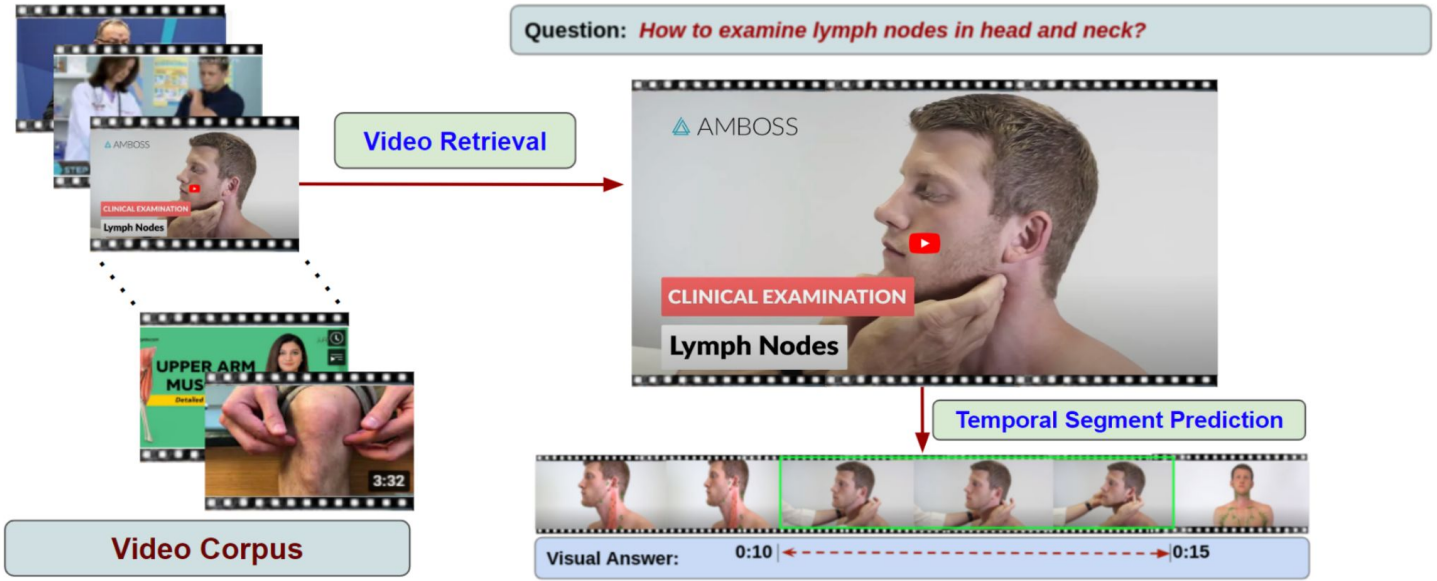
UNCWAI at MedVidQA 2023: T5 Model for Video Temporal Segment Prediction

Long Qi, Owen Deen, ZhiFei Xie, Xingyu Cui, and Gulustan Dogan

Outline

- Overview of the MedVidQA Task 1
- Data Collection and Preprocessing
- Summary of Approaches:
  - Cosine Similarity of Video Ranking
  - T5 Model implementation
  - BigBird Model comparison
- Key Results: Performance Metrics of Different Runs

Overview of the MedVidQA 2023 Task 1



Data Collection


```
{
        "sample_id": 2711,
        "question": "How to perform chin tucks to treat neck pain?",
        "answer_start": "00:07",
        "answer_end": "01:13",
        "answer_start_second": 7,
        "answer_end_second": 73,
        "video_length": 186,
        "video_id": "h5MvX50zTLM",
        "video_url": "https://www.youtube.com/watch?v=h5MvX50zTLM"
    }
```


