Attentive User-Engaged Adversarial Neural Network for Community Question Answering

Yuexiang Xie,[1,2] **Ying Shen,**[1] **Yaliang Li,**[2] **Min Yang,**[3] **Kai Lei**[*,1,4]

[1]Shenzhen Key Lab for Information Centric Networking & Blockchain Technology (ICNLAB), SECE, Peking University
[2]Alibaba Group [3]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences
[4]PCL Research Center of Networks and Communications, Peng Cheng Laboratory, Shenzhen, China
xieyx@pku.edu.cn, {shenying, leik}@pkusz.edu.cn,
yaliang.li@alibaba-inc.com, min.yang@siat.ac.cn

Abstract

We study the community question answering (CQA) problem that emerges with the advent of numerous community forums in the recent past. The task of finding appropriate answers to questions from informative but noisy crowdsourced answers is important yet challenging in practice. We present an Attentive User-engaged Adversarial Neural Network (AUANN), which interactively learns the context information of questions and answers, and enhances user engagement with the CQA task. A novel attentive mechanism is incorporated to model the semantic internal and external relations among questions, answers and user contexts. To handle the noise issue caused by introducing user context, we design a two-step denoise mechanism, including a coarse-grained selection process by similarity measurement, and a fine-grained selection process by applying an adversarial training module. We evaluate the proposed method on large-scale real-world datasets SemEval-2016 and SemEval-2017. Experimental results verify the benefits of incorporating user information, and show that our proposed model significantly outperforms the state-of-the-art methods.

Introduction

Community Question Answering (CQA) has gained increasing popularity in the recent past. More and more people would like to use web forums such as Yahoo![1], Stack Overflow[2] and Quora[3] to freely ask questions or provide answers to others. Because of the uneven quality of answers, it takes much effort to go through a set of candidate answers and find answers that might be useful. We consider the problem of automatically finding an appropriate answer to a given question with assistance from community question answering forums which have accumulated a large number of questions.

The major challenge of this task is how to utilize the user context information that reflects user's expertise and hobby,

---

[1]https://answers.yahoo.com
[2]https://www.stackoverflow.com
[3]https://www.quora.com

Category: Doha Shopping
Q1: What is best mall in Doha to buy good furniture? Where are best furniture stores and showrooms. Thank you :) (Posted by Angela Dondi in Dec, 2012)
A1: There are several; my Favorite is Pan Emirates @ Salwa Road. (Posted by Yasir_Kh in Jan, 2013)
A2: Best of the best is haraj mall in najemma. (Posted by anonymous in Dec, 2012)

Category: Sports in Qatar
Q2: Anybody know any evening swimming classes offered in Doha for adults only; maybe 3 classes per week possibly. Thank you. (Posted by Lucky_gal in Jan, 2013)
A3: Dana club also offer swimming classes for adults and kidz. You can adjust your own schedule and number of classes as well. (Posted by Yasir_Kh in Jan, 2013)
A4: Aspire club is conducting swimming class for adults... u have to register online in their site to get the appointment. (Posted by hindiwatan in Jan, 2013)

Table 1: Examples of CQA Forums

as well as the contextual information about question and answer sentences. Typical examples of CQA forums are shown in Table 1. The CQA web forums usually have few restrictions, which leads to several differences between CQA task and traditional QA task. First of all, these forums contain not only numerous question and answer sentences, but also metadata such as user identifier, question category and date of posting, which provides researchers with the potential to make improvement in the CQA task (Wen et al. 2018). Secondly, the questions and answers in CQA are usually multiple sentences and can contain much noise (e.g., abbreviations, misspelled words and emoticons), while traditional QA task deals with relatively simpler and cleaner sentences. Thirdly, for questions in a community forum, the received answers vary a lot in terms of content and quality. Accordingly, we can conclude that CQA task contains more noise in questions and answers, but can provide access to other metadata, such as user information.

Recent researches (Wu, Xu, and Houfeng 2018; Shao et al. 2019) regard the CQA task as a text matching prob-

lem of question and answer sentences. Inspired by the success of deep learning in Natural Language Processing (NLP), a variety of deep neural network based architectures have been proposed for CQA (Zhang et al. 2017; Yang et al. 2019). The existing methods are focused on capturing the hidden semantic relationship between question and answer context, but ignore the plentiful information of the user context from CQA forums that reflects user's expertise and hobby. Note that in practice, when judging the answer quality, people are influenced in a way by relevant information from historical answers. For example, when finding an appropriate answer to a question about "cars", people might scan answer provider's historical contents to judge how professional and reliable he/she is in "car" field. Based on common knowledge, experts would probably contribute more good answers (Xiang et al. 2016). Thus we believe that CQA models can obtain further development by introducing user's historical answers.

However, since user's historical answers might come from several questions of different topics, directly introducing user's historical answers into CQA models can cause severe noise issues. As shown in Table 1, **A1** and **A3** are provided by the same user (Yasir_Kh), but they might have no effect for each other in sentences matching because of totally different topics. Before taking advantage of the plentiful information of user history answers, we need to filter out the irrelevant information.

