THROUGH THE LENS OF NEURAL NETWORK: ANALYZING NEURAL QA MODELS VIA QUANTIZED LATENT REPRESENTATION

Anonymous authors
Paper under double-blind review

ABSTRACT

In recent years, deep learning models remain *black boxes*, where the decision-making process is still opaque to humans. In this work, we try to explore the probabilities of understanding how machine *thinks* when doing question-answering tasks. In general, words are represented by continuous latent representations in the neural-based QA models. Here we train the QA models with discrete latent representations, so each word in the context is also a token in the model. In this way, we can know what a word sequence in the context looks like through the lens of the QA models. We analyze the QA models trained on QuAC (Question Answering in Context) and CoQA (A Conversational Question Answering Challenge) and organize several rules the models obey when dealing with this kind of QA task. We also find that the models maintain much of the original performance after some hidden layers are quantized.

1 INTRODUCTION

Over the past few years, more and more people have been enchanted at the power of deep neural networks; they try to apply various networks on different fields such as reading comprehension, computer vision, image recognition, speech recognition, etc. Nowadays, there are many outstanding deep learning-based models capable of yielding human-comparable results on some benchmark corpora. For instance, those leading models on SQuAD (Rajpurkar et al., 2018) and CoQA (Reddy et al., 2019) have already passed the baseline of human performance; however, their mechanisms remain *black boxes* for us since nobody knows how they make a decision. When models make the wrong prediction, we can merely do some fine-tuning or use more training data. Most of the time, we do not know what modifications to make for better performance.

In general, the latent representation in a deep neural network can be regarded as what a machine thinks when solving tasks. People have already begun to elaborate on the meaning of latent representation (Doshi-Velez & Kim, 2017; Montavon et al., 2018); however, only those deep learning researchers are able to understand and utilize the results. In this work, we seek the possibility for not only further realizing how the machine thinks but making deep neural networks much more comprehensible for all people.

In recent years, conversational question answering is found to be pretty challenging for existing deep neural networks. Given one context and a series of questions, machines sometimes should output two different answers to the same question because, in these datasets, output answers according to not only the context and question but also those questions that are asked in the previous time steps. Besides, a model should handle the co-reference issue when trying to obtain correct answers. For instance, it should know the relation between those pronouns and people, food, location, etc. In order to do so, a model has to store these kinds of information in its latent representations. We focus on analyzing the neural QA models trained on this kind of task.

In this work, we demonstrate how to *visualize* the original latent vectors as more meaningful representations, which can be easily understood by humans. Originally, latent representations lie in a continuous high dimensional space, and it is pretty hard for humans to know what they mean. Therefore, to simplify it, we propose a vector quantized method, which can quantize latent representations into discrete vectors. Although, in general, vector quantization is usually applied to computer vision or speech recognition rather than natural language processing, we believe that this architecture is quite suitable for *visualizing* the latent representations. After converting the hidden representations into discrete vectors, humans can read the articles through the lens of deep neural networks. Generally, humans are more comfortable with these discrete tokens since words and sentences in all languages are in this form.

One of the reasoning results is shown in Figure 1. Each color represents a codeword. The rare codewords are with the red tone, while the common ones are with the blue tone. Within the same context, a machine uses different colors to represent the same word. This is because even for the same word, it should contain different meanings at different time steps. For instance, ”Richie” is colored as orange in Figure 1(a) and as blue in Figure 1(b). By contrast, different words can use the same color, such as those dark blue tokens in Figure 1. Note that the spans inside green boxes are the ground truth, while the ones with underline are model outputs.

After coloring all tokens, we organize several rules which machine follows based on the experiment result. First of all, we have found that for these deep neural networks, they do not need that a massive size of codebook to achieve good performance. On average, a neural network can use only 20 or so vectors to cover almost all information. This striking fact is based on very little performance degradation. In the second place, we have found the meaning of some codewords, some of which are nonsense, whereas some of which preserve a lot of information. However, even for those nonsense words at the current time step, it may be very meaningful at other time steps, depending on the question machine is asked. Apart from this, we successfully identify those tokens that are responsible for marking the boundary of answer spans, which is supposed to be the key element of most QA models.

2 Related work

Deep learning for reading comprehension has attracted much attention (Bordes et al., 2015; Kumar et al., 2015; Xiong et al., 2016; Hermann et al., 2015; Shih et al., 2015). Generally, conventional reading comprehension systems are designed as a cascade of some components (syntactic parser, semantic parser, etc.), while in the case of end-to-end reading comprehension model, a neural network which takes a story and some questions as input and an answer as output is directly learned from training data, making it feasible to jointly learn components of conventional QA system.

In addition, the mechanisms such as hopping and attention are widely used in reading comprehension models so as to model the deduction process (Weston et al., 2014). In a dynamic memory network (DMN) (Kumar et al., 2015), questions trigger an iterative attention process which allows the model to condition its attention on the inputs and the result of previous iterations. After reasoning over multiple supporting facts, Neural Reasoner (Peng et al., 2015) is able to find an answer to a given question. The Recurrent Entity Network (EntNet) (Henaff et al., 2017) is the first method to solve all the tasks in the 10k training examples setting of bAbI (Weston et al., 2015). The Query-Reduction Network (QRN) (Seo et al., 2016b) effectively handles both short-term and long-term sequential dependencies in order to reason over multiple facts. To achieve complex relational reasoning, new models are proposed (Bansal et al., 2017; Santoro et al., 2017; Palm et al., 2018; Pavez et al., 2018). Recent attention mechanisms and network architectures have been shown to be quite helpful for SQuAD (Xiong et al., 2017; Seo et al., 2016a; Wang et al., 2017; Hu et al., 2017; Huang et al., 2017). Also, on the SQuAD leaderboard[1], deep learning-based models are competitive with human performance and even surpass it (Yang et al., 2019; Liu et al., 2019; Devlin et al., 2018). In addition to exploiting only plain context-sensitive features such as character or word embedding, Zhang et al. (2019a) proposes to incorporate explicit contextual semantics in order to promote natural language understanding. SG-Net (Zhang et al., 2019b) is proposed to guide machines by incorporating syntactic constraints into their attention mechanisms.

Interpretation of neural networks sometimes serve as preliminary works to some tasks such as model compression (Voita et al., 2019) or Machine Translation (Voita et al., 2018). Among all interpretation methodologies, visualization has been the most intuitive and widely accepted in a plethora of such works. Most of them relate components of a neural network such as activations (Kádár et al.,

[1] https://rajpurkar.github.io/SQuAD-explorer/

(a) The 1st question turn (What shows did she appear on?)

(b) The 2nd question turn (What was the show about?)

(c) The 9th question turn (Is there anything else interesting?)

Figure 1: Vector quantized latent representations for different question turns.

2016; Liu et al., 2018), weights of neurons (Ding et al., 2017) or attention heads (Bahdanau et al., 2015) to their corresponding positions in the inputs or outputs. Saliency based on some importance measure is quite common as well. Voita et al. (2018) calculated the importance of each attention head and related them to the syntactic role each head is assumed to take by experiment. Feng et al. (2018) assigned each word in an input context an importance calculated from gradient change by removing the word. Nevertheless, most of these interpretation measures yield continuous values and thus are hard to analyze and visualize directly, leaving the reasoning process of deep neural network still remain unexploited.

3 METHODOLOGY

In this kind of task, inputs should contain a context $\mathbb{C}$ and some related questions $\mathbb{Q}$, both of which are in natural language. Here is an example in Figure 1. For the same context, models will be asked several different yet related questions, such as the first question in Figure 1(a) and the second one in Figure 1(b). Models should know what *show* is in the second question after answering the first question. For the purpose of understanding the reasoning process, we replace one of any layers that should be analyzed with a vector quantized (VQ) layer. In fact, for any methods that can quantize latent space into discrete one should work. Here we simply choose the one used in VQ-VAE (van den Oord et al., 2017). The idea of our model structure is shown in Figure 2. VQ layer should have the exact same shape with the layer that we need to analyze, then with an extra codebook, we are able to replace all latent vectors with quantized ones. In addition, replacing different layers with VQ ones will have different influences on performance; we will discuss this phenomenon based on the experiments on SDNet in section 5.1.

As for the gradient issue of VQ layer, we simply copy the gradients from the next layer's input to the previous layer's output. Then, the final loss for this model structure should include 3 terms: QA loss, VQ (Vector Quantization) loss, and commitment loss. While VQ loss helps the embedding vector move toward the previous layer's output, commitment loss makes sure the previous layer commits to an embedding.

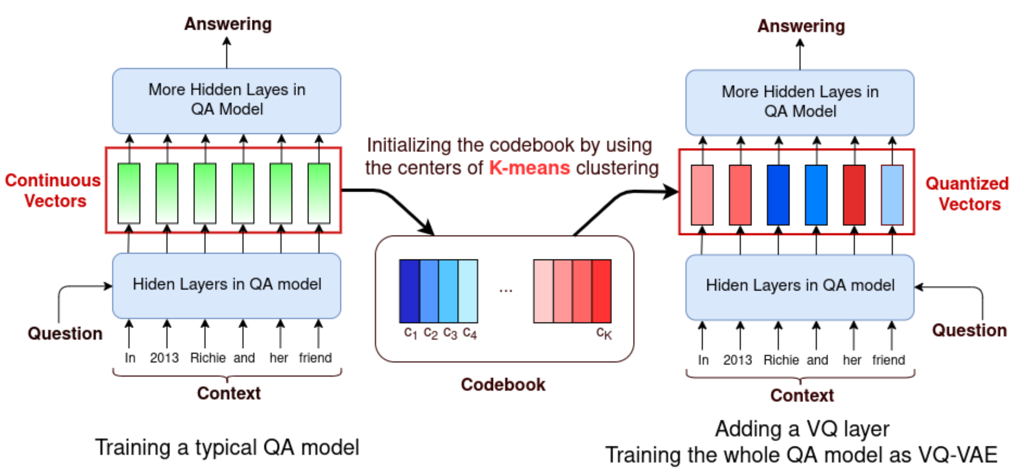


Figure 2: Vector quantized model.

4 QUESTION-ANSWERING MODEL

In this paper, we use FlowQA (Huang et al., 2018) and SDNet (Zhu et al., 2018) as the example of QA models to be studied. We modify both of them then train and evaluate on QuAC (Choi et al., 2018) and CoQA (Reddy et al., 2019) separately. These two datasets are quite similar, both are conversational question answering datasets. Given one context and a series of questions, models are asked to answer them chronologically; they have to answer them based on the previous questions being asked. One example is shown in Figure 1.

We train a bunch of FlowQA models with the same architecture for the purpose of organizing some general reasoning rules. On the other hand, we implement two different versions of SDNet. We put the VQ layer in different positions (SDNet(v1) and SDNet(v2)) so as to observe the reasoning process respectively. We will elaborate on the differences between these two versions in section 5.1.

The network architecture and all details of FlowQA are in Appendix A.1 and the two different network architectures of SDNet are shown in Appendix A.2 respectively.

5 EXPERIMENTS

5.1 DENSITY OF CODEBOOKS

The final performance of two models with different codebook size, K, is listed in Table 1. Note that we will not use the latent vectors of those models with gray color or do the further discussion and experiment in the following analysis. And, we also try to boost the performance of VQ models. Firstly, we randomly sample 1M vectors from the reproduced model, then we apply the K-means algorithm to it. Then, we use these 16 vectors to initialize our codebook and find that it does help boost the performance. Nonetheless, after comparing all the hidden representations with the original $K = 16$ model, we can hardly find any differences between them. Consequently, most of our experiments are based on the vanilla VQ model without K-means vector initialization. Moreover, for the sake of analyzing easily, we try to choose K as small as possible.

