Appendix for Paper 3558

1 Case Study for Category Classification

We list a QA pair from SemEval-2015 (in *Visas and Permits category*) that is correctly predicted by ECQA-EKM but incorrectly predicted by ECQA-EKM without the question categorization task (denoted as ECQA-EKM w/o Category), and visualize the attention scores of both models in Figure 1. We observe that ECQA-EKM can assign higher scores to the important information in the question and answer such as "sponsor", "easy", "approve", "immigration" guiding by the category (i.e., *Visas and Permits*) of the input question. However, ECQA-EKM w/o Category cannot recognize some important information about *Visas and Permits* (e.g., "sponsor", "approve") and gives wrong prediction since there are only a few overlapped information in word level between the question and the answer.

Question: requirements for a single mom to sponsor a nanny? I am curious to find out if single working mothers are allowed to sponsor or hire nannies here in Qatar. Responses will be appreciated. Thanks!
Positive Answer: yea its easy just get your divorced certificate and your ID and go to the immigration office and they will let u know if they approve it or not within two weeks and u have to decide first which country u want the maid from.

(a) The attention weights by ECQA-EKM (b = 1).

Question: requirements for a single mom to sponsor a nanny? I am curious to find out if single working mothers are allowed to sponsor or hire nannies here in Qatar. Responses will be appreciated. Thanks!
Positive Answer: yea its easy just get your divorced certificate and your ID and go to the immigration office and they will let u know if they approve it or not within two weeks and u have to decide first which country u want the maid from.

(b) The attention weights by ECQA-EKM without query categorization (b = 1).

Figure 1: The attention weights of an QA pair chosen from SemEval-2015, whose category label is *Visas and Permits*.

2 Large-scale Chinese Law Dataset (LawQA)

We elaborate the process of generating our LawQA dataset in this section. Firstly, we collect a large pool of law related QA pairs with categorical information from the forum[1]. All the questions asked by netizens will be answered by the licensed lawyers. The questions are divided into ten categories, including Violence, Traffic accident, Medical accident, Consumer Rights, Property disputes, Bank, Criminal defense, Divorce, Inheritance and Labor Contract. Then, we remove the redundant QA pairs, and set the minimum length of question and answer to be 14 characters, to avoid the vagueness in the text. Our resized QA dataset contains 10 balanced categories with 100,000 questions. Since one question may have multiple answers, we have a clean QA dataset with overall 272,416 positive QA pairs. Table 1 shows one pair of LawQA.

To build the training set for answer selection, we manually collect negative samples by randomly selecting one answer form another category to form the negative sample for each QA pair (positive sample). Finally, we have a 544,832 QA pairs for training. In terms of testing set, for each distinct question, we set the candidate pool to be 100. Finally, we select 1000 pairs as validation set, 2000 pairs as test set, and the remaining data is used for training.

The experimental results are summarized in Table A.2. From Table A.2, we observe that our model performs significantly better than the compared methods on the large-scale LawQA dataset.

[1] http://china.findlaw.cn/

继承 Inheritance	我和丈夫的财产，丈夫前妻的儿子有权继承我和丈夫的财产吗? Does the son of my husband's ex-wife have the right to inherit property from me and my husband?	
	你好，你丈夫前妻的儿子若是你丈夫与前妻的共同子女，有权继承你丈夫的遗产。 If the son of your husband's ex-wife is also your husband's son, he will have the right to inherit legacy from your husband.	

Table A.1: An example of LawQA pair

Method	Accuracy	F1 score
CNN	67.1	65.9
LSTM	71.2	70.5
Bi-LSTM-attention	72.5	73.4
CNN-LSTM-CRF	73.1	72.3
AP-LSTM	73.4	73.0
AI-CNN	75.6	74.9
ECQA-EKM	**77.9***	**76.7***

Table A.2: Quantitative evaluation results (%) on LawQA. Numbers with * mean that improvement from our model is statistically significant over the baseline methods (t-test, p-value < 0.05).

3 Statistical Significance Tests

In this experiment, we also perform the statistical significance tests. The experimental results are summarized in Table A.3 and Table A.5, which show that our model is statistically significantly better the compared baseline methods on both datasets (t-test, p-value < 0.05).

Method	Accuracy	F1 score
JAIST	79.1	78.96
KeLP	81.96	80.73
CNN	77.33	76.92
LSTM	76.21	75.15
Bi-LSTM-attention	81.12	79.09
BGMN	81.24	80.22
CNN-LSTM-CRF	82.15	81.33
AP-LSTM	79.45	79.06
AI-CNN	83.06	81.92
ECQA-EKM	**86.93***	**85.05***

Table A.3: Quantitative evaluation results (%) on SemEval-2015. Numbers with * mean that improvement from our model is statistically significant over the baseline methods (t-test, p-value < 0.05).