In this paper, we exploit user information in the CQA forums and gather users' historical answers to build user contexts in addition to questions and answers. User's hobby and expertise are captured to reflect and boost the performance of CQA task. A novel attentive mechanism is incorporated to model the semantic internal and external relations among questions, answers and user contexts. In order to alleviate the noise issue caused by adding plentiful but noisy information from user's historical answers, we design a two-step denoise mechanism. Specifically, a coarse-grained selection process by similarity measurement is applied to screen out helpful user's historical answers. Then an adversarial training module is designed to improve the ability of filter out the irrelevant information, which is considered as a fine-grained selection process.

The main contributions of our work can be summarized as follows:

- We propose a novel Attentive User-engaged Adversarial Neural Network (AUANN), which learns the context semantic relationship of questions and answers by applying attention mechanism, and enhances user engagement with the CQA task.

- In order to alleviate the noise issue caused by introducing user's historical answers, we design a two-step denoise mechanism. It consists of a coarse-grained selection process by calculating similarity, and a fine-grained selection process by applying an adversarial training module to model user's historical answers.

- We evaluate our model on two widely-adopted public datasets SemEval-2016 and SemEval-2017. Compared to the state-of-the-art CQA methods, our model achieves

better performance. Experimental results verify the benefits of incorporating user information.

Problem Definition

Given a question $Q = \{q_t\}_{t=1}^{L_Q}$ and a candidate answer $A = \{a_t\}_{t=1}^{L_A}$, where q_t and a_t denote the t-th word in question and answer respectively, and L_Q and L_A are the lengths of sentences, we formulates CQA as a binary classification problem (Nakov et al. 2017; Wu, Xu, and Houfeng 2018). The information of the user who submits the question or provides an answer is also considered when modeling. A user context $U = \{U_i\}_{i=1}^N$ contains N historical answers of the user who provides the answer A. The task of CQA is to classify the tuple (Q, A, U) as "Good" (answer A is an appropriate answer to question Q) or "Bad" (answer A is not an appropriate answer to question Q).

Methodology

An overview of the proposed Attentive User-engaged Adversarial Neural Network (AUANN) is illustrated in Figure 1. To be specifics, we first employ Bidirectional Long Short-Term Memory (Bi-LSTM) to encode the input sentences. Then we adopt a novel attentive mechanism to model the semantic internal and external relationships among questions, answers and user contexts. A two-step denoise mechanism is designed to alleviate noise issue and model user context.

Input Encoder

Word-Level Embedding   Given a sequence of words, we firstly convert them into pre-trained word-level embeddings with Glove (Pennington, Socher, and Manning 2014) as initialization. The word-level embedding projects the semantically related words into close proximity of the same embedding space according to the word co-occurrence statistics. Formally we obtain the representations of question, answer and user context as:

$$
\begin{aligned}
E^Q &= \{e_t^Q\}_{t=1}^{L_Q} \in \mathbb{R}^{d_e \times L_Q}, \\
E^A &= \{e_t^A\}_{t=1}^{L_A} \in \mathbb{R}^{d_e \times L_A}, \\
E^{U_i} &= \{e_t^{U_i}\}_{t=1}^{L_{U_i}} \in \mathbb{R}^{d_e \times L_{U_i}},
\end{aligned} \tag{1}
$$

where d_e is the embedding size and U_i denotes one of the user's historical answers, i.e., $U_i \in U$.

We feed embedding vectors E into a highway network encoder layer (Srivastava, Greff, and Schmidhuber 2015) with shared weights, which allows the neural networks to have several paths for information flowing across different layers and to make optimization of deep neural networks easier:

$$
\widetilde{H} = \text{Highway}(E). \tag{2}
$$

Bidirectional Long Short-Term Memory   Then we adopt the basic Bidirectional Long Short-Term Memory (Bi-LSTM) to encode the input sequences, which has been successfully applied to a wide range of NLP tasks (McCann et al. 2017; Shah and Barber 2018). The Bi-LSTM layer contains two sub-networks for the head-to-tail and the tail-to-head contexts, respectively. Thus it can not only capture the
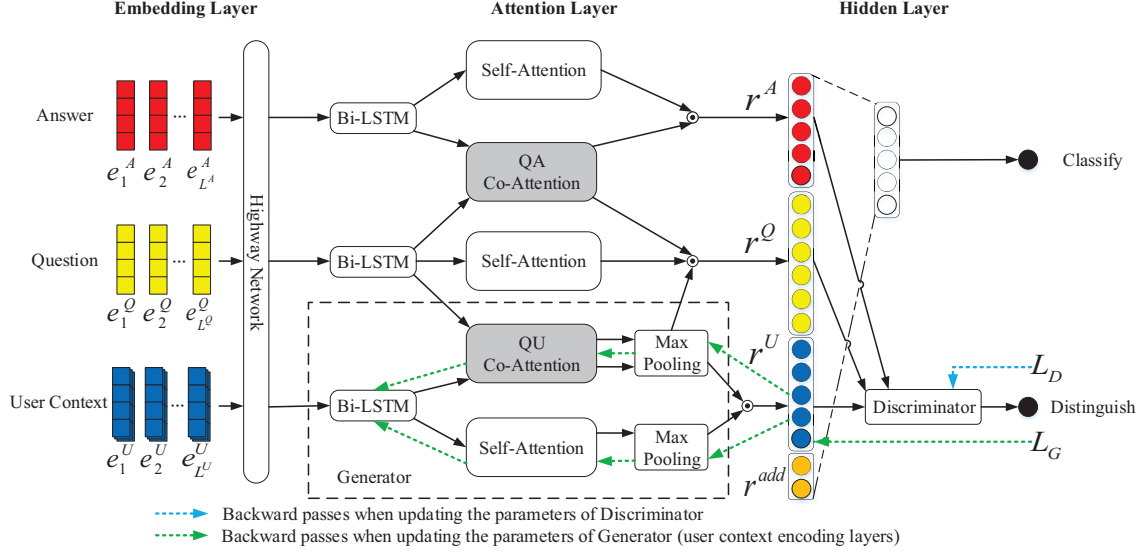
Figure 1: Attentive User-engaged Adversarial Neural Network

