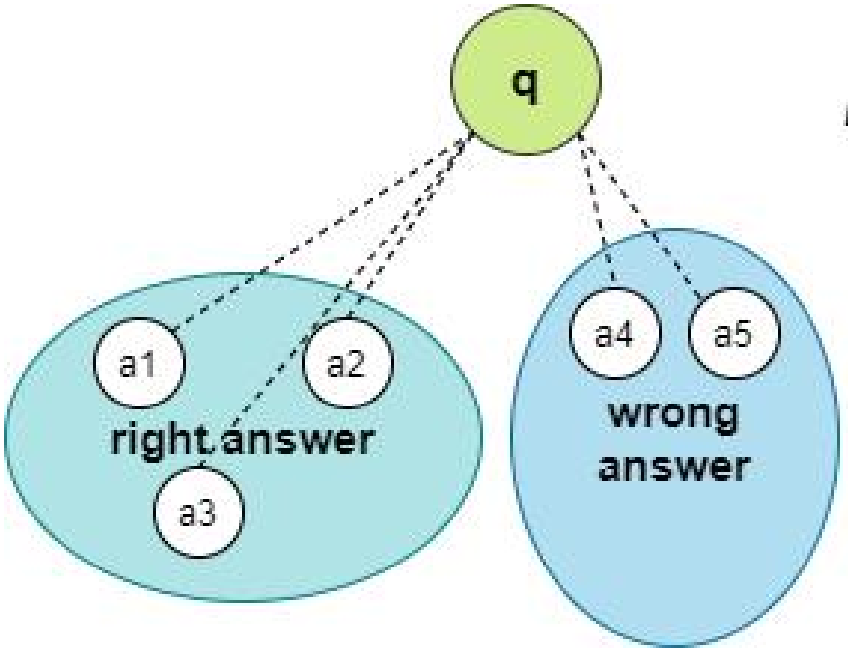
Lightweight Multiple Perspective Fusion with Information Enriching for BERT-based Answer Selection

