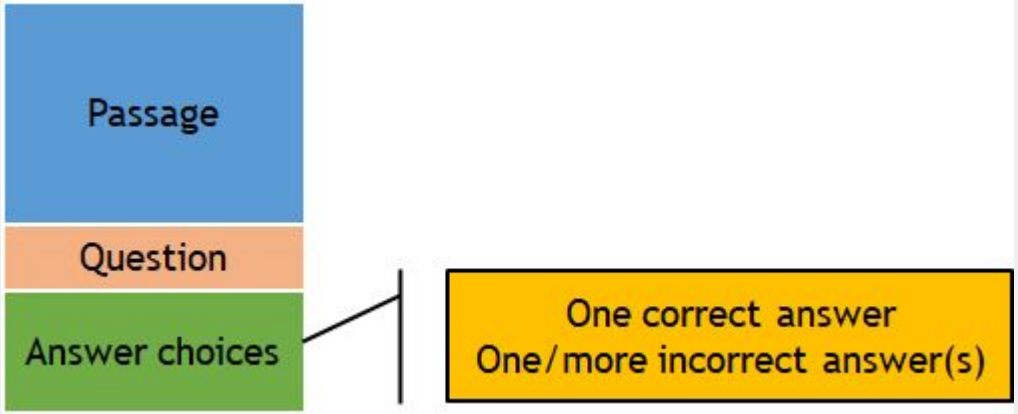
CIKM 2020

Learning to Distract: A Hierarchical Multi-Decoder Network for Automated Generation of Long Distractors for Multiple-Choice Questions for Reading Comprehension

Kaushal Kumar Maurya and Maunendra Sankar Desarkar
Department of Computer Science and Engineering
Indian Institute of Technology Hyderabad (IITH)

Task Definition

Distractor Generation: The task of generating *incorrect options* for reading comprehension MCQ.



Task: Generation of long, coherent and grammatically correct wrong options given a triplet $<$*article*, *question*, *correct answer*$>$.

Task Definition

Distractor Generation: The task of generating *incorrect options* for reading comprehension MCQ.

Task: Generation of long, coherent and grammatically correct wrong options given a triplet $<$*article*, *question*, *correct answer*$>$.

Considerations while generating distractors

Generated distractors

- Should be *in the context* of the question
- Should be *semantically related* to the answer

- Should not be *semantically equivalent* to the answer
- Should not be *exactly* same with each other
- Should not be *very* different from each other.

- Generate all distractors *simultaneously*

Applications:

1. The distractor generation system can be utilized for educational purposes in language *learning assessment*

2. As reverse task, the system can also be used to automatically create *annotated* datasets to push research in reading comprehension and Q&A systems [1]

3. The variant of the proposed model can be used to generate different utterances in *conversational* systems

Problems with Existing Approaches:

1. Existing methods use Jaccard similarity over a pool of candidate distractors to sample the distractors.
   - This often makes the generated distractors *too* obvious or *not* relevant to the question context [2].

2. Some approaches did not consider the answer in the model.
   - This caused the generated distractors to be either *answer-revealing* or semantically *equivalent* to the answer [3].

Imitating Human Approach:

Two step approach:

1. Search for article sentences that are in context with the question

2. Avoid sentences which are semantically equivalent to the answer.

The resultant sentences are potential candidates for distractor generation.

Our Contribution:

1. We propose a novel **H**ierarchical **M**ulti-**D**ecoder **Net**work (HMD-Net) to tackle the task of automated distractor generation

2. We release a new high-quality distractor generation dataset RACE++ DG prepared from RACE++ dataset by leveraging contextual similarities

3. We introduce a novel *dis-similarity* loss in HMD-Net for distractor generation and a new BERT [4] cosine similarity (BERT-CS) based metric for automated evaluation.