information from past contexts, but also access future contexts:

$$H = \text{Bi-LSTM}(\widetilde{H}) = \{h_t\}_{t=1}^L \in \mathbb{R}^{d_h \times L}, \qquad (3)$$

where L is the length of sequence and d_h is the hidden size of Bi-LSTM. The output at timestep t is represented by h_t. Finally we obtain H^Q, H^A, and $\{H^{U_i}\}_{i=1}^N$ for question, answer and user context, respectively.

Attention Layer

Semantic self-attention and co-attention mechanism are adopted to focus on the important and useful part of question, answer and user context, from the internal and interact perspective, respectively.

Semantic Self-Attention The basic idea behind the semantic self-attention mechanism in our model is to perceive the importance of different words of a sequence based on its semantic information. We first apply a mean-pooling operation to obtain the context semantic information for each sequence, which is added to the sequence representation obtained from input encoder:

$$\hat{H} = [H; \text{Mean-Pooling}(H)] \in \mathbb{R}^{(d_h+1) \times L}, \qquad (4)$$

where [;] is the concatenate operation. Formally, the sentence representation under semantic self-attention is given by:

$$\alpha = \{\alpha_t\}_{t=1}^L = \text{softmax}(u_s^\top \tanh(W_s \hat{H})), \qquad (5)$$

$$r_s = \sum_{t=1}^L \alpha_t h_t, \qquad (6)$$

where $W_s \in \mathbb{R}^{d_h \times (d_h+1)}$ and $u_s \in \mathbb{R}^{d_h}$ are parameters to be learned. We apply semantic self-attention to question, answer and user context and obtain their representations r_s^Q,

r_s^A, and $\{r_s^{U_i}\}_{i=1}^N$ respectively. Finally, we integrate features of user context by a max-pooling operation:

$$r_s^U = \max_{1<i<N} r_s^{U_i}. \qquad (7)$$

Co-Attention We apply co-attention mechanism (Deng et al. 2018) to calculate interactive attention between each QA pair, which enables QA pairs to be aware of their semantic potential relationships. Specifically, We first compute attention matrix M_c:

$$M_c = (H^Q)^\top W_c H^A \in \mathbb{R}^{L_Q \times L_A}, \qquad (8)$$

where $W_c \in \mathbb{R}^{d_h \times d_h}$ is the parameter matrix to be learned. Then column-wise and row-wise max-pooling are applied on M_c to generate the context-based co-attention weights M_c^Q and M_c^A for question and answer separately:

$$M_c^Q = \max_{1<t<L_A} M_c \in \mathbb{R}^{L_Q}, \quad M_c^A = \max_{1<t<L_Q} M_c^\top \in \mathbb{R}^{L_A}. \qquad (9)$$

We obtain the question and answer representations under co-attention mechanism by:

$$r_c^Q = \sum_{t=1}^{L_Q} \beta_t^Q h_t^Q, \text{ with } \beta^Q = \{\beta_t^Q\}_{t=1}^{L_Q} = \text{softmax}(M_c^Q),$$

$$r_c^A = \sum_{t=1}^{L_A} \beta_t^A h_t^A, \text{ with } \beta^A = \{\beta_t^A\}_{t=1}^{L_A} = \text{softmax}(M_c^A). \qquad (10)$$

We regard each user's historical answer $U_i \in U$ as an independent context, and conduct the same operation over each pair of question and historical answer (Q, U_i) to learn N representations under co-attention mechanism. Finally, comprehensive representation under co-attention of

question and user context are retained by applying a max-pooling layer among the N results, expressed by $\hat{r}_c^Q$ and r_c^U. Through the concatenate operation $[;]$, the representations under co-attention of question, answer and user context are given by:

$$r^Q = [r_s^Q; r_c^Q; \hat{r}_c^Q], \quad r^A = [r_s^A; r_c^A], \quad r^U = [r_s^U; r_c^U]. \tag{11}$$

Denoise Mechanism

In order to alleviate the noise issue caused by introducing user's historical answers, we design a two-step denoise mechanism, including a coarse-grained selection process by similarity measurement and a fine-grained selection process by applying an adversarial training module.

Similarity Measurement We adopt the coarse-grained selection process during the data preparation to directly filter out the irrelevant content in user's historical answers. Specifically, we aim to select N historical answers that is the most similar to the answer A according to the similarity calculated by Eq.12:

$$\text{similarity}(A, U_i) = \frac{\left(\max_{1<t<L_A} E^A \right) \cdot \left(\max_{1<t<L_{U_i}} E^{U_i} \right)}{\left| \max_{1<t<L_A} E^A \right| \left| \max_{1<t<L_{U_i}} E^{U_i} \right|}, \tag{12}$$

where E^A and E^U indicate the word-level embedding representations of the given answer and the user's historical answer. Then we feed the selected contents into our model.