Sun Yat-sen University

Yu Gu, Meng Yang, Peiqin Lin

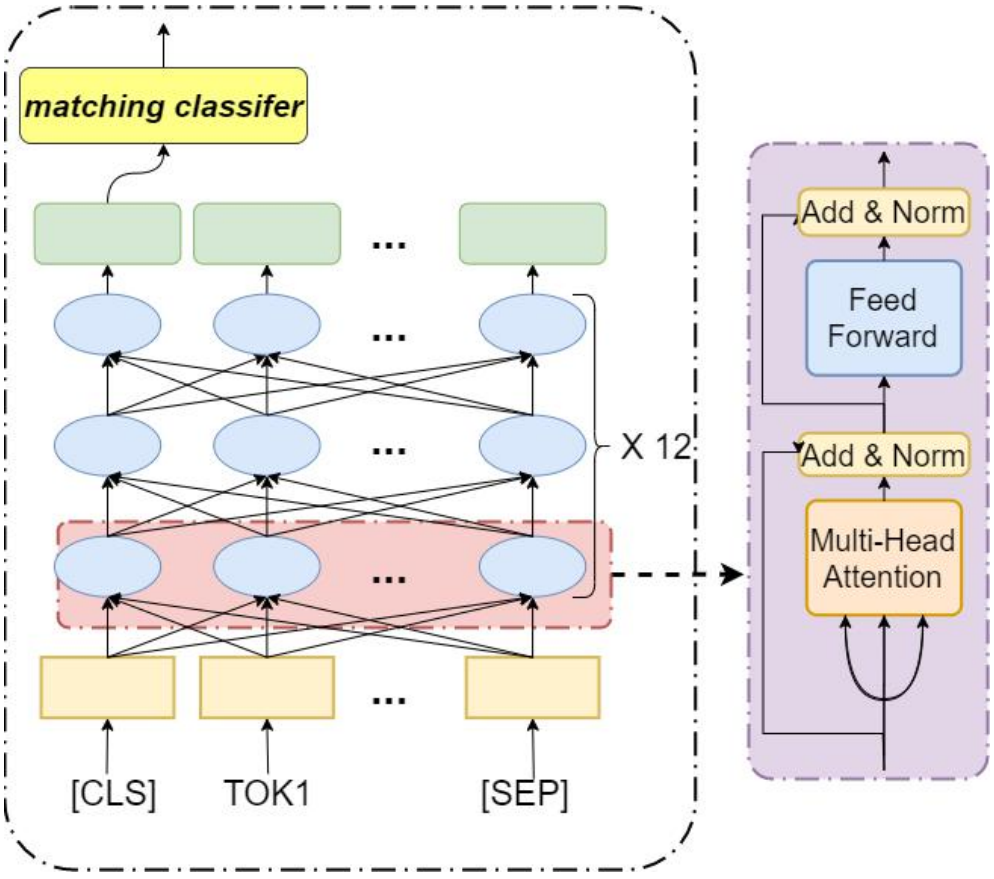
Task description: Answer Selection



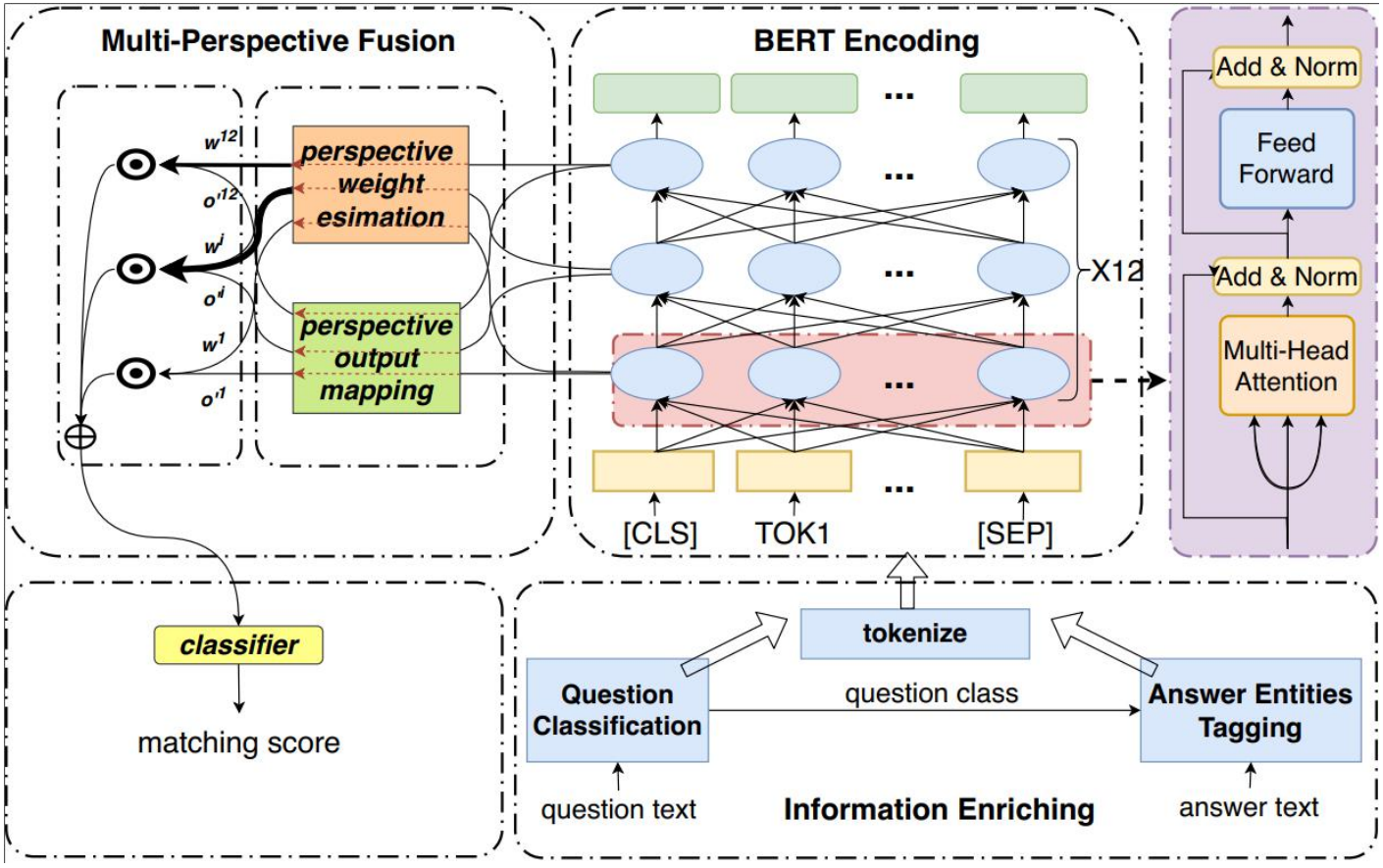
match and rank

q-a1
q-a2
q-a5
q-a4
q-a3

Baseline Method



Method Framework



Information Inriching

A wrong case of BERT

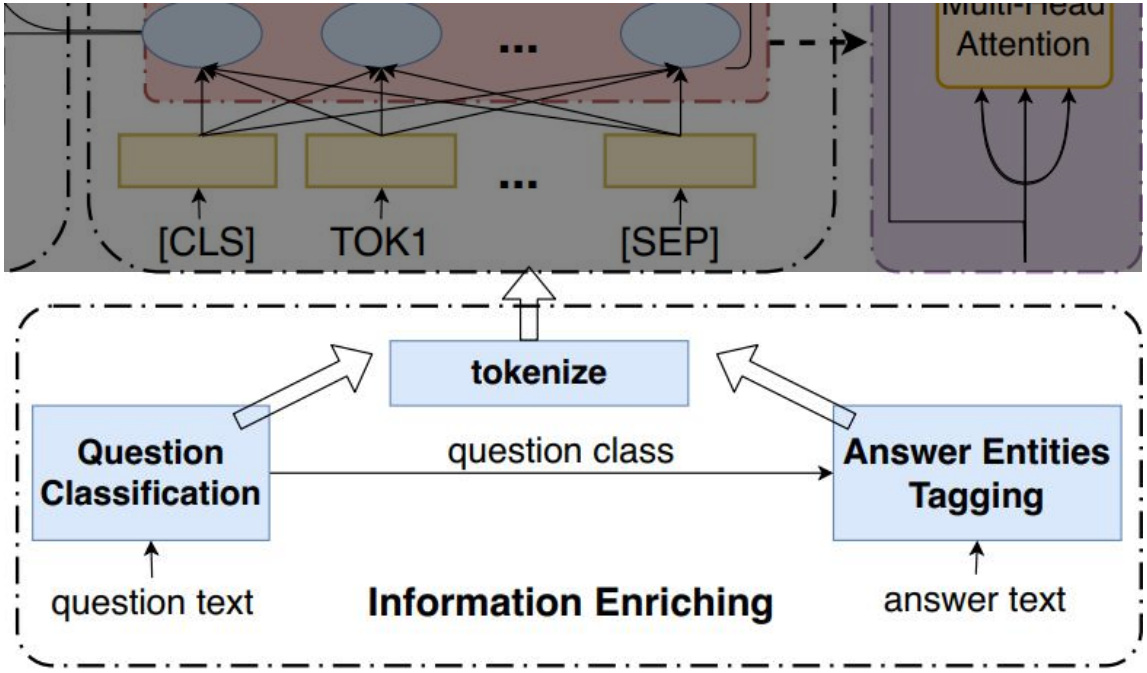
Question	How African Americans were immigrated to the US?"
Right Answer	As such, African immigrants are to be distinguished from African American people, the latter of whom are descendants of mostly West and Central Africans who were involuntarily brought to the United States by means of the historic Atlantic slave trade ."
Wrong Answer	From the Immigration and Nationality Act of 1965 to 2007, an estimated total of 0.8 to 0.9 million Africans immigrated to the United States, accounting for roughly 3.3% of total immigration to the United States during this period."

How (Manner)

When (Time)

How many (Quantity)

Information Inriching



Information Inriching

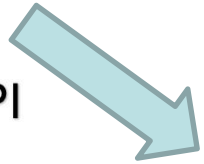
before	Question	How big is bmc software in houston, tx".
	Answer:	Employing over 6,000, BMC is often credited with pioneering the BSM concept as a way to help better align IT operations with business needs.".
after	Question	How big is bmc software in houston, tx (number, volume size)"
	Answer:	Employing [over 6,000](number, cardinal), BMC is often credited with pioneering the BSM concept ...

