Examples of Questions and Answers

Prof. Luca Cagliero
Dipartimento di Automatica e Informatica
Politecnico di Torino

Examples of closed questions

Question nr. 1

Which of the following statements holds true for the BERT pre-training phase?

- *It uses a context window to define positive examples.*
- *Self-supervised training with masked language modeling and next sentence prediction is used in the training phase.*
- *The model is trained using Next Word Prediction task*
- *It is trained by corrupting text with an arbitrary noising function, and learning a model to reconstruct the original text.*
- *None of the above.*

Question nr. 1

Which of the following statements holds true for the BERT pre-training phase?

- *It uses a context window to define positive examples.*
- ***Self-supervised training with masked language modeling and next sentence prediction is used in the training phase.***
- *The model is trained using Next Word Prediction task*
- *It is trained by corrupting text with an arbitrary noising function, and learning a model to reconstruct the original text.*
- *None of the above.*

Question nr. 2

The Negative Sampling

- *Can be used for bootstrapping supervised classification models.*
- *Can be used to avoid data overfitting.*
- *Can be used to find negative words in sentiment analysis.*
- *Can be used to select negative examples while training a Word2Vec model.*
- *None of the above.*

Question nr. 2

The Negative Sampling

- *Can be used for bootstrapping supervised classification models.*
- *Can be used to avoid data overfitting.*
- *Can be used to find negative words in sentiment analysis.*
- ***Can be used to select negative examples while training a Word2Vec model.***
- *None of the above.*

Question nr. 3

Considering the one-hot encoding representation

- *Each textual unit is represented by a dense vector consisting of real-valued elements*
- *Each textual unit is represented by a sparse vector consisting of boolean elements.*
- *Each textual unit is weighted by its number of occurrences in the input corpus.*
- *It varies according to the size of the context window.*
- *None of the above.*

Question nr. 3

Considering the one-hot encoding representation

- *Each textual unit is represented by a dense vector consisting of real-valued elements*
- ***Each textual unit is represented by a sparse vector consisting of boolean elements.***
- *Each textual unit is weighted by its number of occurrences in the input corpus.*
- *It varies according to the size of the context window.*
- *None of the above.*

Question nr. 4

Considering the HITS algorithm, during the hub update step

- *Each node relevance score is normalized by the number N of nodes.*
- *Each link relevance score is normalized by the number N of nodes.*
- *For each node, the hub score is the sum of the authority scores of each node that it points to.*
- *For each node, the hub score is the sum of the hub scores of each node that points to it.*
- *None of the above.*

Question nr. 4

Considering the HITS algorithm, during the hub update step

- *Each node relevance score is normalized by the number N of nodes.*
- *Each link relevance score is normalized by the number N of nodes.*
- ***For each node, the hub score is the sum of the authority scores of each node that it points to.***
- *For each node, the hub score is the sum of the hub scores of each node that points to it.*
- *None of the above.*

Question nr. 5

In the self-attention mechanism implemented in BERT

- *The attention score of a reference token is computed separately for each token in the sequence, including the reference token itself.*
- *The attention score of a reference token is computed separately for each token in the sequence, except for the reference token itself.*
- *The attention score of a reference token is computed by attending only part of the sequence according to the positional encoding.*
- *The attention score of special tokens (e.g.,* SEP*) is computed by attending only other special tokens (e.g.,* CLS*).*
- *None of the above.*

Question nr. 5

In the self-attention mechanism implemented in BERT

- ***The attention score of a reference token is computed separately for each token in the sequence, including the reference token itself.***
- *The attention score of a reference token is computed separately for each token in the sequence, except for the reference token itself.*
- *The attention score of a reference token is computed by attending only part of the sequence according to the positional encoding.*
- *The attention score of special tokens (e.g.,* SEP*) is computed by attending only other special tokens (e.g.,* CLS*).*
- *None of the above.*

Question nr. 6

Rule-based Named Entity Recognition is not enhanced by

- *Syntactic Parsing.*
- *POS tagging.*
- *Lexical databases.*
- *Domain-specific ontological models.*
- *Stemming.*

Question nr. 6

Rule-based Named Entity Recognition is not enhanced by

- *Syntactic Parsing.*
- *POS tagging.*
- *Lexical databases.*
- *Domain-specific ontological models.*
- ***Stemming.***

Question nr. 7

Which of the following is a valid definition of n-gram?

