



BERT-BASED DISTRACTOR GENERATION FOR SWEDISH READING COMPREHENSION QUESTIONS USING A SMALL-SCALE DATASET

Dmytro Kalpakchi and Johan Boye

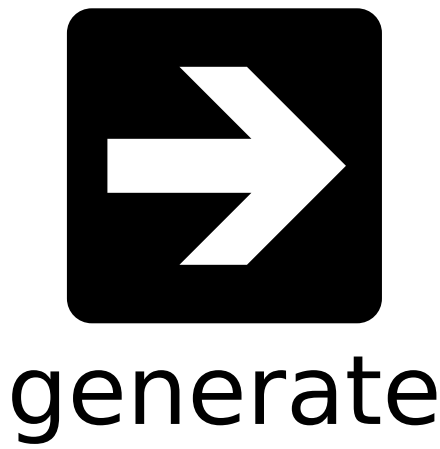
THE DISTRACTOR GENERATION PROBLEM

BASE TEXT T
[...] Medical checks may be needed if the risk assessment shows that the work may involve special risks. Medical check of minors is needed when the risk assessment identifies risks for the minor. Some examples include heavy lifting in healthcare or opening and closing a store if the minor is mature for these tasks. In addition to what is required by other regulations, one needs to focus on the minor's physical and mental development. You can read more in the regulations on medical checks in working life (AFS 2019: 3). If you, as an employer, employ children under the age of 13, you must apply for a permit. This applies, for example, when children under the age of 13 are to perform and rehearse in cultural or artistic activities and at sports and advertising events, such as artists, extras or the like. [...]

QUESTION (STEM) Q
When must the employer apply for a permission to employ a person?

ANSWER (KEY) A
a) if a person is under 13 y.o.

Input data



generate

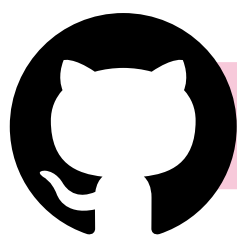
DISTRACTORS DX
b) if a person is under 18 y.o.
c) if a person has no parents
d) if a person is under 12 y.o.
e) when the stars are not right

Output data
 good!
 might be true
 not wrong
 obviously wrong

- What is a good distractor?**
- Wrong
 - Not obviously wrong!
 - Varied
 - Formulated similarly to the key

DATA

The new SweQUAD-MC dataset - now with distractors! 1190 multiple choice questions (MCQs) in **Swedish**. Each MCQ has base text, stem, key **and at least 2 distractors**.