Information Inriching

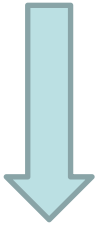
How big is bmc software in houston, tx

How big is bmc software in houston, tx (number, volume size)

QC API



NUM,volsize
number, volume size



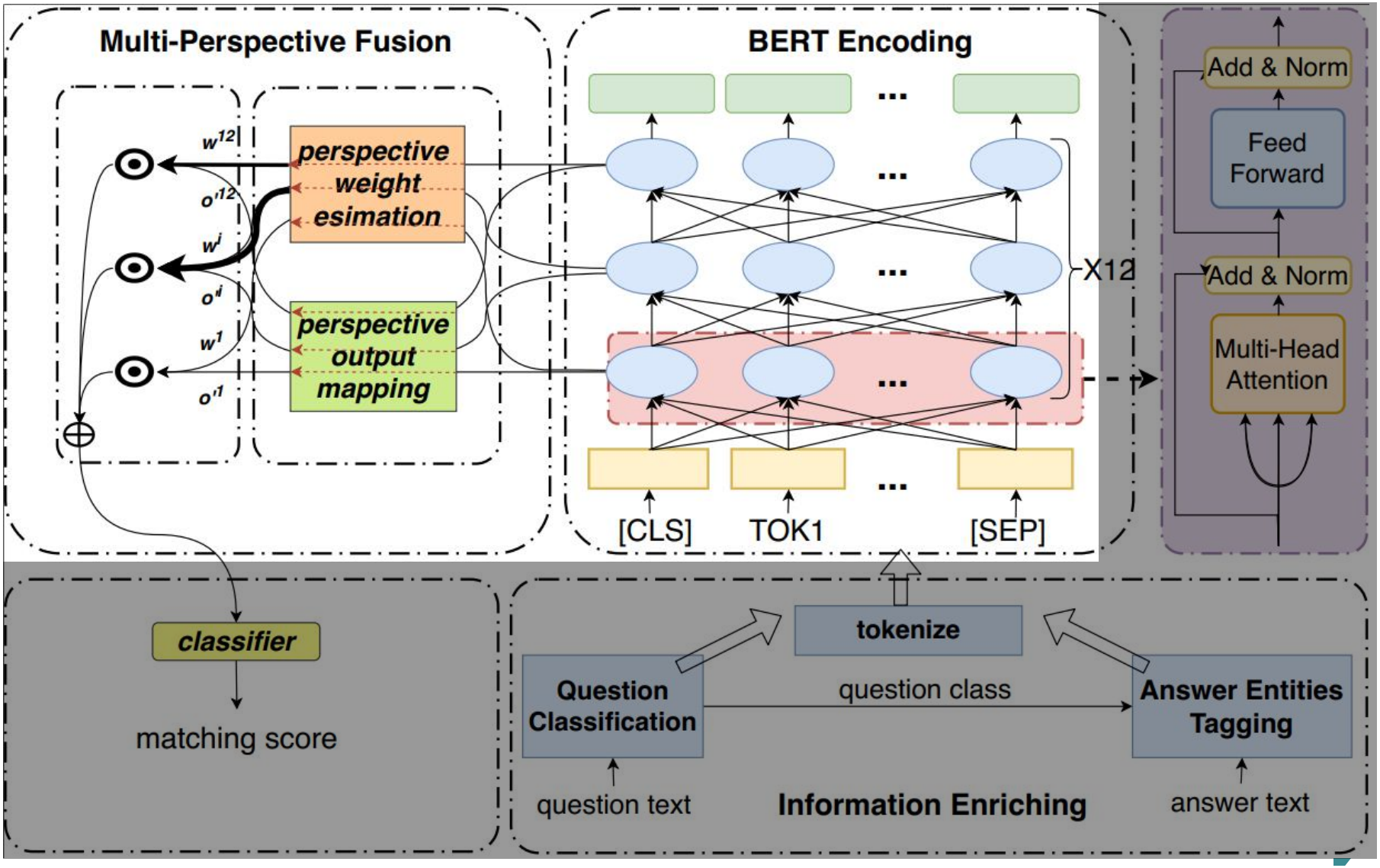
NER tools

over 6,000

