AIMS
African Master's in Machine Intelligence

# Deep NLP Part III – Machine Reading, Question Answering and Dialog

Final Exam

**Instructor:** Antoine Bordes (Facebook AI Research)
**Assistants:** Angela Fan, Louis Martin (Facebook AI Research)

May 28, 2021

**Name:**
**Email:**

---

1. The exam is 2 hours long (but was originally intended for 1h30, you can hand it in early).
2. Write clearly! If an instructor cannot understand your answer, it is not an answer.
3. Send your exam to both angelafan@fb.com and louismartin@fb.com.

**Distribution of Marks**

| Section | Points | Score |
|---|---|---|
| 1. Machine Reading | 25 | |
| 2. Open domain Question Answering | 25 | |
| 3. Dialog (Bonus) | 5 | |
| **Total** | 55 | |

**Machine Reading (25 points)**

1. Give three reasons why artificial data is useful for Machine Reading? (3 points)
   - Machine Reading systems are hungry for data.
   - Machine reading systems need a certain type of data where answers can be found within the given passage text.
   - Contain enough noise for your model to learn.

2. Give three limitations of artificial data? (3 points)
   - Hard to generalize.
   - Prune to overfitting.
   - Usually short and doesn't match real-life data.

3. Give an example of a Winograd Schema (3 points)

4. How is answer predicted in the bAbI dataset? Which training loss can used for training on it? (3 points)
   The answer is found in the given phrases. Cross entropy can be used as a training loss

5. How is an answer predicted in the SQuAD dataset? Which training loss can used for training on it? (3 points)
   The answer can be retrieved from the given phrases. Cross entropy can be used as a training loss.
   Note:
   In SQuAD 2.0, some questions have no answers.

6. What does the acronym BiLSTM stand for? (1 point)
   **Bi**-directional **L**ong **S**hort-**T**erm **M**emory.

7. Give three reasons why pretrained contextualized word embeddings (ELMo/BERT) work so well. (3 points)
   - Since they capture the context of the words.
   - They are usually trained on a huge amount of data.
   - They grasp the meaning of the word within the given context. So, similar words appears to have similar embeddings.

8. Give three current challenges of Machine Reading systems (3 points)
   - Interpretability: current end-to-end systems are hard to be interpreted.
   - Scalability: Usually these models are massive and it's not that easy

to scale them.
- Clean Data: The amount of clean, structured, general data is not enough for these models.

9. What is the task of coreference resolution? (3 points)

Co-reference resolution is the task at which you refer a pronoun like “him”, “her”.. etc. to the main object.

**Open domain Question Answering (25 points)**

BestPhoneEver.com is a successful phone company that sells a lot of mobile phones. However, those phones have issues regularly and BestPhoneEver.com receives so many calls that their customer service cannot answer all of them conveniently. They hire you to build a question answering system to help the customer service agents to answer more quickly.

You have access to:
- a collection of 10k questions and their answers given by the customer service in the past
- a database that groups all the data on each phone and plan sold by the company
- a set of 5k documents that describe multiple bugs that occurred to all recent phones in the last couple of years

Describe the system that you would propose. Think at how it is conceived, which data is used, how the system is evaluated, etc.

Regarding the data, probably I will use the following:
- The 10k collection of questions.
- The 5k documents that describe common bugs.

Next step is to structure these data:
- clean the data from any noise
- structure the 10k collection of questions into a (question, answer) pairs.

I will create a retrieval system that:
- Given any new question, it will compare the similarity between that new question and the other questions.
- get the most similar question
- return the answer of that question.

If the answer is not known, it the user about asks more information.

**Dialog (bonus 5pts)**

10. What are the main differences between a goal-oriented and a chit-chat dialog system? (2 points)

| | **Goal-oriented** | **Chit-chat** |
|---|---|---|
| **domain** | specific-domain | open-domain |
| **evaluation** | goal is achieved | how long was the conversation |
| **Model** | Rule-based (mostly) | end-to-end (mostly) |
| | | |

11. What typical characteristics of dialog make it difficult to train models from data and evaluate them? (3 points)

- Consistency issues.
- There are multiple right responses for the same query.
- If the answer is not known, it the user about asks more information.