SEOUL NATIONAL UNIVERSITY
Department of Electrical and Computer Engineering

Learning to Rank Question-Answer Pairs using Hierarchical Recurrent Encoder with Latent Topic Clustering

Seunghyun Yoon, Joongbo Shin, and Kyomin Jung

Goal of our Research

- We propose a novel end-to-end neural architecture for ranking answers from candidates.
- We introduce a Hierarchical Recurrent Dual Encoder (HRDE) model to effectively calculate the affinity among question-answer pairs to determine the ranking. It prevents performance degradations in understanding longer texts while other recurrent neural networks suffer.
- We propose a Latent Topic Clustering (LTC) module to extract latent information from the target dataset, and apply these additional information in end-to-end training.
- Extensive experiments are conducted to investigate efficacy and properties of the proposed model. Our proposed model outperforms previous state-of-the-art methods in the Ubuntu Dialogue Corpus and Samsung QA Corpus.

(Q) how do i set a timer of clock in applications and development for samsung galaxy s4 mini?

(A) 1 from within the clock application, tap timer tab. 2 tap the hours, minutes, or seconds field and use the on-screen keypad to enter the hour, minute, or seconds. the timer plays an alarm at the end of the countdown. 3 tap start to start the timer. 4 tap stop to stop the timer or reset to reset the timer and start over. 5 tap restart to resume the timer counter.

Model

- Hierarchical Recurrent Dual Encoder (HRDE) divides long sequential text data into small chunk such as sentences, and encodes the whole text from word-level to chunk-level by using two hierarchical level of RNN architecture.

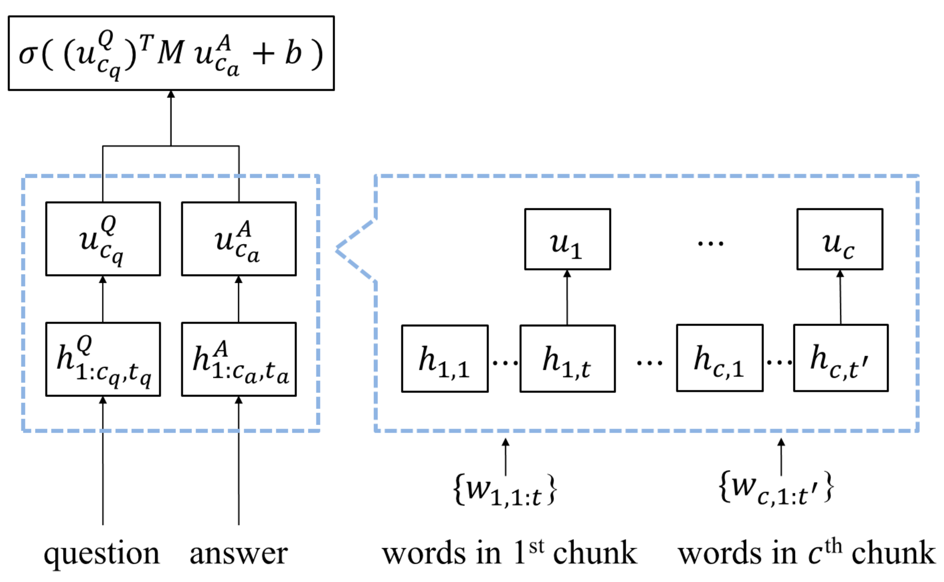

Figure 1: Diagram of the HRDE model.

$$h_{c,t} = f_\theta(h_{t-1}, w_{c,t}),$$
$$u_c = g_\theta(u_{c-1}, h_c),$$

f_θ, g_θ : the RNN function in hierarchical architecture
$h_{c,t}$: word-level RNN's hidden status at t^{th} word in c^{th} chunk
$w_{c,t}$: t^{th} word in c^{th} chunk
u_c : chunk-level RNN's hidden state at c^{th} chunk

$$p(\text{label}) = \sigma\left(\left(h^Q_t\right)^T M\, h^A_t + b \right), \quad \mathcal{L} = -\log \prod_{n=1}^{N} p(\text{label}_n | h^Q_{n,t}, h^A_{n,t})$$

- Latent Topic Clustering (LTC) module groups the target data to help the neural network find the true-hypothesis with more information from the topic cluster in end-to-end training.