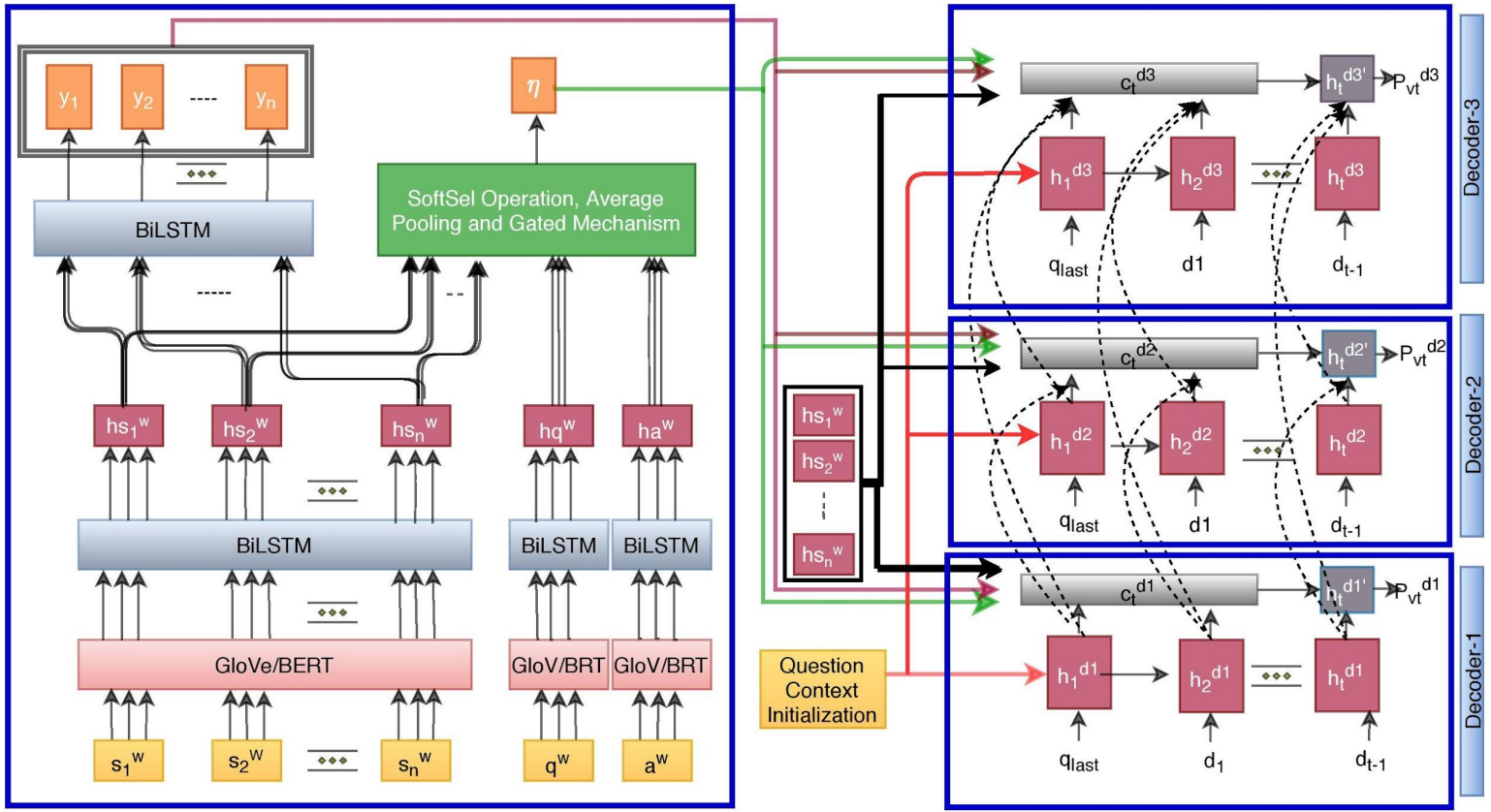
Problem Statement:

Our aim to generate D_1 , D_2 and D_3 given the triplet $<S, Q, A>$

$$D_i = \arg\max_{\bar{D}_i} logP(\bar{D}_i|S, Q, A; \theta_i)$$

Where, $P(\bar{D}_i|S, Q, A; \theta_i)$ conditional log-likelihood of ith distractor

Architectural Diagram of HMD-Net:



Distractor Hierarchical Encoder

Distractor Multi-Decoder

Distractor Hierarchical Encoder:

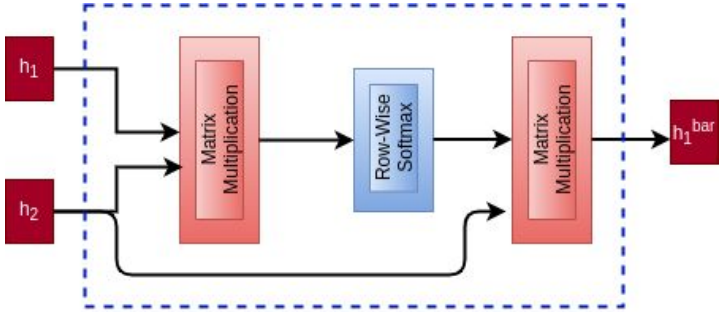
Input: Triplet <article, question, answer>
Output: Article sentence representation, Each token representation of components of triplet, and softsel matching score (SSMS)

Steps:

1. Softsel operation [4] for Evidence Encoding
2. Average pooling (AP)
3. Gated Contextual Representation(RCR) [4]
4. Softsel Matching Score

Softsel Operations:

1. It encodes the most relevant aspects of a sequence to another sequence.
2. The input to SoftSel operation are two sequences, and output is an encoded sequence



SoftSel Operation

Here, h_1 and h_2 are input sequence. For example h_1 can be sequence of question tokens representation

