

Supplementary Material: NumNet: Machine Reading Comprehension with Numerical Reasoning

Qiu Ran^{1*}, Yankai Lin^{1*}, Peng Li¹, Jie Zhou¹, Zhiyuan Liu²

¹Pattern Recognition Center, WeChat AI, Tencent Inc, China

²Department of Computer Science and Technology, Tsinghua University, Beijing, China
Institute for Artificial Intelligence, Tsinghua University, Beijing, China

State Key Lab on Intelligent Technology and Systems, Tsinghua University, Beijing, China
{soulcaptran, yankailin, patrickpli, withtomzhou}@tencent.com
liuzy@tsinghua.edu.cn

A Baseline Enhancements

The major enhancements leveraged by our implemented NAQANet+ model include:

(1) “real number”: Unlike NAQANet only considers integer numbers, we also consider real numbers.

(2) “richer arithmetic expression”: We conceptually append an extra number “100” to the passage to support arithmetic expressions like “100-25”, which is required for answering questions such as “*How many percent were not American?*”.

(3) “passage-preferred”: If an answer is both a span of the question and the passage, we only propagate gradients through the output layer for processing “Passage span” type answers.

(4) “data augmentation”: The original questions in the DROP dataset are generated by crowd-sourced workers. For the comparing questions which contain answer candidates, we observe that the workers frequently only change the incorrect answer candidate to generate a new question. For example, “*How many from the census is bigger: Germans or English?*” whose golden answer is “Germans” is modified to “*How many from the census is bigger: Germans or Irish?*”. This may introduce undesired inductive bias to the model. Therefore, we propose to augment the training dataset with new questions automatically generated by swapping the candidate answers, e.g., “*How many from the census is bigger: English or Germans?*” is added to the training dataset.

We further conduct ablation studies on the enhancements. And the validation scores on the development set are shown in Table 1. As can be seen from Table 1:

(1) The uses of real number and richer arithmetic expression are crucial for answering numerical questions: both EM and F1 drop drastically by

up to 15 – 21 points if they are removed.

(2) The passage-preferred strategy and data augmentation are also necessary components that contribute significant improvements for those comparing questions.

*indicates equal contribution

Method	Comparison		Number		ALL	
	EM	F1	EM	F1	EM	F1
NAQANet+	69.11	75.62	66.92	66.94	61.11	64.54
- real number	66.87	73.25	45.82	45.85	47.82	51.22
- richer arithmetic expression	68.62	74.55	52.48	52.51	52.02	55.32
- passage-preferred	64.06	72.34	66.46	66.47	59.64	63.34
- data augmentation	65.28	71.81	67.05	67.07	61.21	64.60

Table 1: Baseline enhancements ablation.