AAAI 2019, HONOLULU, U.S.A.

Multi-Task Learning with Multi-View Attention for Answer Selection and Knowledge Base Question Answering

Yang Deng, Yuexiang Xie, Yaliang Li, Min Yang, Nan Du, Wei Fan, Kai Lei, Ying Shen

School of Electronics and Computer Engineering, Peking University Shenzhen Graduate School
Tencent Medical AI Lab
Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences

Presenter: **Yang Deng** (ydeng@pku.edu.cn)
Center for Internet Research and Engineering , Peking University (PKU)
Jan 29th, 2019. Honolulu, U.S.A.

北京大学 PEKING UNIVERSITY

Outline

- Introduction
  - Answer Selection & Knowledge Base Question Answering
  - Multi-Task Learning
- Methodology
  - Problem Definition
  - Multi-Task Learning for Question Answering
  - Multi-View Attention Scheme
- Experiment
  - Multi-Task Learning Results
  - Ablation Analysis of Multi-View Attention
  - Case Study of Multi-View Attention
- Summary

Introduction

Answer Selection

- Problem

  Given a question, answer selection aims to pick out the correct answer from a set of candidates.

- Current Study

  Deep Learning Model, Attention Mechanism, External Knowledge

Knowledge Base Question Answering (KBQA)

- Problem

  Given a question, KBQA aims to pick out a fact from a given knowledge base to answer the question.

- Current Study

  Semantic Parsing, Deep Learning Model, Contextual Information

Introduction

Multi-Task Learning

- Multi-task learning aims to jointly learn different related tasks.

- Motivations
  - both answer selection and KBQA can be regarded as a ranking problem, with one at text-level while the other at knowledge-level.
  - these two tasks can benefit each other: answer selection can incorporate the external knowledge from knowledge base, while KBQA can be improved by learning contextual information from answer selection.