github.com/dkalpakchi/SweQUAD-MC

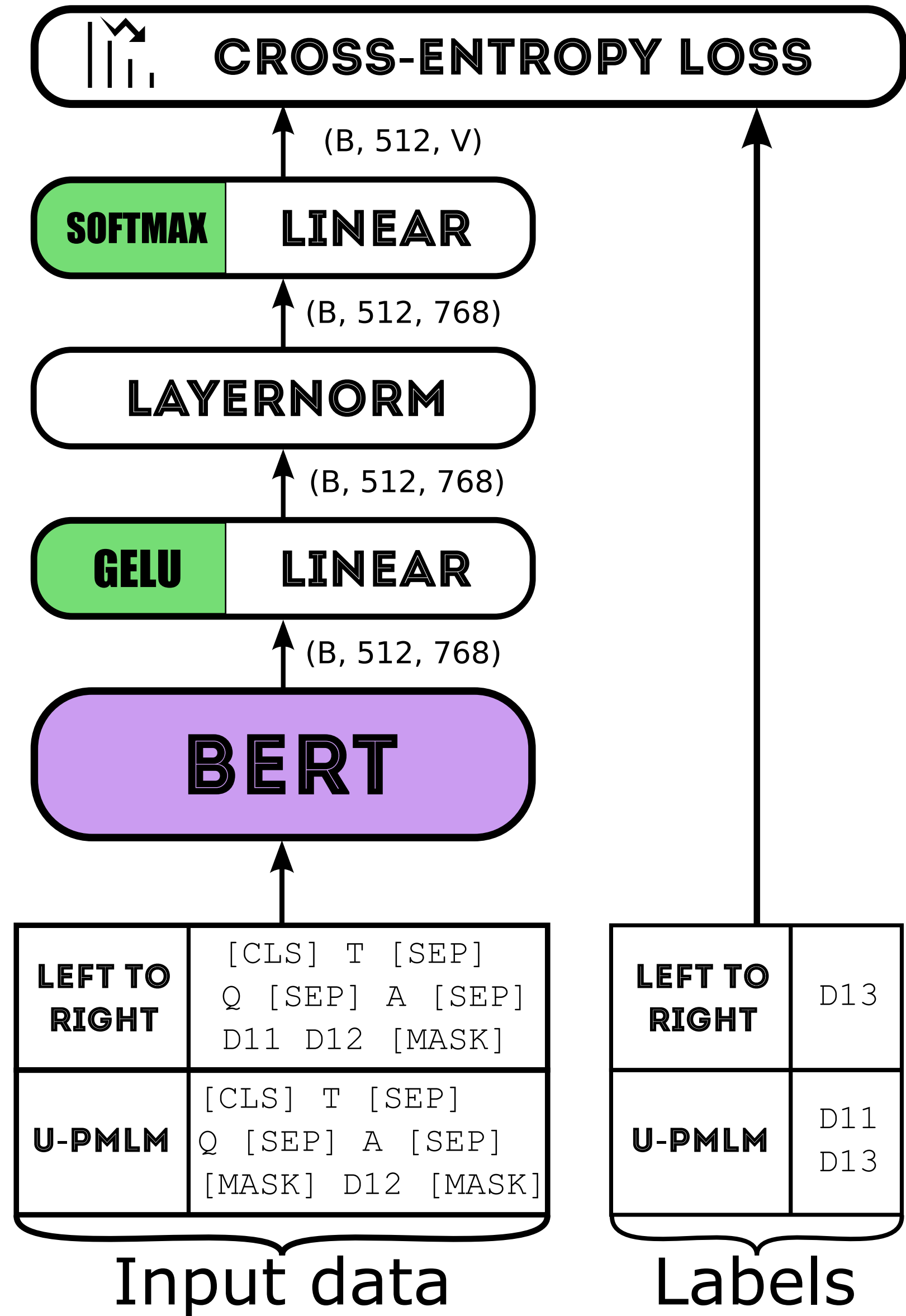
100x smaller than SQuAD!

WHAT IS A DATAPOINT?

CTX = T_{384} [SEP] Q [SEP] A T_{384} - the first 384 tokens of T

Input for left-to-right variant	Target
[CLS] CTX [SEP] [MASK]	D11
[CLS] CTX [SEP] D11 [MASK]	D12
[CLS] CTX [SEP] D11 D12 [MASK]	[SEP]
[CLS] CTX [SEP] D11 D12 [SEP] [MASK]	D21
[CLS] CTX [SEP] D11 D12 [SEP] D21 [MASK]	D22
[CLS] CTX [SEP] D11 D12 [SEP] D21 D22 [MASK]	D23
[CLS] CTX [SEP] D11 D12 [SEP] D21 D22 D23 [MASK]	[SEP]
Input for u-PMLM variant	Target(s)
[CLS] CTX [SEP] D11 [MASK]	D12
[CLS] CTX [SEP] [MASK] D12	D11
[CLS] CTX [SEP] D11 D12 [SEP] D21 [MASK] [MASK]	D22, D23
[CLS] CTX [SEP] D11 D12 [SEP] D21 [MASK] D23	D22
[CLS] CTX [SEP] D11 D12 [SEP] [MASK] D22 [MASK]	D21, D23

PROPOSED MODEL



HUMAN EVALUATION (U-PMLM VARIANT)

Student's perspective

Basic idea: if students can guess the correct answer (from 4 options) without reading the text, distractors are likely not good!