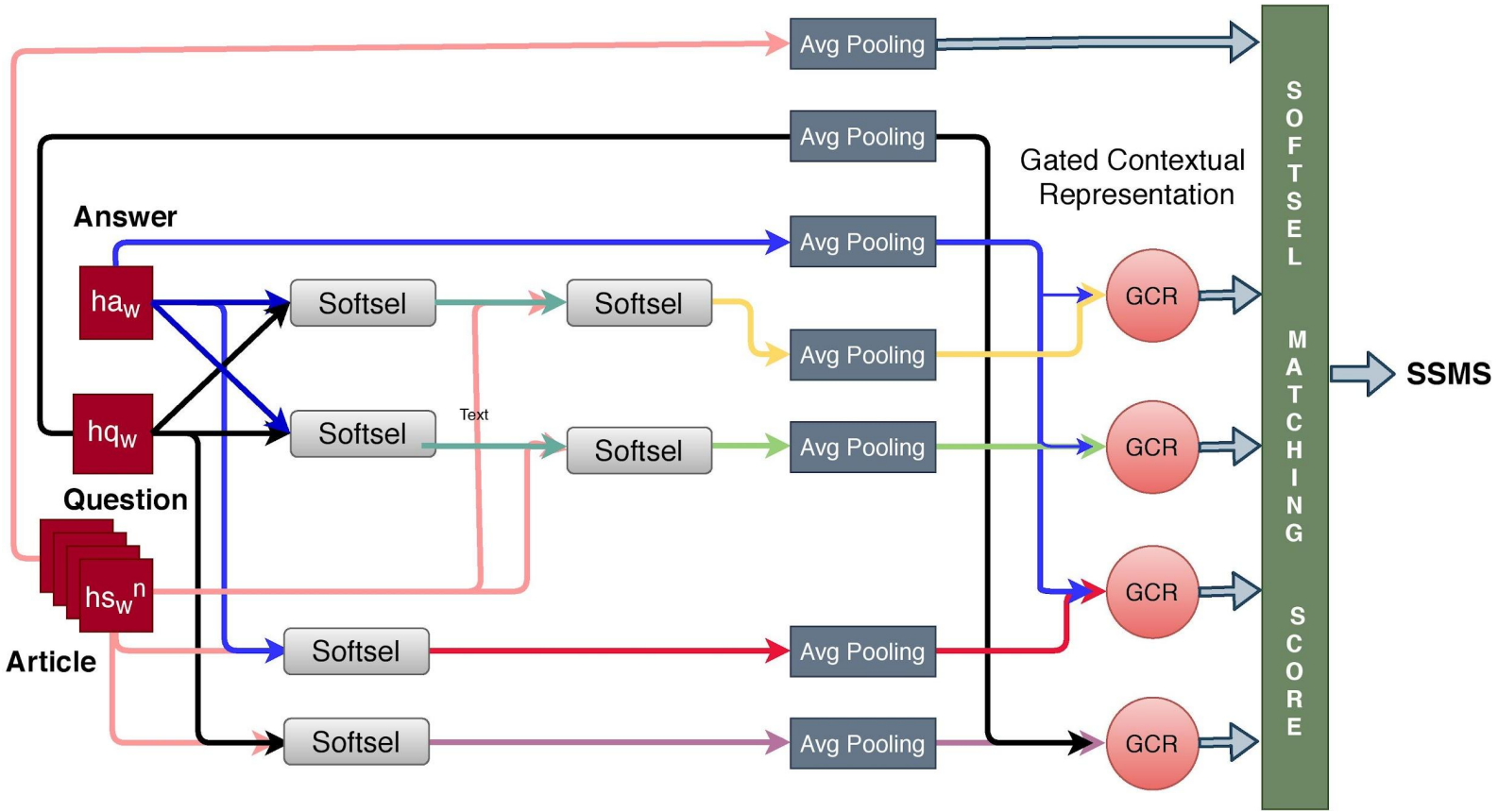
Three Steps:

1. **Cartesian Similarity:** It is obtained for given two input sequences h_1 and h_2 across all possible states (i.e. token's representation)

2. **Row-wise Softmax:** Applied *softmax* over rows of cartesian similarity scores.

3. **Weighted Sum:** A *weighted sum* of second sequence h_2 is encoded at given state of first sequence h_1. For given state of first sequence this representation encodes the *most influential* parts of the second sequence.

Encoder Flow Diagram:



Softsel Matching Score:

$$m_i = \lambda_q s_i^T W_z q_F - \lambda_a (s_i^T W_z a_F + s_i^T W_z q a_F + s_i^T W_z a q_F) + b_z$$