Methodology

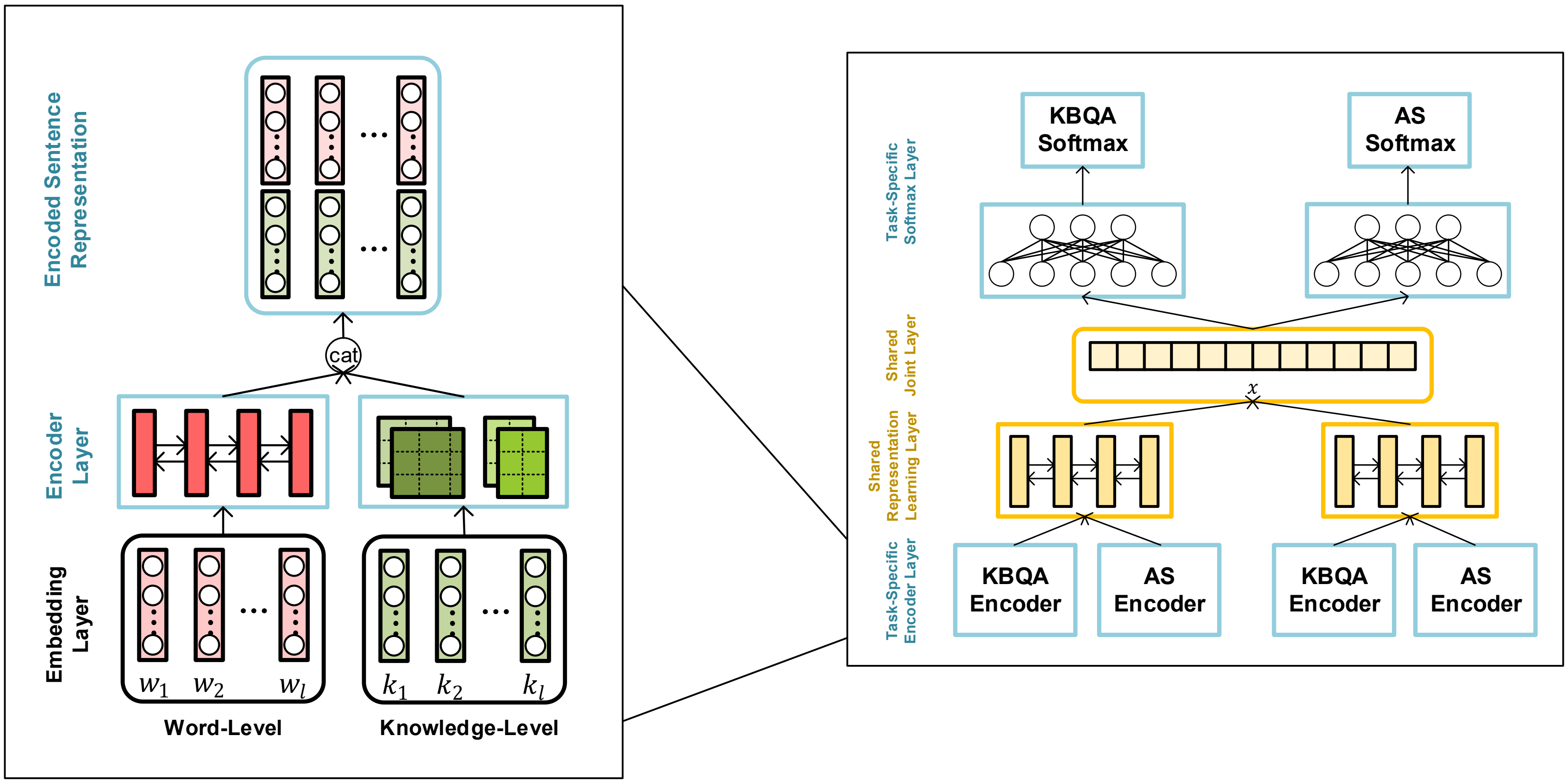
Problem Definition

- Ranking Problem: Given a question $q_i \in Q$, the task is to rank a set of candidate answer sentences or facts $a_i \in A$.
- Inputs:
  - a word sequence $W = \{w_1, w_2, \ldots, w_L\}$
  - a knowledge sequence $K = \{\mathrm{k}_1, \mathrm{k}_2, \ldots, \mathrm{k}_L\}$
  - D_t as the t-th preprocessed task dataset with N samples: $D\_t = \{(W_{q_i}^t, K_{q_i}^t, W_{a_i}^t, K_{a_i}^t, Y_i^t)\}_{i=1}^{N_t}$, where Y_i^t denotes the label of i-th QA pair in t-th task.
- Outputs:
  - a relevancy score $f(q, a) \in [0, 1]$ for each question-answer pair

		Answer Selection	KBQA
word	Q	what was johnny appleseed 's real name ?	what is the name of a track created by katy perry ?
	A	john chapman , aka american folk hero johnny appleseed .	katy perry music artist track witness
knowledge	Q	johnny_appleseed	katy_perry
	A	john_chapman, johnny_appleseed	katy_perry, music.artist.track, witness

