Machine Learning for Education Laboratory
ML4ED

BERT goes to EPFL: MCQ prediction with a Muppet twist!

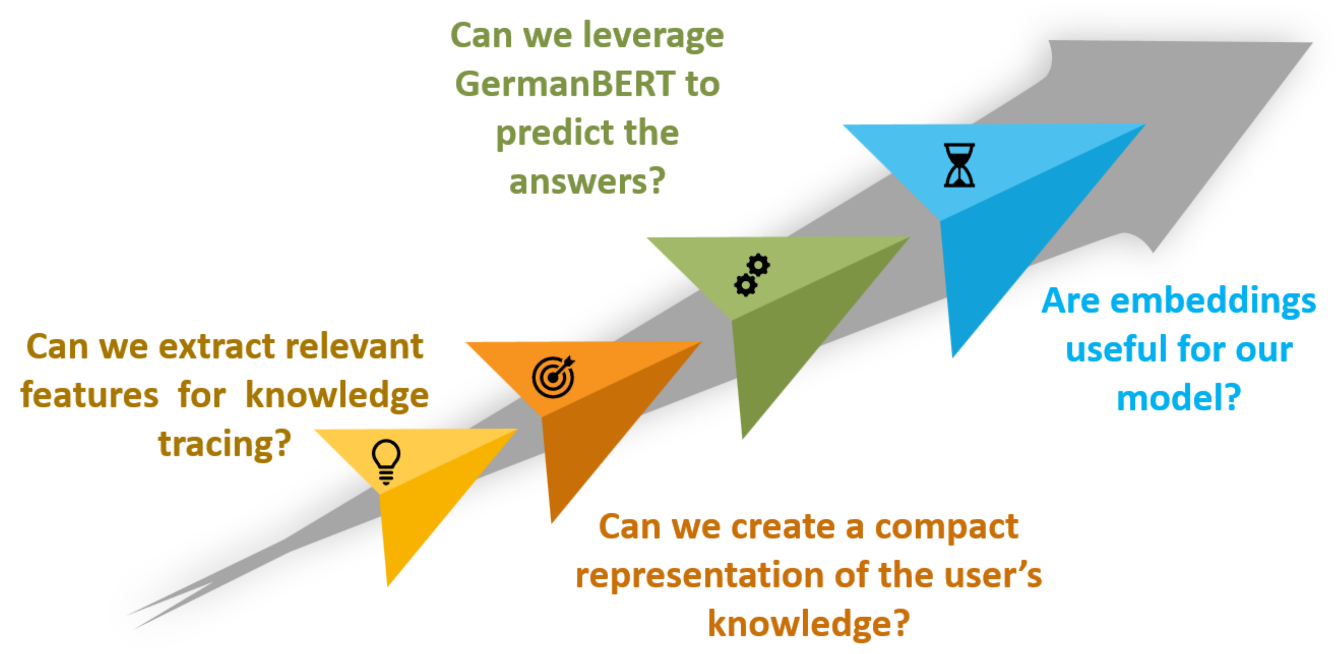
Answer Forecasting: ML-Driven Predictions in Lernnavi