- Obtain subtitles of video id using the python library, youtube-transcript-api

video_id	question	timestamps	subtitles
h5MvX50zTLM	How to perform chin tucks to treat neck pain?	(7, 73)	for the top three exercises for neck.pain and posture exercise number one is.the chin tuck
h5MvX50zTLM	How to perform scapular retractions to treat neck pain?	(74, 132)	called chin tuck 10 second hold and.second exercise is scapular retraction.what that reall
h5MvX50zTLM	How to perform corner stretches to treat neck pain?	(132, 181)	times if youre feeling brave you can do.it up to 20 times third exercise for top.three exerci
Se3kf5X4PCc	How can I do hamstring stretches for getting rid of lower back pain?	(56, 72)	
Se3kf5X4PCc	How can I do stretching of piriformis muscles to get rid of lower back pain?	(75, 93)	
Se3kf5X4PCc	How to do knee to opp. shoulder to loosen lower back and hip tightness?	(96, 112)	
UZqktEPlxTU	How do you put on a clavicle brace?	(38, 105)	and remove this brace This will help prevent you from further injuring.yourself especially if
_N89T_Yqu68	How to assess the supraclavicular lymph nodes	(40, 57)	assessing the supraclavicular lymph.nodes they are as the name would suggest.above th
_N89T_Yqu68	How to examine the submental lymph nodes?	(58, 107)	and the lymph nodes have been.exaggerated and can be seen here the.submental lymph
_N89T_Yqu68	How to examine the cervical lymph nodes?	(108, 131)	again examined carefully with the.fingers bilaterally the cervical lymph.nodes are located i
_N89T_Yqu68	How to locate and examine the pre and post auricular lymph nodes?	(132, 154)	and are examined with the flaps of the.fingers in this similar Z shape the pre.Oracle lymph
_N89T_Yqu68	How to examine the occipital lymph nodes	(155, 170)	lymph nodes are actually located behind.the Pinner and are examined so the final.set of l
g-gNQPyxU4c	How to avoid snoring?	(370, 405)	all he has just a little thin piece of tape. Its about a quarter inch wide.and maybe an inch l
VHN5zPaw96w	How to use an incentive spirometer to maintain respiratory fitness?	(52, 77)	Saras going to exhale naturally and place the device in her mouth and inhale taking.a nice
PdMJAhD7NEw	How to fix chronic neck tension while lying down on the ground?	(113, 277)	exercises to actually permanently fix.this problem okay so as I said at the.outset you can
6kQEDRQdJZ8	How to identify if you may have a pinched nerve?	(28, 171)	Sure. So first lets talk about the symptoms.What are the things youre gonna feel when.yo
6kQEDRQdJZ8	How to identify if you may have Saturday night palsy?	(409, 448)	Sure. The other thing we thought.we talked about a couple fun ones right.Brad theres so
6kQEDRQdJZ8	How to identify if you may have honeymoon palsy?	(449, 469)	You dont even know what that would be.But another one would be.honeymoon P honeyr
6kQEDRQdJZ8	How to identify if you may have handcuff neuropathy?	(470, 480)	That also causes a palsy in your arm.Another one is could be related.as a handcuff neuro
6kQEDRQdJZ8	How to identify if you may have crutch palsy?	(481, 514)	they can actually put pressure on the nerves.And the last one were gonna mention is.is a
6kQEDRQdJZ8	How to prevent your foot from falling asleep while you drive?	(552, 574)	The other one that is a common one that is.youre driving along.This is why you shouldnt v
0dr5yuoBOF4	How do i apply pressure dressing on my leg?	(206, 234)	all right so were gonna say that the.injury is here to my upper extremity.this is where the w

Data Preprocessing

- The youtube-transcript-api package gave us time stamps for key phrases.
- After removing special characters and turning text to lowercase, we stored those timestamp values with each phrase pair.
- We wanted to enhance data uniformity and reduce complexity for the model.

$$IOU = \frac{prediction - and - gold}{prediction - or - gold}$$

T5 Model

- One challenge we faced in the training stage, was the limit to the number of input tokens that T5 could receive. The maximum was 700 tokens, yet subtitles could exceed that.

