Politecnico di Torino
Dipartimento di Automatica e Informatica
Deep NLP (01VIXSM)
Written Exam

April 11, 2022

Name: ____
Surname: ____
Student ID: ____

Exam rules:

- The present exam consists of 6 pages (including this cover page) and 7 questions overall. Any inconsistencies/printing errors in the written exam content must be reported to the teacher *at the beginning* of the exam.
- Exam duration: *60 minutes.*
- Withdraw is allowed only *at the end* of the exam.
- The exam is *closed-book*. Electronic devices, mobile phones, smart watches, and extra papers (even blank papers) are *not allowed.*
- Closed-ended questions: cross the right answer (just one) at pag. 2. Wrong or misszing answers to closed-ended questions will receive *no penalty.*
- Open questions: write your answers below the text of the question. If you need more space please use the last page (i.e., pag. 6) and/or the back side of the paper.

Evaluation grid

Question:	1	2	3	4	5	6	7	Total
Points:	1	1	1	1	5	5	6	20
Score:								

1. (1 point) Supervised domain adaptation entails
   - fine-tuning a model trained on open-domain data for a specific task.
   - exploiting in-domain data to specialize models trained on open-domain data.
   - applying a fine-tuned model to out-domain data.
   - transferring a domain-specific model from one task to another.
   - enriching in-domain data with open-domain data.
2. (1 point) What of the following statement holds **False** for word embedding models?
   - FastText encodes subwords.
   - GloVe exploits occurrence-based statistics.
   - GloVe do not support domain adaptation.
   - WordVec relies on a Deep Feed-forward Neural Network.
   - Word2Vec supports domain adaptation.
3. (1 point) The intent of an AI chatbot cannot
   - be associated with multiple actions.
   - be derived from the entities in the request.
   - be part of a RASA story.
   - be derived from the associated action.
   - be derived from external sources.
4. (1 point) Shallow parsing
   - does not depend on the language.
   - is associated with Named Entity Recognition.
   - does not require Named Entity Disambiguation.
   - is required for intent detection.
   - is mandatory for syntactic parsing.

5. (5 points) Explain the **aspect-based summarization** task:

   1. Formalize the task.
   2. Clarify the main differences with general-purpose summarization.
   3. Provide at least two practical examples of application of aspect-based summarization.

5.1) Given a set of documents D covering different aspects A, produce a summary S of D covering each of the aspects in A. Each aspect-specific summary should contain all the key-information from the documents in D related to the corresponding aspect.

5.2) The main difference with general-purpose summarization is that each summary is specific to an aspect. This allows for a more targeted and concise summary. Such an additional constraint is commonly not enforced by general-purpose summarizers. The summarization process is also more fine-grained, as it needs to account for the different aspects that might be present in the documents.

5.3) Example 1: Summarization of topical news (e.g., Covid-19). Examples of topics: vaccine, restrictions, policies and regulations, etc.

Example 2: Summarization of lecture notes (e.g., databases). Examples of topics: SQL, DB models, etc.

Example 3: Summarizing customer reviews of products, in order to identify key areas of improvement for the product;

6. (5 points) Considering the **attention mechanism**.
    1. Explain the intuition behind the attention score in modern transformer models (e.g., BERT).
    2. Define the main functions used to compute the attention.
    3. Analyze the complexity with the text length.

6.1) The attention mechanism is a technique that mimics cognitive attention. It is a technique to compute a weight assigned to each element of a sequence, so that the model can focus on specific parts of the input when predicting a token. The intuition is that the model can learn which elements in the input are more important for the prediction by assigning higher weights to them. It is implemented by using a query-key mechanism similar to those applied for query-driven content retrieval in key-value DBs. Specifically, it mimics the retrieval of a value for a query based on a key in a database.

Each token in the input sequence is represented by three different vectors: query, key and value. Each token in the input sequence uses its query vector to compute dot product with all the value vectors of the input sequence (including itself). This way, each token can know the contribution of all the other tokens to its prediction. Then, the score is normalized and the softmax function is applied to compute attention coefficients. To obtain the output representation of each token, the value vectors are multiplied by the attention coefficients.

6.2) The formula to compute the attention scores as introduced in transformer model is reported below:

$$\text{Attention}(Q, K, V) = \text{softmax}(\frac{QK^T}{\sqrt{n}})V \tag{1}$$

Where Q is the query matrix, K is the key matrix, V is the value matrix, and n is the length of the sequence.

6.3) The attention mechanism has quadratic complexity with respect to the sequence length $o(n^2)$ because the attention is computed n times for each token in the sequence (where n is the sequence length).

7. (6 points) Given a financial news you want to detect the relations between different text spans. Specifically, the goal is to *identify and extract*, within a sentence or a longer text block contained in the news content, the causal elements and the consequential ones. Assume that they consist of distinct cause-effect pairs.

   Example:

   - **Original news content:** Zhao found himself 60 million yuan indebted after losing 9,000 BTC in a single day (February 10, 2014)
   - **Extracted cause:** losing 9,000 BTC in a single day (February 10, 2014)
   - **Extracted effect:** Zhao found himself 60 million yuan indebted.

   Assume that a ground-truth consisting of 1000 ⟨content, cause, effect⟩ triples are given.

   1. Design a complete NLP pipeline for accomplishing the task. For each pipeline step clarify the goal, the algorithms used, and the main settings.
   2. Describe the metrics used to evaluate the system performance and explain their meaning.

   Sketch of the pipeline

   - Language identification and machine translation (if need be)
   - Selection of a machine learning model able to classify single tokens. In this case, it could be useful to select a pre-trained encoder model (e.g., BERT).
   - Text tokenization using a tokenizer selected according to the previous step.
   - Training/Fine-tuning a token classification model (similar to NER).
   - Examples of annotations used in the training phase:
     - B-CAUSE: begin token for the clause
     - I-CAUSE: token inside a clause (including final token)
     - B-EFFECT: begin token for the effect
     - I-EFFECT: token inside a effect (including final token)
     - OTHER: token outside clause and effect
   - Running the fine-tuned model on inference model (e.g., BERT) to detect cause/effect snippets.

   7.2) The following two metrics can be used to evaluate the proposed system:

   - *ROUGE-based scores* can be used to identify the overlap between the system and reference text snippets.
   - *Exact match* can be used to compute the fraction of cause/effect that are correctly identified.

This page is intentionally left blank to accommodate work that wouldn't fit elsewhere and/or scratch work.