UNIVERSITÄT
ZU KÖLN

Recap and Consequences

HS In Context Learning (ICL) (Summer term 2024)

Nils Reiter,
nils.reiter@uni-koeln.de

July 11, 2024

INSTITUT FÜR
DIGITAL HUMANITIES
UNIVERSITÄT ZU KÖLN

IDH Summer Party 2024

Do. 11.07., 19 Uhr

Uniwiese, zwischen Mensa und IDH

"Digital art of a group of students having a barbecue and enjoying life"

Section 1

Recap

What Happened So Far

- 16.05. ICL Overview
- 06.06. Manual Template Engineering
- 13.06. Automated Template Engineering
- 20.06. Answer Space Design Methods
- 27.06. Prompt Ensembling
- 04.07. Prompt Augmentation

What Happened So Far

- 16.05. ICL Overview
- 06.06. Manual Template Engineering
- 13.06. Automated Template Engineering
- 20.06. Answer Space Design Methods
- 27.06. Prompt Ensembling
- 04.07. Prompt Augmentation
- A developing field
  - No clear nomenclature
  - Liu et al. (2023): One attempt to structure things (and a good one, I think)
  - Experiments rarely focus on a single aspect
  - Experimental setup difficult to control

Prompting Scenarios

- Interactive in a chat bot: 🡅Manual prompt engineering
  - Direct use and implicit validation
  - Results don’t have to be perfect to be useful
  - Users make connections and fill holes
  - Strategies involve different components (e.g., examples)

Prompting Scenarios

- Interactive in a chat bot: ⬆Manual prompt engineering
  - Direct use and implicit validation
  - Results don't have to be perfect to be useful
  - Users make connections and fill holes
  - Strategies involve different components (e.g., examples)
- 'Batch use' for automatic classification
  (i.e., use LLM-prompting to analyse large quantities of data)
  - Builds on top of traditional ML applications and assumptions
  - No immediate validation during application, therefore evaluation on test set
  - Subsequent applications rely on measured correctness

Prompting Scenarios

- Interactive in a chat bot: ↑Manual prompt engineering
  - Direct use and implicit validation
  - Results don't have to be perfect to be useful
  - Users make connections and fill holes
  - Strategies involve different components (e.g., examples)
- 'Batch use' for automatic classification
  (i.e., use LLM-prompting to analyse large quantities of data)
  - Builds on top of traditional ML applications and assumptions
  - No immediate validation during application, therefore evaluation on test set
  - Subsequent applications rely on measured correctness
- LLM-prompting likely well suited for "human-in-the-loop" approaches
- But interactively developed prompts likely do not generalize well

Prompting Steps

- Prompt template: $t =$ `[X] Overall, it was a [Z] movie.`
  - ↑Template engineering: Choose among alternative formulations (e.g., `[X] The movie was [Z].`)
  - ↑Prompt augmentation: Add additional contexts to the prompt (e.g., `[X1] The movie was [Z1]. [X2] The movie was [Z2]. [X] the movie was [Z].`)

Prompting Steps

- Prompt template: $t =$ `[X] Overall, it was a [Z] movie.`
  - 🡅Template engineering: Choose among alternative formulations (e.g., `[X] The movie was [Z].`)
  - 🡅Prompt augmentation: Add additional contexts to the prompt (e.g., `[X1] The movie was [Z1]. [X2] The movie was [Z2]. [X] the movie was [Z].`)
- Three steps
  - Apply template $f(x, t) =$ `I love this movie. Overall, it was a [Z] movie.`
  - Answer search: Select the best z to fill in the template
    - 🡅Answer space design: Define potential answers
    - ⚠Different options 🡇
  - Answer mapping: Map most probable answer z to output y
    - 🡅Answer space design: ...and how to map them

Prompting Steps

- Prompt template: $t =$ `[X] Overall, it was a [Z] movie.`
  - 🡅Template engineering: Choose among alternative formulations (e.g., `[X] The movie was [Z].`)
  - 🡅Prompt augmentation: Add additional contexts to the prompt (e.g., `[X1] The movie was [Z1]. [X2] The movie was [Z2]. [X] the movie was [Z].`)
- Three steps
  - Apply template $f(x, t) =$ `I love this movie. Overall, it was a [Z] movie.`
  - Answer search: Select the best z to fill in the template
    - 🡅Answer space design: Define potential answers
    - ⚠Different options 🡇
  - Answer mapping: Map most probable answer z to output y
    - 🡅Answer space design: ...and how to map them
- 🡅Prompt ensembling
  - Do everything with multiple prompts
  - Options to combine their answers (e.g., majority vote)

Getting LLM Answers

- Two options
  1. Let the model generate something, map it onto the target label (answer mapping)
     - Sometimes difficult to restrict output to defined vocabulary, need to interpret model output (which is yet another NLP task)
     - E.g., asking the model to only produce a single token

Getting LLM Answers

- Two options
  1. Let the model generate something, map it onto the target label (answer mapping)
     - Sometimes difficult to restrict output to defined vocabulary, need to interpret model output (which is yet another NLP task)
     - E.g., asking the model to only produce a single token
  2. Ask the model for the label with the highest probability
     - Easier task
     - Labels are pre-defined, but best label may not be what the model would have produced

Getting LLM Answers

- Two options
  1. Let the model generate something, map it onto the target label (answer mapping)
     - Sometimes difficult to restrict output to defined vocabulary, need to interpret model output (which is yet another NLP task)
     - E.g., asking the model to only produce a single token
  2. Ask the model for the label with the highest probability
     - Easier task
     - Labels are pre-defined, but best label may not be what the model would have produced
- Often underspecified in research literature!
- Huggingface blog post: huggingface.co

Section 2

Consequences

Discussion Groups

- How does prompting (as a machine learning paradigm) in interactive and batch use change the way things are done (in your opinion and according to what we know now)
  - in natural language processing
  - in the humanities
  - in academia in general
  - in industry (IT/other)
  - in the (German/Western) society
- What will remain the same after all?
- Which new possibilities are opened up? Which activity/method goes away?
- What will become easier, what will become harder?
- What do we need to find out next?

Discussion Groups

- How does prompting (as a machine learning paradigm) in interactive and batch use change the way things are done (in your opinion and according to what we know now)
  - in natural language processing
  - in the humanities
  - in academia in general
  - in industry (IT/other)
  - in the (German/Western) society
- What will remain the same after all?
- Which new possibilities are opened up? Which activity/method goes away?
- What will become easier, what will become harder?
- What do we need to find out next?

Procedure

- Split up into groups of 3-4 people
- Discuss for about 30 minutes, take notes in Google Slides ⟶
- Present in plenary session