Methodology

- Multi-Task QA network (MTQA-net)
  - Task-specific Encoder Layer



Methodology

Multi-Task Learning for Question Answering

- Shared Representation Learning Layer

$$S_q = \textbf{Bi-LSTM}(H_q); \quad S_a = \textbf{Bi-LSTM}(H_a).$$

$$s_q = Average(S_q),\ s_a = Average(S_a).$$

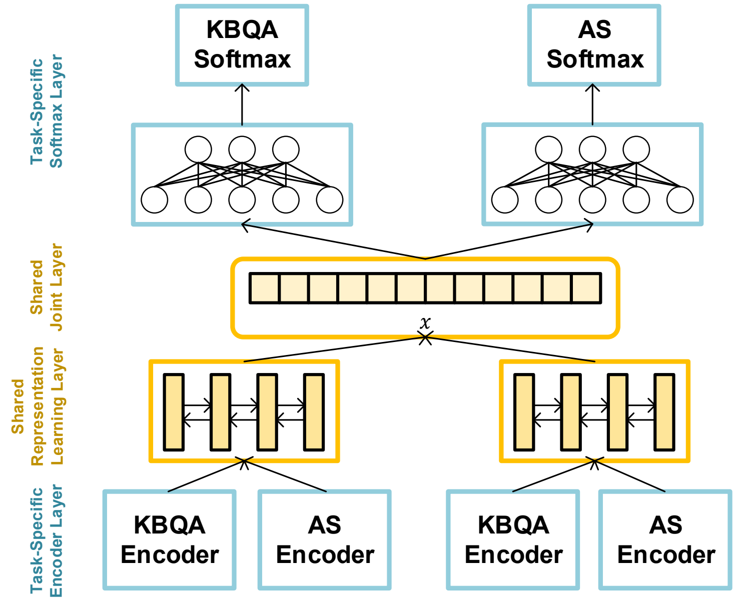
$$x = [s_q, s_a, x_{ol}].$$

- Task-specific Softmax Layer

$$p^{(t)} = \text{softmax}\left(W_s^{(t)} x + b_s^{(t)}\right)$$

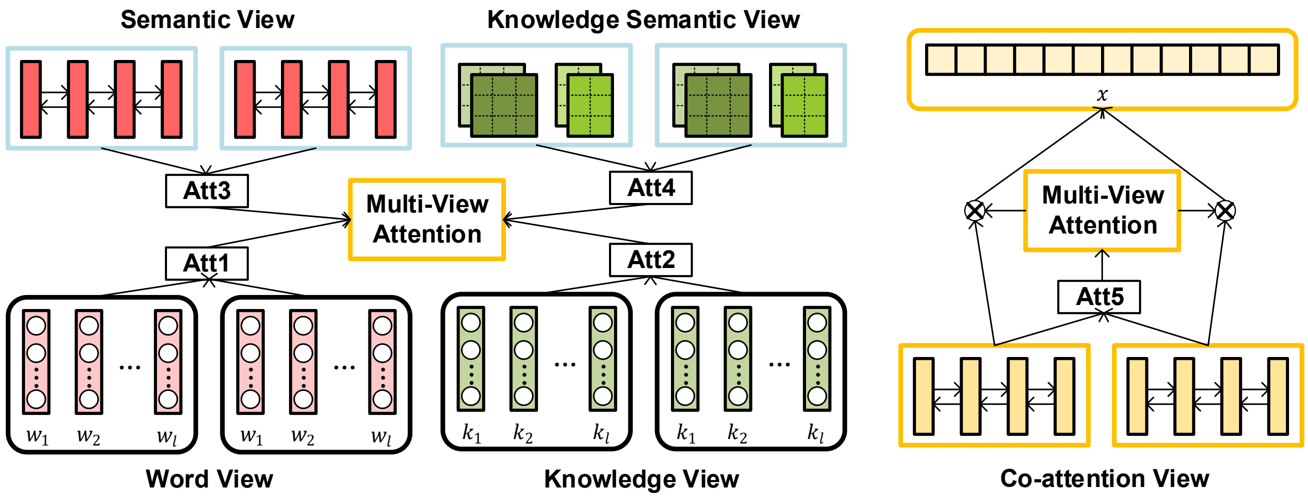
- Multi-Task Learning

$$L = -\sum_{t=1}^{T} \lambda_t \sum_{i=1}^{N_t} \left[y_i^{(t)} \log p_i^{(t)} + \left(1 - y_i^{(t)}\right) \log\left(1 - p_i^{(t)}\right) \right]$$



Methodology

Multi-View Attention Scheme



- Word View & Knowledge View & Co-attention View
  - Two-way attention
- Semantic View & Knowledge Semantic View
  - Semantic information as attention source
- Multi-View Attention Sharing
  - Two-way attention

Experiment

Datasets

	Dataset	#Question (train/dev/test)	#QA Pairs (train/dev/test)
Answer Selection	Yahoo QA	50098/6289/6283	253K/31K/31K
	TREC QA	1229/82/100	53417/1148/1517
KBQA	SimpleQuestions	71038/10252/20464	571K/80K/164K
	WebQSP	3067/-/1632	302K/-/160K

