIPTION

Users that experts at searching are often able to find better search results by using creative reasoning, critical thinking, query reformulation using knowledge of the search engine query syntax, and synthesis of the results. Such interactive web search is an example of learning from interaction, characterized by exploration and delayed rewards. With this intuitive understanding of search, reinforcement learning techniques can be applied to search deeper for better, more accurate, better supported answers by automatic interaction with the search engine (e.g., automatic answer evaluation and query refinement), in ways that mimic web search as conducted by human experts. The approach described herein addresses the problem of the complexity/expressivity of the action space by defining it as a grammar that can express a vast range of types of actions, from simple linear generation of queries, to queries that utilize special search operators and also by allowing expression of templates of actions (e.g., find another answer, find a given answer in another document).

This disclosure describes techniques to improve the quality of search results by using a search agent that is based on machine learning. The search agent is configured to automatically formulate new queries in a sequence, analyze intermediate results, and choose the best result as per the chosen success metric. For example, a user may enter the query “deepest lake in the northwest usa” via a search engine or other interface.

Instead of returning a set of results as search engines typically do, the search agent as described herein is configured to perform a deeper search that optionally includes query reformulation, to identify a better result. For example, a default set of results from a search engine may return pages or websites related to “great slave lake,” which while the largest lake in the Northwest, is actually in Canada, and therefore, doesn’t match the user’s information need. On the other hand, with a search agent as described herein, query reformulation strategies, e.g., “deepest lake in the northwest usa -canada” (where the search engine syntax of the minus sign is utilized to exclude results that meet certain criteria) can yield websites for “crater lake” which is the correct result.

The search agent can be implemented as a modular structure composed of several machine-learned components, and can be trained using a combination of supervised learning and reinforcement learning techniques (e.g., following the approach described in [2]). Formulation of new queries is performed using a bidirectional encoder transformer model (e.g. BERT) that encodes the search state. A Context Free Grammar is used that defines a structure action space that matches the search engine query syntax and allows flexible representations of different types of query templates, a Monte Carlo Tree Search (MCTS) that generates valid search queries, and enables pre-training the agent with supervised learning. The search agent can use the full query language implemented by the search engine, including search operators such as “-”, as described in the above example.

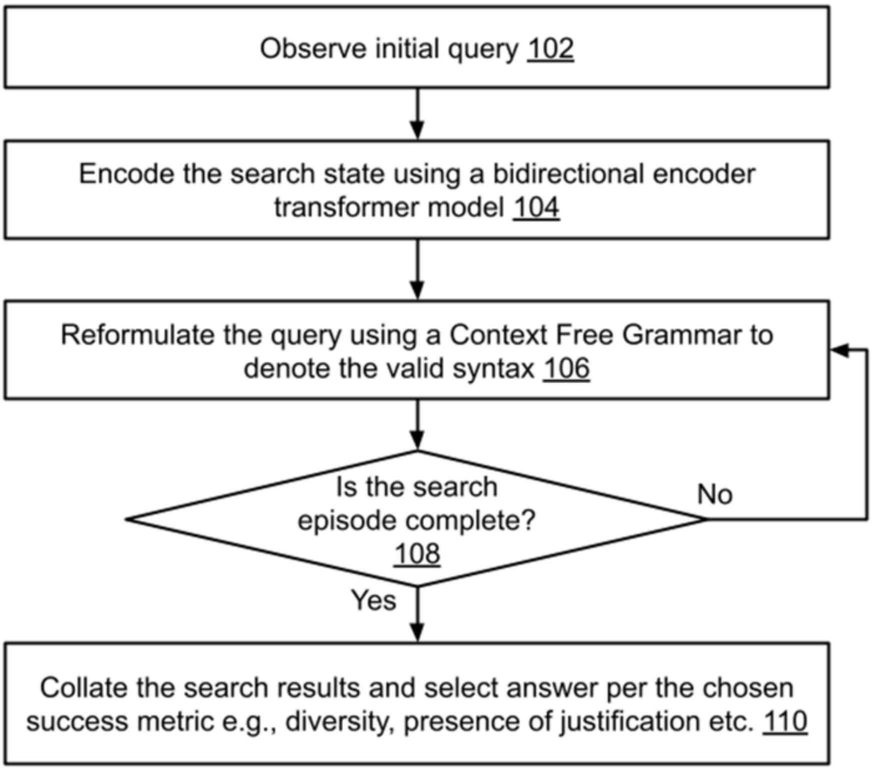

Fig. 1: Process to Improve Search Results using Reformulated Queries

Fig. 1 illustrates an example process to utilize a machine-learned agent web search agent to return search results, per techniques described in this disclosure. With user permission, an initial query is observed (102). The search state is encoded using a bidirectional encoder transformer model such as BERT (104). A Context Free Grammar (CFG) is used to define the structure of the action space for the search engine query syntax. The query is reformulated, e.g., using a modified Monte Carlo Tree Search (MCTS) technique that uses the Context Free Grammar to generate only valid query templates (106). This step is repeated until the search episode is complete, e.g., the complete tree has been traversed (108). The search agent can also be pre-trained using a supervised learning technique using past search results. The search results are collated and a particular answer is selected based on the chosen success metric e.g., diversity, presence of justification etc. (110).

In addition to improving search results delivered by a search engine, the described techniques can also be used to provide search results that satisfy other qualitative properties besides relevance. This can be achieved by designing success metrics (rewards) that take into account factors such as diversity of results, the presence of justification or support for a given answer, etc. With user permission, the described techniques can also provide responses to user queries provided to a digital assistant. If the user permits, the results can also include information from user data such as calendars, apps, etc.

The formulation of the action space as a grammar allows a broad and expressive range of new queries that the search agent can compose. Query reformulations can range from simple reformulations, to queries that contain special search operators, and query macros that define special high-level actions (e.g., search for a different result). The use of a grammar to constrain the action space ensures that the queries are guaranteed to be well-formed. This is particularly advantageous for systems (e.g., a virtual assistant) that may rely on some form of semantic parsing to execute the query. Even on web search, the search agent can make use of expressive power-user tools, such as special search operators. The grammar action space makes the query planning phase efficient since exploration of ill-formed queries is avoided and the model does not need to learn the specific sequences of actions that are valid or not valid, which is a key element for learning to be successful.

The described techniques can be implemented within a search engine and can provide improved results by using the search agent to reformulate or deepen user queries, e.g., in near real-time. Further, the described techniques can also be used to generate data (e.g., high quality results for specific queries) that a search engine or virtual assistant can utilize when providing responses to the user.

CONCLUSION

This disclosure describes the use of machine learning techniques, including supervised learning and reinforcement learning to train a search agent to search deeper for better, more accurate, better supported answers by interacting with the search engine. The interaction mimics strategies utilized by human experts to carry out accurate web search. The search agent can be modular, and to provide answers to a user query, performs operations such as formulation of new queries in a sequence, analysis of intermediate results, and selection of results based on a chosen success metric that can take into account factors such as accuracy, diversity, presence of justification, etc.

REFERENCES

1. Yuan, Xingdi, Jie Fu, Marc-Alexandre Cote, Yi Tay, Christopher Pal, and Adam Trischler. "Interactive machine comprehension with information seeking agents." *arXiv preprint arXiv:1908.10449* (2019).
2. Mastering Atari, Go, Chess and Shogi by Planning with a Learned Model available online at https://deepmind.com/research/publications/Mastering-Atari-Go-Chess-and-Shogi-by-Planning-with-a-Learned-Model