It is obvious that even for a very small K such as 4, both QA models can still maintain good performance. In the past, most applications of VQ-VAE are about image, audio, and video, few people have ever imagined that it can be applied to QA models, even for such small K. It is worth noting that the performance of SDNet(v2) is still very good when $K = 1$. $K = 1$ implies that each context word uses the same codeword, so the only way for the model to achieve such good performance is mainly because of the residual layers designed in vanilla SDNet. In SDNet(v2), the VQ layers are residual layers, whose outputs are combined with their inputs, including word embedding and BERT embedding, leading to information propagating to the next layers.

Table 1: Performance of two vector quantized models. The light cyan row stands for the vanilla model without vector quantization, while those cells with gray color means that they are under the 90% performance of vanilla model. The pink row is the model with K-means initialization.

	F1 Score		
K	FlowQA on QuAC	SDNet(v1) on CoQA	SDNet(v2) on CoQA
Reproduce	63.641	76.439	76.439
512	59.232	72.426	74.575
128	59.419	72.797	74.318
64	59.352	69.789	74.182
32	59.350	70.456	73.671
16	59.268	71.128	73.320
8	59.082	70.999	72.409
4	57.267	71.752	71.692
2	54.974	53.546	71.684
1	35.095	16.119	68.759
16(K-means)	62.566	-	-

We have found that even for a large K, models do not actually need that many codewords. In average, for these two kinds of QA reasoning tasks, models only need a small capacity to reach a quite good performance. When we say that only 24 codewords are used, it means that to evaluate the whole validation data, the model only utilizes 24 codewords in its codebook. For example, for the $K = 512$ FlowQA model, it actually utilize 24 codewords when predicting.

For Model 1, Model 2, and Model 3 of FlowQA, it means that we use the same model structure yet train three individual models. Besides, in SDNet(v1) we put one VQ layer at the rear part of the model, while in SDNet(v2) we put two VQ layers in the middle part of the model. As for Layer 1 and Layer 2 of SDNet(v2), it means that we replace two hidden layers in one model with two VQ layers yet sharing the same codebook.

Other implementation details can be found in Appendix A.2.

Table 2: Actually used codewords. The gray cells mean that models use all codewords in their codebooks.

	FlowQA			SDNet	SDNet(v2)	
K	Model 1	Model 2	Model 3	Model 1	Model 1	
Layer					Layer1	Layer2
512	24	30	28	23	169	346
128	25	24	32	23	36	94
64	23	30	30	18	20	53
32	23	32	32	17	10	24
16	15	16	15	16	10	16
8	8	-	-	8	6	8

5.2 Meaning of nonsense words

We try to visualize the reasoning process of machines by coloring its latent representation (codeword) and organize some rules. One of the examples with different time steps is shown in Figure 1(a), Figure 1(b), and Figure 1(c). In the figures, different colors represent different codewords. As for other time steps and yet another example, we put them in Appendix B.5.

After observing tons of the colored figures and the statistical analysis, we define the first two high frequency, which cover 65% for all codewords, as nonsense words (Dark Blue and Sky Blue). These two codewords account for a huge portion in all size codebooks; therefore, we assume that it should be more meaningless than other rare codewords.

Also, we can easily notice that the most colorful part is usually the model output span, and this colorful segment is slowly shifted forward over time; the beginning of the context turns into dark blue and sky blue in the last question turns. This is because, for these two datasets we use, the positions of answers have a pretty strong relation with question turns. For instance, the answer to the first question should appear at the beginning of context, whereas the answer to the last question should appear at the end part of the context. We will discuss this phenomenon in the section 5.3.

In Table 3, ”Overall” column represents the percentage of these two codewords, ”Model Output” means that the percentage of these two codewords appear within model output spans, and ”Subject of Model Output” is the percentage of these nonsense words appearing within the subjects of model output spans. We put the other three tables of SDNet in Appendix B.2. The most interesting thing of these tables is that although model output spans usually do not contain those two nonsense words, their subjects actually do contain numerous nonsense words. We conclude that it is because, at the layer we quantize, subjects do not really matter to models. For FlowQA here, we set the quantized layer at the rear part, and at this layer, the model should have already known the exact answer span. Thus, other grammatical information such as object, location, adjective,...etc, may be more crucial than subject. Most questions in these 2 datasets are about co-reference between different subjects. As long as models know the correct answer position, subjects do not play a part anymore.

Table 3: Nonsense Words in FlowQA.

FlowQA			
K	Overall	Model Output	Subject of Model Output
512	0.4312	0.2396	0.4893
128	0.3960	0.2336	0.4454
64	0.4003	0.2271	0.4622
32	0.4084	0.2405	0.4812
16	0.4893	0.2898	0.5448
8	0.5763	0.4156	0.6261
Average	0.4503	0.2744	0.5082

5.3 Position and portion of nonsense words

In the previous section, we have mentioned that we consider the most two frequent codewords as nonsense words. In this part, we use another viewpoint to analyze it. In general, the higher a codeword's percentage, the less meaning it preserves. In addition, in section 5.2 we have mentioned that the colorful segments which might contain answers will gradually shift forward as time goes on. In Figure 3, code 0 always stands for the rarest codeword whereas the largest code number is the most common one. We can see that those crucial codewords certainly shift forward as time goes by; on the contrary, nonsense words shift backward. Note that the y-axis is the normalized position of context. For each word in a context, the normalized position stands for the position of that word divided by the length of the context. For example, normalized position$= 0, 0.5, 1$, means that the word is the first, middle, and the last word of the context. For other K we put them in Appendix B.1.

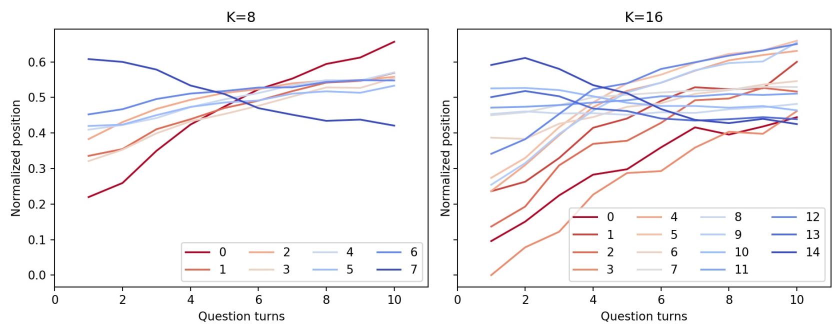


Figure 3: Average position of different codewords in FlowQA (K=8, 16).

5.4 Part-of-speech

We find that codewords may contain some syntactic and positional information. Such phenomenon is observed in Figure 5 (for the other K, see Appendix B.4). In the case of QuAC, whose answers tend to be sentences, input words with POS tags such as *VERB*, *NOUN*, and *PROPN* (proper noun), are critical for answering questions. For words with such POS tags, if they are unrelated to the current question, the model will assign them nonsense codewords (say, some of the most frequent) for masking, so as not to mislead the later hidden layers. That is why the critical words' tags, *VERB*, *NOUN*, and *PROPN*, dominate almost all columns including nonsense words, code 0, code 1, and code 2.

Moreover, the most important in Figure 4 and Figure 5 is that there are usually one or two codewords possibly representing the start position and the end position for the answer, which may be due to both models' nature as pointer networks. For the example in Figure 4, code 9 and code 7 take responsibility for this task. In order to extract answers, models always output the probability of start and end positions. Now we may hypothesize that at least in the quantized layer, the model has already tried to mark the boundary of answers.

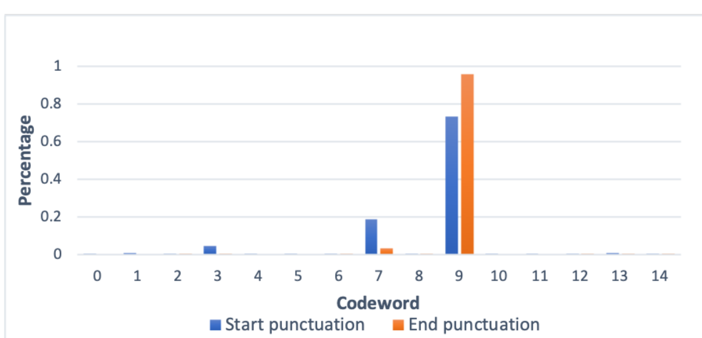


Figure 4: Codewords distribution on start/end punctuation of FlowQA (K=16).

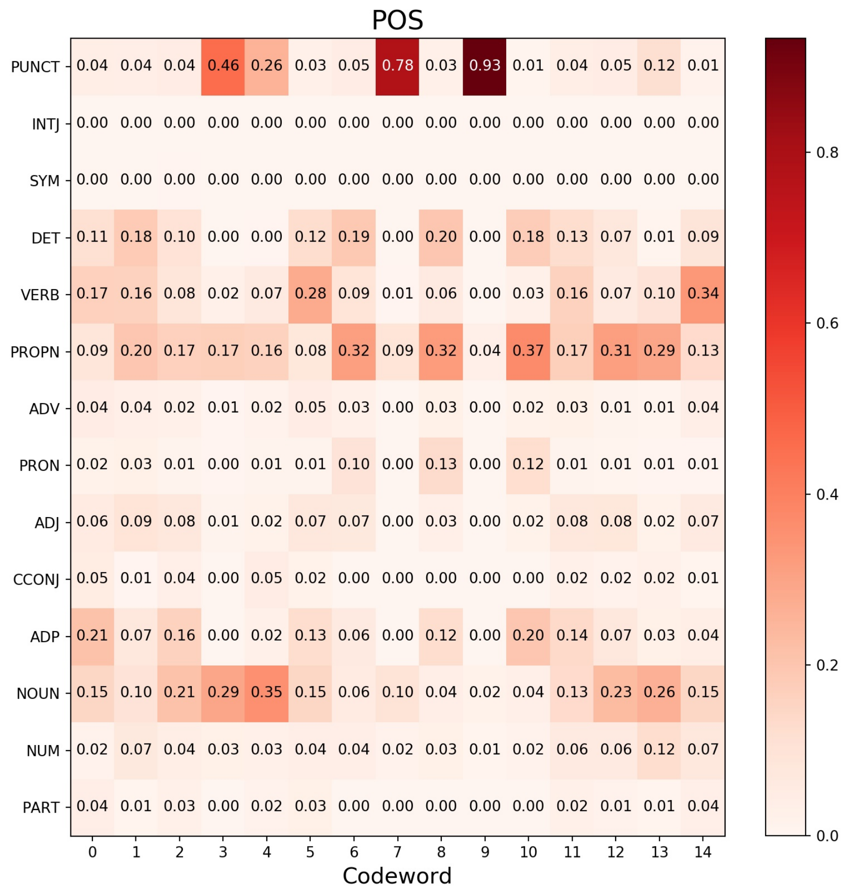


Figure 5: The frequency of POS tags in each codeword of FlowQA (K=16). Note that for all heat maps in this work, the codewords are ordered by their frequecy in the latent representation, from the most frequent to the rarest. For each codeword in hidden layer, we can find the corresponding word token in input layer. Then we use that word to generate POS tags.

