Multimodal Machine Learning Lab: Literature Research

October 2024

This collection will, for each paper, give a short description of the paper's contributions plus a short explanation why this is relevant for the research objective. Literature research helps to identify

- research that solves the same problem
- research that identifies related problems
- research that shows aspects that might be related or useful
- state-of-the-art methods and
- future research directions.

The central research question is:

What are the central components of a system that allows users to navigate the infinite index while supporting them in creative tasks?

A selection of topics is given below; further topics may be added if sensible.

1 Traditional Information Retrieval Foundations

Traditional IR foundations regarding the fields of creative search, serendipity, diversity, etc., which might also be applied for query suggestion.

2 IR Methods to Search Within Image Spaces

Methods from information retrieval that are applied in image search to use queries (or other modalities) to retrieve images, including reranking approaches.

3 Human-Computer Interfaces

Modalities for communicating human intent or information need to the computer and navigating the infinite index. This might include gestures, eye tracking and

analyzing the time spent watching or interacting. Interfaces for computational creativity are particularly relevant.

4 Dialogues

Using dialogues might help to get the information need and the already existing information from the users. Similar problems exist when generating explanation.

5 Probability-Defined Surprise

Defining metrics that describe information value and surprise of information nuggets is important for a directed generation. It is also relevant how these metrics can be tailored to a specific user and how the relevant parameters can be obtained from the user.

6 Real-Time Image Generation

Approaches for generating images in real time, in particular using Stable Diffusion, use techniques like quantization, approximation or low-resolution images. Some approaches are compiled in `x-stable-diffusion`.[1]

7 Prompt Engineering Assistance

Helping users in generating prompts by rewriting existing prompts or generating prompts from existing images. The equivalent IR term is query rewriting.

8 Image Manipulation Techniques

Beyond simple prompt engineering, there are tools to assist users in modifying generated images in a directed manner. This might also include finetuning the generator.

[1] `https://github.com/stochasticai/x-stable-diffusion`

9 Input Modalities for Conditioning Diffusion Models

Different modalities (beyond text) can be used to condition diffusion models. This might allow users to better control the generation. Approaches that condition on embedding spaces other than CLIP are particularly interesting.

10 Image Generation Evaluation

Evaluating generated images according to various criteria: Traditionally, generated images are evaluated on how realistic they look. For this project, it may also be relevant to evaluate user surprise and prompt alignment.

11 Output Modalities for Exploration of the Prompt Space

The generated images can be displayed in a simple grid, but other representations may also be helpful. These include representations such as animations, infinite zooms (in a latent space), 360° *views, 3D spaces, or other interactive representations.*

12 Analyzing the CLIP Space

Understanding the embedding space implied by CLIP is important to restrict the used embedding vectors to admissible values and to generate high quality images.

13 Navigating Through CLIP Space

Modifying CLIP values, e.g., through optimization is the foundation of navigating the infinite index.

14 Symbolic Images

Using symbolic images in a multimodal context (text and images) including stock image datasets and applications, especially for evaluation scenarios.

15 Incubator

This section can be used temporarily for (as of yet) uncategorized finds.