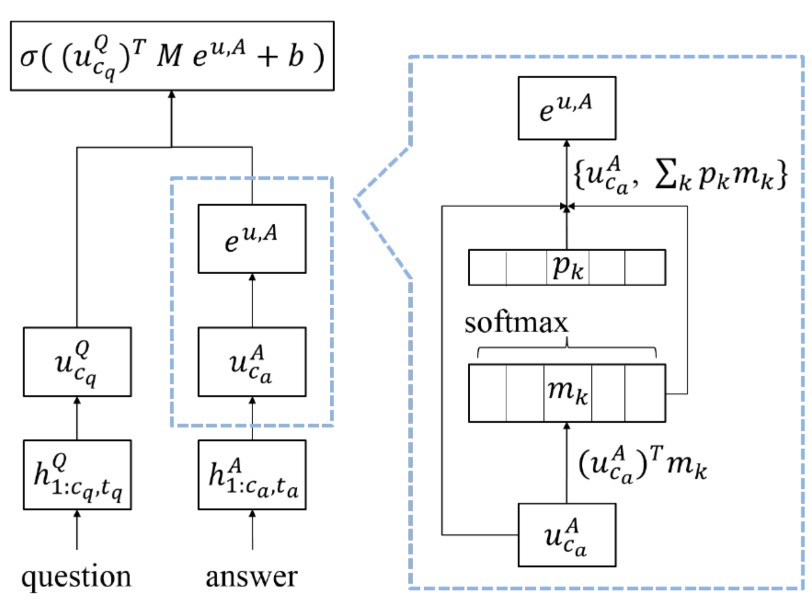

Figure 2: Diagram of the HRDE-LTC.

$$p_k = softmax\left((\boldsymbol{x})^T m_k \right),$$
$$\boldsymbol{x}_k = \sum_{k=1}^{K} p_k m_k,$$
$$\boldsymbol{e} = concat\,(\boldsymbol{x}, \boldsymbol{x}_k),$$

p_k : similarity between the $\boldsymbol{x}$ and each latent topic vector m_k
$\boldsymbol{x}_k$: summing over m_k weighted by the p_k
$\boldsymbol{e}$: final vector with latent topic information added

Dataset

- Ubuntu Dialogue Corpus V1/V2 Preprocessing the Ubuntu Chat Logs, which refer to a collection of logs from the Ubuntu-related chat room for solving problem in using the Ubuntu system.
- Samsung QA Corpus Question and answer pair dataset related to an actual user's interaction with the consumer electronic product domain from crowd QA web.

Dataset	# Samples			Message (Avg.)			Response (Avg.)		
	Train	Val.	Test	# tokens	# groups	# tokens /group	# tokens	# groups	# tokens /group
Ubuntu-v1	1M	35,609	35,517	162.47 ±132.47	8.43 ±6.32	20.14 ±18.41	14.44 ±13.93	1	-
Ubuntu-v2	1M	19,560	18,920	85.92 ±74.71	4.95 ±2.98	20.73 ±20.19	17.01 ±16.41	1	-
Samsung QA	81,808	10,000	10,000	12.84 ±6.42	1	-	173.48 ±192.12	6.09 ±5.58	29.28 ±31.91

Table 1: Properties of the Ubuntu and Samsung QA dataset. The message and response are {context}, {response} in Ubuntu and {question}, {answer} in the Samsung QA dataset. Standard deviations are shown below each average value.