6 CONCLUSION

In this paper, we propose *vector quantized layer* (VQ layer) on QA models, a novel network architecture for the QA task. Based on this idea, we are able to understand how machine *thinks* essentially without losing much performance. For the VQ layer we use in this work, it can be put on any layers of any QA models as long as we need to analyze it. After coloring those quantized vectors, it is pretty easy for people to observe the action of different models. Besides, based on the experiments, we have found some interesting phenomena. For the pointer networks we utilize in this paper, we successfully find that models themselves indeed use some codewords to mark the boundary of answer spans. Also, for some keywords such as *VERB*, *NOUN*, and *PROPN*, the models will use some nonsense tokens to mask them so as not to mislead the latter layers. Finally yet importantly, QA models actually do not need that huge capacity to reach high performance. That is, it can merely use 20 or so codewords in a layer to propagate all information.

REFERENCES

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015. URL http://arxiv.org/abs/1409.0473.

Trapit Bansal, Arvind Neelakantan, and Andrew McCallum. Relnet: End-to-end modeling of entities and relations. In *AKBC*, 2017.

A. Bordes, N. Usunier, S. Chopra, and J. Weston. Large-scale simple question answering with memory networks. *CoRR*, abs/1506.02075, 2015.

Eunsol Choi, He He, Mohit Iyyer, Mark Yatskar, Wen-tau Yih, Yejin Choi, Percy Liang, and Luke Zettlemoyer. Quac: Question answering in context. *arXiv preprint arXiv:1808.07036*, 2018.

Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*, 2014.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

Yanzhuo Ding, Yang Liu, Huanbo Luan, and Maosong Sun. Visualizing and understanding neural machine translation. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1150–1159, Vancouver, Canada, July 2017. Association for Computational Linguistics. doi: 10.18653/v1/P17-1106. URL https://www.aclweb.org/anthology/P17-1106.

Finale Doshi-Velez and Been Kim. Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*, 2017.

Shi Feng, Eric Wallace, Mohit Iyyer, Pedro Rodriguez, Alvin Grissom II, and Jordan L. Boyd-Graber. Right answer for the wrong reason: Discovery and mitigation. *CoRR*, abs/1804.07781, 2018. URL http://arxiv.org/abs/1804.07781.

Mikael Henaff, Jason Weston, Arthur Szlam, Antoine Bordes, and Yann LeCun. Tracking the world state with recurrent entity networks. In *ICLR*, 2017.

K. M. Hermann, T. Kociský, E. Grefenstette, L. Espeholt, W. Kay, M. Suleyman, and P. Blunsom. Teaching machines to read and comprehend. *CoRR*, abs/1506.03340, 2015.

Minghao Hu, Yuxing Peng, and Xipeng Qiu. Reinforced mnemonic reader for machine comprehension. *CoRR, abs/1705.02798*, 2017.

Hsin-Yuan Huang, Chenguang Zhu, Yelong Shen, and Weizhu Chen. Fusionnet: Fusing via fully-aware attention with application to machine comprehension. *arXiv preprint arXiv:1711.07341*, 2017.

Hsin-Yuan Huang, Eunsol Choi, and Wen-tau Yih. Flowqa: Grasping flow in history for conversational machine comprehension. *arXiv preprint arXiv:1810.06683*, 2018.

Ákos Kádár, Grzegorz Chrupala, and Afra Alishahi. Representation of linguistic form and function in recurrent neural networks. *CoRR*, abs/1602.08952, 2016. URL http://arxiv.org/abs/1602.08952.

A. Kumar, O. Irsoy, J. Su, J. Bradbury, R. English, B. Pierce, P. Ondruska, I. Gulrajani, and R. Socher. Ask me anything: Dynamic memory networks for natural language processing. *CoRR*, abs/1506.07285, 2015.

Nelson F. Liu, Omer Levy, Roy Schwartz, Chenhao Tan, and Noah A. Smith. Lstms exploit linguistic attributes of data. *CoRR*, abs/1805.11653, 2018. URL http://arxiv.org/abs/1805.11653.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*, 2019.

Grégoire Montavon, Wojciech Samek, and Klaus-Robert Müller. Methods for interpreting and understanding deep neural networks. *Digital Signal Processing*, 73:1–15, 2018.

Rasmus Berg Palm, Ulrich Paquet, and Ole Winther. Recurrent relational networks. In *submitted to NIPS*, 2018.

Juan Pavez, Hector Allende, and Hector Allende-Cid. Working memory networks: Augmenting memory networks with a relational reasoning module. In *ACL*, 2018.

Baolin Peng, Zhengdong Lu, Hang Li, and Kam-Fai Wong. Towards neural network-based reasoning. *CoRR*, abs/1508.05508, 2015.

Matthew E Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. Deep contextualized word representations. *arXiv preprint arXiv:1802.05365*, 2018.

Pranav Rajpurkar, Robin Jia, and Percy Liang. Know what you don't know: Unanswerable questions for squad. *arXiv preprint arXiv:1806.03822*, 2018.

Siva Reddy, Danqi Chen, and Christopher D Manning. Coqa: A conversational question answering challenge. *Transactions of the Association for Computational Linguistics*, 7:249–266, 2019.

Adam Santoro, David Raposo, David G.T. Barrett, Mateusz Malinowski, Razvan Pascanu, Peter Battaglia, and Timothy Lillicrap. A simple neural network module for relational reasoning. In *NIPS*, 2017.

Minjoon Seo, Aniruddha Kembhavi, Ali Farhadi, and Hannaneh Hajishirzi. Bidirectional attention flow for machine comprehension. *arXiv preprint arXiv:1611.01603*, 2016a.

Minjoon Seo, Sewon Min, Ali Farhadi, and Hannaneh Hajishirzi. Query-reduction networks for question answering. *arXiv preprint arXiv:1606.04582*, 2016b.

Kevin J. Shih, Saurabh Singh, and Derek Hoiem. Where to look: Focus regions for visual question answering. *CoRR*, abs/1511.07394, 2015.

Aaron van den Oord, Oriol Vinyals, et al. Neural discrete representation learning. In *Advances in Neural Information Processing Systems*, pp. 6306–6315, 2017.

Elena Voita, Pavel Serdyukov, Rico Sennrich, and Ivan Titov. Context-aware neural machine translation learns anaphora resolution. *CoRR*, abs/1805.10163, 2018. URL `http://arxiv.org/abs/1805.10163`.

Elena Voita, David Talbot, Fedor Moiseev, Rico Sennrich, and Ivan Titov. Analyzing multi-head self-attention: Specialized heads do the heavy lifting, the rest can be pruned. *CoRR*, abs/1905.09418, 2019. URL `http://arxiv.org/abs/1905.09418`.

Wenhui Wang, Nan Yang, Furu Wei, Baobao Chang, and Ming Zhou. Gated self-matching networks for reading comprehension and question answering. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pp. 189–198, 2017.

Jason Weston, Sumit Chopra, and Antoine Bordes. Memory networks. *arXiv preprint arXiv:1410.3916*, 2014.

Jason Weston, Antoine Bordes, Sumit Chopra, Alexander M Rush, Bart van Merriënboer, Armand Joulin, and Tomas Mikolov. Towards ai-complete question answering: A set of prerequisite toy tasks. *arXiv preprint arXiv:1502.05698*, 2015.

C. Xiong, S. Merity, and R. Socher. Dynamic memory networks for visual and textual question answering. *CoRR*, abs/1603.01417, 2016.

Caiming Xiong, Victor Zhong, and Richard Socher. Dynamic coattention networks for question answering. In *ICLR*, 2017.

Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Ruslan Salakhutdinov, and Quoc V Le. Xlnet: Generalized autoregressive pretraining for language understanding. *arXiv preprint arXiv:1906.08237*, 2019.

Zhuosheng Zhang, Yuwei Wu, Hai Zhao, Zuchao Li, Shuailiang Zhang, Xi Zhou, and Xiang Zhou. Semantics-aware bert for language understanding. *arXiv preprint arXiv:1909.02209*, 2019a.

Zhuosheng Zhang, Yuwei Wu, Junru Zhou, Sufeng Duan, and Hai Zhao. Sg-net: Syntax-guided machine reading comprehension. *arXiv preprint arXiv:1908.05147*, 2019b.

Chenguang Zhu, Michael Zeng, and Xuedong Huang. Sdnet: Contextualized attention-based deep network for conversational question answering. *arXiv preprint arXiv:1812.03593*, 2018.

A IMPLEMENTATION DETAILS

A.1 FLOWQA

Our modified FlowQA is shown in Figure 6. In this task, each context and a series of questions are first encoded with GloVe[2] and Elmo (Peters et al., 2018). Then, we keep both the bidirectional GRU (Chung et al., 2014) layers, which are used to encode context, and unidirectional GRU layers, which are used to propagate information between different questions, intact. Then, we add an extra codebook with size K and use it to quantize the output of the last flow layer. With the help of this quantization mechanism, we are able to know the reasoning process of this model more easily.

Thus, the output shape of the previous layer is $[L_C, N_Q, H]$, where L_C stands for the max length of context, N_Q is the number of question turns, and H represents the hidden dimension. Then for each H-dimensional vector, we can always find one vector in the codebook, which is composed of K H-dimensional codewords, to replace it, letting it be the input of the next layer.

Finally, as for the value of β in VQ-VAE, we assign 1.0 on it instead of the default setting 0.25.

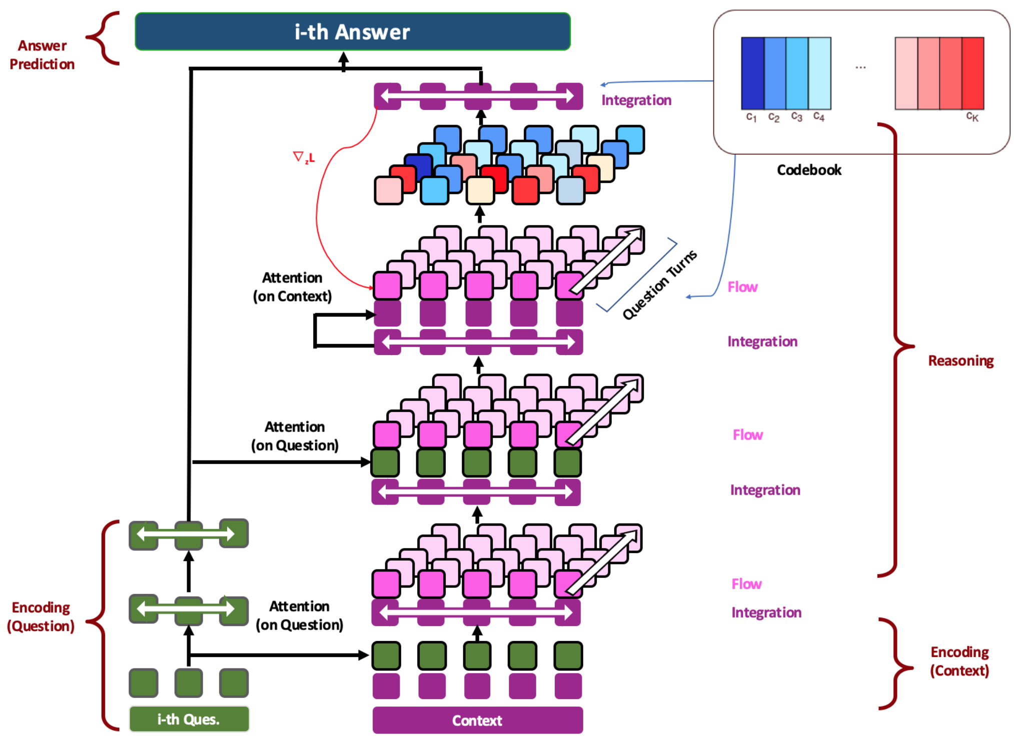


Figure 6: FlowQA with vector quantized layer.