Method	Accuracy	F1 score	MAP
JAIST	74.50	62.16	77.56
Kelp	75.11	64.36	79.19
CNN	73.23	64.92	76.21
LSTM	72.84	64.15	76.08
Bi-LSTM-attention	74.55	69.82	77.31
BGMN	74.06	69.39	77.10
CNN-LSTM-CRF	75.18	70.04	77.45
AP-LSTM	75.47	71.72	77.12
AI-CNN	76.30	72.75	79.17
ECQA-EKM	**78.96***	**74.85***	**82.18***

Table A.4: Quantitative evaluation results (%) on SemEval-2016. Numbers with * mean that improvement from our model is statistically significant over the baseline methods (t-test, p-value < 0.05).

Method	Accuracy	F1 score	MAP
JAIST	73.78	68.04	87.24
Kelp	73.89	69.87	88.43
BGMN	74.75	75.39	87.68
CNN	73.22	72.14	86.21
LSTM	74.05	73.45	86.28
Bi-LSTM-attention	76.60	74.82	88.05
BGMN	74.75	75.39	87.68
CNN-LSTM-CRF	77.18	77.04	87.66
AP-LSTM	77.64	76.82	87.82
AI-CNN	78.24	77.75	88.33
ECQA-EKM	**81.98***	**80.19***	**89.85**

Table A.5: Quantitative evaluation results (%) on SemEval-2017. Numbers with * mean that improvement from our model is statistically significant over the baseline methods (t-test, p-value < 0.05).

4 Initial Knowledge Representation Learning with Attention

We perform entity mention detection by n-gram matching and provide a set of top-N entity candidates from KB for each entity mention in the input. The embeddings of entities in KB are learned by DeepWalk (Perozzi et al., 2014). Formally, we present candidate entities for the entity mention at time step t as $\{ent_{t1}, ent_{t2}, ..., ent_{tK}\} \in \mathbb{R}^{K \times d_{\text{kb}}}$, where d_{kb} is the dimension of the entity embedding in KB. We design a context-guided attention mechanism to learn the knowledge representation of each entity mention in the input by congregating the embeddings of the corresponding candidate entities in the KB.

$$\tilde{e}_t = \sum_{i=1}^{K} \alpha_{ti} ent_{ti} \tag{1}$$

$$\alpha_{ti} = softmax(\rho(ent_{ti}, \mu(H^d))) \tag{2}$$

$$\rho(ent_{ti}, \mu(H^d)) = \tanh(W_{kb} ent_{ti} + W_c \mu(H^d) + b_{kb}) \tag{3}$$

where $\tilde{e}_t$ is the knowledge representation for the t-th entity in the document, μ is the mean pooling operation; W_{kb} and W_c are weight matrices to be learned; b_{kb} is the bias term; α_{ti} denotes the context guided attention weight over the i-th candidate entity embedding ent_{ti}. With this context-guided attention mechanism defined by Equations 1-3, we can obtain the initial knowledge representations of input document and query, denoted as $\tilde{e}^q = [\tilde{e}^q_1, \ldots, \tilde{e}^q_n]$ and $\tilde{e}^a = [\tilde{e}^a_1, \ldots, \tilde{e}^a_m]$ respectively.

After obtaining the knowledge representation for each entity mention in the document, a CNN layer is then employed to capture the local n-gram information and learn a higher level knowledge representation $Z \in \mathbb{R}^L$ (L is the number of hidden states of CNN) for each question and answer. Formally, we learn the knowledge representations for the question Q and answer A as follows:

$$Z^q = \textbf{CNN}(\tilde{e}^q), \quad Z^a = \textbf{CNN}(\tilde{e}^a) \tag{4}$$

We refer the interested readers to (Kim, 2014) for the implementation details of CNN in text modeling.

The hidden states Z^q and Z^a are inputted into a mean-pooling layer to obtain the initial knowledge representations of the question and the answer, denoted as K^q and K^a respectively:

$$K^q = \sum_{i=1}^{L_q} Z^q_i / L_q, \quad K^a = \sum_{i=1}^{L_a} Z^a_i / L_a \tag{5}$$

where L_q and L_a are the numbers of the hidden states of the question and answer, respectively.

4.1 New Experiment Results

Table A.6 demonstrates the experimental results of ECQA-EKM by using attention mechanism to learn the initial knowledge representation (denoted as ECQA-EKM-new) on SemEval-2015 dataset (there is no time to re-run the experiments on SemEval-2016 and SemEval-2017). ECQA-EKM-new achieves slightly better results than ECQA-EKM.

Method	Accuracy	F1 score
ECQA-EKM	**86.93**	**85.05**
ECQA-EKM-new	**87.18**	**85.37**

Table A.6: Quantitative evaluation results of ECQA-EKM and ECQA-EKM-new on SemEval-2015.

5 The Effect of Parameter K

K is the number of entity candidates from KB for each entity mention in the input document. In this paper, we investigate the effect of K by varying its value from 1 to 10 with step size 1. We report the experimental results on SemEval-2015 dataset in Table A.7. We can achieve best results when $K = 5$ on SemEval-2015. As K increases from 1 to 10, the accuracy and F1 scores increase sharply till an optimal value (when $K = 5$), after which it decreases slightly.