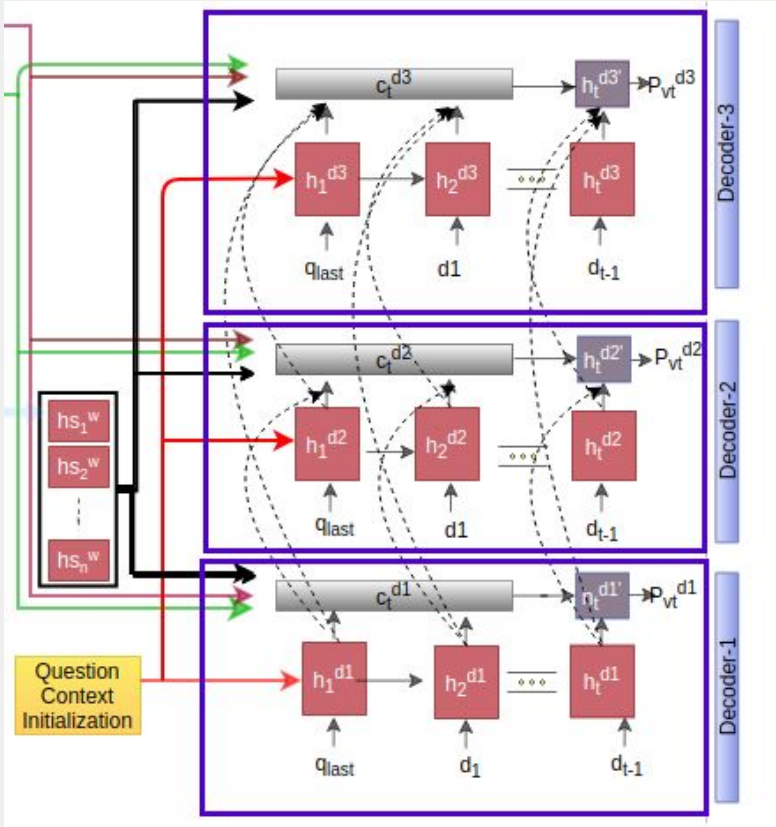
- We used three evidence encoding representations for answer and one for question to ensure that the generated distractors should not be semantically equivalent to the answer
- m_i is score for i^{th} sentence of the article which indicate that how potential i^{th} sentence is for distractor generation

Distractor Hierarchical Multi-Decoder:

1. **Input:** SSMS and article
2. **Output:** Three distractors
3. Utilized hierarchical articale sentence representation

Question Context Initialization:

1. Used a separate uni-directional LSTM layer to encode the question
2. Use the last token of question i.e., q_{last}
3. Employed final cell state and hidden state of LSTM to initialize each decoder



1. For given decoder, the learned and penalise the attention scores for other decoders. For example attention equation for decoder three is:

$$att3 = att3 - (\lambda_1 * att1) - (\lambda_2 * att2)$$

2. Used three attention scores
   i). α: standard word-attention scores
   ii). β: Sentence attention score
   iii). η: SSMS

final attention score $$\bar{\alpha}_{i,j}^{d_k} = \frac{\alpha_{i,j}^{d_k} \beta_i^{d_k} \eta_i}{\sum_{i,j} \alpha_{i,j}^{d_k} \beta_i^{d_k} \eta_i}$$

Dis-Similarity Function [3]:

1. Feed ground truth distractor to uni-directional LSTM and find last hidden state h_{gt}
2. Collected the last hidden state representation from each decoder i.e. h_{d1}, h_{d2} and h_{d3}
3. Find a cosine similarity score

$$ds_i = \cos(h_{gt}, h_{di})$$

Training Loss:

$$L = \sum_{k=1}^{3} \left(- \sum_{D_k \in V} logP(D_k | S, Q, A; \theta_k) - \lambda_{ds} * (1 - ds_k) \right)$$

Datasets:

1. Used two datasets: RACE DG [2] and RACE ++ DG
2. RACE DG was available where RACE++ DG has been prepared by us
3. RACE++ DG preparation Steps

i). Removed distractors like distractors like 'all of the above,' 'option a and option b are correct, etc.

ii). Distractor, question, and answer should have a minimum length of three.

iii). Removed questions with fill in the blanks are at the beginning or in the middle of the question.

iv). Used BERT cosine similarity to extract semantically relevant triplet and distractors

Parameters	RACE DG	RACE++ DG
Total no. of train samples	96501	135321
Total no. of dev samples	12089	16915
Total no. of test samples	12284	16915
Avg. article length (tokens)	342.0	342.3
Avg. question length	9.76	10.9
Avg. answer length	8.63	8.00
Avg. distractor length	8.48	7.68
Avg. sentences length (in article)	19.9	19.6
Avg. no. of distractors per triplet	2.1	2.3

Sample Data Records (From RACE DG):

Article-1	Article-2
The healthy habits survey shows that only about one-third of american seniors have correct habits. Here are some findings and expert advice. 1. how many times did you brush your teeth yesterday? Finding: a full 33 % of seniors brush their teeth only once a day. step: remove the 300 types of bacteria in your mouth each morning with a battery operated toothbrush. Brush gently for 2 minutes, at least twice a day. 2. how many times did you wash your hands or bathe yesterday? Finding: seniors, on average, bathe fewer than 3 days a week. And nearly 30% wash their hands only 4 times a day- half of the number doctors recommend. step: We touch our faces around 3,000 times a day- often inviting germs to enter our mouth, nose, and eyes. Use toilet paper to avoid touching the door handle. And, most important, wash your hands often with hot running water and soap for 20 seconds. 3. how often do you think about fighting germs? Finding: Seniors are not fighting germs as well as they should. step: be aware of germs. Do you know it is not your toilet but your kitchen sponge that can carry more germs than anything else? to kill these germs, keep your sponge in the microwave for 10 seconds.	At east china university of science and technology, students will get a coupon if they eat up their food. Students can collect coupons and exchange them for small gifts, such as books, magazines, mobile phone covers and hand warmers. It 's been such a surprise, said liang zahaoyun, 19, a student at the university in shanghai. It has given us one more motivation to finish our food. The measure is part of a national eat-up campaign which is organized by students to deal with food waste on campuses. Why only on campuses, you might ask? Because according to a report by xinhna news agency, students waste twice as much food as the national average. The campaign on campus food waste is receiving attention across the country. The aim of the campaign is not only to encourage students to finish their food. We hope it can also encourage students to choose a more environment - friendly and healthy lifestyle, said tao siliang, secretary of the youth league committee at shanghai university. But some school food is poorly prepared, so students do not like to finish it all. Some schools have taken notice of this and they are taking measures to improve it. I 'm glad that we've reduced food waste since the' eat-up 'campaign began. But if we call on students to waste less food, we should also improve the service and food standard on campuses. Said tao.
Question: Doctors suggest that people should wash their hands **Answer:** Eight times a day **Distractor:** Four times a day	**Question:** The best title for this passage may be **Answer:** Eat - up campaign on campus **Distractor:** reduce waste on campus