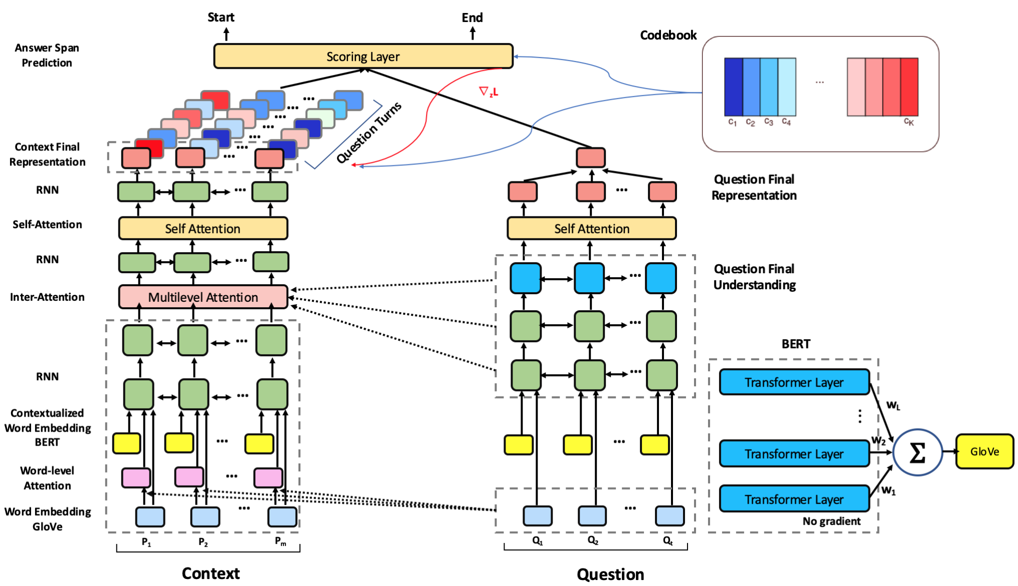
A.2 SDNET

Our modified 2 versions of SDNet are shown in Figure 7(a) and Figure 7(b) respectively. In this task, each context and a series of questions are encoded with Glove and BERT (Devlin et al., 2018). Instead of using another dimensional flow to propagating information between different questions, SDNet here utilizes the power of attention. We keep all layers intact except for the *Context Final Representation* layer. Then a size K codebook is also added here. Note that for the original model (Zhu et al., 2018), the author sets the number of the question turns as batch size for each training iteration and we simply follow this setting.

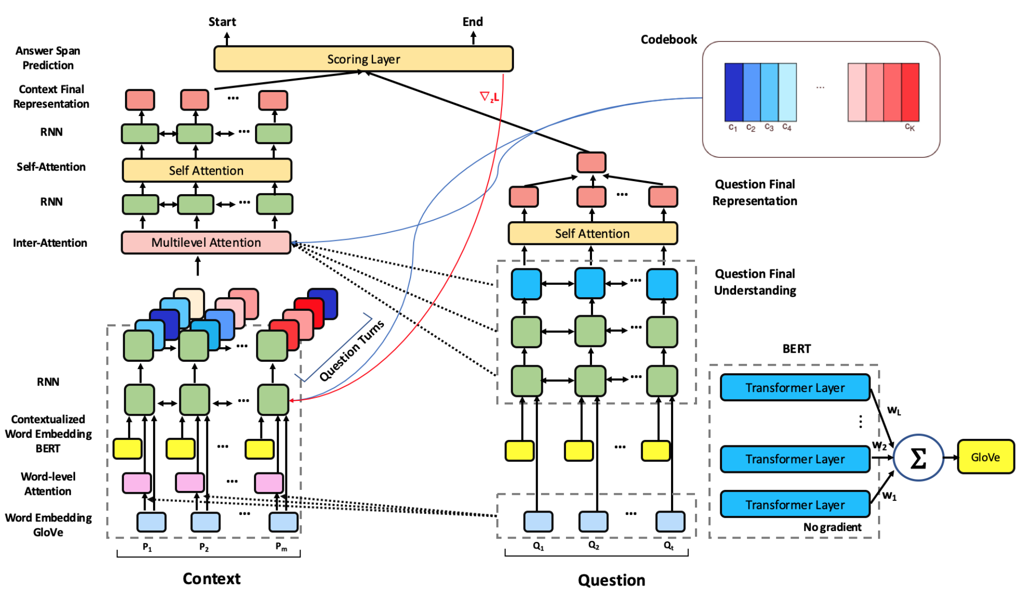
[2]https://github.com/stanfordnlp/GloVe

As a result, the output shape of the previous RNN layer should be $[L_C, N_Q, H]$, where L_C stands for the max length of context, N_Q stands for the number of question turns, and H represents the hidden dimension.

Afterward, merely do the same operations as we did for FlowQA: finding one vector in the codebook to replace the original one, feeding them into the next layer, then dealing with the gradient issue. Lastly, we also use QA loss, VQ loss, and commitment loss to train this vector quantized SDNet. Note that the value of β is set to 1.0 here as well.



(a) Version 1



(b) Version 2

Figure 7: SDNet with vector quantized layer.

B Extra Experiment Results

B.1 Average Position of Different Codewords in FlowQA

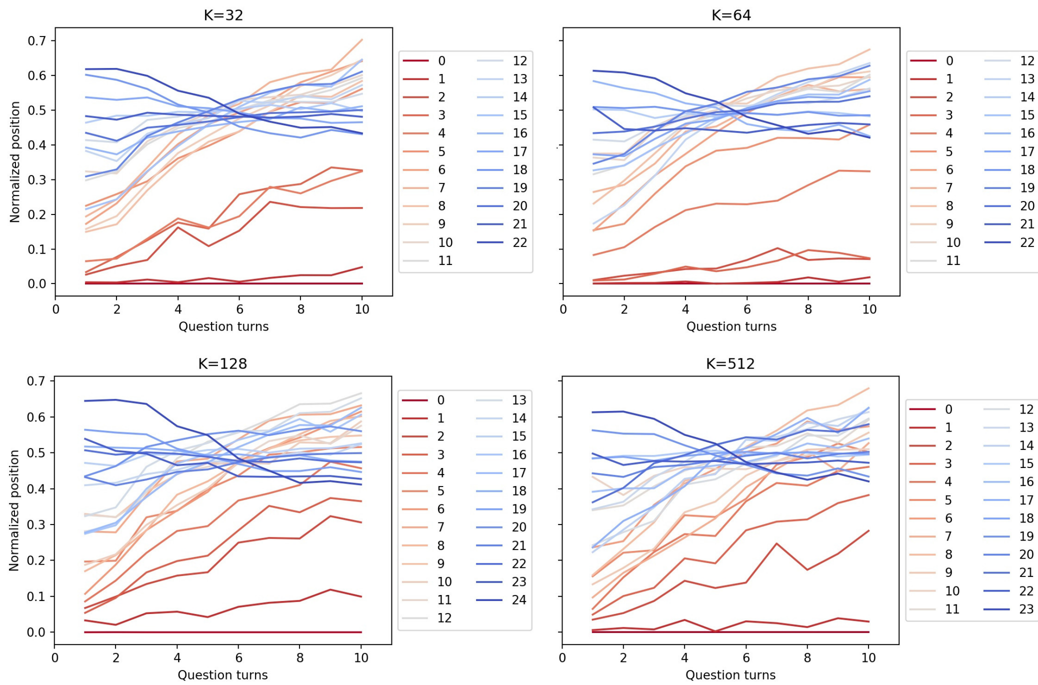


Figure 8: Average position of different codewords in FlowQA.

B.2 Nonsense words in SDNet(v1) and SDNet(v2)

Table 4: Nonsense words in SDNet(v1).

SDNet(v1)			
K	Overall	Model Output	Subject of Model Output
512	0.6996	0.5947	0.9654
128	0.5661	0.5139	0.5776
64	0.7849	0.5321	0.7633
32	0.6623	0.4490	0.6930
16	0.6578	0.5139	0.7456
8	0.8905	0.5367	0.8458
4	0.9626	0.5947	0.9654
Average	0.7463	0.5189	0.7538

B.3 Answerability

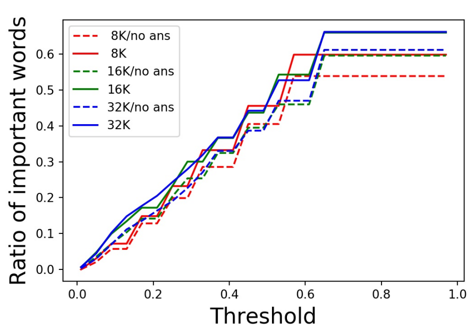
In QuAC, models are asked to whether the question is answerable or not, just like those in SQuAD 2.0 (Rajpurkar et al., 2018). We found that typically, rare codewords' percentages when the question is not answerable is lower than the percentages when the question is answerable. We use Figure 9 to show. The lines represent answerable questions while dotted lines are for not answerable ones. As for the meaning of the threshold on the horizontal axis, it means that the percentage we define codewords as a meaningful codeword . For instance,1.0 means that all codewords are meaningful.

Table 5: Nonsense words in SDNet(v2) layer 1.

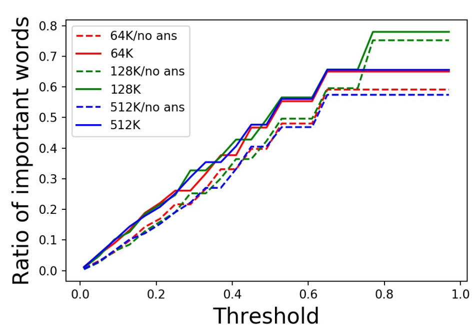
SDNet(v2) layer 1			
K	Overall	Model Output	Subject of Model Output
512	0.0518	0.0536	0.2183
128	0.1316	0.1656	0.4106
64	0.1984	0.2115	0.4731
32	0.4373	0.3807	0.5752
16	0.5208	0.5707	0.5255
8	0.9998	1.0	1.0
4	0.9999	0.9999	0.9994
Average	0.4771	0.4831	0.6003

Table 6: Nonsense words in SDNet(v2) layer 2.