Adversarial Training Inspired by Generative Adversarial Network (GAN) (Goodfellow et al. 2014) and the great success of adversarial training for domain adaptation (Chen and Cardie 2018) and multi-task learning (Liu, Qiu, and Huang 2017), we propose an adversarial training module to alleviate the noise issue in CQA. The item "noise" refers to the irrelevant information in user context to the given question. Because user context consists of user's historical answers from several questions with different topics, it contains mixture contents and need to be denoised. The main idea of our adversarial training module is to force the encoding layers to learn relevant information from user context, which can be regarded as a fine-grained selection process.

In the proposed adversarial training module, we design a discriminator D, and regard the user context encoding layers as a generator G, as shown in Figure 1. The discriminator D takes the final representations of answer r^A and user context r^U as input and attempts to distinguish them based on their relevance to the given question. Accordingly, we consider the final representation of question r^Q as one of the inputs to D and concatenate r^Q and r^A, r^Q and r^U before we feed them into D. The generator G is trained to extract features from user context that are relevant to the given question, or in other words, to confuse D.

Specifically, we adopt the relativistic standard GAN (RSGAN) (Jolicoeur-Martineau 2019) instead of standard GAN to better fit our goal of improving the correlation between r^A

and r^U. The discriminator in RSGAN is proposed to measure the difference degree between two inputs, which refers the difference between answer and user context in our case. In our study, according to the given question, the discriminator aims to estimate the probability that the given answer is more relevant to the question than the user context to the question. The discriminator is given by:

$$D(\hat{r}) = \text{sigmoid}(C([r^Q; r^A]) - C([r^Q; r^U])), \tag{13}$$

where $C(\cdot)$ is the non-transformed layer. The output value of $C(\cdot)$ can be interpreted as how realistic the input data is in Jolicoeur-Martineau (2019), which denotes the relevance in our model. Formally, the RSGAN loss functions for discriminator and generator is defined as:

$$\mathcal{L}_D = -\mathbb{E}_{(r^Q, r^A, r^U)}[\log(D(\hat{r}))], \tag{14}$$
$$\mathcal{L}_G = -\mathbb{E}_{(r^Q, r^A, r^U)}[\log(1 - D(\hat{r}))]. \tag{15}$$

We minimize $\mathcal{L}_D$ and $\mathcal{L}_G$ alternately, and the former updates the parameters of D while the latter acts on generator G. As shown in Figure 1, the dotted arrows indicate the passes of backward propagation. In our study, when the model reaches the Nash equilibrium, i.e., $D(\hat{r}) = 1 - D(\hat{r})$ (Goodfellow et al. 2014), the discriminator D cannot distinguish between r^A and r^U, indicating that the Generator G is capable enough to extract relevant features. Therefore, we believe that the features learned from user context are relevant enough to the question and our model has screened out the helpful parts from the added user information.

The architecture of discriminator D is relatively flexible and can be designed by practitioners to suit different tasks and inputs. In our study, we use a multi-layer perceptron (MLP) with dropout mechanism to alleviate over-fitting.

Hidden Layer and Output Layer

An additional features $r_{add} = [\gamma_1, \gamma_2]$ are exploited in our overall architecture. To be specifics, the first value $\gamma_1 \in \{0, 1\}$ indicates whether the given answer and question are provided by the same user, and the second value $\gamma_2 \in \{0, 1\}$ denotes whether the answer provider is anonymous. The intuition is that answers written by the question providers are often meaningless because they contain dialogues or detailed inquiries. In addition, anonymous answers are often of poor quality according to statistics (Zhang et al. 2017). These additional features are generated from the metadata (i.e., user information) without using any other tools or methods, which proves again that user information can improve the overall performance.

Thus, the final input of the hidden layer can be given as $[r^A; r^Q; r^U; r_{add}]$, and its output then goes through a softmax layer for binary classification. The loss function for the overall model is:

$$\mathcal{L} = -\sum_i [y_i \log p_i + (1 - y_i) \log(1 - p_i)] + \lambda \|\theta\|_2, \tag{16}$$

where p is the output of the softmax layer and y is the label. $\lambda \|\theta\|_2$ denotes the L^2 regularization. During the training process, the classification loss is used to train the overall CQA model, while the RSGAN objective is used to train the discriminator and the generator. The dotted arrows shown in Figure 1 indicate the passes of backward propagation.

	SemEval-2016 (train/dev/test)	SemEval-2017 (train/dev/test)
The Number of Questions	4,880/244/327	5,207/244/293
The Number of Answers	36,198/2,440/3,270	39,468/2,440/2,930
Average Length of Questions	43/47/48	43/47/54
Average Length of Answers	38/36/37	38/36/40

Table 2: Statistics of SemEval-2016 Task 3 and SemEval-2017 Task 3

Experiment

We conduct the experiment on two widely-adopted public community question answering datasets from SemEval-2016 Task 3 (Nakov et al. 2016) and SemEval-2017 Task 3 (Nakov et al. 2017). We use the English data and the statistics of datasets are summarized in Table 2.