Methods Compared :

1. Sequence-to-sequence [6] model
2. Hierarchical to Encode-Decoder (HRED) model [7]
3. Hierarchical Static Attention (HSA) model [2]
4. Hierarchical Co-Attention (HCA) model [3]
5. Static Attn + Multi-Decoder (SMD) model
6. Encoder of HMD-Net + Decoder of HSA (EHMD+DHSA) model

Note: Three other models are implemented by adding linguistic features (LF) and BERT

1. HMD-Net+LF
2. HMD-Net+BERT
3. EHMD+DHSA+BERT

Evaluation Metrics :

Automatic Evaluation Metrics

1. BLEU 1-4 [8]
2. ROUGE-L [9]
3. METEOR [10]
4. Embedding Average [11]
5. Greedy Match [12]
6. Vector Extrema Score [13]
7. BERT-CS

Manual Evaluation Metrics

1. Comparative Study
2. Quantitative Study
   i). Grammatical Correctness (how grammatically correct the distractors are?)
   ii). Distractibility *(how confusing the distractors are?)*

Automatic Evaluation Results (on RACE DG) :

	Models	BLEU-1	BLEU-2	BLEU-3	BLEU-4	ROUGE-L	METEOR	Embd Avg	G. Match	Ext.Score	BERT-CS
1st	Seq2Seq [17]	25.28	12.43	7.12	4.52	13.58	-	-	-	-	-
	HRED* [25]	27.96	14.41	9.05	6.35	14.68	-	-	-	-	-
	HSA* [5]	28.18	14.57	9.19	6.43	14.89	-	-	-	-	-
	HCA [31]	28.65	15.15	9.77	7.01	15.39	-	-	-	-	-
	EHMD+DHSA	28.25	14.52	9.34	6.66	24.03	10.76	0.569 ± 0.00006	2.530 ± 0.0004	0.357 ± 0.00005	0.813
	SMD	28.78	15.60	10.12	7.26	25.59	11.22	0.574 ± 0.0006	2.585 ± 0.0004	0.362 ± 0.00005	0.817
	HMD-Net	29.26	16.16	10.16	**7.66**	25.78	11.58	0.582 ± 0.00006	2.619 ± 0.0004	0.367 ± 0.00005	0.818
	HMD-Net+ LF	29.80	16.31	10.64	7.57	26.31	11.56	0.581 ± 0.00006	2.629 ± 0.0004	0.367 ± 0.00005	0.823
	EHMD+DHSA+BERT	29.44	16.02	10.06	6.6	25.04	11.08	0.586 ± 0.00005	2.610 ± 0.0004	0.364 ± 0.00005	0.823
	HMD-Net+ BERT	**30.99**	**17.30**	**11.09**	7.52	**26.50**	**12.07**	**0.591 ± 0.00005**	**2.667 ± 0.0004**	**0.370 ± 0.00005**	**0.823**
2nd	Seq2Seq [17]	25.13	12.02	6.56	3.93	13.20	-	-	-	-	-
	HRED* [25]	27.85	13.39	7.89	5.22	14.48	-	-	-	-	-
	HSA* [5]	27.85	13.41	7.87	5.17	14.41	-	-	-	-	-
	HCA [31]	27.29	13.57	8.19	5.51	14.85	-	-	-	-	-
	EHMD+DHSA	27.41	13.47	7.96	5.27	22.75	10.41	0.563 ± 0.00006	2.455 ± 0.0004	0.352 ± 0.00005	0.812
	SMD	28.17	14.62	8.96	6.00	24.15	10.82	0.570 ± 0.00006	2.519 ± 0.0004	0.355± 0.00005	0.814
	HMD-Net	28.84	15.06	9.29	6.37	24.79	11.15	0.580± 0.00006	2.591 ± 0.0004	0.364± 0.00005	0.818
	HMD-Net + LF	29.19	15.33	9.34	6.23	24.90	11.27	0.583 ± 0.00006	2.595 ± 0.0004	0.366 ± 0.00005	0.820
	EHMD+DHSA+BERT	30.16	15.9	9.68	6.19	24.05	11.29	0.583 ± 0.00005	2.535 ± 0.0003	0.359 ± 0.00004	0.823
	HMD-Net + BERT	**30.93**	**16.89**	**10.64**	**7.10**	**25.76**	**11.96**	**0.595 ± 0.00005**	**2.646 ± 0.0004**	**0.368 ± 0.00005**	**0.826**
3rd	Seq2Seq [17]	25.34	11.53	5.94	3.33	13.23	-	-	-	-	-
	HRED* [25]	26.73	12.55	7.21	4.58	14.86		-	-	-	-
	HSA* [5]	26.93	12.62	7.25	4.59	14.72	-	-	-	-	-
	HCA [31]	26.64	12.67	7.42	4.88	15.08	-	-	-	-	-
	EHMD+DHSA	26.93	12.97	7.32	4.56	22.31	10.29	0.560 ± 0.00005	2.416 ± 0.0003	0.352 ± 0.00005	0.811
	SMD	27.50	13.69	7.90	5.01	23.38	10.39	0.562 ± 0.00006	2.463 ± 0.0004	0.350 ± 0.00005	0.813
	HMD-Net	27.64	13.98	8.22	5.33	23.42	10.53	0.572 ± 0.00006	2.526 ± 0.0004	0.356 ± 0.00005	0.815
	HMD-Net + LF	29.09	14.64	8.63	5.60	24.63	10.99	0.580 ± 0.00005	2.540 ± 0.0004	0.360 ± 0.00005	0.819
	EHMD+DHSA+BERT	29.62	15.47	9.52	6.18	23.93	11.27	0.585 ± 0.00005	2.513 ± 0.0003	0.359 ± 0.00004	0.823
	HMD-Net + BERT	**29.70**	**15.95**	**9.74**	**6.21**	**24.91**	**11.37**	**0.584 ± 0.00005**	**2.614 ± 0.0004**	**0.363 ± 0.00005**	**0.824**