SDNet(v2) layer 2			
K	Overall	Model Output	Subject of Model Output
512	0.0405	0.0347	0.1404
128	0.0988	0.0951	0.1589
64	0.1989	0.1535	0.2857
32	0.3071	0.2298	0.3397
16	0.6608	0.3514	0.6287
8	0.7080	0.5498	0.7998
4	0.8090	0.8019	0.7966
Average	0.4033	0.3166	0.4450



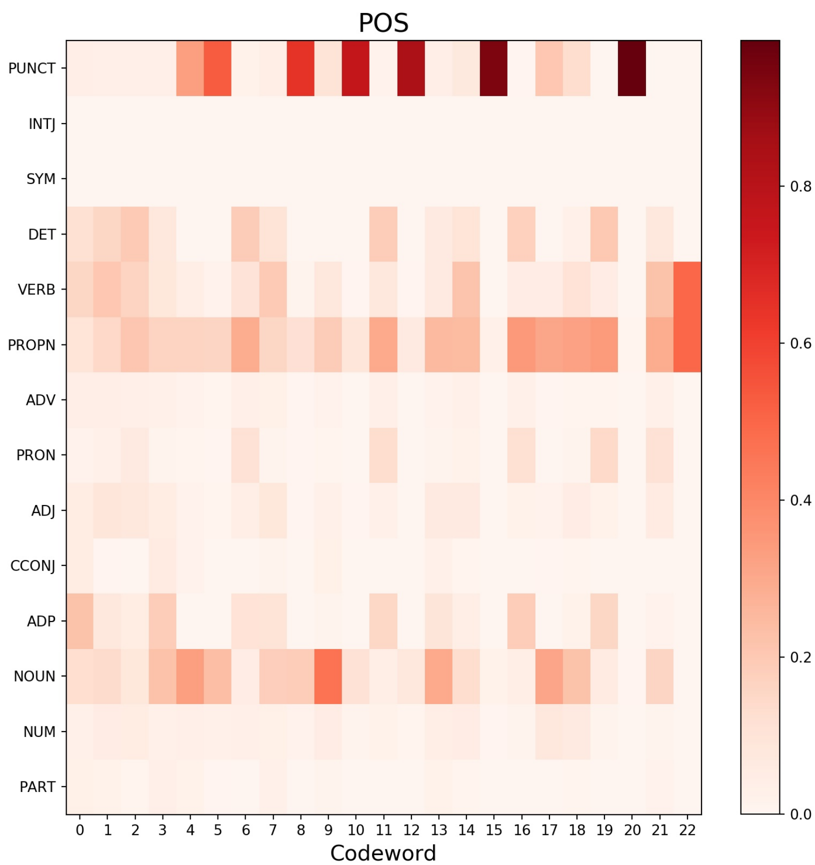
(a) Smaller codebook size



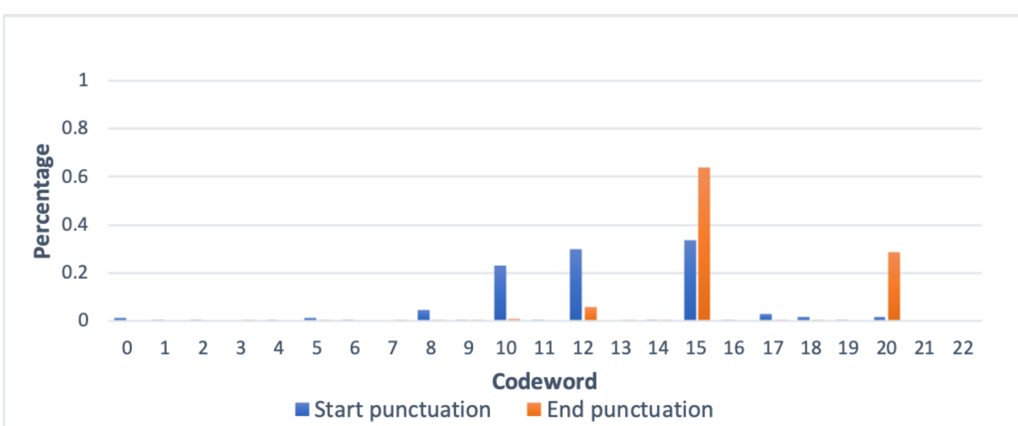
(b) Bigger codebook size

Figure 9: Influence of important words on answerability.

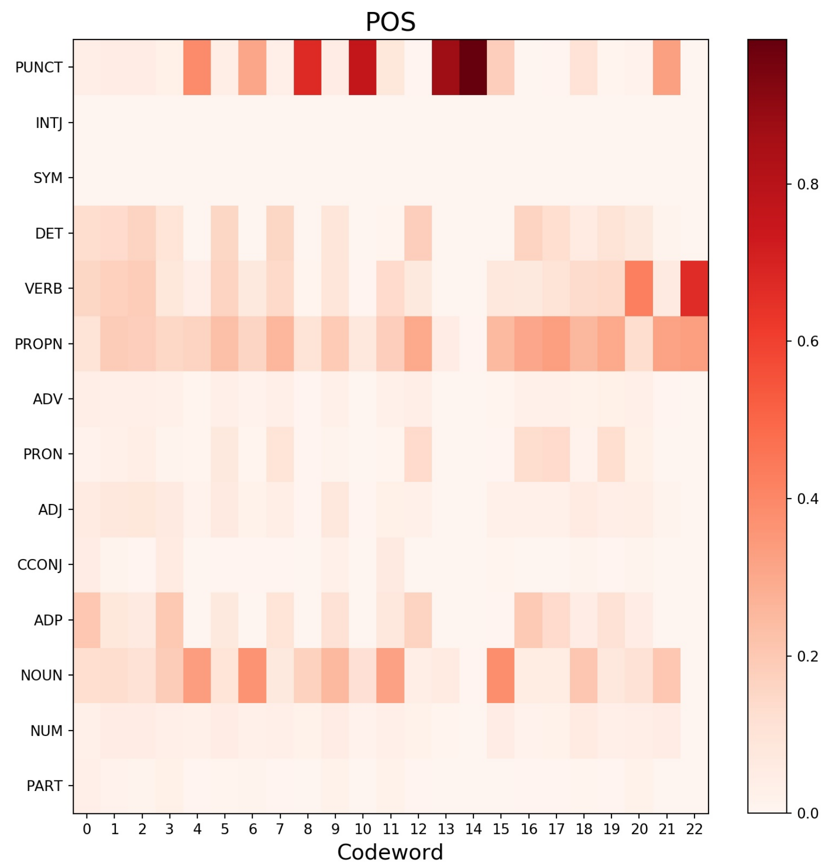
B.4 PART-OF-SPEECH



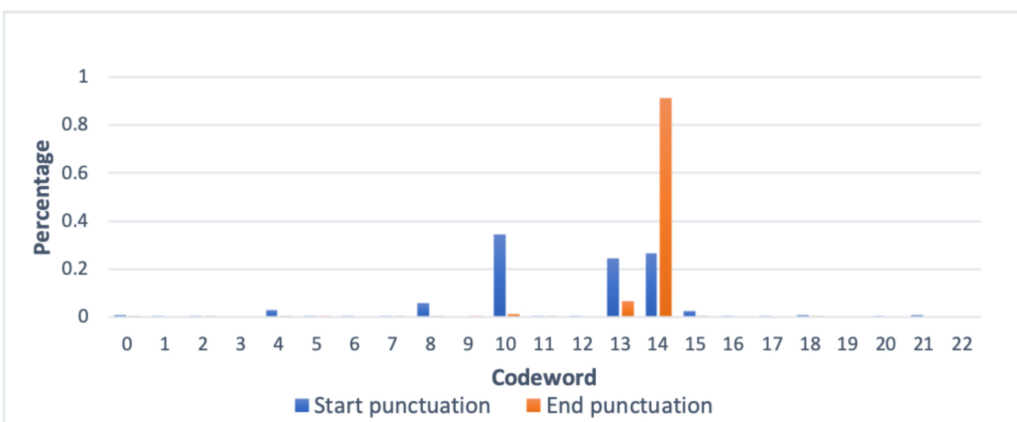
(a) POS-codeword heat map (K=32)



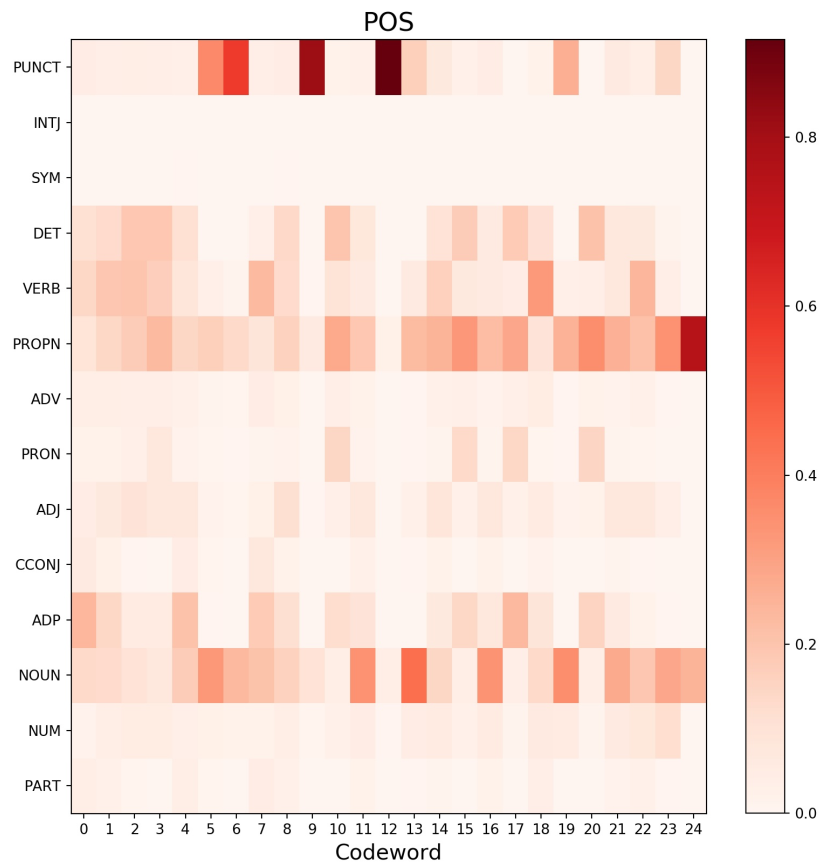
(b) Codewords distribution (K=32)



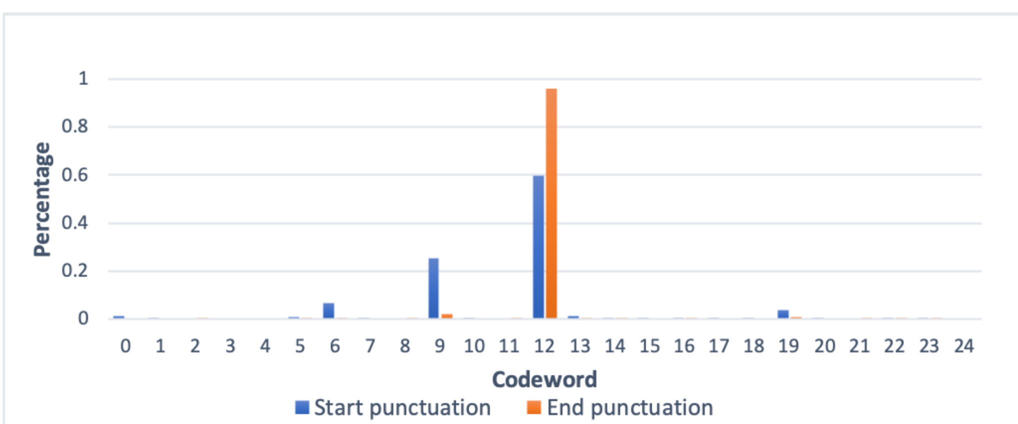
(c) POS-codeword heat map (K=64)



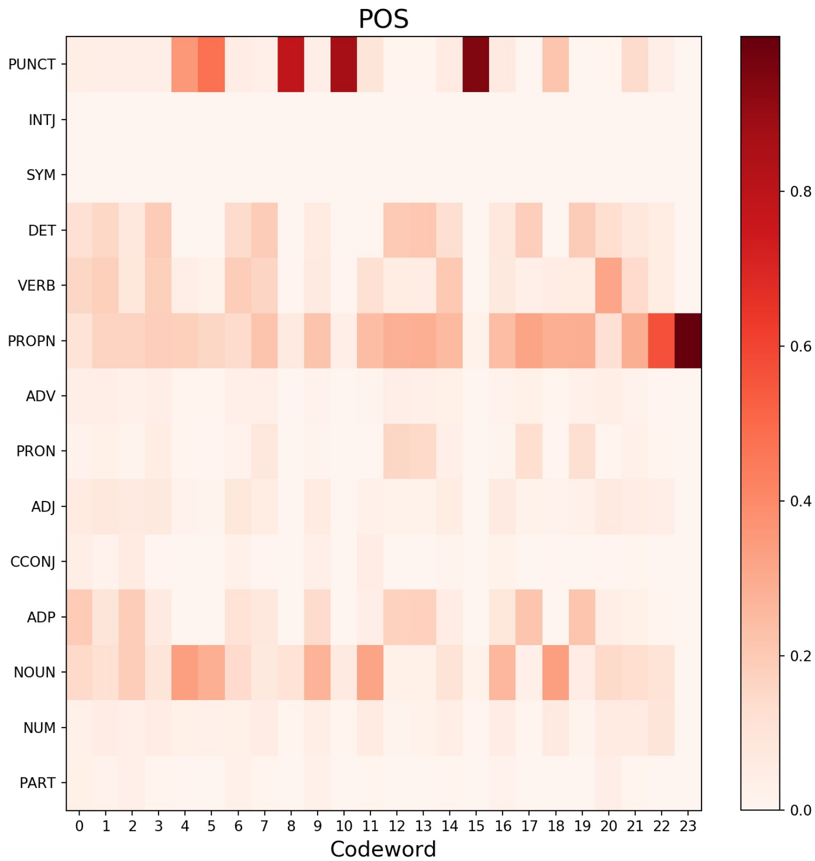
(d) Codewords distribution (K=64)