Formally: \bar{N}_s - average number of correctly answered MCQs

$H_0 : \bar{N}_s = 25\%$ One-sample $N = 54$ **Goal:**
 $H_1 : \bar{N}_s \neq 25\%$ two-tailed $\alpha = 0.05$ show that data
t-test $\beta = 0.05$ supports H_0

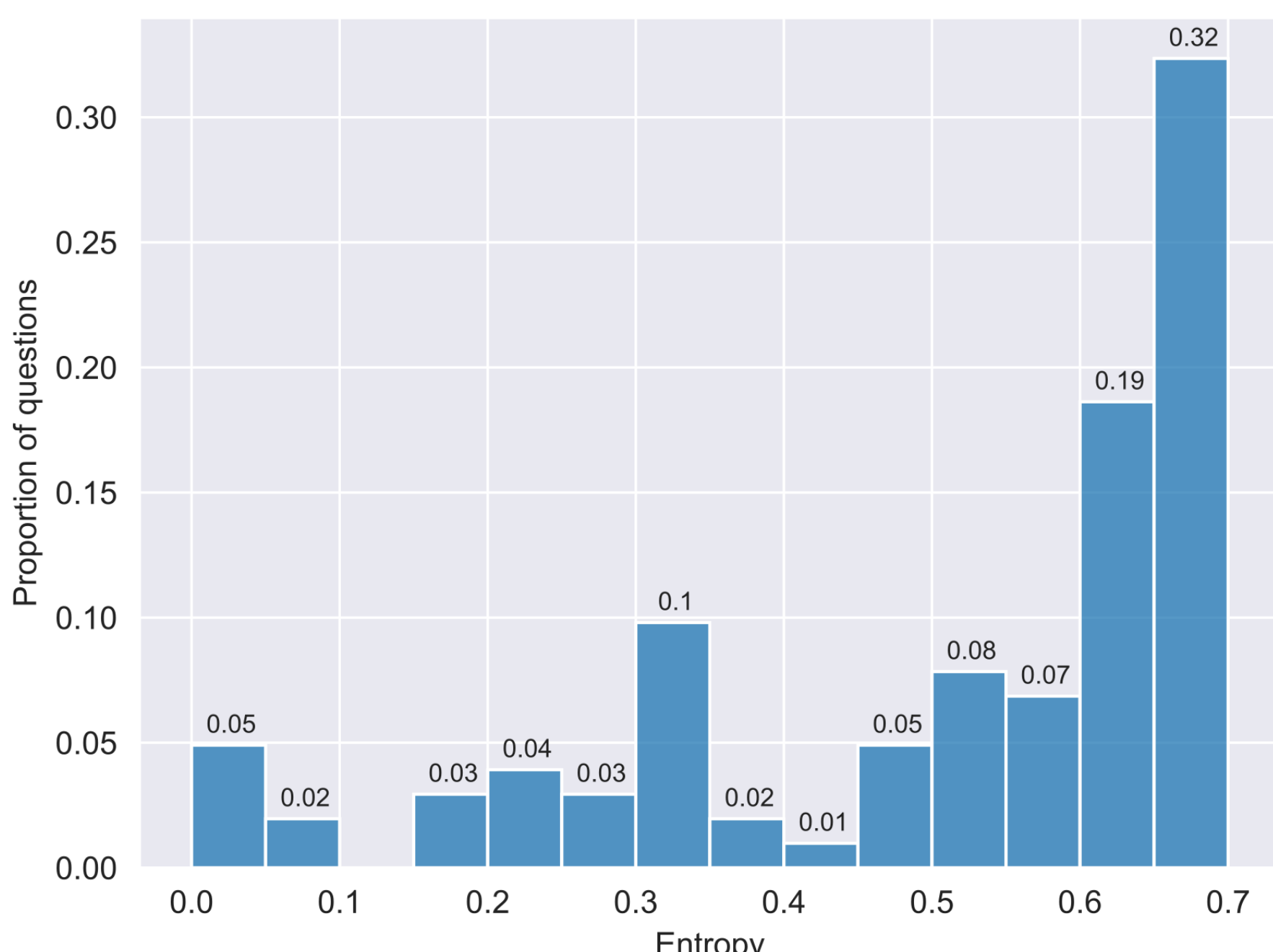
Data showed...

$\bar{N}_s \approx 61\%$

$(p < 0.05)$

... a clear support for H_1

We quantify distraction quality by entropy H between $P(A)$ and $P(D1 \cup D2 \cup D3)$
 $0 \leq H \leq 0.69$



Teacher's perspective

Basic idea: present 5 teachers MCQs with generated distractors and ask them which ones they would choose and why.

1.47 out of 3 distractors per MCQ were accepted by teachers!

Reasons for rejecting distractors