Automatic Evaluation Results (on RACE++ DG) :

	Models	BLEU-1	BLEU-2	BLEU-3	BLEU-4	ROUGE-L	METEOR	Embd Avg	G. Match	Ext.Score	BERT-CS
1st	EHMD+DHSA	32.68	18.62	12.41	8.96	31.23	12.3	0.5713 ± 0.00005	2.4006 ± 0.0003	0.3649 ± 0.00004	0.8395
	SMD	33.18	18.45	11.43	7.36	**32.48**	12.53	0.5854 ± 0.00005	2.62 ± 0.0003	0.3770 ± 0.00004	0.8424
	HMD-Net	33.37	18.61	11.64	7.66	32.29	12.49	0.5877 ± 0.00005	2.6689 ± 0.0003	0.3766 ± 0.00004	0.8419
	HMD-Net+ LF	33.45	18.81	11.87	7.93	32.18	12.54	0.5826 ± 0.00005	2.6641 ± 0.0003	0.3762 ± 0.00004	0.8429
	EHMD+DHSA+BERT	33.57	19.38	12.79	8.96	31.81	12.44	0.5848 ± 0.00004	2.6624 ± 0.0003	0.3710 ± 0.00004	0.8444
	HMD-Net+ BERT	**34.58**	**20.26**	**13.54**	**9.66**	32.24	**12.85**	**0.5939 ± 0.00005**	**2.6904 ± 0.0003**	**0.3782 ± 0.00004**	**0.8452**
2nd	EHMD+DHSA	31.46	16.5	10.1	6.65	28.69	11.58	0.5656 ± 0.00005	2.3023 ± 0.0003	0.3540 ± 0.00004	0.8371
	SMD	32.42	17.29	10.36	6.54	30.41	12.09	0.5774 ± 0.00005	2.5257 ± 0.0003	0.3684 ± 0.00004	0.8422
	HMD-Net	33.99	17.62	10.42	6.45	30.49	12.23	0.5839+/- 0.00005	2.5756 ± 0.0003	0.3709 ± 0.00004	0.8423
	HMD-Net + LF	33.26	18.03	10.79	6.81	**31.01**	12.37	0.5846 ± 0.00005	2.6099 ± 0.0003	0.3763 ± 0.00004	0.8426
	EHMD+DHSA+BERT	33.47	18.83	12.28	8.5	29.51	12.21	0.5838 ± 0.00004	2.6248 ± 0.0002	0.3642 ± 0.00004	0.8425
	HMD-Net + BERT	**34.01**	**19.53**	**12.83**	**9.02**	30.86	**12.51**	**0.5953 ± 0.00004**	**2.6554 ± 0.0003**	**0.3763 ± 0.00004**	**0.8430**
3rd	EHMD+DHSA	31.27	15.85	9.38	6.05	27.67	11.49	0.5698 ± 0.00005	2.2752 ± 0.0002	0.3533 ± 0.00004	0.8362
	SMD	31.73	16.39	9.42	5.72	29.85	11.73	0.5794 ± 0.00005	2.483 ± 0.0003	0.3682 ± 0.00004	0.8376
	HMD-Net	32.14	16.67	9.55	5.69	29.75	11.95	0.5864 ± 0.00004	2.5196 ± 0.0003	0.3683 ± 0.00004	0.8369
	HMD-Net + LF	31.89	16.89	9.85	6.07	29.75	11.95	0.5736 ± 0.00005	2.5819 ± 0.0003	0.3653 ± 0.00004	0.8380
	EHMD+DHSA+BERT	33.26	18.59	12.05	8.32	29.12	12.14	0.5817 ± 0.00004	2.5675 ± 0.0002	0.3635 ± 0.00004	0.8401
	HMD-Net + BERT	**33.29**	**18.84**	**12.28**	**8.52**	**29.87**	**12.17**	**0.5881 ± 0.00005**	**2.6214 ± 0.0002**	**0.3690 ± 0.00004**	**0.8400**