Value of K	Accuracy	F1 score
1	83.24	81.74
2	84.78	83.36
3	86.16	84.22
4	86.89	85.03
5	86.93	85.05
6	86.93	85.04
7	86.91	85.01
8	86.85	84.97
9	86.62	84.68
10	86.34	84.45

Table A.7: Experimental results of ECQA-EKM on SemEval-2015 by varying the value of K.

6 Case Study for "How to" Questions

In Table A.8, we list a "how to" question that is incorrectly predicted by AI-CNN (the strongest baseline) but correctly predicted by ECQA-EKM. We observe that AI-CNN tend to assign a higher score to the negative answer than the positive answer, since the negative answer is more similar to the given question at the word level. However, with the background knowledge from KB, ECQA-EKM can correctly identify the positive answer based on the relative facts contained in the KB such as (labor, related_to, work) and (permit, related_to, ticket of leave).

Question	How to go home without exit permit. My sponsor is not willing to give me exit permit in the fear of my not return. I am about 3 years at Qatar; and I want to go home on a vacation before Eid. Is there any way to go home without the exit permit from sponsor? I want to go home before Eid; no matter what does it cost; even cancellation of my visa. Thank you for your advice.
Positive answer	You have 7 days... You left it a bit late to resign; no? If you are due to go home they have to give you a ticket. If not; then go to the Labour Department and report him.
Negative answer	My sponsor is not willing to let me go in any way until all of my projects are finished. Can my embassy or the labor department help?

Table A.8: Example of one "how to" question and its two answers from SemEval-2017.

7 The Effect of Parameter T

T is the number of classifiers in our ensemble learning. In this paper, we investigate the effect of T by varying its value from 1 to 5 with step size 1. We report the experimental results on SemEval-2015

dataset in Table A.9. As T increases from 1 to 3, the accuracy and F1 scores increase till $T = 3$, after which the accuracy becomes stable.

Value of K	Accuracy	F1 score
1	84.28	82.45
2	86.17	83.93
3	86.93	85.05
4	86.98	85.07
5	86.99	85.06

Table A.9: Experimental results of ECQA-EKM on SemEval-2015 by varying the value of T.

8 The overall architecture of ECQA-EKM

Figure 2 illustrates the overview architecture of ECQA-EKM, which consists of four modules. In *interactive knowledge-aware representation learning*, we first learn the initial representations (including the initial context representations and initial knowledge representations) of the given question and answer. Then, we combine the strengths of the context and the knowledge representations by extracting interactive information between them via a multi-head interactive attention network. *Text categorization* module assigns a category label to the given question, and the category information is then fed into the question representation to learn a category-specific text encoder. In *community question answering* module, we construct multiple CQA classifiers and ensemble their results as the final prediction result to more effectively and robustly solve the CQA problem. ECQA-EKM is trained via a *multi-task learning module* by regarding community question answering as the primary task and question categorization as the auxiliary task.

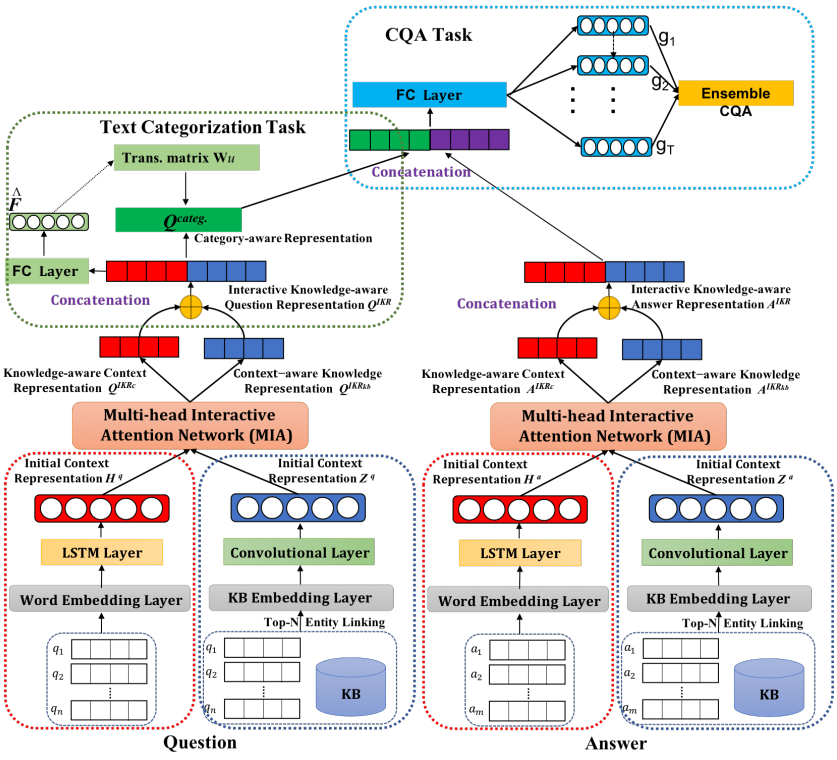

Figure 2: The overview architecture of ECQA-EKM.

References

Yoon Kim. 2014. Convolutional neural networks for sentence classification.

Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, pages 701–710.