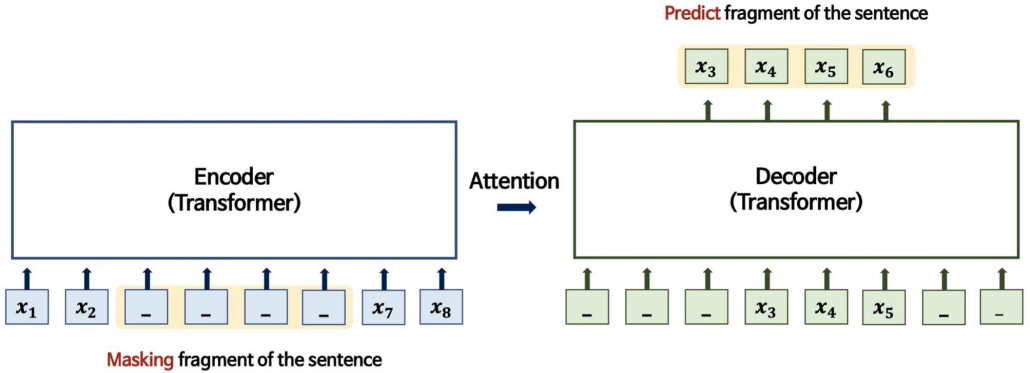
if token of context bigger than 700 we will truncate the text

```
for indx, row in train.iterrows():
    if row["len"]>700:
        train.loc[indx,"context"]=train.loc[indx,"answer"]
```

```
for indx, row in val.iterrows():
    if row["len"]>700:
        val.loc[indx,"context"]=val.loc[indx,"answer"]
```

T5 Model

- T5 stands for “text-to-text transfer transformer” model and was developed by Google research.
- It is designed to handle NLP tasks by converting all tasks to a text-to-text format.
- T5 is based on the transformer architecture and uses self-attention mechanisms to process input text.
- Encoder processes the input and decoder generates output text.



T5 Model

- Benefits of T5 - every task is treated as text generation: question-answering, summarization, translation, etc
- Model is pre trained on large corpus of text, learning language structures and patterns.
- T5 models come in different sizes - small to extra-large.
- The larger variants generally offer better performance at increased computational requirements.

Model size variants

Model	Parameters	# layers	d_{model}	d_{ff}	d_{kv}	# heads
Small	60M	6	512	2048	64	8
Base	220M	12	768	3072	64	12
Large	770M	24	1024	4096	64	16
3B	3B	24	1024	16384	128	32
11B	11B	24	1024	65536	128	128

Test 1 - T5 small with 30 epochs

Model Prediction:
 <pad> Would be if you had a wound the first step youre gonna do is put pressure on it.If its a large wound youre g
onna put your entire palm of your hand and all of your weight on the wound.Were going to put a lot of pressure and
hopefully that stops the bleed.If it stops the bleed you need to maintain the pressure on the wound until help arri
ves.If that doesnt stop the bleed oure gonna move to packing the wound with gauze.And typically were gonna place it
in the handle.And thats getting to the source of the bleeding.So were going to just keep packing more and more and
it could be a deep wound.So were going to just keep packing more and more and more and it could be a deep wound.So
were going to just keep packing more and more and more and it could be a deep wound.So were going to just keep pack
ing more and

Gold Answer:
 If its a large wound youre gonna put your entire palm of your hand and all of your weight on the wound.Were going
to put a lot of pressure and hopefully that stops the bleed.If it stops the bleed you need to maintain the pressure
on the wound until help arrives.If that doesnt stop the bleed were gonna move to packing the wound with gauze.And t
ypically we can use a hemostatic gauze or any gauze

