Multi-label Aspect Classification on Question-Answering Text with Contextualized Attention-based Neural Network

Hanqian Wu[1, 2, *], **Shangbin Zhang**[1, 2], **Jingjing Wang**[3], **Mumu Liu**[1, 2], **Shoushan Li**[3]

[1]School of Computer Science and Engineering, Southeast University, China

[2]Key Laboratory of Computer Network and Information Integration of Ministry of Education, Southeast University, China

[3] NLP Lab, School of Computer Science and Technology, Soochow University, China

hanqian@seu.edu.cn, ternencewind@outlook.com, djingwang@gmail.com, liudoublemu@163.com, lishoushan@suda.edu.cn

Introduction

- In this paper, we model QA aspect classification (QA-AC) task as a multi-label classification problem where each QA style review is explicitly mapped to multiple aspect categories instead of only one aspect category.
- We propose a contextualized attention-based neural network approach to capture both the contextual information and the QA matching information inside QA style reviews for the task of QA-AC. Specifically, we first propose two aggregating strategies to integrate multi-layer contextualized word embeddings of the pre-trained language representation model so as to capture contextual information. Then, we propose a bidirectional attention layer to capture the QA matching information.
- Empirical results demonstrate that our proposed model outperforms several state-of-the-art baselines by larger margins on QA style reviews.

Aspect Category: Performance, Battery

Battery Performance

Question: *亲们这款手机待机时间长吗？性能如何？谢谢了！*

Answer: *老人机足够了，待机时间一般。*

Performance Battery

<EN>

Question: *Does this phone have a long standby time? What about its performance? Thank you!*

Answer: *as a geriatric cellular phone is enough, the standby time is just so so.*

Fig. 1. A translated example of QA style reviews from an e-commerce website.

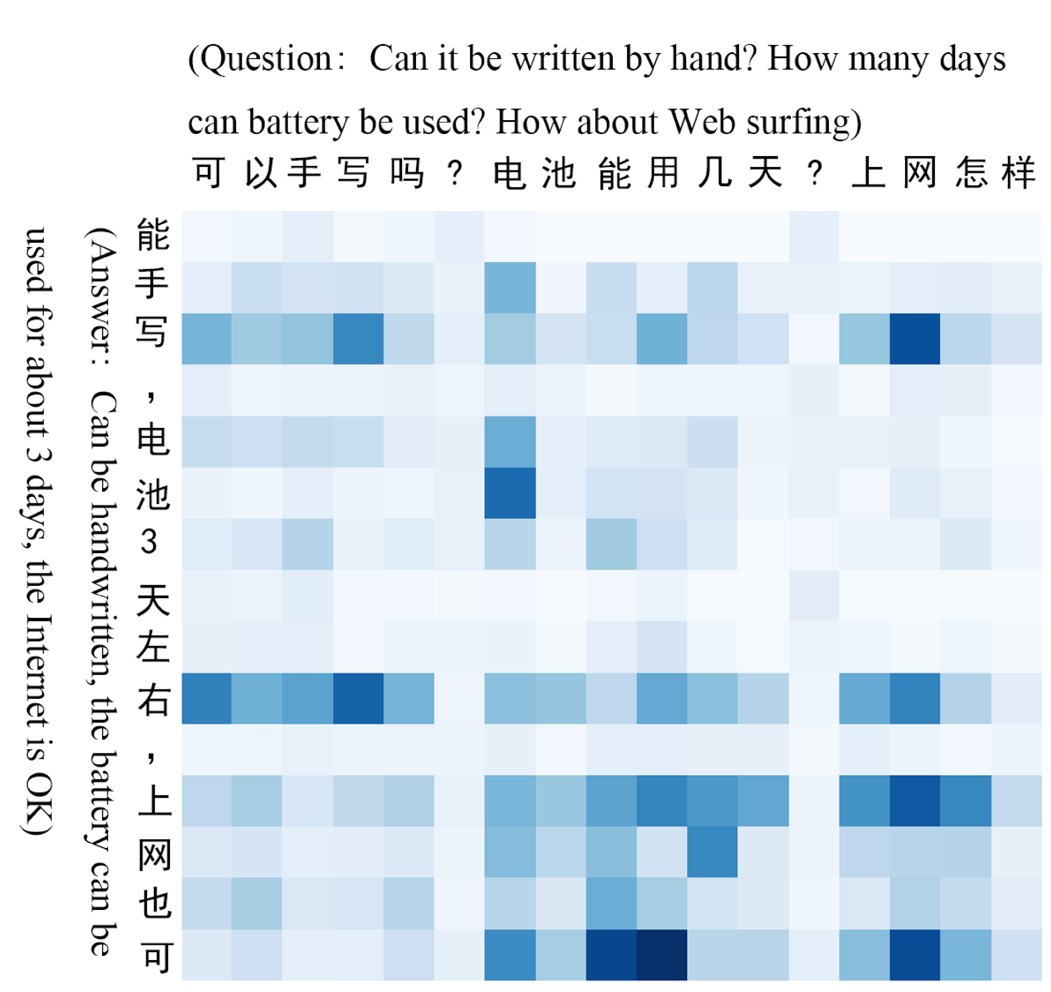

Fig. 4. An attention visualization with the aspects of “Function” and “Battery”.