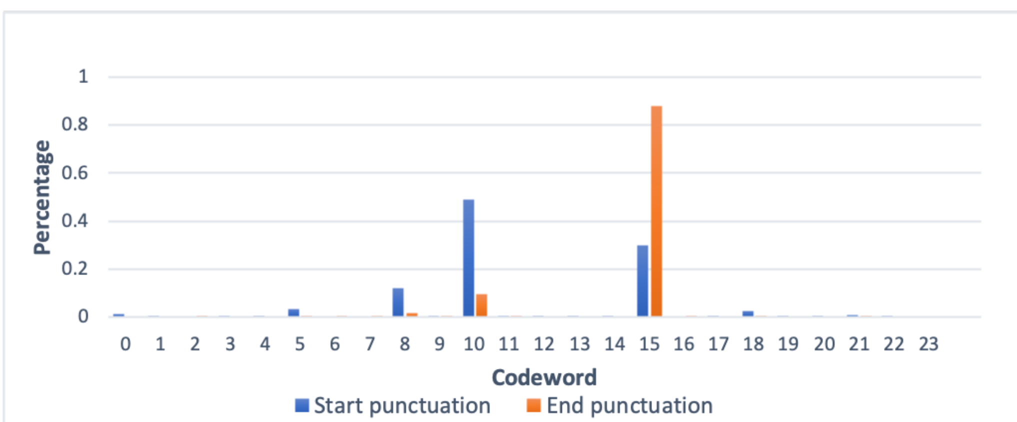
(e) POS-codeword heat map (K=128)



(f) Codewords distribution (K=128)



(g) POS heat map



(h) K=512

Figure 10: Codebooks distribution on start/end punctuation in FlowQA (K=32, 64, 128, 512).

B.5 VISUALIZATION EXAMPLES OF VECTOR QUANTIZED LATENT REPRESENTATIONS

In 2003 , Richie and her then - best friend Paris Hilton starred in the reality series The Simple Life , in which they lived for a month with a family in the rural community of Altus , Arkansas . The series premiere debuted on Fox on December 2 , 2003 , to 13 million viewers , increasing Fox 's reach among viewers 18 - 49 . The series continued for a second and third season . Fox subsequently canceled the show after a dispute between Hilton and Richie , but it was aired by E ! Entertainment Television for its fourth season , drawing nearly one million viewers and was followed by a relatively unsuccessful fifth season . Complications arose during production of the fifth season , with both Richie and Hilton facing major charges for DUI , and at risk of serving jail time . Hilton was sentenced to 23 days in jail , though this was after production of the season had finished . Richie accepted a plea agreement and was sentenced to four days in jail , again after production had wrapped . Despite talks of a sixth season , the series finished its run at the end of the moderately unsuccessful fifth season . In 2005 , Richie made her acting debut in the comedic drama Kids in America . Richie has guest starred in television series including Eve , Six Feet Under and American Dreams , 8 Simple Rules for Dating My Teenage Daughter , and playing the recurring character Heather Chandler on Chuck . In July 2008 , plans were announced to turn Richie 's 2005 novel , The Truth About Diamonds , into a television series . On January 1 , 2010 , it was announced that Richie would appear as a guest judge on Project Runway . In early 2016 , Richie auditioned for a role in the NBC sitcom , Great News , created by Tracey Wigfield and produced by Wigfield and Tina Fey , which is set in the world of television news . Although Richie 's audition was not successful , she was later brought in to replace Kimrie Lewis - Davis in the supporting role of Portia , an anchorwoman at the news channel depicted in the series . The show 's producers had initially cast Lewis - Davis as Portia , but decided to go in a different direction with the character after seeing the results of the pilot episode . When speaking about casting Richie in her first scripted series - regular television role , Fey explained : " She 's a really funny person and has an instant likability . Those are instincts you ca n't teach . People have it or not . She has good timing without really pushing it . " CANNOTANSWER

(a) The 1st question turn (What shows did she appear on?)

In 2003 , Richie and her then - best friend Paris Hilton starred in the reality series The Simple Life , in which they lived for a month with a family in the rural community of Altus , Arkansas . The series premiere debuted on Fox on December 2 , 2003 , to 13 million viewers , increasing Fox 's reach among viewers 18 - 49 . The series continued for a second and third season . Fox subsequently canceled the show after a dispute between Hilton and Richie , but it was aired by E ! Entertainment Television for its fourth season , drawing nearly one million viewers and was followed by a relatively unsuccessful fifth season . Complications arose during production of the fifth season , with both Richie and Hilton facing major charges for DUI , and at risk of serving jail time . Hilton was sentenced to 23 days in jail , though this was after production of the season had finished . Richie accepted a plea agreement and was sentenced to four days in jail , again after production had wrapped . Despite talks of a sixth season , the series finished its run at the end of the moderately unsuccessful fifth season . In 2005 , Richie made her acting debut in the comedic drama Kids in America . Richie has guest starred in television series including Eve , Six Feet Under and American Dreams , 8 Simple Rules for Dating My Teenage Daughter , and playing the recurring character Heather Chandler on Chuck . In July 2008 , plans were announced to turn Richie 's 2005 novel , The Truth About Diamonds , into a television series . On January 1 , 2010 , it was announced that Richie would appear as a guest judge on Project Runway . In early 2016 , Richie auditioned for a role in the NBC sitcom , Great News , created by Tracey Wigfield and produced by Wigfield and Tina Fey , which is set in the world of television news . Although Richie 's audition was not successful , she was later brought in to replace Kimrie Lewis - Davis in the supporting role of Portia , an anchorwoman at the news channel depicted in the series . The show 's producers had initially cast Lewis - Davis as Portia , but decided to go in a different direction with the character after seeing the results of the pilot episode . When speaking about casting Richie in her first scripted series - regular television role , Fey explained : " She 's a really funny person and has an instant likability . Those are instincts you ca n't teach . People have it or not . She has good timing without really pushing it . " CANNOTANSWER

(b) The 2nd question turn (What was the show about?)

In 2003 , Richie and her then - best friend Paris Hilton starred in the reality series The Simple Life , in which they lived for a month with a family in the rural community of Altus , Arkansas . The series premiere debuted on Fox on December 2 , 2003 , to 13 million viewers , increasing Fox 's reach among viewers 18 - 49 . The series continued for a second and third season . Fox subsequently canceled the show after a dispute between Hilton and Richie , but it was aired by E ! Entertainment Television for its fourth season , drawing nearly one million viewers and was followed by a relatively unsuccessful fifth season . Complications arose during production of the fifth season , with both Richie and Hilton facing major charges for DUI , and at risk of serving jail time . Hilton was sentenced to 23 days in jail , though this was after production of the season had finished . Richie accepted a plea agreement and was sentenced to four days in jail , again after production had wrapped . Despite talks of a sixth season , the series finished its run at the end of the moderately unsuccessful fifth season . In 2005 , Richie made her acting debut in the comedic drama Kids in America . Richie has guest starred in television series including Eve , Six Feet Under and American Dreams , 8 Simple Rules for Dating My Teenage Daughter , and playing the recurring character Heather Chandler on Chuck . In July 2008 , plans were announced to turn Richie 's 2005 novel , The Truth About Diamonds , into a television series . On January 1 , 2010 , it was announced that Richie would appear as a guest judge on Project Runway . In early 2016 , Richie auditioned for a role in the NBC sitcom , Great News , created by Tracey Wigfield and produced by Wigfield and Tina Fey , which is set in the world of television news . Although Richie 's audition was not successful , she was later brought in to replace Kimrie Lewis - Davis in the supporting role of Portia , an anchorwoman at the news channel depicted in the series . The show 's producers had initially cast Lewis - Davis as Portia , but decided to go in a different direction with the character after seeing the results of the pilot episode . When speaking about casting Richie in her first scripted series - regular television role , Fey explained : " She 's a really funny person and has an instant likability . Those are instincts you ca n't teach . People have it or not . She has good timing without really pushing it . " CANNOTANSWER

(c) The 3rd question turn (Was it popular?)

In 2003 , Richie and her then - best friend Paris Hilton starred in the reality series The Simple Life , in which they lived for a month with a family in the rural community of Altus , Arkansas . The series premiere debuted on Fox on December 2 , 2003 , to 13 million viewers , increasing Fox 's reach among viewers 18 - 49 . The series continued for a second and third season . Fox subsequently canceled the show after a dispute between Hilton and Richie , but it was aired by E ! Entertainment Television for its fourth season , drawing nearly one million viewers and was followed by a relatively unsuccessful fifth season . Complications arose during production of the fifth season , with both Richie and Hilton facing major charges for DUI , and at risk of serving jail time . Hilton was sentenced to 23 days in jail , though this was after production of the season had finished . Richie accepted a plea agreement and was sentenced to four days in jail , again after production had wrapped . Despite talks of a sixth season , the series finished its run at the end of the moderately unsuccessful fifth season . In 2005 , Richie made her acting debut in the comedic drama Kids in America . Richie has guest starred in television series including Eve , Six Feet Under and American Dreams , 8 Simple Rules for Dating My Teenage Daughter , and playing the recurring character Heather Chandler on Chuck . In July 2008 , plans were announced to turn Richie 's 2005 novel , The Truth About Diamonds , into a television series . On January 1 , 2010 , it was announced that Richie would appear as a guest judge on Project Runway . In early 2016 , Richie auditioned for a role in the NBC sitcom , Great News , created by Tracey Wigfield and produced by Wigfield and Tina Fey , which is set in the world of television news . Although Richie 's audition was not successful , she was later brought in to replace Kimrie Lewis - Davis in the supporting role of Portia , an anchorwoman at the news channel depicted in the series . The show 's producers had initially cast Lewis - Davis as Portia , but decided to go in a different direction with the character after seeing the results of the pilot episode . When speaking about casting Richie in her first scripted series - regular television role , Fey explained : " She 's a really funny person and has an instant likability . Those are instincts you ca n't teach . People have it or not . She has good timing without really pushing it . " CANNOTANSWER

(d) The 4th question turn (How many seasons did it have?)