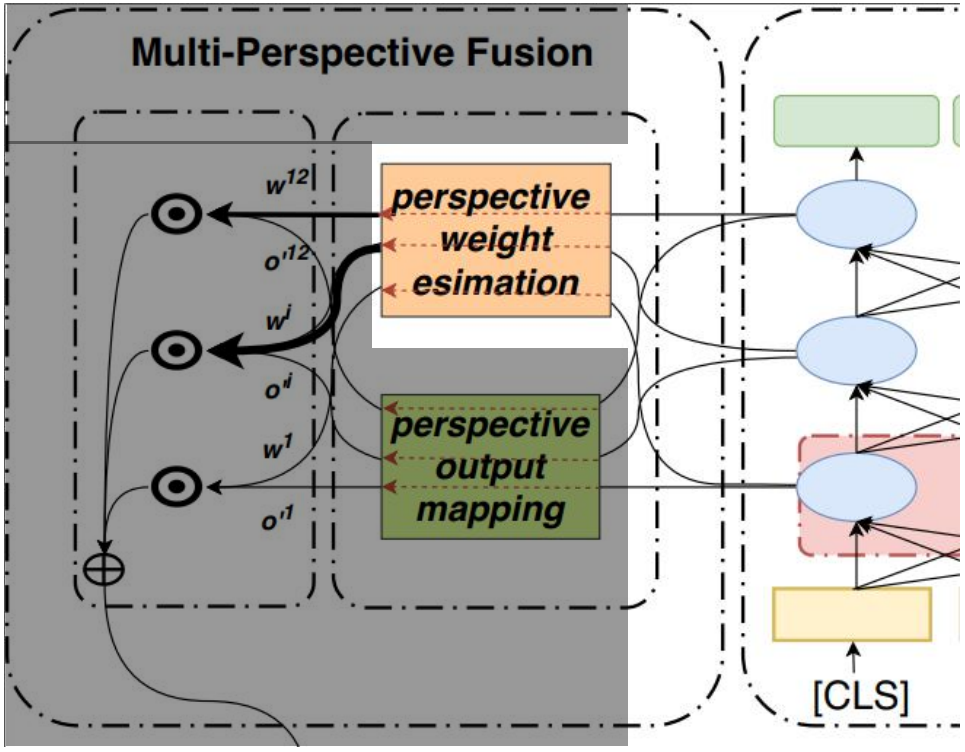
Employing **over 6,000**, **BMC** is often credited with pioneering the BSM concept as a way to help better align IT operations with business needs.".

Employing [over 6,000] (number, cardinal),
BMC is often credited with pioneering the BSM concept ...