Empirical Results

- Comparison with the state-of-the-art methods

Model	Ubuntu-v1			
	1 in 2 R@1	1 in 10 R@1	1 in 10 R@2	1 in 10 R@5
TF-IDF [1]	0.659	0.410	0.545	0.708
CNN [2]	0.848	0.549	0.684	0.896
LSTM [2]	0.901	0.638	0.784	0.949
CompAgg [3]	0.884	0.631	0.753	0.927
BiMPM [4]	0.897	0.665	0.786	0.938
RDE	0.898 ±0.002	0.643 ±0.009	0.784 ±0.007	0.945 ±0.002
RDE-LTC	0.903 ±0.001	0.656 ±0.003	0.794 ±0.003	0.948 ±0.001
HRDE	0.915 ±0.001	0.681 ±0.001	0.820 ±0.001	0.959 ±0.001
HRDE-LTC	**0.916** ±0.001	**0.684** ±0.001	**0.822** ±0.001	**0.960** ±0.001

Model	Ubuntu-v2			
	1 in 2 R@1	1 in 10 R@1	1 in 10 R@2	1 in 10 R@5
LSTM [1]	0.869	0.552	0.721	0.924
RNN [5]	0.907 ±0.002	0.664 ±0.004	0.799 ±0.004	0.951 ±0.001
CNN [5]	0.863 ±0.003	0.587 ±0.004	0.721 ±0.005	0.907 ±0.003
RNN-CNN [5]	0.911 ±0.001	**0.672** ±0.002	0.809 ±0.002	0.956 ±0.001
Attention [6] (RNN-CNN)	0.903 ±0.002	0.653 ±0.005	0.788 ±0.005	0.945 ±0.002
CompAgg [3]	0.895	0.641	0.776	0.937
BiMPM [4]	0.877	0.611	0.747	0.921
RDE	0.894 ±0.002	0.610 ±0.008	0.776 ±0.006	0.947 ±0.002
RDE-LTC	0.899 ±0.002	0.625 ±0.004	0.788 ±0.004	0.951 ±0.001
HRDE	0.914 ±0.001	0.649 ±0.001	0.813 ±0.001	0.964 ±0.001
HRDE-LTC	**0.915** ±0.002	0.652 ±0.003	**0.815** ±0.001	**0.966** ±0.001

Models [1-6] are from (Lowe et al., 2015; Kadlec et al., 2015; Wang and Jiang, 2016; Wang et al.,2017; Baudis et al.,2016; Tan et al., 2015), respectively.

Model	Samsung QA			
	1 in 2 R@1	1 in 10 R@1	1 in 10 R@2	1 in 10 R@5
TF-IDF	0.939	0.834	0.897	0.953
RDE	0.978 ±0.002	0.869 ±0.009	0.966 ±0.003	0.997 ±0.001
RDE-LTC	0.981 ±0.002	0.880 ±0.009	0.970 ±0.003	0.997 ±0.001
HRDE	0.981 ±0.002	0.885 ±0.011	0.971 ±0.004	0.997 ±0.001
HRDE-LTC	**0.983** ±0.002	**0.890** ±0.010	**0.972** ±0.003	**0.998** ±0.001

Table 2: Model performance results for the Ubuntu-v1 dataset (left-top), Ubuntu-v2 dataset (right) and Samsung QA dataset (left-bottom), respectively.