In our setting, the sentences are tokenized using NLTK[4]. Pre-trained GloVE embeddings of 300 dimensions are adopted as word-level embeddings. Out-of-vocabulary words are set by randomly sampling values from the standard normal distribution. The max length of question, answer and user context are set to 40, 30 and 30 respectively. The hidden size of LSTM is set to 128. The hidden size of discriminator is set to 128. We use Adam optimizer for optimization with learning rate 5e-4. The model parameters are regularized by L^2 regularization with the strength of 1e-5. The hyperparameter N is tuned in the set $\{3,4,5\}$.

Model Comparisons

We compare our model with baselines in terms of F1 Score, Accuracy and Mean Average Precision (MAP). For a perfect ranking, a system has to place all "Good" answers above "Bad" answers under each given question.

Several state-of-the-art baselines are adopted for comparison. Baselines include top systems from the SemEval-2016 task 3 and SemEval-2017 task 3 (Kelp, ConvKN, etc.) and neural network based models (ARC-II, AI-CNN, etc.):

- **AP** (Santos et al. 2016): It proposes a two-way attention mechanism named Attentive Pooling (AP) that can project the paired inputs into the same latent space and jointly learn a similarity measure for sentence matching.

- **ARC-I, ARC-II** (Hu et al. 2014): It proposes novel deep convolutional network architectures ARC-I and ARC-II for sentences matching. The models can not only combine the hierarchical sentence by composition and pooling operation, but also capture of different levels matching patterns.

- **ConvKN** (Barrón-Cedeno et al. 2016): It combines convolutional tree kernels with convolutional neural networks and additional manually designed features. ConvKN achieves the best result among several participants evaluated by accuracy on SemEval-2016 Task 3.

- **AI-CNN** (Zhang et al. 2017): It proposes to treat different text segments differently and designs a novel attentive interactive neural network to focus on those segments that can aid answer selection in CQA.

---

[4]http://www.nltk.org/

- **IIT-UHH** (Nandi et al. 2017): It proposes a rich feature-based Support Vector Machine (SVM) system that integrates textual, semantic and domain specific features, word embedding and topic-modeling features to solve the CQA task.

- **Beihang-MSRA** (Feng et al. 2017): In addition to using traditional NLP features such as TF-IDF, it also introduces a neural network-based matching model that enables the model to measure text similarity beyond the dictionary.

- **Kelp** (Filice et al. 2016; 2017): It proposes a supervised kernel-based framework to automatically learn syntactic/semantic patterns useful from training question-answer pairs for recognizing answers. Kelp achieves the best result among several participants on SemEval-2016 Task 3 and SemEval-2017 Task 3, evaluated by MAP.

- **ECNU** (Wu and Lan 2016; Wu et al. 2017): It trains a classifier with traditional NLP features and learns the question-comment representation via CNN and Bi-LSTM. ECNU achieves the best result among several participants on SemEval-2016 Task 3 and SemEval-2017 Task 3 evaluated by F1 score, and the best result evaluated by accuracy on SemEval-2017 Task 3.

- **UIA-LSTM-CNN** (Wen et al. 2018): It proposes a hybrid attention mechanism to model question-answer pairs and exploits user information for answer selection in CQA.

Table 3 reports the experimental results on SemEval-2016 and SemEval-2017. In general, our model outperforms the state-of-the-art methods by a noticeable margin on both datasets. For instance, on the SemEval-2016 dataset, the proposed method improves about 2.32%, 0.42% and 1.43% on the metrics over these baselines. Kelp, IIT-UHH and ECNU achieve good results, indicating that feature engineering can play an important role in the CQA. Compared with AP and AI-CNN , our model utilizes plentiful user information. When ranking a set of answers, our model considers not only the semantic similarity between the QA pairs, but also the hidden semantic relationships with user context. Therefore, our model consistently outperforms AP and AI-CNN by a noticeable margin. Compared with UIA-LSTM-CNN model that utilizes user's historical answers without filtering and therefore introduces irrelevant information, our model uses an adversarial training module to reduce noise (i.e., irrelevant information) and thus achieves the better result.

Ablation Study

For a thorough comparison, we report the ablation test to analyze the improvements contributed by each part of our

Model	SemEval-2016			SemEval-2017		
	Accuracy	MAP	F1 score	Accuracy	MAP	F1 score
AP (Santos et al. 2016)†	75.47	77.12	71.72	-	-	-
ARC-I (Hu et al. 2014)†	74.07	77.05	69.50	-	-	-
ARC-II (Hu et al. 2014)†	75.26	77.98	71.64	-	-	-
ConvKN (Barrón-Cedeno et al. 2016)	75.54	77.66	66.16	-	-	-
DFFN (Suggu et al. 2016)	76.67	_82.34_	66.22	-	-	
AI-CNN (Zhang et al. 2017)	76.87	80.14	_73.03_	-	-	-
IIT-UHH (Nandi et al. 2017)	-	-	-	72.70	86.88	73.94
Beihang-MSRA (Feng et al. 2017)	-	-	-	51.98	88.24	68.40
Kelp (Filice et al. 2016; 2017)	75.11	79.19	64.36	73.89	_88.43_	69.87
ECNU (Wu and Lan 2016; Wu et al. 2017)	74.31	77.28	66.72	_78.43_	86.72	_77.67_
UIA-LSTM-CNN (Wen et al. 2018)	_78.17_	80.86	68.44	77.13	87.96	76.45
AUANN (ours)	**80.49**	**82.76**	**74.46**	**78.46**	**89.59**	**79.81**
(1) w/o user context	77.73	81.00	72.47	77.27	88.02	77.98
(2) w/o semantic self-attention	78.17	81.27	72.72	76.66	88.85	77.15
(3) w/o co-attention	77.37	81.74	73.32	76.11	88.20	77.20
(4) w/o added features	78.20	81.38	73.45	77.02	87.60	77.46
(5) w/o similarity measurement	76.54	80.33	72.59	76.31	88.71	77.65
(6) w/o discriminator	76.09	79.86	71.89	76.89	87.78	78.14

