MOSIG – MSIAM – 2019-2020 Information Access and Retrieval – GBX9MO23
Georges Quénot – Philippe Mulhem – Jean-Pierre Chevallet
3 February 2020 – 9h00-11h00 (9:00am-11:00am) – 2 hours

Course materials, the two papers related to the examination, personal notes, and calculators (without network capabilities) are allowed.

The examination consists in questions related to two scientific papers and/or to the contents of the course:

[1] Jianfeng Dong, Xirong Li, Chaoxi Xu, Shouling Ji, Yuan He, Gang Yang, Xun Wang, "Dual Encoding for Zero-Example Video Retrieval", the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 9346-9355.

[2] Hamed Zamani, Mostafa Dehghani, W. Bruce Croft, Erik Learned-Miller, and Jaap Kamps. "From Neural Re-Ranking to Neural Ranking: Learning a Sparse Representation for Inverted Indexing". In Proceedings of the 27th ACM International Conference on Information and Knowledge Management (CIKM), 2018. ACM, New York, NY, USA, 497-506.

Please use separate examination sheets for questions related to paper [1] and questions related to paper [2].

In the following questions, we expect real explanations with details, and not only excerpts from the papers. You should spend about 5 minutes per question and we expect concise answers.

Questions related to paper 1.

Q1.1: a) What are the two main methods for zero-example video retrieval and how do they work? b) What are their respective main advantages and disadvantages?

Q1.2: When compared with the previous approaches, what are the two main innovations introduced in the proposed work?

Q1.3: a) What is the goal of the common space learning component? b) How many learnable parameters does it have (excluding those in the batch normalization layers)?

Q1.4: In the dual encoding network, which architectural components are common between the text and the video branches and which are different and how do they differ?

Q1.5: a) Are the performance measures used in tables 1 and 2 consistent? b) Are all comparisons significant?

Q1.6 a) What is an ablation study? b) What do you conclude from the ablation study presented in section 5.1? c) Are all levels really helpful in the final fusion?

Q1.7: Considering the results presented on TRECVid, which factors other than the networks architecture and the training procedure have a significant influence on the overall system performance?

Q1.8: In the experiments on image-to-text and text-to-image, how is the system adapted in order to work with still images instead of videos?

Q1.9: What is the difference between mean (or average) and max pooling, in their formulation and in their expected effect?

Questions related to paper 2.

Q2.1: Explain in detail the principle of the inverted files in information retrieval. Why are they useful? What are their advantages (compared to not using any inverted file) regarding memory usage, query processing speed?

Q2.2: Compared to not using any inverted file on a classical Vector Space model using cosine matching function, does an inverted file implementation provides exactly the same results than a non-inverted file implementation? Explain why (using the cosine formula) and examples with cosine values.

Q2.3: Classically, approaches that use embedding of word used in information retrieval are based on a re-ranking process (i.e., the embeddings are used only to rerank the top-k results provided by a classical IR matching like BM25 for instance).

a) Explain why "simple" approaches that represent documents using the embeddings of their words are not able to cope with the retrieval on large corpus of documents.
b) Explain why reranking is then a solution to cope with this problem.

Q2.4: Explain in detail the goal of the paper, the difference with the approaches considered in Q3 above, and shortly the approach provided.

Q2.5: Explain in detail (with examples) the role of the di1, qi, di2 for the learning the inverted file (subsection 3.3. of the paper). What do we expect from the "Retrieval Objective"?

Q2.6: Explain in detail why the authors have to comment a lot (in subsection 3.3) about the loss function related to sparsity estimation with the L0 and L1 norms.

Q2.7: The lambda parameter for the global Loss function (formula (6) n subsection 3.3) is used to control the importance of the sparsity. Explain if using only lambda has some advantage or drawback compared to weighting by (1-alpha) the hinge loss weight and by alpha for the sparsity?

Q2.8: Explain in detail of the (classical) problem of training data is tackled by the paper?

Q2.9: Experiments (section 4):

c) Explain why the authors consider several evaluations collections to test their approach?
d) Explain why the proposal achieves lower results for ClueWeb than for the Robust.