- Degradation Comparison for Longer Texts

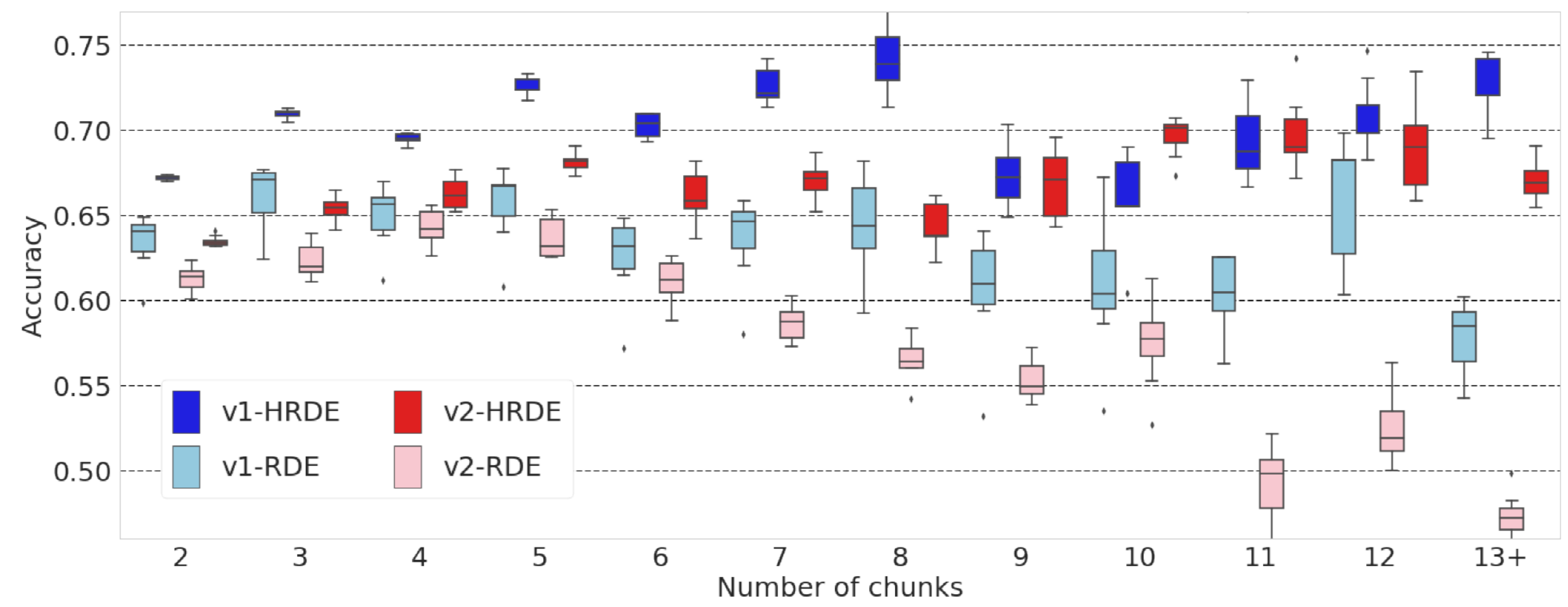

Figure 3: The HRDE and RDE model performance comparisons for the number-of-chunk in the Ubuntu dataset.

- Comprehensive Analysis of Latent Topic Clustering

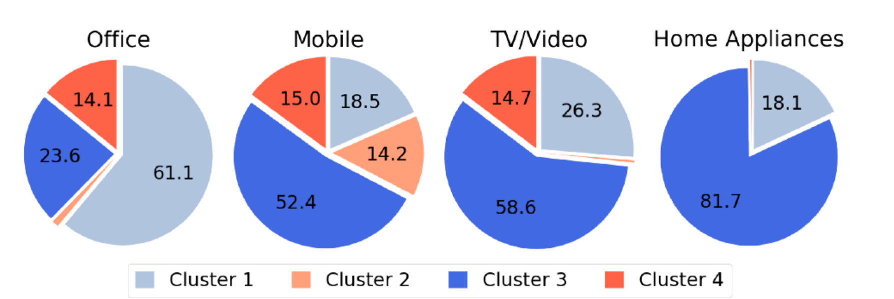

Figure 4: Examples of the cluster proportions for 4 real *categories*.

Cluster	Example
1	How to adjust the brightness on the s**d300 series monitors
2	How do I reject an incoming call on my Samsung Galaxy Note 3?
3	How should I clean and maintain the microwave?
4	How do I connnect my surround sound to this TV and what type of cables do I need

Table 3: Example sentences for each cluster.

code, data, paper → http://david-yoon.github.io