Table 3: Performance of different approaches on the SemEval-2016 Task 3 and SemEval-2017 Task 3 datasets. Models marked with † are reported directly from (Zhang et al. 2017).

model: (1) **w/o user context**: We discard the user context in our model. (2) **w/o semantic self-attention**: We remove the semantic self-attention in our model, which is designed to perceive the importance of different words of a sequence based on its semantic information. (3) **w/o co-attention**: We discard the co-attention in our model, which is proposed to calculate semantic potential relations between question, answer and user context. (4) **w/o added features**: We replace the additional features $r_{add} \in \mathbb{R}^2$ in hidden layer by zero padding. (5) **w/o similarity measurement**: We replace the coarse-grained selection process with random selection. (6) **w/o discriminator**: We remove the discriminator in our model and no longer use adversarial training (i.e., the fine-grained selection process) to model user context. The parameters of user context encoding layers are updated only depend on the classification error.

The results are listed in Table 3. Compared AUANN with (1) **w/o user context**, we can conclude that adding user's historical answers (i.e., user contexts) can improve the performance of CQA task. However, such contexts that contain irrelevant information might sometimes play as a negative role, which makes (6) **w/o discriminator** less performing than (1) **w/o user context**. We introduce user context in basic model (6) without applying advertise training module (i.e., the fine-grained denoised process). Our model outperforms (5) **w/o similarity measurement** and (6) **w/o discriminator** by a great margin, demonstrating the contribution of the two-step denoise mechanism adopted in screening informative parts of user context.

From Table 3 we can also see that using semantic self-attention and co-attention both contribute to model performance. The reason is that attentive mechanism can weight

Question: Hi, can anybody tell me how I could get 3 visit VISA for my family? Suggestion for hotel names and contact numbers will be highly appreciated. Thank you!

Positive Answer: At the immigration department. Prepare the application and documents and submit them.
User's Historical Answer: It depends on your VISA. If you entered Qatar with the Visit VISA, you don't need an exit permit. If you came on Business VISA and stay more than 1 month, you will need an exit permit.

Negative Answer: ...I decided to renew my passport. The problem is my renewed passport NO. will not be the same with my passport NO. indicated in my VISA. Would the Embassy still visa-stamp on my new passport.

Table 4: Case Study

the importance of different words or sentences based on their own contexts (i.e., semantic self-attention) or interactive semantic similarity (i.e., co-attention).

Case study

We list a typical example from SemEval-2017 Task 3 dataset in Table 4, which is incorrectly predicted by well-performed models ECNU and IIT-UHH according to the public results, but correctly predicted by our model. We observe that ECNU and IIT-UHH tends to assign a higher score to the Negative Answer than the Positive Answer because the Negative Answer is more similar to the given Question at the
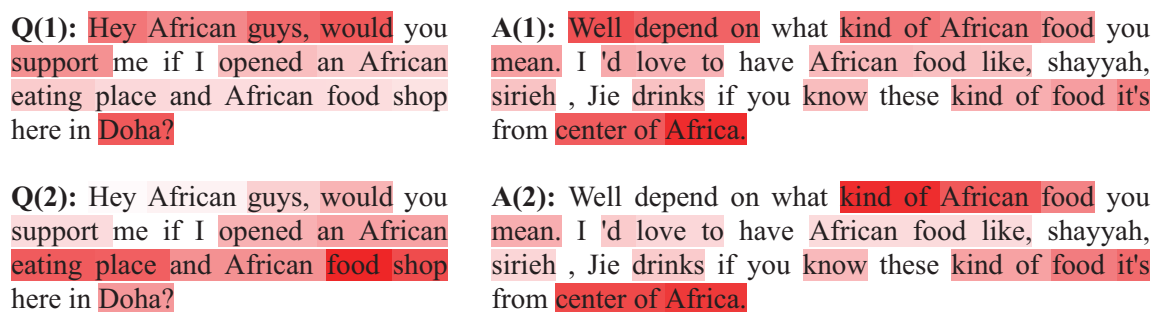
Q(1): Hey African guys, would you support me if I opened an African eating place and African food shop here in Doha?	**A(1):** Well depend on what kind of African food you mean. I 'd love to have African food like, shayyah, sirieh , Jie drinks if you know these kind of food it's from center of Africa.
Q(2): Hey African guys, would you support me if I opened an African eating place and African food shop here in Doha?	**A(2):** Well depend on what kind of African food you mean. I 'd love to have African food like, shayyah, sirieh , Jie drinks if you know these kind of food it's from center of Africa.