Multi-perspective fusion



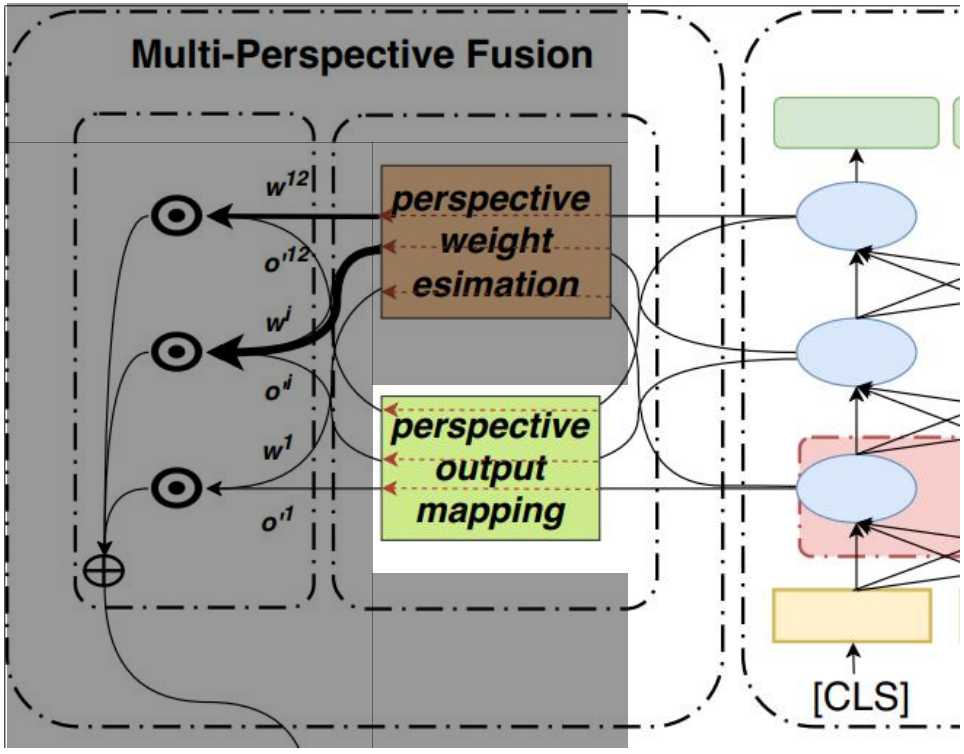
Multi-perspective fusion



$$\boldsymbol{w}' = f_w(\boldsymbol{E}_{CLS})$$

$$\boldsymbol{w} = softmax(\boldsymbol{w}')$$

Multi-perspective fusion



$$\boldsymbol{V}_{CLS}^{i} = tanh(\boldsymbol{W}_p \boldsymbol{E}_{CLS}^{i} + \boldsymbol{b}_p)$$

$$\boldsymbol{o} = \sum_{i=0}^{l} w_i \boldsymbol{V}_{CLS}^{i}$$

Implementation

Weight estimation function：linear layer

$$f_w(\boldsymbol{E}_{CLS}) = \boldsymbol{W}_l \boldsymbol{E}_{CLS} + \boldsymbol{b}_l$$

Loss function : margin loss

$$L = \max(0, m - \mathrm{sim}(q, a+) + \mathrm{sim}(q, a-))$$

Experiments

Datasets

Dataset	Train	Dev	Test
TrecQA	1229	65	68
WikiQA	873	126	243
YahooQA	50112	6289	6283
SemEvalcQA-16	4879	244	327
SemEvalcQA-17	4879	244	293

Experiments

Evaluation Metrics

$$MAP(Q) = \frac{1}{Q}\sum_{j=1}^{|Q|}\frac{1}{m_j}\sum_{k=1}^{m_j} Precision(R_{jk})$$

$$MRR(Q) = \frac{1}{|Q|}\sum_{j=1}^{|Q|}\frac{1}{rank_j}$$

where Q is the set of questions, m_j is the number of relevant answers to q_j, R_{jk} is a list of candidate answers that contains top k relevant answers, $Precision$ is a function that measures the ratio of the number of relevant answers to the total candidate answers, $rank_j$ is the first relevant answer rank for q_j.

