Exploiting Numerical-Contextual Knowledge to Improve Numerical Reasoning in Question Answering

Jeonghwan Kim, Junmo Kang, Kyung-min Kim, Giwon Hong, Sung-Hyon Myaeng

The North American Chapter of the Association for Computational Linguistics (NAACL), 2022

IR&NLP Lab

Motivation

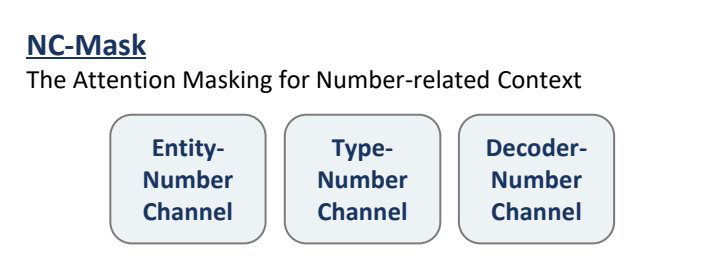
	Model Prediction	Ground Truth
Question : "Is John *older* or *younger*?" **Passage** : "John is in his *80s* and the other person is in her *90s*."	*older* ✗	*younger* ✓
Question : "Which age group makes up the *largest* percentage of population?" **Passage** : "14.3% ... are in their *20s*, ... 14.9% in their *40s* and 13.0% are in their *50s*."	*50s* ✗	*40s* ✓

In the first example, the model interprets **90s** and **80s** as **YEAR type** numbers, instead of the contextually correct **AGE type**

Thus, **identifying the correct number types are essential** to derive the correct answer in a numerical reasoning over text setting

Overview

Pre-trained language models (PLM) exhibit a general tendency to **overly rely on parametric knowledge** (i.e., knowledge acquired and stored in the parameters)



This phenomenon can also be evidenced in question answering (QA) tasks like **numerical reasoning over text**[1]

We propose a simple yet effective **regularized attention masking scheme** to alleviate the over-reliance issue and exploit the much-relevant contextual knowledge

Preliminary Study

Q) What kind of parametric knowledge reside within these PLM embeddings?

1. Sample **number embeddings (e.g., 2018)** from BERT-base
2. Use **FAISS to retrieve top-5 tokens** with highest cos. sim. score

What is inside the number embeddings?	
2018	currently, October, 1990s, July, 19th
50	1950s, various, significant, ...
114	12th, 11th, 14th, 15th, ...
11	11th, 12th, 10th, 8th, ...

Mostly pre-existing **DATE & TIME related knowledge** acquired from the pre-training and finetuning steps

Such **parametric knowledge influence how the numbers are treated and interpreted** within the given context

Approach

An attention masked QA model that leverages relevant context information flow to interpret numbers in text

→ **NC-BERT (Numerical-Contexutal BERT)**

NC-BERT consists of:

1. **Attention mask (NC-Mask)**
2. **A regularizer to retain numeracy** within the number embeddings.

NC-Mask

The Attention Masking for Number-related Context

Entity-Number Channel

$$\alpha = softmax\left(A_E \odot \frac{QK^T}{\sqrt{d_k}}\right)V$$

Gathers number-related **entity information**

Question: "How many employees work at Johnny's?"

Passage: "... There are around 73 employees working at Johnny's ..."

Type-Number Channel

$$\beta = softmax\left(A_T \odot \frac{QK^T}{\sqrt{d_k}}\right)V$$

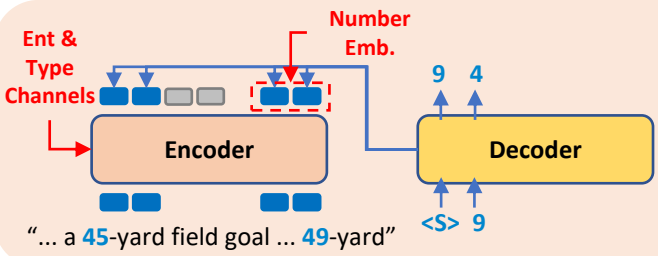
Gathers number-related surrounding token information that **defines the number type**

Question: "How many employees work at Johnny's?"

Passage: "... There are around 73 employees working at Johnny's ..."

Decoder-Number Channel

Reduces number-irrelevant context info.



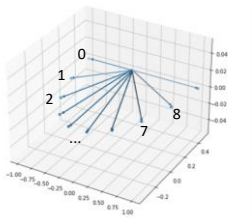
NC-Mask alone is not enough; the masking scheme overwrites the pre-existing numeracy (i.e., magnitude) information → **Numeracy Dilution Issue**

Adopting DICE Regularization to Retain Numeracy

DICE[2] regularization preserves the numeracy information within the number embeddings

$$L_{DICE} = \left\| 2\frac{|x-y|}{|x|+|y|} - d_{cos}(\boldsymbol{v_x}, \boldsymbol{v_y}) \right\|_2$$

$$L = L_{span} + L_{decoder} + L_{DICE}$$



Experiments & Results

Model	Number	Date	All
	F1	F1	F1
GenBERT	75.21	56.37	72.30
+ Entity-Num	76.24	56.33	72.61
+ Type-Num	75.30	**56.59**	72.31
+ Decoder-Num	75.37	55.98	72.34
NC-Mask	**76.89**	**56.32**	**72.65**
NC-Mask + DICE	**77.72**	**56.31**	**73.59**
GenBERT + DICE	76.12	55.98	72.38

Entity-Num & Type-Num channel improves both the Number and Date-type question accuracy → **Channels are complementary**

DICE-reg. alone does not improve numerical reasoning much → **NC-Mask & DICE reg. are interdependent to one another**

P) The total number of active military personnel in the Croatian Armed Forces stands at 14,506 and 6,000 reserves working in various service branches of the armed forces. In May 2016, Armed Forces had ...

Q) How many active military personnel and reserve are in the Croatian Armed Forces?

NC-BERT

14,506 + 6,000 = 20,506

Original

14,506 + ? = 14,506

References

[1] Talmor et al., 2020; *oLMpics – On What Language Model Pre-training Captures*, TACL 2020
[2] Sundararaman et al., 2020; *Methods for Numeracy-Preserving Word Embeddings*, EMNLP 2020