Manual Evaluation Approach:

1. **Comparative Study:** *Which of the proposed models is performing the best?*
    i). 30 annotators, 3 annotator-sets (each size 10), and 120 questions from 40 articles (three questions from each article)
    ii). Along with article and question four model outputs are given (SMD, HMD-Net, HMD-Net+LF, and HMD-Net+BERT)
    iii). Annotators to select the most closest distractible answer
2. **Quantitative Study:** *What is the quality of the generated text?*
    i). Considered large evaluation dataset over six models
    ii). 14 annotators, 2 annotator-sets (each size 7), and 350 questions from 117 articles (approx three questions from each article)
    iii). Models are: SMD, HMD-Net, HMD-Net+LF, EHMD+DSHA, EHMD+DSHA+BERT and HMD-Net+BERT
    iv). Rate grammatical correctness and distractibility on scale of 1-5 (1 is very bad and 5 is very good)

Manual Evaluation Results:

Models	Annot-set1	Annot-set2	Annot-set3
SMD	24	27	25
HMD-Net	25	26	27
HMD-Net + LF	34	30	33
HMD-Net + BERT	37	37	35

Comparative Study Results

	Models	Annot-set1	Annot-set2
GC	EHMD+DHSA	4.007	3.298
	SMD	4.058	3.894
	HMD-Net	3.780	3.747
	HMD-Net+LF	4.061	3.988
	EHMD+DHSA+BERT	4.054	**4.071**
	HMD-Net+BERT	**4.155**	3.982
DA	EHMD+DHSA	2.431	2.557
	SMD	2.567	2.457
	HMD-Net	2.522	2.491
	HMD-Net+LF	2.680	2.560
	EHMD+DHSA+BERT	2.661	**2.752**
	HMD-Net+BERT	**2.752**	2.634

Quantitative Study Results

Model Components and Output Verifications

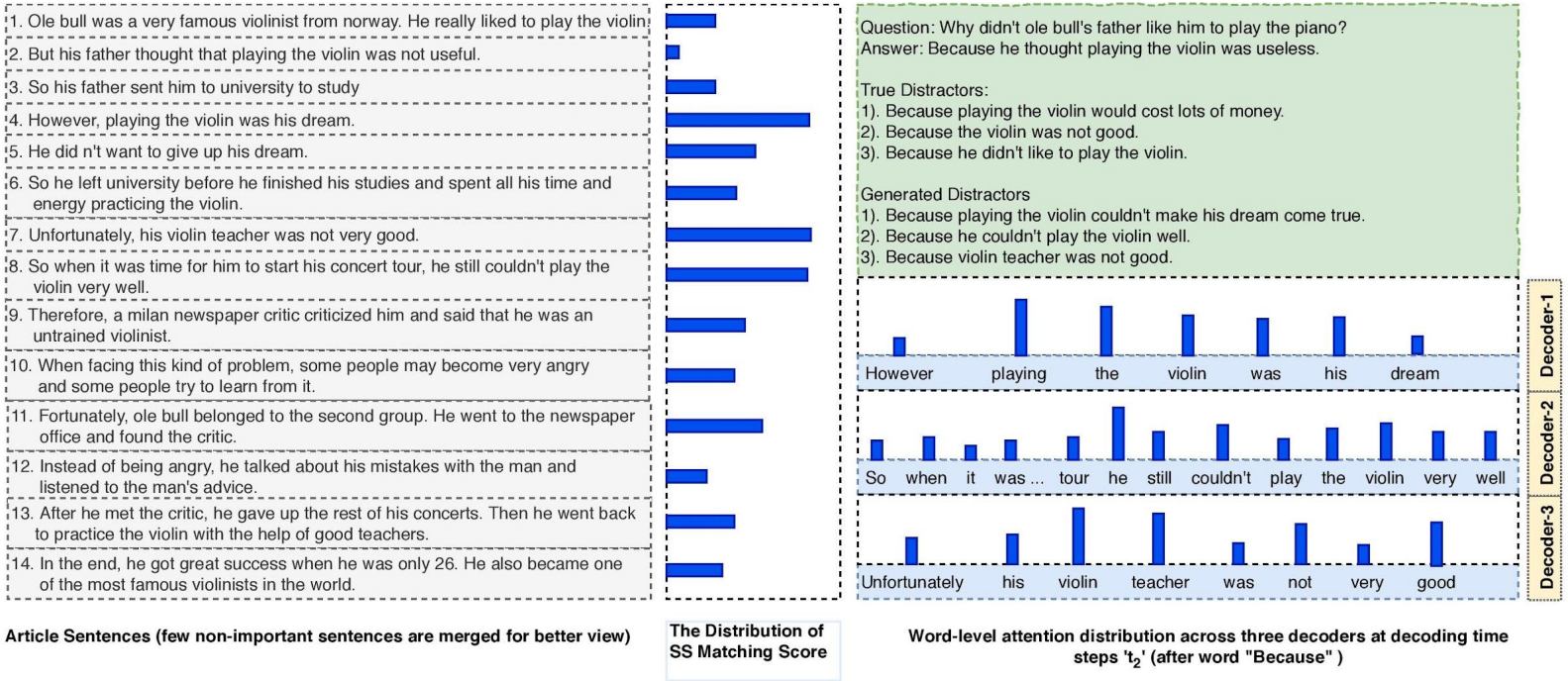
Models	BLEU-1	BLEU-2	BLEU-3	BLEU-4	ROUGE-L	METEOR
Full HMD-Net Model	29.26	16.16	10.16	7.66	25.78	11.58
Without EEL (with CR)	28.60	15.17	9.68	6.90	25.15	10.96
Without CR (with EEL)	29.15	15.71	10.20	7.27	25.62	11.40
*Without CER	28.86	15.46	9.96	7.15	25.38	11.11
Without h_aq & h_qa	28.92	15.89	10.39	7.50	25.56	11.44
Without DSL	28.35	15.24	9.91	7.05	25.39	10.96
With last two Question tokens in QCI	20.90	9.38	5.45	3.50	17.45	8.24

