Xin-yu Duan, Si-liang Tang, Sheng-yu Zhang, Yin Zhang, authorcr Zhou Zhao, Jian-ru Xue, Yue-ting Zhuang, Fei Wu, 2018. Temporality-enhanced knowledge memory network for factoid question answering. *Frontiers of Information Technology & Electronic Engineering*, 19(1): 104-115.
https://doi.org/10.1631/FITEE.1700788

Temporality-enhanced knowledge memory network for factoid question answering

Key words: Question answering; Knowledge memory; Temporality interaction

Contact: Si-liang TANG
E-mail: siliang@zju.edu.cn
ORCID: http://orcid.org/0000-0001-6837-2634

Introduction

- The existing approaches focus merely on exploiting the semantic relevance between individual sentence with its corresponding answer and attempt to discover a better embedding space to perform the semantic classification.

- Different from traditional content-based methods, a novel architecture TE-KMN is proposed to leverage auxiliary knowledge that relates to each reading sentence and the temporal cues in a sequence of descriptive sentences with respect to a given question.

Temporality-enhanced knowledge memory network on this work

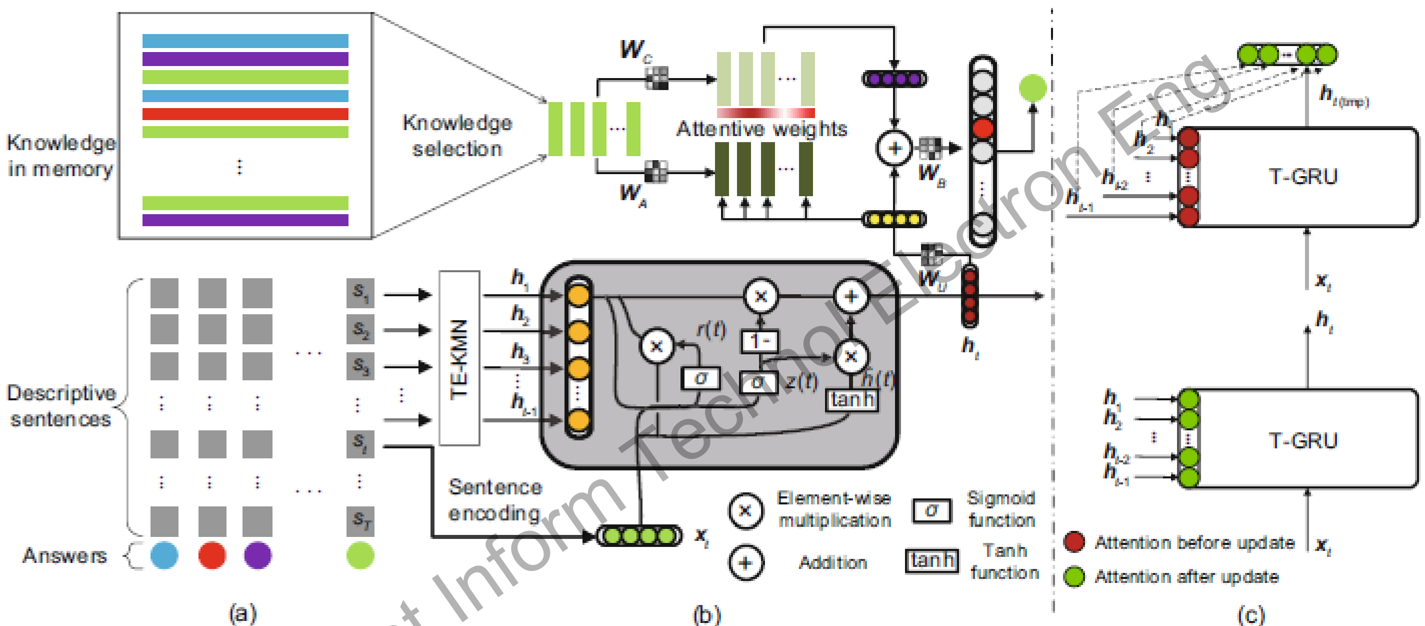

Fig. 1 Overview of our proposed temporality-enhanced knowledge memory network (TE-KMN)
(a) The lower left is the training data. For each answer, there are several ordered descriptive sentences which gradually describe the different aspects of one answer (i.e, one entity). The ordered ahead sentences remark fewer and harder cues with respect to a given answer, and the later sentences are 'giveaways'. The upper left is a knowledge memory consisting of auxiliary external knowledge that will be used during reading each sentence. In this study, the knowledge is from Wikipedia. (b) Given an answer, assume there are T ordered sentences in total describing different aspects of this answer (i.e., $s_1, s_2, \ldots, s_T$). Each sentence s_i is encoded to a distributed vector x_i. The hidden representation h_i $(1 \leq i \leq t-1)$ of each sentence will influence the learning of hidden representation h_t of x_t. The hidden representation h_t is then transmitted to the knowledge memory network to be further handled by the relevant auxiliary knowledge that relates to sentence s_t. (c) The update of the attention mechanism in T-GRU. The upper right is the T-GRU with attention before update. The lower right is the same T-GRU with attention after update. Parameters other than α_i remain intact during the update. $h_1, h_2, \ldots, h_{t-1}$ are the hidden outputs from the T-GRUs that read the previous sentences. The red and green solid circles represent the values of α_i before and after update, respectively. References to color refer to the online version of this figure

Algorithm

Algorithm 1 TE-KMN for quiz bowl

Input: Quiz bowl question set Q, answer set A, knowledge set K, selected knowledge C, and the number of iterations m.