Luca Zunino, Elena Grazia Gado, Tommaso Martorella

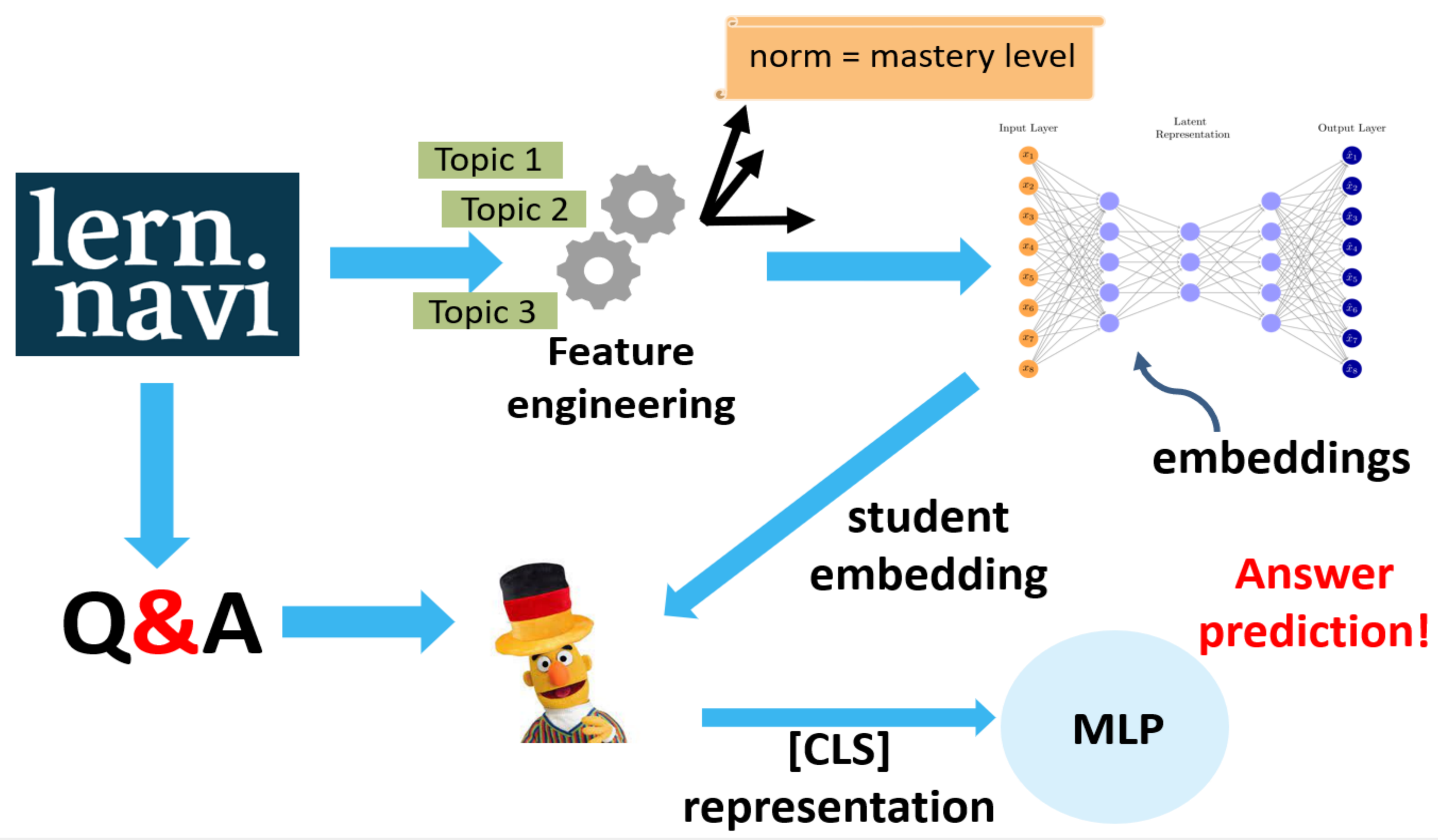
Machine Learning for Behavioral Data (MLBD) Course, EPFL, 2023

1 INTRODUCTION + RESEARCH QUESTIONS

Motivation:
ML model to predict how a student will respond to a given question to enable tailored learning experiences