Figure 2: An Example of the Visualization of Attention

word level, such as "passport" and "VISA". Different from them, our model can capture user's expertise in applying for VISA by introducing user's historical answer (i.e., user context), and further learns the hidden semantic relationships between question-answer pair by attention mechanism. Thus our model can obtain the better result.

Qualitative Evaluation

To better understand the ability of our model, we randomly choose one QA pair from SemEval-2017 and visualize the attention scores predicted by our model in Figure 2. Specifically, **Q(1)** and **Q(2)** denote the attention of question over answer and user context, while **A(1)** and **A(2)** are the semantic self-attention and co-attention of answer over question, respectively. The color depth indicates the importance degree of the words, the darker the more important.

From the visualization of **Q(1)** and **Q(2)**, we can observe that both cases are paid attention to words that are contextually related to the answer, such as "African", "food", "Doha". Differently, in question's attention over user context **Q(2)**, words that are closely related to user information (i.e., hobby, occupation) are assigned with higher weights, including "eating place" and "food shop". The difference of attention value between **Q(1)** and **Q(2)** indicates that user context can provide potentially useful information to our model.

Similar situation appears in **A(2)** and **A(1)**. Contextually related words such as "food" and "drinks" are assigned higher attention weight by both semantic self-attention and co-attention mechanism. Compared with answer's co-attention over question **A(2)**, answer's semantic self-attention **A(1)** focuses on its own information contained in sentence (e.g., "depend on"), which is hard to provide sufficient knowledge to the entire model.

Related Work

Current CQA task aims to automate the process of finding appropriate answers to questions in a community-created discussion forum (Nakov et al. 2016; 2017). Recent researches (Wu, Xu, and Houfeng 2018; Yang et al. 2019; Shao et al. 2019) often regard the CQA task as a text matching problem and have proposed various deep learning networks to learn the semantic representation of question-answer pairs. The problem of CQA consisting of multiple sentences has attracted the attention of some re-

searchers (Zhang et al. 2017; Wen et al. 2018; Suggu et al. 2016; Zhang et al. 2018). The recent study (Zhang et al. 2017) believes that different text segments may have different effects on the answer selection and adopts a novel attention mechanism to measure the importance of each segment. Researcher (Wu, Xu, and Houfeng 2018) notices the difference between the subject (summary of the main points of the question) and the body (elaboration on the subject in detail) of question, and treats them separately by using an orthogonal decomposition strategy to decompose each question from similarity and disparity. (Suggu et al. 2016) combines the advantages of both hand-crafted features and deep learning based systems. These methods typically regard CQA as a multi-sentence QA task and use an attention-based model to weight different parts of the sentence, which has been proven to be successful in a wide range of NLP tasks (Vaswani et al. 2017; Li et al. 2019; Liang et al. 2019). Although these methods can play a role, they do not take full advantage of the user information in CQA.

On the other hand, some studies aim to enhance the association between question and answer (Fang et al. 2016; Tay et al. 2017). The recent study (Fang et al. 2016) encodes the social interaction cues in community to boost CQA task, in addition to the content of question-answer pairs. However, using information from such heterogeneous social networks may be noisy and useless. Our task differs from the existing methods in that they exclusively consider the interrelations between QA pairs rather than alleviate the noise issue and capture useful knowledge from user context.

Conclusions

In this paper, we study the task of finding appropriate answers in the context of the community question answering (CQA). We present an Attentive User-engaged Adversarial Neural Network (AUANN). A two-step denoise mechanism is designed to overcome noise issue caused by adding plentiful user context information. It consists of a coarse-grained selection process by similarity measurement, and a fine-grained selection process by applying an adversarial training module to model user's historical answer. We propose a novel attentive mechanism to capture the semantic internal and external relations among questions, answers and user contexts. Results on two large-scale datasets demon-

strate the effectiveness of considering the user information, and show that our model performs significantly better than several state-of-the-art baseline methods.

In the future, we are going to introduce the social relationship of users in CQA forums to further capture user's hobby and expertise according to their community.

Acknowledgments

References

Barrón-Cedeno, A.; Da San Martino, G.; Joty, S.; Moschitti, A.; Al-Obaidli, F.; Romeo, S.; Tymoshenko, K.; and Uva, A. 2016. Convkn at semeval-2016 task 3: Answer and question selection for question answering on arabic and english fora. In *SemEval@ACL*, 896–903.

Chen, X., and Cardie, C. 2018. Multinomial adversarial networks for multi-domain text classification. In *NAACL*, 1226–1240.

Deng, Y.; Xie, Y.; Li, Y.; Yang, M.; Du, N.; Fan, W.; Lei, K.; and Shen, Y. 2018. Multi-task learning with multi-view attention for answer selection and knowledge base question answering. In *AAAI*.

Fang, H.; Wu, F.; Zhao, Z.; Duan, X.; Zhuang, Y.; and Ester, M. 2016. Community-based question answering via heterogeneous social network learning. In *AAAI*.

Feng, W.; Wu, Y.; Wu, W.; Li, Z.; and Zhou, M. 2017. Beihangmsra at semeval-2017 task 3: A ranking system with neural matching features for community question answering. In *SemEval@ACL*, 280–286.

