

Recurrent Networks in AQA Answer Selection

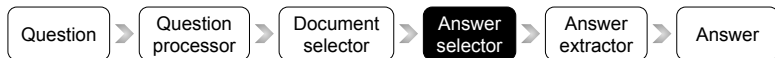
Radoslav Sabol, Marek Medved', Aleš Horák

Natural Language Processing Centre
Faculty of Informatics, Masaryk University

December 7, 2018

AQA

AQA system

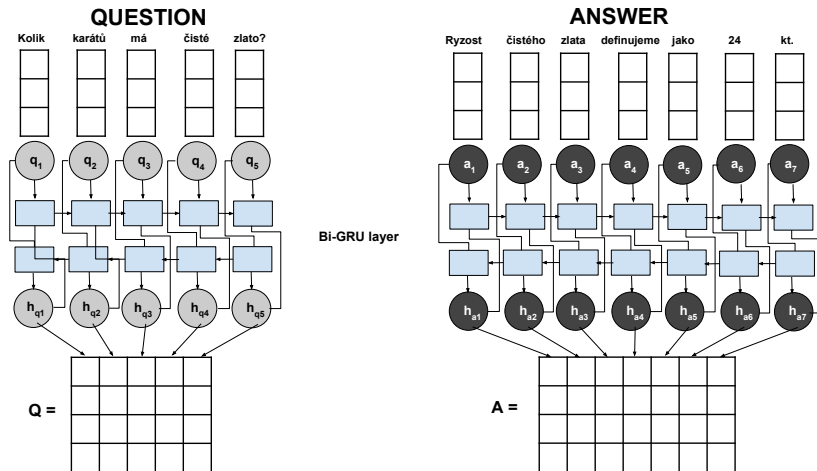


Answer Selection

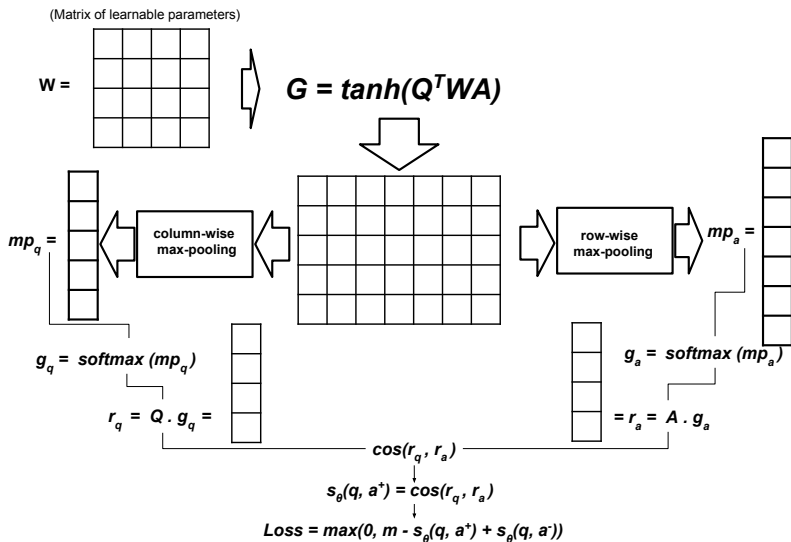
Answer Selection Task

question and set of possible answers are given
sorting answers by score
score reflects the probability of correct answer

Answer Selection Architecture



Answer Selection Architecture



Training and Configuration

Dataset

SQAD v2.1 - 8566 questions

Train set - 50%

Validation set - 10%

Test set - 40%

Training

25 epochs

50 randomly chosen answers for each question

$$\text{Loss} = \max\{0, m - s_{\theta}(q, a^{+}) + s_{\theta}(q, a^{-})\}$$

m - constant margin (0.2)

s_{θ} is the cosine similarity as computed by the network with parameters θ , q is the input question and a^{+}/a^{-} are the positive/negative answers.