In 2003 , Richie and her then - best friend Paris Hilton starred in the reality series The Simple Life , in which they lived for a month with a family in the rural community of Altus , Arkansas . The series premiere debuted on Fox on December 2 , 2003 , to 13 million viewers , increasing Fox 's reach among viewers 18 - 49 . The series continued for a second and third season . Fox subsequently canceled the show after a dispute between Hilton and Richie , but it was aired by E ! Entertainment Television for its fourth season , drawing nearly one million viewers and was followed by a relatively unsuccessful fifth season . Complications arose during production of the fifth season , with both Richie and Hilton facing major charges for DUI , and at risk of serving jail time . Hilton was sentenced to 23 days in jail , though this was after production of the season had finished . Richie accepted a plea agreement and was sentenced to four days in jail , again after production had wrapped . Despite talks of a sixth season , the series finished its run at the end of the moderately unsuccessful fifth season . In 2005 , Richie made her acting debut in the comedic drama Kids in America . Richie has guest starred in television series including Eve , Six Feet Under and American Dreams , 8 Simple Rules for Dating My Teenage Daughter , and playing the recurring character Heather Chandler on Chuck . In July 2008 , plans were announced to turn Richie 's 2005 novel , The Truth About Diamonds , into a television series . On January 1 , 2010 , it was announced that Richie would appear as a guest judge on Project Runway . In early 2016 , Richie auditioned for a role in the NBC sitcom , Great News , created by Tracey Wigfield and produced by Wigfield and Tina Fey , which is set in the world of television news . Although Richie 's audition was not successful , she was later brought in to replace Kimrie Lewis - Davis in the supporting role of Portia , an anchorwoman at the news channel depicted in the series . The show 's producers had initially cast Lewis - Davis as Portia , but decided to go in a different direction with the character after seeing the results of the pilot episode . When speaking about casting Richie in her first scripted series - regular television role , Fey explained : " She 's a really funny person and has an instant likability . Those are instincts you ca n't teach . People have it or not . She has good timing without really pushing it . " CANNOTANSWER

(e) The 5th question turn (Did she have any other shows?)

In 2003 , Richie and her then - best friend Paris Hilton starred in the reality series The Simple Life , in which they lived for a month with a family in the rural community of Altus , Arkansas . The series premiere debuted on Fox on December 2 , 2003 , to 13 million viewers , increasing Fox 's reach among viewers 18 - 49 . The series continued for a second and third season . Fox subsequently canceled the show after a dispute between Hilton and Richie , but it was aired by E ! Entertainment Television for its fourth season , drawing nearly one million viewers and was followed by a relatively unsuccessful fifth season . Complications arose during production of the fifth season , with both Richie and Hilton facing major charges for DUI , and at risk of serving jail time . Hilton was sentenced to 23 days in jail , though this was after production of the season had finished . Richie accepted a plea agreement and was sentenced to four days in jail , again after production had wrapped . Despite talks of a sixth season , the series finished its run at the end of the moderately unsuccessful fifth season . In 2005 , Richie made her acting debut in the comedic drama Kids in America . Richie has guest starred in television series including Eve , Six Feet Under and American Dreams , 8 Simple Rules for Dating My Teenage Daughter , and playing the recurring character Heather Chandler on Chuck . In July 2008 , plans were announced to turn Richie 's 2005 novel , The Truth About Diamonds , into a television series . On January 1 , 2010 , it was announced that Richie would appear as a guest judge on Project Runway . In early 2016 , Richie auditioned for a role in the NBC sitcom , Great News , created by Tracey Wigfield and produced by Wigfield and Tina Fey , which is set in the world of television news . Although Richie 's audition was not successful , she was later brought in to replace Kimrie Lewis - Davis in the supporting role of Portia , an anchorwoman at the news channel depicted in the series . The show 's producers had initially cast Lewis - Davis as Portia , but decided to go in a different direction with the character after seeing the results of the pilot episode . When speaking about casting Richie in her first scripted series - regular television role , Fey explained : " She 's a really funny person and has an instant likability . Those are instincts you ca n't teach . People have it or not . She has good timing without really pushing it . " CANNOTANSWER

(f) The 6th question turn (Was she a main role?)

In 2003 , Richie and her then - best friend Paris Hilton starred in the reality series The Simple Life , in which they lived for a month with a family in the rural community of Altus , Arkansas . The series premiere debuted on Fox on December 2 , 2003 , to 13 million viewers , increasing Fox 's reach among viewers 18 - 49 . The series continued for a second and third season . Fox subsequently canceled the show after a dispute between Hilton and Richie , but it was aired by E ! Entertainment Television for its fourth season , drawing nearly one million viewers and was followed by a relatively unsuccessful fifth season . Complications arose during production of the fifth season , with both Richie and Hilton facing major charges for DUI , and at risk of serving jail time . Hilton was sentenced to 23 days in jail , though this was after production of the season had finished . Richie accepted a plea agreement and was sentenced to four days in jail , again after production had wrapped . Despite talks of a sixth season , the series finished its run at the end of the moderately unsuccessful fifth season . In 2005 , Richie made her acting debut in the comedic drama Kids in America . Richie has guest starred in television series including Eve , Six Feet Under and American Dreams , 8 Simple Rules for Dating My Teenage Daughter , and playing the recurring character Heather Chandler on Chuck . In July 2008 , plans were announced to turn Richie 's 2005 novel , The Truth About Diamonds , into a television series . On January 1 , 2010 , it was announced that Richie would appear as a guest judge on Project Runway . In early 2016 , Richie auditioned for a role in the NBC sitcom , Great News , created by Tracey Wigfield and produced by Wigfield and Tina Fey , which is set in the world of television news . Although Richie 's audition was not successful , she was later brought in to replace Kimrie Lewis - Davis in the supporting role of Portia , an anchorwoman at the news channel depicted in the series . The show 's producers had initially cast Lewis - Davis as Portia , but decided to go in a different direction with the character after seeing the results of the pilot episode . When speaking about casting Richie in her first scripted series - regular television role , Fey explained : " She 's a really funny person and has an instant likability . Those are instincts you ca n't teach . People have it or not . She has good timing without really pushing it . " CANNOTANSWER

(g) The 7th question turn (What other shows did she play?)

In 2003 , Richie and her then - best friend Paris Hilton starred in the reality series The Simple Life , in which they lived for a month with a family in the rural community of Altus , Arkansas . The series premiere debuted on Fox on December 2 , 2003 , to 13 million viewers , increasing Fox 's reach among viewers 18 - 49 . The series continued for a second and third season . Fox subsequently canceled the show after a dispute between Hilton and Richie , but it was aired by E ! Entertainment Television for its fourth season , drawing nearly one million viewers and was followed by a relatively unsuccessful fifth season . Complications arose during production of the fifth season , with both Richie and Hilton facing major charges for DUI , and at risk of serving jail time . Hilton was sentenced to 23 days in jail , though this was after production of the season had finished . Richie accepted a plea agreement and was sentenced to four days in jail , again after production had wrapped . Despite talks of a sixth season , the series finished its run at the end of the moderately unsuccessful fifth season . In 2005 , Richie made her acting debut in the comedic drama Kids in America . Richie has guest starred in television series including Eve , Six Feet Under and American Dreams , 8 Simple Rules for Dating My Teenage Daughter , and playing the recurring character Heather Chandler on Chuck . In July 2008 , plans were announced to turn Richie 's 2005 novel , The Truth About Diamonds , into a television series . On January 1 , 2010 , it was announced that Richie would appear as a guest judge on Project Runway . In early 2016 , Richie auditioned for a role in the NBC sitcom , Great News , created by Tracey Wigfield and produced by Wigfield and Tina Fey , which is set in the world of television news . Although Richie 's audition was not successful , she was later brought in to replace Kimrie Lewis - Davis in the supporting role of Portia , an anchorwoman at the news channel depicted in the series . The show 's producers had initially cast Lewis - Davis as Portia , but decided to go in a different direction with the character after seeing the results of the pilot episode . When speaking about casting Richie in her first scripted series - regular television role , Fey explained : " She 's a really funny person and has an instant likability . Those are instincts you ca n't teach . People have it or not . She has good timing without really pushing it . " CANNOTANSWER

(h) The 8th question turn (Is she still acting?)

In 2003 , Richie and her then - best friend Paris Hilton starred in the reality series The Simple Life , in which they lived for a month with a family in the rural community of Altus , Arkansas . The series premiere debuted on Fox on December 2 , 2003 , to 13 million viewers , increasing Fox 's reach among viewers 18 - 49 . The series continued for a second and third season . Fox subsequently canceled the show after a dispute between Hilton and Richie , but it was aired by E ! Entertainment Television for its fourth season , drawing nearly one million viewers and was followed by a relatively unsuccessful fifth season . Complications arose during production of the fifth season , with both Richie and Hilton facing major charges for DUI , and at risk of serving jail time . Hilton was sentenced to 23 days in jail , though this was after production of the season had finished . Richie accepted a plea agreement and was sentenced to four days in jail , again after production had wrapped . Despite talks of a sixth season , the series finished its run at the end of the moderately unsuccessful fifth season . In 2005 , Richie made her acting debut in the comedic drama Kids in America . Richie has guest starred in television series including Eve , Six Feet Under and American Dreams , 8 Simple Rules for Dating My Teenage Daughter , and playing the recurring character Heather Chandler on Chuck . In July 2008 , plans were announced to turn Richie 's 2005 novel , The Truth About Diamonds , into a television series . On January 1 , 2010 , it was announced that Richie would appear as a guest judge on Project Runway . In early 2016 , Richie auditioned for a role in the NBC sitcom , Great News , created by Tracey Wigfield and produced by Wigfield and Tina Fey , which is set in the world of television news . Although Richie 's audition was not successful , she was later brought in to replace Kimrie Lewis - Davis in the supporting role of Portia , an anchorwoman at the news channel depicted in the series . The show 's producers had initially cast Lewis - Davis as Portia , but decided to go in a different direction with the character after seeing the results of the pilot episode . When speaking about casting Richie in her first scripted series - regular television role , Fey explained : " She 's a really funny person and has an instant likability . Those are instincts you ca n't teach . People have it or not . She has good timing without really pushing it . " CANNOTANSWER

(i) The 9th question turn (Is there anything else interesting?)

Figure 11: Example 1 of FlowQA.

Feller was elected the inaugural president of the Major League Baseball Players ' Association in 1956 . As president , he appeared before Congress to speak about baseball 's reserve clause . Feller was the first player to get a franchise to agree to a share of game receipts when he was the starting pitcher for Indians ' games . He was also the earliest player to incorporate himself (as Ro - Fel , Inc.) . He was also one of the first players to work for the right of a player to enter free agency . Feller and his wife , Virginia Winther , had three sons , Steve , Martin , and Bruce . The couple divorced in 1971 ; from the divorce settlement , Virginia received the house she and Feller had built . Virginia died on May 6 , 1981 in her home in Shaker Heights , Ohio . In retirement , Feller lived with his second wife , Anne Feller , in Gates Mills , a suburb of Cleveland . Feller is credited with being the first baseball star to sign autographs at baseball memorabilia conventions , and was such a frequent guest at such events that one ESPN writer speculated that he may have signed more autographs than any other person . In June 2009 , at the age of 90 , Feller was one of the starting pitchers at the inaugural Baseball Hall of Fame Classic , which replaced the Hall of Fame Game at Cooperstown , New York . Feller was treated for leukemia in August 2010 . By October , Feller was fitted with a pacemaker and was diagnosed with pneumonia and thrush , an infection of the mucous membrane lining the mouth and throat . He was transferred on December 8 from the Cleveland Clinic to hospice care . On December 15 , Feller died of complications from leukemia at 92 . CANNOTANSWER

(a) The 1st question turn (What was he doing in his later life?)