Output: Given each sentence of a quiz bowl question, predict the answer.

Initialize all parameters of TE-KMN
for $i = 1$ to m **do**
  for $q \in Q$ **do**
    $C = \varnothing$
    for $s \in q$ and s is a sentence of q **do**
      C += knowledgeselection(s)
    end for
    for $s \in q$ and s is a sentence of q **do**
      $h_t = \text{T-GRU}(h_1, h_2 \ldots, h_{t-1}, x_i)$
      $b_t = \text{KMN}(h_t, C)$
      Calculate the loss l_t with b_t
    end for
    Accumulate the training loss
    Update parameters by the stochastic gradient descent method
  end for
end for

Attention mechanism in T-GRU

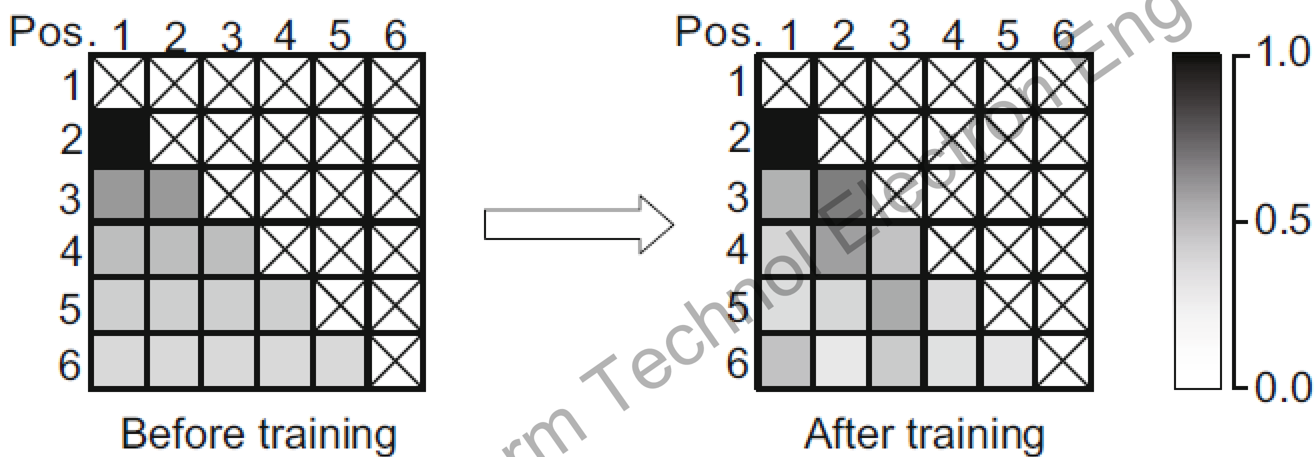

Fig. 2 Update of attention in a sequence of ordered sentences shown in Table 1

Pos. represents the position of each sentence in the question. Each row exhibits the attention changes of one sentence with each of its previous sentences. The darker color indicates a greater ratio of the corresponding attention

Measurement results (1)

Table 3 Accuracy of obtaining the right answer given different numbers of first sentences of each quiz bowl question

Number of first sentences given	Accuracy (%)								
	BOW	QANTA	DAN	HCNN	GRU	BiLSTM	TE-KMN^{-K}	TE-KMN^{-T}	TE-KMN
1	34.24	43.21	43.83	44.93	40.76	50.53	41.33	63.52	**66.07**
2	50.31	56.93	57.51	56.42	51.56	62.04	52.50	67.61	**70.77**
3	51.12	57.26	57.93	59.19	53.11	63.27	54.74	69.44	**72.91**
4	51.40	57.45	58.02	60.25	54.21	63.61	56.32	69.93	**73.94**
5	51.72	57.59	58.11	61.44	55.46	64.02	57.70	70.22	**74.46**

Table 4 Accuracy with different ratios of training data

Ratio of training data	Accuracy (%)								
	BOW	QANTA	DAN	HCNN	GRU	BiLSTM	TE-KMN^{-K}	TE-KMN^{-T}	TE-KMN
25%	22.38	30.75	32.50	31.94	26.17	33.31	26.40	46.22	**47.70**
50%	31.42	42.88	43.59	46.09	39.00	46.85	39.94	57.03	**60.41**
75%	40.37	49.93	51.78	53.62	46.23	55.20	47.04	63.59	**67.00**
100%	51.91	57.64	58.22	61.51	55.53	64.19	57.85	70.35	**74.80**

Measurement results (2)

Table 5 Accuracy of obtaining the right answer when using different knowledge memories

Number of first sentences given	Accuracy (%)			
	TE-KMN^{-K}	TE-KMN^{wiki}	TE-KMN^{train}	TE-KMN
1	41.33	42.08	64.58	**66.07**
2	52.50	53.28	68.71	**70.77**
3	54.74	55.66	70.58	**72.91**
4	56.32	58.60	71.45	**73.94**
5	57.70	61.31	72.00	**74.46**

Conclusions

- We have introduced a neural network method that incorporates the temporal cues in a sequence of ordered sentences with respect to a given question and auxiliary external knowledge to boost the performance in answering quiz bowl questions.

- Empirical results on the NAQT quiz bowl dataset show that TE-KMN outperforms state-of-the-art methods, which proves that temporal cues and auxiliary external knowledge contribute to the performance of our proposed model.