Training

Stochastic Gradient Descent (SGD)

Adaptive learning rate

$$\lambda_t = \frac{\lambda}{t}$$

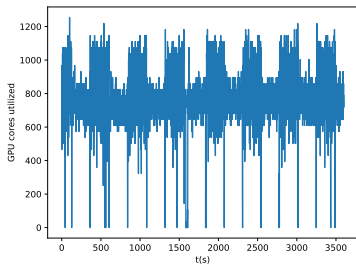
λ_t - learning rate at epoch t

t - current epoch

λ - initial learning rate

GPU Utilization

a) one model



b) two models at the same time

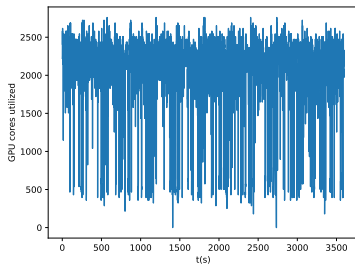


Figure: GPU core utilization for one hour of training. Maximum number of working GPU cores is 3584.

CPU running time for epoch – approx. 3 hours

GPU running time for epoch – 380 seconds, 28× improvement

Results

Table: The results for combinations of hyperparameter values

Output size	Dropout	Learning rate λ	Test set (in %)
260	0.2	0.05	61.89
		0.1	65.47
		0.2	68.29

Table: Comparison with previous version

Pos.	old answer selection No. in %	current answer selection No. in %
1	56.53	66.03

Answer selection results for particular question-answer types

q-a type	total	first pos.	accuracy
LOCATION - OTHER	2	2	100.00%
ADJ_PHRASE - PERSON	2	2	100.00%
LOCATION - ORGANIZATION	1	1	100.00%
DATETIME - NUMERIC	1	1	100.00%
ABBREVIATION - ABBREVIATION	32	28	87.50%
LOCATION - LOCATION	600	464	77.33%
ENTITY - ORGANIZATION	81	60	74.07%
DATETIME - DATETIME	737	538	73.00%
VERB_PHRASE - YES_NO	374	254	67.91%
ENTITY - OTHER	151	101	66.89%
ADJ_PHRASE - ENTITY	3	2	66.67%
PERSON - PERSON	406	265	65.27%

Answer selection results for particular question-answer types

q-a type	total	first pos.	accuracy
ADJ_PHRASE - OTHER	86	56	65.12%
ENTITY - TEACHING	40	26	65.00%
NUMERIC - NUMERIC	364	228	62.64%
ENTITY - PERSON	8	5	62.50%
ENTITY - ENTITY	412	249	60.44%
LOCATION - ENTITY	5	3	60.00%
CLAUSE - OTHER	82	38	46.34%
OTHER - OTHER	5	2	40.00%
CLAUSE - ENTITY	11	2	18.18%
PERSON - ORGANIZATION	1	0	0.00%
ENTITY - LOCATION	2	0	0.00%
CLAUSE - ORGANIZATION	2	0	0.00%
sum	3408	2327	68.29%

Example (Correctly answered question, Type: PERSON - PERSON)

Question: Jak nazval Kolumbus obyvatele ním objevené země?

Chosen answer: Kolumbus nevěděl, že objevil pro Evropu nový kontinent a obyvatele objevených zemí nazval Indios (španělský výraz pro Indy).

- **SCORE = 0.71075**

Candidate answers that ended up on 2nd and 3rd place:

Kolumbus nebyl prvním Evropanem , který navštívil Ameriku – o pět století dříve jejich břehů dosáhla norská expedice vedená Leifem Erikssonem , která založila kolonii na dnešním Newfoundlandu.

- **SCORE = 0.60405**

V roce 1538 nazval Gerhard Mercator Amerikou celý kontinent .

- **SCORE = 0.53724**

Example (Incorrectly answered question, Type: VERB_PHR - YES_NO)

Question: Má pivo příznivé účinky?

Chosen incorrect answer: Současně výrazně podporuje chuť k jídlu , což může vést při nestřídmé konzumaci pokrmů k nárůstu tělesné hmotnosti .

- **SCORE = 0.66412**

Right answer (that ended up in 5th position):

Uvádí se, že konzumace piva má příznivé účinky na dobrou náladu, podporu krevního oběhu, snížení rizika srdečních příhod a působí proti vysokému krevnímu tlaku.

- **SCORE = 0.64472**

Future work

Future work

further improve current answer selection module
connect answer selection module to current pipeline
train model for phrase similarity