Our Model

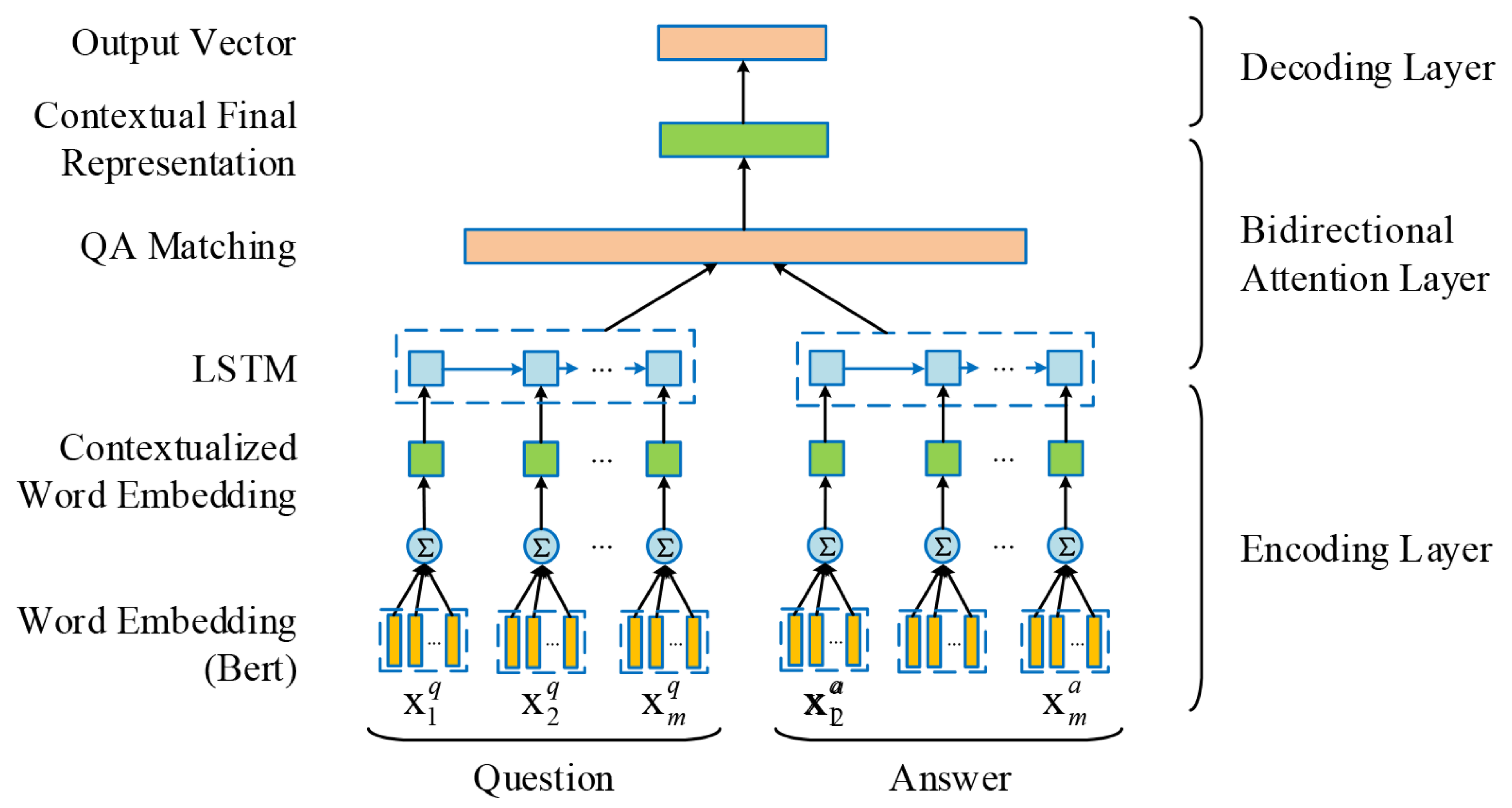

Fig. 2. The overall structure of our model.

