Final Exam

601.467/667 Introduction to Human Language Technology

Fall 2022
Johns Hopkins University
Co-ordinator: Philipp Koehn

14 December 2022

Complete all questions.

Use additional paper if needed.

Time: 75 minutes.

Name of student: _____

Q1. Question Answering *20 points*

1. Question Answering (QA) and Machine Reading Comprehension (MRC) are both fields that are concerned with building automatic systems that can provide answers to questions. Please explain the main difference between the two fields, in terms of motivation or problem setup. [10pts]

   Answer: MRC tasks are designed to test the capabilities of reading and reasoning and serves as a benchmark for AI research. QA focuses more on the end-user, so the goal is to to have something that is useful. As a result, MRC is usually restricted to one document where the answer is present, to be read in depth; QA exploits multiple knowledge sources and employs multiple technologies such as IR, NER, and Parsing to get the correct answer.

2. One of the outputs of the Question Analysis component of a QA system is the answer type. Please indicate what are reasonable answer types for the following two questions, and why is accurate answer type prediction important to the overall QA system. [10pts]

   - Question A: Who invented the dish washer?
   - Question B: What is the population of Maryland?

   Answer: The answer type for Questions A might be PERSON, or any hyponym like INVENTOR, SCIENTIST, ENGINEER. The answer type for Question B might be NUMBER. Answer types are important to help QA systems narrow down the search space of potential answers. For example, for Question B, any answer that is not a number would be definitely incorrect.

Q2. Digital Humanities *20 points*

1. Which is the more general model architecture: CNN or GCN? [5pts]

   GCN

2. Explain how the less-general architecture identified in the previous question is related to the more-general one [15pts]

   A convolutional neural network (CNN) is a graph neural network (GCN) in which the nodes in the graph correspond to elements in a grid structure (like pixels) and the edges correspond to positions in the grid relative to the given element.

Q3. Clinical NLP *20 points*

1. What are the three ERAs of Clinical NLP research and systems? [10pts]

   Rule-based, statistical, language models

2. Name 2 common clinical NLP tasks that are not typically considered in generic NLP pipelines? [10pts]

   Two of the following: Section segmentation, Negation, de-identification, NER for medications/symptoms/comorbidities/treatments, patient similarity, automated coding, concept linking, phenotyping

Q4. Ethical Problems *20 points*

1. Describe these two Ethical Principles in AI: Autonomy and Beneficence. [10pts]

   Autonomy: *Humans should choose how and whether to delegate decisions to AI systems, to accomplish human-chosen objectives*
   Beneficence: *AI should promote well-being, preserve dignity, and sustain the planet "The development of AI should ultimately promote the well-being of all sentient creatures," We should "ensure that AI technologies benefit and empower as many people as possible" "AI technology must be in line with ensuring the basic preconditions for life on our planet, continued prospering for mankind and the preservation of a good environment for future generations."*

2. Describe what Institutional Review Boards are and what is their mission. [10pts]

   The IRB is "responsible for protecting the rights and welfare of the human subjects of research conducted by faculty and staff at the Institutions". The IRB evaluates the ethical aspects of human subject research The board usually requires the investigators to inform the participants about the research in which they are involved

Q5. Large Language Models *20 points*

1. Select all the answer(s) to fill in the blank (. . . . .) in each item. [12 pts, 3 pts each]

   (a) Given scaling trends of the past few years, the performance of models grow . . . . . with model size. ☐ logarithmically ☐ linearly ☐ exponentially

   (b) . . . . . is *not* an aspect of scaling large language models.
   ☐ Model size ☐ Data size ☐ Transformer architecture

   (c) . . . . . is an argument for the infeasibility of scale due to limited computing.
   ☐ Advances in computing hardware are much slower than the trends for scaling language models.
   ☐ Advances in parallel computing can support the fast pace of scaling models.
   ☐ Scaling language models continues to incur a lot of costs (monetary, carbon footprint, computing resources, etc.)
   ☐ Scaling models might reduce the overall costs: the availability of a few large models may prevent the cost of building many smaller ones.

   (d) . . . . . is an argument that "data" should not be a bottleneck for scaling language models.
   ☐ There size of the internet is consistently growing.
   ☐ There size of Wikipedia is consistently growing.
   ☐ One can mine data from other modalities (e.g., text data mined from videos).
   ☐ Even with limited data, we can use it more effectively to get more gains.

(a) The performance of models grows logarithmically with model sizes. (b) Based on the discussion in class, data size and model size are aspects of scaling trends. (c) The first and the third statements are arguments for the infeasibility of scaling due to limited computing resources. (d) All the provided answers support the argument that "data" is not a bottleneck for scaling language models.

2. Answer the following questions in a few sentences (no more than 5 sentences for each) [8 pts, 4 pts each]

   (a) Explain what long tail of problems in natural language is (provide an example).

   Not all natural language instances have the same difficulty. Some sentences or tasks that frequently appear in our daily discourse, belong to the "head" of a hypothetical distribution of language tasks. In contrast, because of the combinatorial nature of concepts or ideas, there are many *sentences or tasks that are rare in our discourse even though – the "long tail" of the hypothetical distribution of language tasks.*
   Here are a few examples:

   i. *Doing basic mathematical operations are a lot more common among small numbers (e.g., "sum of 5 and 2") than large numbers (e.g., "sum of 523,235 and 278,057"). Note the space of small numbers is a lot smaller than the space of large numbers.*
   ii. *In the context of machine translation, there are few rich-resource languages such as English or Spanish. However, there are plenty of other languages that suffer from limited resources.*
   iii. *In the context of self-driving cars, driving in large streets on a bright day is an easier challenge for models given their prevalence in say, California. However, driving during a storm is not that frequent and hence a more challenging task.*

   (b) Explain how the long tail of problems in natural language poses a challenge to language models.

   Language models are empowered by absorbing massive amounts of patterns from their massive pre-training data. Instances in the "head" of the distribution are generally frequent and hence easier to tackle for language models. However, language models struggle with the "tail" of the distribution as they are most tasks often have infinite many rare instances in the "tail" as they are infrequent and extremely large.