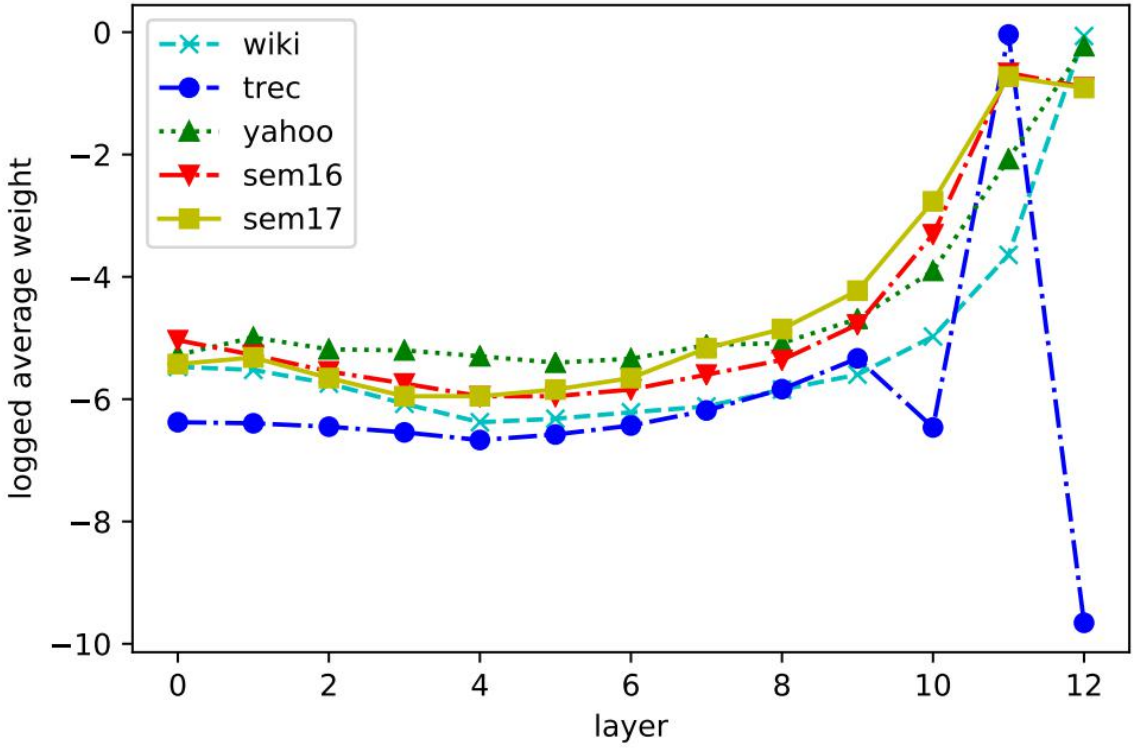
Experiments

Main Results

	Wiki		Trec		Yahoo		Semeval-16		Semeval-17		Mean	
	MAP	MRR	MAP	MRR	MAP	MRR	MAP	MRR	MAP	MRR	MAP	MRR
AP-CNN [15]	0.689	0.696	0.753	0.851								
Multihop-Sequential-LSTM [21]	0.722	0.738	0.813	0.893	-	-	-	-	-	-	-	-
BiMPM [27]	0.718	0.731	0.802	0.875	-	-	-	-	-	-	-	-
KeLP [6]	-	-	-	-	-	-	-	-	0.884	0.928	-	-
HyperQA [20]	0.712	0.727	0.770	0.825	0.683	0.801	-	-	-	-	-	-
MVFNN [17]	0.746	0.758	-	-	-	-	0.800	0.8718	-	-	-	-
Comp-Clip (LM + LC + TL) [14]	0.834	0.848	0.875	0.940	-	-	-	-	-	-	-	-
CNN + SRewrite + GAN [12]	0.737	0.751	0.788	0.842	-	-	-	-	-	-	-	-
BERT-Cross-Bi-nGRUR [18]	0.760	0.773	0.813	0.880	0.806	0.886	-	-	-	-	-	-
Base + RNN [5]	0.784	0.801	0.872	0.899	-	-	-	-	-	-	-	-
BERTSel (our reproduced [2])	0.807	0.821	0.877	0.923	0.966	0.966	0.825	0.898	0.903	0.945	0.876	0.911
BERTSel (Mult-Int) [11]	0.793	0.811	0.893	0.934	0.969	0.969	0.826	0.889	0.897	0.941	0.876	0.908
BERTSel(IE)	0.855	0.870	0.889	0.930	0.963	0.963	0.822	0.895	0.899	0.941	0.886	0.920
BERT-LMPF	0.850	0.866	**0.905**	**0.957**	**0.972**	**0.971**	0.826	0.903	**0.903**	**0.948**	0.891	0.929
BERT-LMPF(IE)	**0.861**	**0.873**	0.903	0.950	0.971	0.971	**0.832**	**0.905**	0.901	0.945	**0.893**	**0.929**

Experiments

the contribution of different layers to the final results



Contribution

Two tricks for applying BERT in Answer Selection:

- Information Enriching integrates Question Category and NER types into BERT when preprocessing

- Multiple perspective fusion fuses the multi-perspective discrimination from different layers inside the BERT.

Thank You