Experiment

Multi-Task Learning Results

Model	Yahoo QA		TREC QA		SimpleQuestions	WebQSP
	P@1	MRR	MAP	MRR	Accuracy	Accuracy
HD-LSTM (Tay et al. 2017)	0.557	0.735	0.750	0.815	-	-
CTRN (Tay, Tuan, and Hui 2018a)	0.601	0.755	0.771	0.838	-	-
HyperQA (Tay, Tuan, and Hui 2018b)	0.683	0.801	0.770	0.825	-	-
KAN(AP-LSTM) (Deng et al. 2018)	0.744	0.840	0.797	0.850	-	-
BiCNN (Yih et al. 2015)	-	-	-	-	0.900	0.777
AMPCNN (Wenpeng et al. 2016)	-	-	-	-	0.913	-
HR-BiLSTM (Yu et al. 2017)	-	-	-	-	0.933	0.825
Multiple View Matching (Yu et al., 2018)	-	-	-	-	0.937	0.854
MTQA-net (STL)	0.737	0.818	0.763	0.832	0.913	0.808
MTQA-net (MTL)	0.752	0.839	0.779	0.841	0.922	0.820
MVA-MTQA-net (STL)	0.806	0.878	0.783	0.838	0.931	0.823
MVA-MTQA-net (MTL)	**0.833**	**0.909**	**0.811**	**0.862**	**0.957**	**0.858**

- MVA-MTQA-net (MTL) outperforms the state-of-the-art results by a noticeable margin on all the datasets
- In both MVA-MTQA-net and its basic model (MTQA-net), multi-task learning (MTL) methods can significantly improve the performance of all four datasets compared with single-task learning (STL)

Experiment

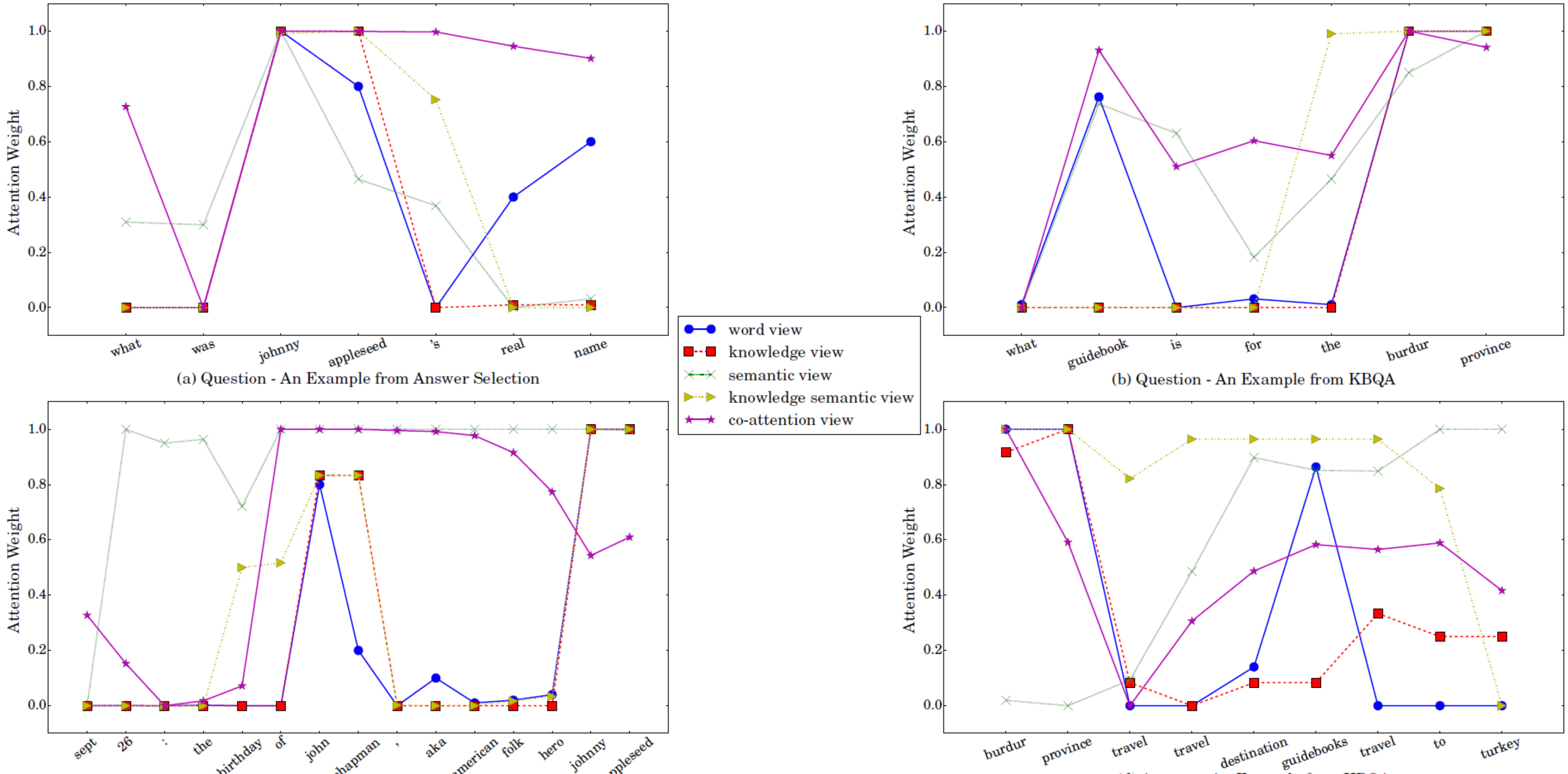
Ablation Analysis of Multi-View Attention