Feller was elected the inaugural president of the Major League Baseball Players ' Association in 1956 . As president , he appeared before Congress to speak about baseball 's reserve clause . Feller was the first player to get a franchise to agree to a share of game receipts when he was the starting pitcher for Indians ' games . He was also the earliest player to incorporate himself (as Ro - Fel , Inc.) . He was also one of the first players to work for the right of a player to enter free agency . Feller and his wife , Virginia Winther , had three sons , Steve , Martin , and Bruce . The couple divorced in 1971 ; from the divorce settlement , Virginia received the house she and Feller had built . Virginia died on May 6 , 1981 in her home in Shaker Heights , Ohio . In retirement , Feller lived with his second wife , Anne Feller , in Gates Mills , a suburb of Cleveland . Feller is credited with being the first baseball star to sign autographs at baseball memorabilia conventions , and was such a frequent guest at such events that one ESPN writer speculated that he may have signed more autographs than any other person . In June 2009 , at the age of 90 , Feller was one of the starting pitchers at the inaugural Baseball Hall of Fame Classic , which replaced the Hall of Fame Game at Cooperstown , New York . Feller was treated for leukemia in August 2010 . By October , Feller was fitted with a pacemaker and was diagnosed with pneumonia and thrush , an infection of the mucous membrane lining the mouth and throat . He was transferred on December 8 from the Cleveland Clinic to hospice care . On December 15 , Feller died of complications from leukemia at 92 . CANNOTANSWER

(b) The 2nd question turn (What did he do as president of the association?)

Feller was elected the inaugural president of the Major League Baseball Players ' Association in 1956 . As president , he appeared before Congress to speak about baseball 's reserve clause . Feller was the first player to get a franchise to agree to a share of game receipts when he was the starting pitcher for Indians ' games . He was also the earliest player to incorporate himself (as Ro - Fel , Inc.) . He was also one of the first players to work for the right of a player to enter free agency . Feller and his wife , Virginia Winther , had three sons , Steve , Martin , and Bruce . The couple divorced in 1971 ; from the divorce settlement , Virginia received the house she and Feller had built . Virginia died on May 6 , 1981 in her home in Shaker Heights , Ohio . In retirement , Feller lived with his second wife , Anne Feller , in Gates Mills , a suburb of Cleveland . Feller is credited with being the first baseball star to sign autographs at baseball memorabilia conventions , and was such a frequent guest at such events that one ESPN writer speculated that he may have signed more autographs than any other person . In June 2009 , at the age of 90 , Feller was one of the starting pitchers at the inaugural Baseball Hall of Fame Classic , which replaced the Hall of Fame Game at Cooperstown , New York . Feller was treated for leukemia in August 2010 . By October , Feller was fitted with a pacemaker and was diagnosed with pneumonia and thrush , an infection of the mucous membrane lining the mouth and throat . He was transferred on December 8 from the Cleveland Clinic to hospice care . On December 15 , Feller died of complications from leukemia at 92 . CANNOTANSWER

(c) The 3rd question turn (What other things was he doing?)

Feller was elected the inaugural president of the Major League Baseball Players ' Association in 1956 . As president , he appeared before Congress to speak about baseball 's reserve clause . Feller was the first player to get a franchise to agree to a share of game receipts when he was the starting pitcher for Indians ' games . He was also the earliest player to incorporate himself (as Ro - Fel , Inc.) . He was also one of the first players to work for the right of a player to enter free agency . Feller and his wife , Virginia Winther , had three sons , Steve , Martin , and Bruce . The couple divorced in 1971 ; from the divorce settlement , Virginia received the house she and Feller had built . Virginia died on May 6 , 1981 in her home in Shaker Heights , Ohio . In retirement , Feller lived with his second wife , Anne Feller , in Gates Mills , a suburb of Cleveland . Feller is credited with being the first baseball star to sign autographs at baseball memorabilia conventions , and was such a frequent guest at such events that one ESPN writer speculated that he may have signed more autographs than any other person . In June 2009 , at the age of 90 , Feller was one of the starting pitchers at the inaugural Baseball Hall of Fame Classic , which replaced the Hall of Fame Game at Cooperstown , New York . Feller was treated for leukemia in August 2010 . By October , Feller was fitted with a pacemaker and was diagnosed with pneumonia and thrush , an infection of the mucous membrane lining the mouth and throat . He was transferred on December 8 from the Cleveland Clinic to hospice care . On December 15 , Feller died of complications from leukemia at 92 . CANNOTANSWER

(d) The 4th question turn (Was he the first to do anything else?)

Feller was elected the inaugural president of the Major League Baseball Players ' Association in 1956 . As president , he appeared before Congress to speak about baseball 's reserve clause . Feller was the first player to get a franchise to agree to a share of game receipts when he was the starting pitcher for Indians ' games . He was also the earliest player to incorporate himself (as Ro - Fel , Inc.) . He was also one of the first players to work for the right of a player to enter free agency . Feller and his wife , Virginia Winther , had three sons , Steve , Martin , and Bruce . The couple divorced in 1971 ; from the divorce settlement , Virginia received the house she and Feller had built . Virginia died on May 6 , 1981 in her home in Shaker Heights , Ohio . In retirement , Feller lived with his second wife , Anne Feller , in Gates Mills , a suburb of Cleveland . Feller is credited with being the first baseball star to sign autographs at baseball memorabilia conventions , and was such a frequent guest at such events that one ESPN writer speculated that he may have signed more autographs than any other person . In June 2009 , at the age of 90 , Feller was one of the starting pitchers at the inaugural Baseball Hall of Fame Classic , which replaced the Hall of Fame Game at Cooperstown , New York . Feller was treated for leukemia in August 2010 . By October , Feller was fitted with a pacemaker and was diagnosed with pneumonia and thrush , an infection of the mucous membrane lining the mouth and throat . He was transferred on December 8 from the Cleveland Clinic to hospice care . On December 15 , Feller died of complications from leukemia at 92 . CANNOTANSWER

(e) The 5th question turn (Did he have other firsts?)

Feller was elected the inaugural president of the Major League Baseball Players ' Association in 1956 . As president , he appeared before Congress to speak about baseball 's reserve clause . Feller was the first player to get a franchise to agree to a share of game receipts when he was the starting pitcher for Indians ' games . He was also the earliest player to incorporate himself (as Ro - Fel , Inc.) . He was also one of the first players to work for the right of a player to enter free agency . Feller and his wife , Virginia Winther , had three sons , Steve , Martin , and Bruce . The couple divorced in 1971 ; from the divorce settlement , Virginia received the house she and Feller had built . Virginia died on May 6 , 1981 in her home in Shaker Heights , Ohio . In retirement , Feller lived with his second wife , Anne Feller , in Gates Mills , a suburb of Cleveland . Feller is credited with being the first baseball star to sign autographs at baseball memorabilia conventions , and was such a frequent guest at such events that one ESPN writer speculated that he may have signed more autographs than any other person . In June 2009 , at the age of 90 , Feller was one of the starting pitchers at the inaugural Baseball Hall of Fame Classic , which replaced the Hall of Fame Game at Cooperstown , New York . Feller was treated for leukemia in August 2010 . By October , Feller was fitted with a pacemaker and was diagnosed with pneumonia and thrush , an infection of the mucous membrane lining the mouth and throat . He was transferred on December 8 from the Cleveland Clinic to hospice care . On December 15 , Feller died of complications from leukemia at 92 . CANNOTANSWER

(f) The 6th question turn (What other things were going on in his later life?)

Feller was elected the inaugural president of the Major League Baseball Players ' Association in 1956 . As president , he appeared before Congress to speak about baseball 's reserve clause . Feller was the first player to get a franchise to agree to a share of game receipts when he was the starting pitcher for Indians ' games . He was also the earliest player to incorporate himself (as Ro - Fel , Inc.) . He was also one of the first players to work for the right of a player to enter free agency . Feller and his wife , Virginia Winther , had three sons , Steve , Martin , and Bruce . The couple divorced in 1971 ; from the divorce settlement , Virginia received the house she and Feller had built . Virginia died on May 6 , 1981 in her home in Shaker Heights , Ohio . In retirement , Feller lived with his second wife , Anne Feller , in Gates Mills , a suburb of Cleveland . Feller is credited with being the first baseball star to sign autographs at baseball memorabilia conventions , and was such a frequent guest at such events that one ESPN writer speculated that he may have signed more autographs than any other person . In June 2009 , at the age of 90 , Feller was one of the starting pitchers at the inaugural Baseball Hall of Fame Classic , which replaced the Hall of Fame Game at Cooperstown , New York . Feller was treated for leukemia in August 2010 . By October , Feller was fitted with a pacemaker and was diagnosed with pneumonia and thrush , an infection of the mucous membrane lining the mouth and throat . He was transferred on December 8 from the Cleveland Clinic to hospice care . On December 15 , Feller died of complications from leukemia at 92 . CANNOTANSWER

(g) The 7th question turn (Was he married for many years?)

Feller was elected the inaugural president of the Major League Baseball Players ' Association in 1956 . As president , he appeared before Congress to speak about baseball 's reserve clause . Feller was the first player to get a franchise to agree to a share of game receipts when he was the starting pitcher for Indians ' games . He was also the earliest player to incorporate himself (as Ro - Fel , Inc.) . He was also one of the first players to work for the right of a player to enter free agency . Feller and his wife , Virginia Winther , had three sons , Steve , Martin , and Bruce . The couple divorced in 1971 ; from the divorce settlement , Virginia received the house she and Feller had built . Virginia died on May 6 , 1981 in her home in Shaker Heights , Ohio . In retirement , Feller lived with his second wife , Anne Feller , in Gates Mills , a suburb of Cleveland . Feller is credited with being the first baseball star to sign autographs at baseball memorabilia conventions , and was such a frequent guest at such events that one ESPN writer speculated that he may have signed more autographs than any other person . In June 2009 , at the age of 90 , Feller was one of the starting pitchers at the inaugural Baseball Hall of Fame Classic , which replaced the Hall of Fame Game at Cooperstown , New York . Feller was treated for leukemia in August 2010 . By October , Feller was fitted with a pacemaker and was diagnosed with pneumonia and thrush , an infection of the mucous membrane lining the mouth and throat . He was transferred on December 8 from the Cleveland Clinic to hospice care . On December 15 , Feller died of complications from leukemia at 92 . CANNOTANSWER

(h) The 8th question turn (Did he re-marry?)

Feller was elected the inaugural president of the Major League Baseball Players ' Association in 1956 . As president , he appeared before Congress to speak about baseball 's reserve clause . Feller was the first player to get a franchise to agree to a share of game receipts when he was the starting pitcher for Indians ' games . He was also the earliest player to incorporate himself (as Ro - Fel , Inc.) . He was also one of the first players to work for the right of a player to enter free agency . Feller and his wife , Virginia Winther , had three sons , Steve , Martin , and Bruce . The couple divorced in 1971 ; from the divorce settlement , Virginia received the house she and Feller had built . Virginia died on May 6 , 1981 in her home in Shaker Heights , Ohio . In retirement , Feller lived with his second wife , Anne Feller , in Gates Mills , a suburb of Cleveland . Feller is credited with being the first baseball star to sign autographs at baseball memorabilia conventions , and was such a frequent guest at such events that one ESPN writer speculated that he may have signed more autographs than any other person . In June 2009 , at the age of 90 , Feller was one of the starting pitchers at the inaugural Baseball Hall of Fame Classic , which replaced the Hall of Fame Game at Cooperstown , New York . Feller was treated for leukemia in August 2010 . By October , Feller was fitted with a pacemaker and was diagnosed with pneumonia and thrush , an infection of the mucous membrane lining the mouth and throat . He was transferred on December 8 from the Cleveland Clinic to hospice care . On December 15 , Feller died of complications from leukemia at 92 . CANNOTANSWER

(i) The 9th question turn (Did he continue working?)

Figure 12: Example 2 of FlowQA.