2 METHODOLOGY



Model architecture

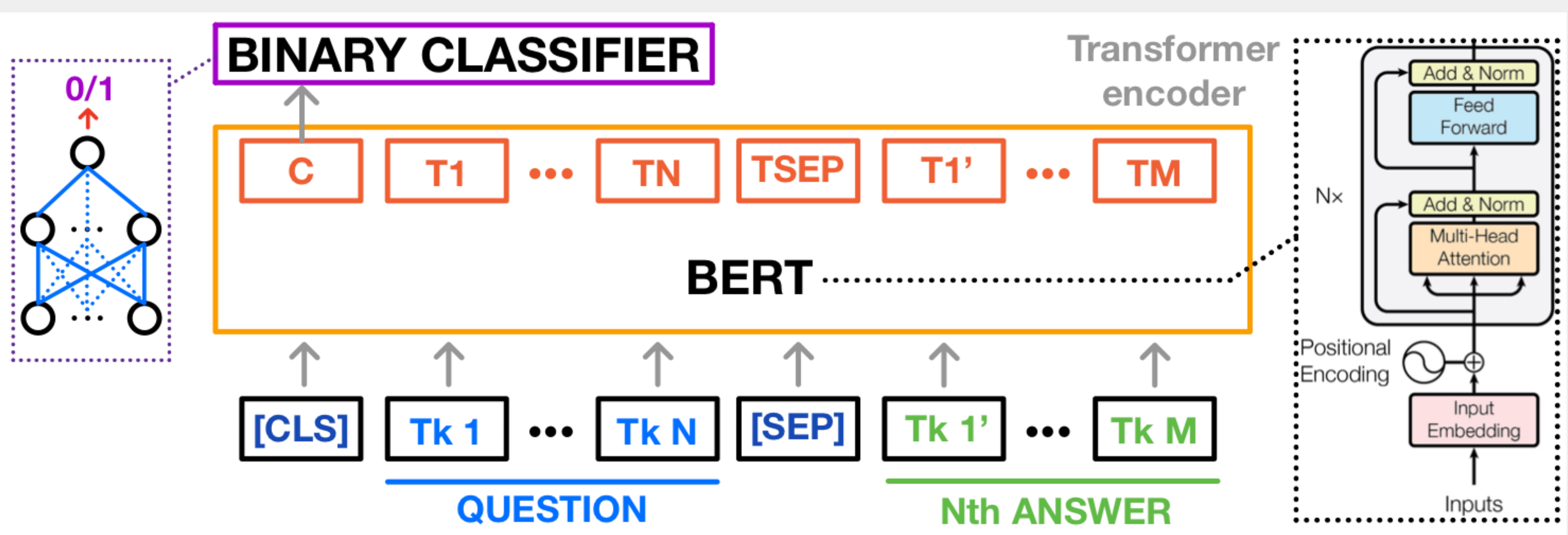


Figure 1. Architecture of the BERT model with a binary classifier head, without student embeddings.

MCQBert1-4: embeddings concatenated/ summed before the classifier/at BERT's input.

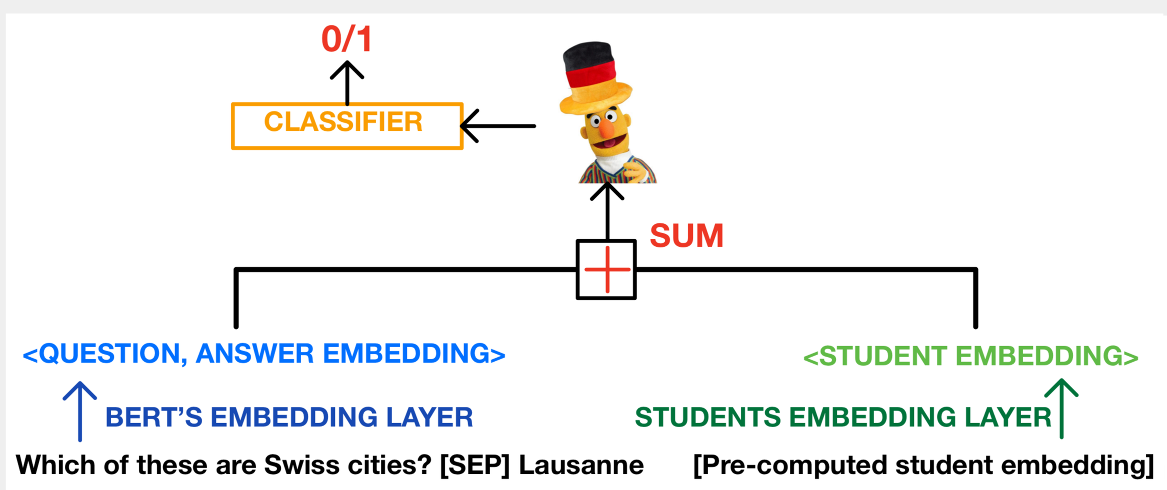


Figure 2. Simplified scheme of MCQBert3.