	Model	Yahoo QA		TREC QA		SimpleQuestions	WebQSP
		P@1	MRR	MAP	MRR	Accuracy	Accuracy
STL	MTQA-net	0.737	0.818	0.763	0.832	0.913	0.808
MTL	MTQA-net	0.752	0.839	0.779	0.841	0.922	0.820
STL	MVA-MTQA-net	0.806	0.878	0.783	0.838	0.931	0.823
	w/o word view	0.792	0.863	0.769	0.834	0.926	0.809
	w/o knowledge view	0.781	0.854	0.761	0.827	0.930	0.818
	w/o semantic view	0793	0.862	0.773	0.837	0.921	0.813
	w/o knowledge semantic view	0.788	0.859	0.762	0.822	0.928	0.814
	w/o co-attention view	0.775	0.850	0.761	0.824	0.917	0.803
MTL	MVA-MTQA-net	**0.833**	**0.909**	**0.811**	**0.862**	**0.957**	**0.858**
	w/o word view	0.824	0.894	0.792	0.854	0.947	0.835
	w/o knowledge view	0.826	0.893	0.796	0.861	0.944	0.844
	w/o semantic view	0.822	0.886	0.789	0.856	0.945	0.836
	w/o knowledge semantic view	0.822	0.890	0.793	0.856	0.944	0.840
	w/o co-attention view	0.811	0.882	0.792	0.847	0.937	0.829

- All kinds of view contribute more or less performance boost to the model.
- Co-attention view attention makes the most contribution to the improvement
- For STL, knowledge and knowledge semantic view attentions are more distinguishable than word view and semantic view in two answer selection tasks, while the word view and semantic attentions contribute more in two KBQA tasks
- For MTL, we observe that each view of attention makes a similar contribution to the improvement in four tasks

Experiment

Case Study of Multi-View Attention



(a) Question - An Example from Answer Selection

(b) Question - An Example from KBQA

(c) Answer - An Example from Answer Selection

(d) Answer - An Example from KBQA

Summary

- We explore multi-task learning approaches for answer selection and knowledge base question answering.
- We propose a novel multi-task learning scheme that leverages multi-view attention mechanism to bridge different tasks.
- Experimental results show that multi-task learning of answer selection and KBQA outperforms state-of-the-art single-task learning methods.

Thanks and QA

Thank you !

Presenter: **Yang Deng** **(ydeng@pku.edu.cn)**
Center for Internet Research and Engineering , Peking University (PKU)
January 29th, 2019. Honolulu, USA.