- *A set of n sentences that are part of a document.*
- *A bag of words describing a document topic.*
- *A contiguous sequence of n characters that are part of a word.*
- *A set of n phonemes that are part of a phrase.*
- *None of the above.*

Question nr. 7

Which of the following is a valid definition of n-gram?

- *A set of n sentences that are part of a document.*
- *A bag of words describing a document topic.*
- ***A contiguous sequence of n characters that are part of a word.***
- *A set of n phonemes that are part of a phrase.*
- *None of the above.*

Examples of open questions

Question nr. 8

Explain the PageRank algorithm

1. *Enumerate and explain the main algorithm steps.*
2. *Exemplify at least one context of usage in NLP.*
3. *Specify the conditions under which PageRank can be applied to a given graph. Explain how to proceed when the conditions are not satisfied.*

Solution 8.1

- *Rule 1. Links from a graph node to itself are ignored.*
- *Rule 2. Multiple outgoing edges from one node to another are treated as a single edge.*
- *Rule 3. Set the same initial value for all nodes (e.g., $P(x) = \frac{1}{N}$)*
- *Apply the Pagerank formula to all vertices at each s*

$$P(x) = \frac{1-\lambda}{N} + \lambda \sum_{y \to x} \frac{P(y)}{out(y)}$$

lambda: damping factor

N: number of vertices

y -> x: node x's incoming edge

Solution 8.2

- *It is used in information retrieval to get an estimation of the overall importance of a node in the graph.*
- *It can be used as additional relevance score for building search engines.*

Solution 8.3

- *The graph must not have sinks (i.e., vertices with only incoming edges).*
- *To solve this problem it is possible to add an outgoing edge from the sink vertex to every other node in the graph.*

Question nr. 9

Explain the Latent Dirichlet Allocation

1. *Elaborate on the steps required to generate an LDA model.*
2. *Describe the Author-Topic Model (ATM) and its similarities/differences with LDA.*
3. *Enumerate <u>at least</u> two practical examples of application of ATM.*

Solution 9.1

- *Generative topic model.*
- *Each word in a document is assumed to be generated either by sampling a topic from a document-specific distribution over topics and by sampling a word from the distribution over words that characterizes that topic.*
- *For each document in the corpus and for each term, a topic is chosen accordingly to the document-topic distribution.*
- *Words are extracted from the input vocabulary V by taking into account the terms probabilities for each given topic in the document mixture.*

Solution 9.2

- *It is a Generative model for documents and extends the Latent Dirichlet Allocation to include authorship information.*
  - *Each author is associated with a multinomial distribution over topics*
  - *Each topic is associated with a multinomial distribution over words*
  - *A document with multiple authors is modeled as a distribution over topics that is a mixture of the distributions associated with the authors*

Solution 9.3

ATM can be used for

- *Who is the most authoritative author on a given topic?*
- *What are the topic covered by a given author?*
- *What is the most authoritative paper of an author?*

Question nr. 10

Elaborate on the Recommendation task

1. *Formulate the task and clarify the main goals.*
2. *Illustrate at least two business scenarios of usage for a recommender system.*
3. *Compare content-based and collaborative filtering systems by highlighting pros and cons of each of the above-mentioned strategies.*

Solution 10.1

- *Let U be a set of users, I be a set of recommendable items, R an ordered set of ratings*
- *The task is to find* $F(\cdot)\colon U \times I \rightarrow R$
- *The goal is to generate user-specific item rankings.*

Solution 10.2

- *NetFlix users receive movie recommendations based on their previous interactions with the platform.*
- *Travelers of a tourism agency receive hotel recommendations based on the census data (e.g., age, gender, job, salary, etc.)*

Solution 10.3

- *Collaborative filtering: recommend to a given user those items that were selected by similar users.*
  - *Pros: no need for content-level explorations (more efficient).*
  - *Cons: popularity bias, cold start.*
- *Content-based approaches: recommend items that are most similar to those previously selected by the same user.*
  - *Pros: solves the cold start and the first rater problems.*
  - *Cons: filter bubble. Need for content-level explorations (more computationally intensive).*

Acknowledgements and copyright license

- Acknowledgements
  - I would like to thank Dr. Moreno La Quatra, who collaborated to the writing and revision of the teaching content

- Affiliation
  - The author and his staff are currently members of the Database and Data Mining Group at Dipartimento di Automatica e Informatica (Politecnico di Torino) and of the SmartData interdepartmental centre
    - https://dbdmg.polito.it
    - https://smartdata.polito.it

Thank you!