3 RESULTS

GermanBERT to LernnaviBERT: finetuning with masking

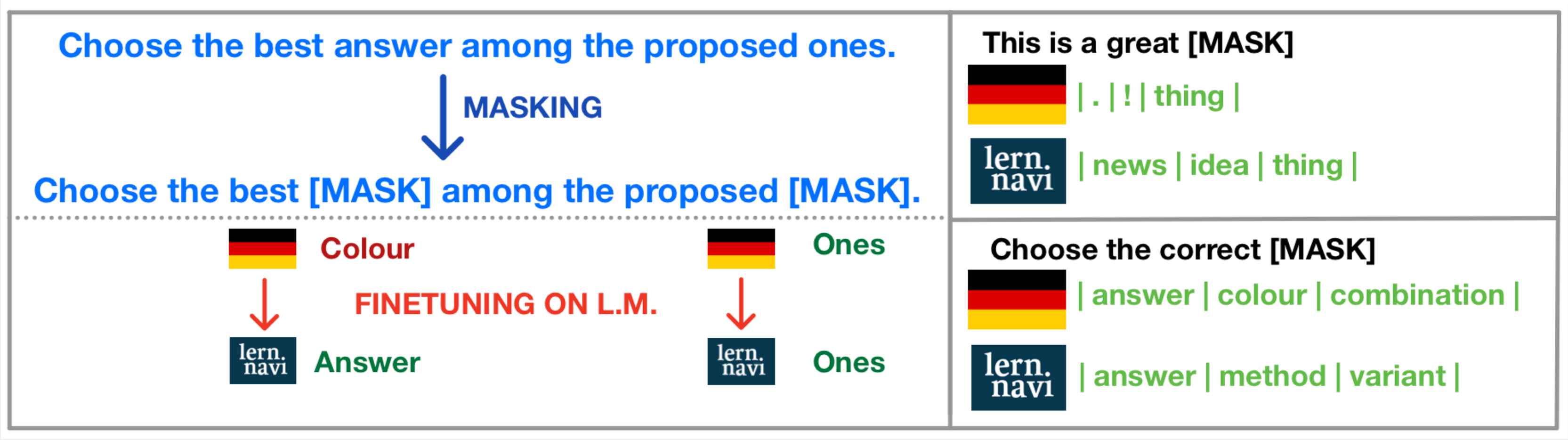


Figure 3. Finetuning GermanBERT on a language modelling task using masking (real examples on the right).

LernnaviBERT finetuning for MCQ answer prediction

Which of these are Swiss cities? A. Lausanne B. Bern C. Paris
Which of these are Swiss cities? [SEP] Lausanne → 1
Which of these are Swiss cities? [SEP] Bern → 0
Which of these are Swiss cities? [SEP] Paris → 1

Figure 4. Procedure used to transform MCQs in a binary classification task. LernnaviBERT is finetuned on predicting the correct answers of MCQs and evaluated both on MCQs never seen before and observed during training.

MCQBert3, new MCQs: average F1 = 0.698, MCC = 0.396
MCQBert3, seen MCQs: average F1 = 0.991, MCC = 0.983

LernnaviBERT finetuning for student answer prediction

Baseline: average F1 = 0.802, MCC = 0.606
MCQBert3: average F1 = 0.805, MCC = 0.612