Filice, S.; Croce, D.; Moschitti, A.; and Basili, R. 2016. Kelp at semeval-2016 task 3: Learning semantic relations between questions and answers. In *SemEval@ACL*, 1116–1123.

Filice, S.; Da San Martino, G.; Moschitti, A.; and . 2017. Kelp at semeval-2017 task 3: Learning pairwise patterns in community question answering. In *SemEval@ACL*, 326–333.

Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2014. Generative adversarial nets. In *NeurIPS*, 2672–2680.

Hu, B.; Lu, Z.; Li, H.; and Chen, Q. 2014. Convolutional neural network architectures for matching natural language sentences. In *NeurIPS*, 2042–2050.

Jolicoeur-Martineau, A. 2019. The relativistic discriminator: A key element missing from standard gan. In *ICLR*.

Li, Z.; Ding, N.; Liu, Z.; Zheng, H.; and Shen, Y. 2019. Chinese relation extraction with multi-grained information and external linguistic knowledge. In *ACL*, 4377–4386.

Liang, D.; Zhang, F.; Zhang, W.; Zhang, Q.; Fu, J.; Peng, M.; Gui, T.; and Huang, X. 2019. Adaptive multi-attention network incorporating answer information for duplicate question detection. In *SIGIR*.

Liu, P.; Qiu, X.; and Huang, X. 2017. Adversarial multi-task learning for text classification. In *ACL*, 1–10.

McCann, B.; Bradbury, J.; Xiong, C.; and Socher, R. 2017. Learned in translation: Contextualized word vectors. In *NeurIPS*, 6294–6305.

Nakov, P.; Màrquez, L.; Moschitti, A.; Magdy, W.; Mubarak, H.; Freihat, A. A.; Glass, J.; and Randeree, B. 2016. SemEval-2016 task 3: Community question answering. In *SemEval@ACL*.

Nakov, P.; Hoogeveen, D.; Màrquez, L.; Moschitti, A.; Mubarak, H.; Baldwin, T.; and Verspoor, K. 2017. Semeval-2017 task 3: Community question answering. In *SemEval@ACL*.

Nandi, T.; Biemann, C.; Yimam, S. M.; Gupta, D.; Kohail, S.; Ekbal, A.; and Bhattacharyya, P. 2017. Iit-uhh at semeval-2017 task 3: Exploring multiple features for community question answering and implicit dialogue identification. In *SemEval@ACL*, 90–97.

Pennington, J.; Socher, R.; and Manning, C. 2014. Glove: Global vectors for word representation. In *EMNLP*, 1532–1543.

Santos, C. d.; Tan, M.; Xiang, B.; and Zhou, B. 2016. Attentive pooling networks. *arXiv preprint arXiv:1602.03609*.

Shah, H., and Barber, D. 2018. Generative neural machine translation. In *NeurIPS*, 1346–1355.

Shao, T.; Cai, F.; Chen, H.; and de Rijke, M. 2019. Length-adaptive neural network for answer selection. In *SIGIR*, 869–872.

Srivastava, R. K.; Greff, K.; and Schmidhuber, J. 2015. Highway networks. *arXiv preprint arXiv:1505.00387*.

Suggu, S. P.; Goutham, K. N.; Chinnakotla, M. K.; and Shrivastava, M. 2016. Hand in glove: Deep feature fusion network architectures for answer quality prediction in community question answering. In *COLING*, 1429–1440.

Tay, Y.; Phan, M. C.; Tuan, L. A.; and Hui, S. C. 2017. Learning to rank question answer pairs with holographic dual lstm architecture. In *SIGIR*, 695–704.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. In *NeurIPS*, 5998–6008.

Wen, J.; Ma, J.; Feng, Y.; and Zhong, M. 2018. Hybrid attentive answer selection in cqa with deep users modelling. In *AAAI*.

Wu, G., and Lan, M. 2016. Ecnu at semeval-2016 task 3: Exploring traditional method and deep learning method for question retrieval and answer ranking in community question answering. In *SemEval@ACL*, 872–878.

Wu, G.; Sheng, Y.; Lan, M.; and Wu, Y. 2017. Ecnu at semeval-2017 task 3: Using traditional and deep learning methods to address community question answering task. In *SemEval@ACL*, 365–369.

Wu, W.; Xu, S.; and Houfeng, W. 2018. Question condensing networks for answer selection in community question answering. In *ACL*, 1746–1755.

Xiang, Y.; Zhou, X.; Chen, Q.; Zheng, Z.; Tang, B.; Wang, X.; and Qin, Y. 2016. Incorporating label dependency for answer quality tagging in community question answering via cnn-lstm-crf. In *COLING*, 1231–1241.

Yang, M.; Chen, L.; Chen, X.; Wu, Q.; Zhou, W.; and Shen, Y. 2019. Knowledge-enhanced hierarchical attention for community question answering with multi-task and adaptive learning. In *IJCAI*, 5349–5355.

Zhang, X.; Li, S.; Sha, L.; and Wang, H. 2017. Attentive interactive neural networks for answer selection in community question answering. In *AAAI*, 3525–3531.

Zhang, H.; Li, Y.; Ma, F.; Gao, J.; and Su, L. 2018. Texttruth: An unsupervised approach to discover trustworthy information from multi-sourced text data. In *KDD*, 2729–2737.