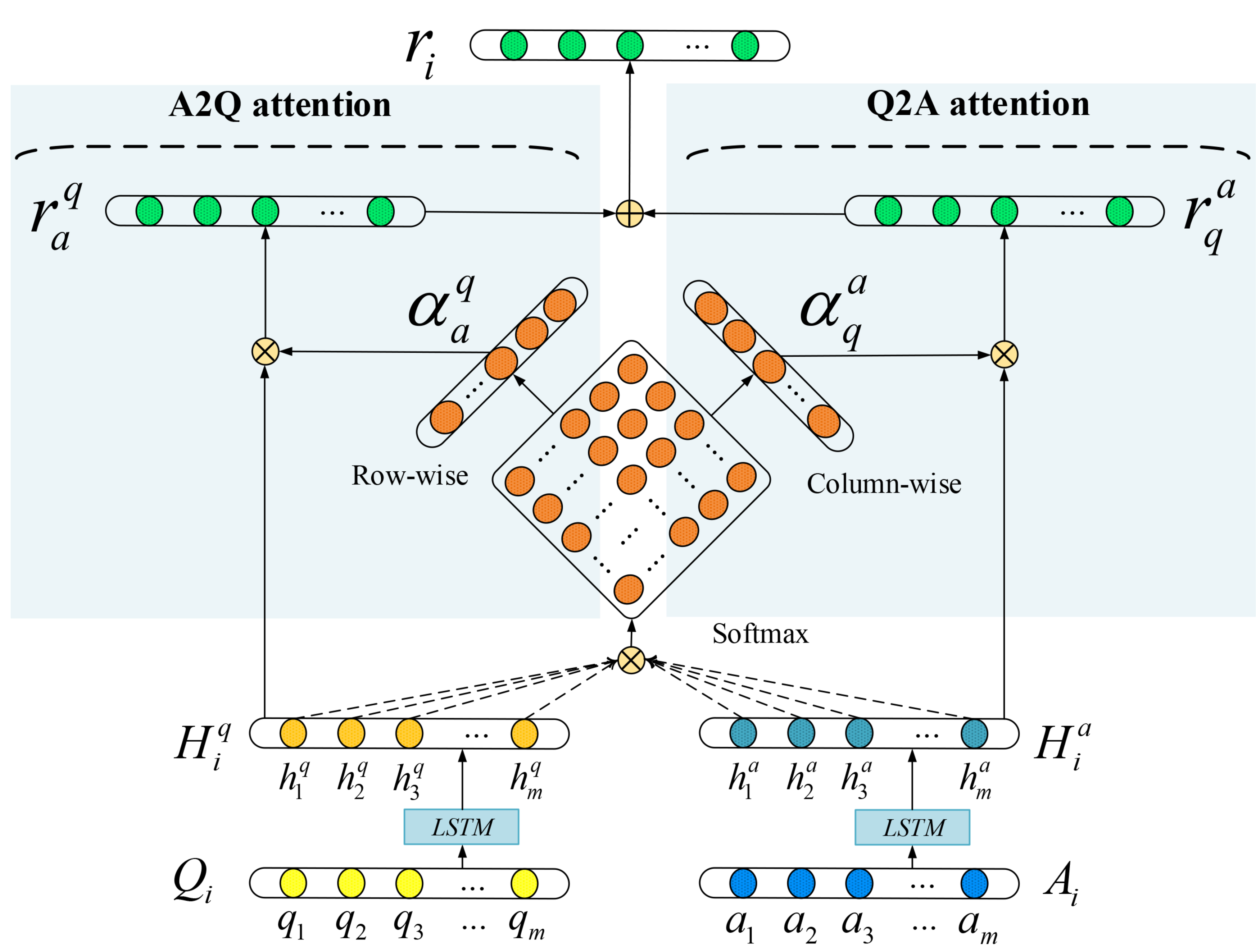

Fig. 3. The structure of Bidirectional Attention Layer.

Experimental Settings

- **DataSets:** We conduct our experiments on the QA style reviews dataset collected from “asking all” in Taobao. The whole corpus has 2580 reviews of electronic appliances and contains seven aspects of products. Each review may express more than one aspect, and we aim to identify the whole aspects list of each review. Notably, the scale of multi-label instances in major aspects exceeds 10%.
- **Training Settings:** In BERT word embeddings, vector dimension is fixed to 768. The max length of the question and the answer are 30, and the initial learning rate is 0:005. Moreover, all models are trained by mini-batch of 32 instances, the dropout rate is 0:5.

Experimental Results

Table. 3. Ablation studies results.

Models	Hamming loss	Accuracy	F1-measure
BANN(BERT)	0.021	0.886	0.935
- LSTM	0.081	0.559	0.606
- Attention	0.029	0.849	0.911
- Q2A Attention	0.082	0.629	0.675
- A2Q Attention	0.031	0.855	0.911
Using weighted summation instead of averaging	**0.019**	**0.896**	**0.942**

Table. 2. Experimental results. The best scores are in bold.

Models	Hamming loss	Accuracy	F1-measure
Binary Relevance [16]	0.064	0.662	0.707
Classifier Chains [16]	0.032	0.836	0.886
LSTM(Word2vec) [3]	0.031	0.843	0.901
BANN(Word2vec)	0.029	0.861	0.904
BANN(BERT)	**0.021**	**0.886**	**0.935**