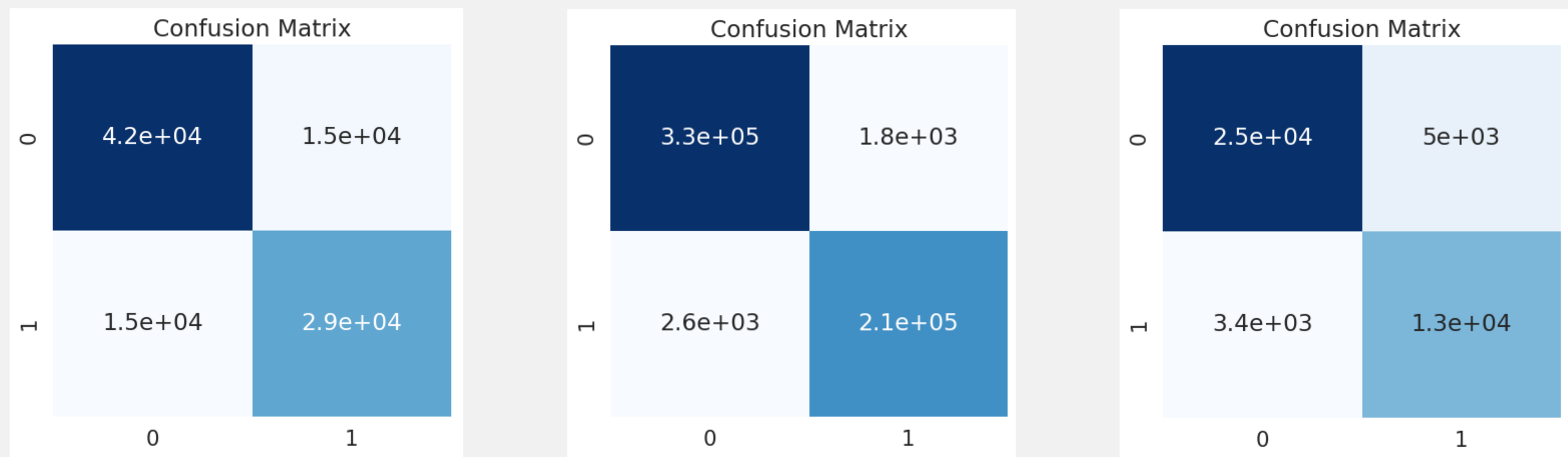
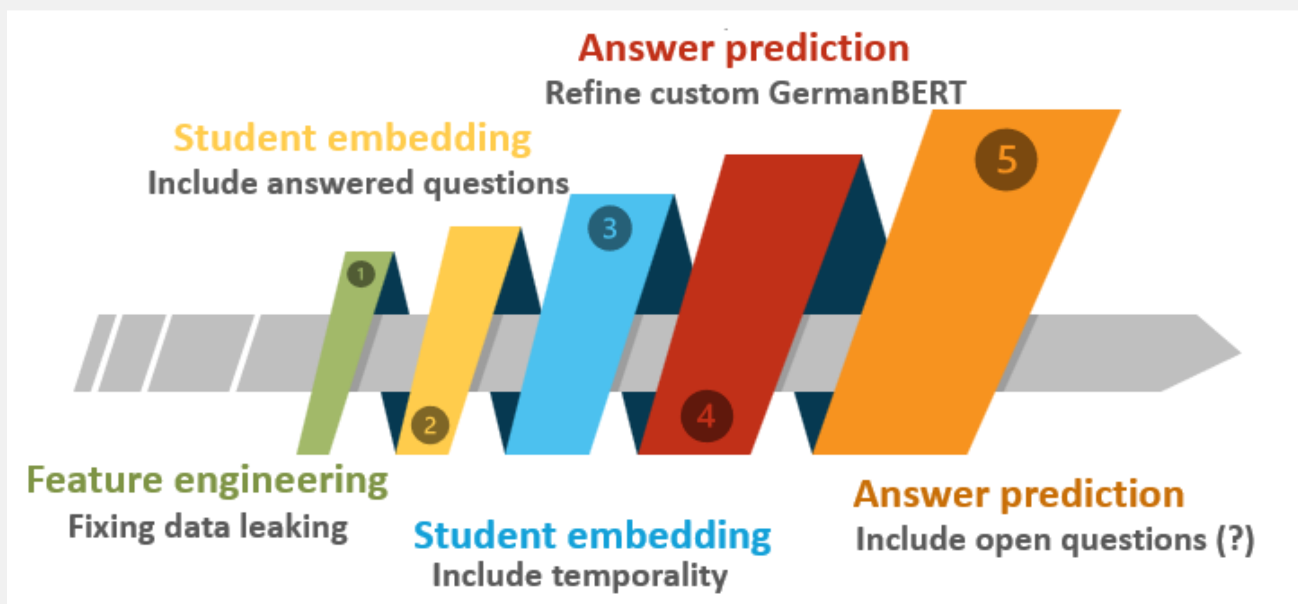


Figure 5. Confusion matrices for MCQBert3. Left/Center: correct answer prediction, MCQs never seen before/already seen. Right: student answer prediction, MCQs never seen before for the student under analysis.

4 CONCLUSIONS AND FUTURE DIRECTIONS

Masking finetuning: perplexity ↓, aligned word predictions.
MCQ finetuning: LernnaviBERT learns correct answers.
Students answers finetuning: effective (high F1 and MCC), student embeddings yield slight improvement over baseline.



REFERENCES

[1] C. Piech, J. Bassen, J. Huang, S. Ganguli, M. Sahami, L. J. Guibas, and J. Sohl-Dickstein, "Deep knowledge tracing", in Advances in Neural Information Processing Systems, vol. 28. Curran Associates, Inc., 2015

[2] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding", 2019.

École polytechnique fédérale de Lausanne