Ablation Study Results

Models	Dist1-Dist2	Dist1-Dist3	Dist2-Dist3
SMD	0.200	0.191	0.216
HMD-Net	0.221	0.210	0.236
HMD-Net + LF	0.215	0.219	0.201
HMD-Net + BERT	0.264	0.251	0.246

Inter-distractor Similarity Test

Case Study: Demo



Article Sentences (few non-important sentences are merged for better view)

The Distribution of SS Matching Score

Word-level attention distribution across three decoders at decoding time steps 't_2' (after word "Because")

Conclusion and Future Work

We presented a data driven approach to generate long and high-quality distractors for reading comprehension MCQ. We exploited the rich interaction among question, answer and passage using SoftSel operation and Gated Mechanism at the encoder side and used three separate decoder in the decoder side.

In future, we will develop an approach where any number of in-context and non-answer- revealing distractors can be generated using a single decoder.

Acknowledgements

We would like to thank all the annotators participated in the manual evaluation process, proofreaders of paper and anonymous reviewers for valuable suggestion. We also like to thank SIGIR for providing “Student Travel Grant” for attending CIKM 2020.

References

1. Liang, Chen, et al. "Distractor generation with generative adversarial nets for automatically creating fill-in-the-blank questions." *Proceedings of the Knowledge Capture Conference*. 2017.
2. Gao, Yifan, et al. "Generating distractors for reading comprehension questions from real examinations." *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 33. 2019.
3. Zhou, Xiaorui, Senlin Luo, and Yunfang Wu. "Co-Attention Hierarchical Network: Generating Coherent Long Distractors for Reading Comprehension." *arXiv preprint arXiv:1911.08648* (2019).
4. Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." *arXiv preprint arXiv:1810.04805* (2018).
5. Min Tang, Jiaran Cai, and Hankz Hankui Zhuo. 2019. Multi-Matching Network for Multiple Choice Reading Comprehension. Proceedings of the AAAI, Honolulu (2019).
6. Thang Luong, Hieu Pham, and Christopher D. Manning. 2015. Effective Approaches to Attention-based Neural Machine Translation. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon, Portugal, 1412–1421.
7. Alessandro Sordoni, Yoshua Bengio, Hossein Vahabi, Christina Lioma, Jakob Grue Simonsen, and Jian-Yun Nie. 2015. A Hierarchical Recurrent Encoder-Decoder for Generative Context-Aware Query Suggestion (CIKM '15). Association for Computing Machinery, New York, NY, USA, 553–562.
8. Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. BLEU: a method for automatic evaluation of machine translation. In Proceedings of the 40th annual meeting on association for computational linguistics. 311–318.
9. Lin, Chin-Yew. "Rouge: A package for automatic evaluation of summaries." *Text summarization branches out*. 2004.
10. Alon Lavie and Michael J Denkowski. 2009. The METEOR metric for automatic evaluation of machine translation. Machine translation 23, 2-3 (2009), 105–115.
11. Vasile Rus and Mihai Lintean. 2012. A comparison of greedy and optimal assessment of natural language student input using word-to-word similarity metrics. In Proceedings of the Seventh Workshop on Building Educational Applications Using NLP. 157–162.
12. John Wieting, Mohit Bansal, Kevin Gimpel, and Karen Livescu. 2016. Towards Universal Paraphrastic Sentence Embeddings. In 4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings, Yoshua Bengio and Yann LeCun (Eds.).
13. Gabriel Forgues, Joelle Pineau, Jean-Marie Larchevêque, and Réal Tremblay. 2014. Bootstrapping dialog systems with word embeddings. In Nips, modern machine learning and natural language processing workshop, Vol. 2.

Thank you!

Questions?