Test 1 - T5 small with 30 epochs

Context:
I'm Lisa Hollister, the director of trauma and acute care surgery for Parkview Health.Today I'm going to show you how to stop the bleed.Stop the bleed is very simple Three steps.The first one is pressure, then wound packing, then a tourniquet.So let's start with pressure.If you come upon a bleeding patient that has a wound the first step you're gonna do is put pressure on it.If it's a large wound, you're gonna put your entire palm of your hand and all of your weight on the wound.We're going to put a lot of pressure and hopefully that stops the bleed.If it stops the bleed you need to maintain the pressure on the wound until help arrives.If that doesn't stop the bleed, we're gonna move to packing the wound with gauze.And typically, we can use a hemostatic gauze or any gauze.or you could use a shirt if you have nothing available.So we're gonna take the gauze and.we're going to pack it inside the wound until you can't pack it anymore.And that's getting to the source of the bleeding.So we're going to just keep packing more and more and more and it could be a deep wound,.so don't be afraid.So once we've gotten this completely packed,.hopefully we can put some pressure on it and that will stop the bleed.So if you're all by yourself, and you need to apply a tourniquet,.Just put the tourniquet on.We're going to tighten it as much as possible.It's a velcro, so super super tight,.Then we're gonna take the handle we're gonna twist it until The bleeding source is stopped.And then we're gonna place it into the handle.We're gonna take the velcro and close it.And we're going to write the time.For courses in our area go to parkviewcom and search "Stop the Bleed"

Question:
How to put gauze inside the wound?

Model Prediction: <pad> were going to pack it inside the wound until you cant pack it anymore.And thats getting to the source of the bleeding.So were going to just keep packing more and more and more and it could be a deep wound.so dont be afraid.So once weve gotten this completely packed.hopefully we can put some pressure on it and that will stop the bleed.So if youre all by yourself and you need to apply a tourniquet.Just put the tourniquet on.Were going to tighten it as much as possible.Its a velcro so super super tight.Then were gonna take the handle were gonna twist it until The bleeding source is stopped</s>

Gold Answer: So were gonna take the gauze and.were going to pack it inside the wound until you cant pack it anymore.And thats getting to the source of the bleeding.So were going to just keep packing more and more and more and it could be a deep wound.so dont be afraid

Loss: 0.10575777292251587

T5 Model Comparisons

Models	**IOU=0.3**	**IOU=0.5**	**IOU=0.7**	**avgIOU**
T5-Small	0.45	0.3	0.2	0.35
T5-Large	0.8	0.5	0.5	0.5877

Table 1:IOU tests on different models.

T5 Training Metrics

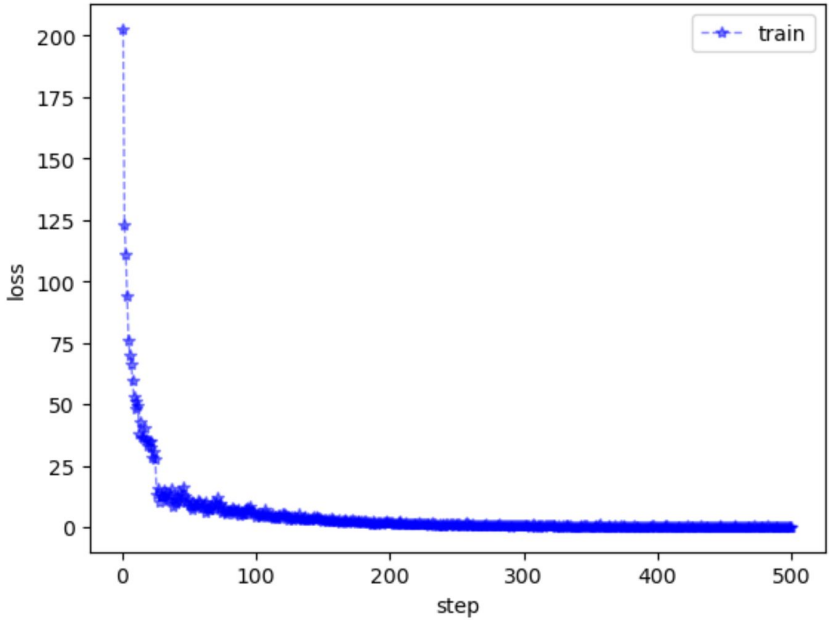


Figure 3:Training loss with respect to each step, each step contains 100 samples.

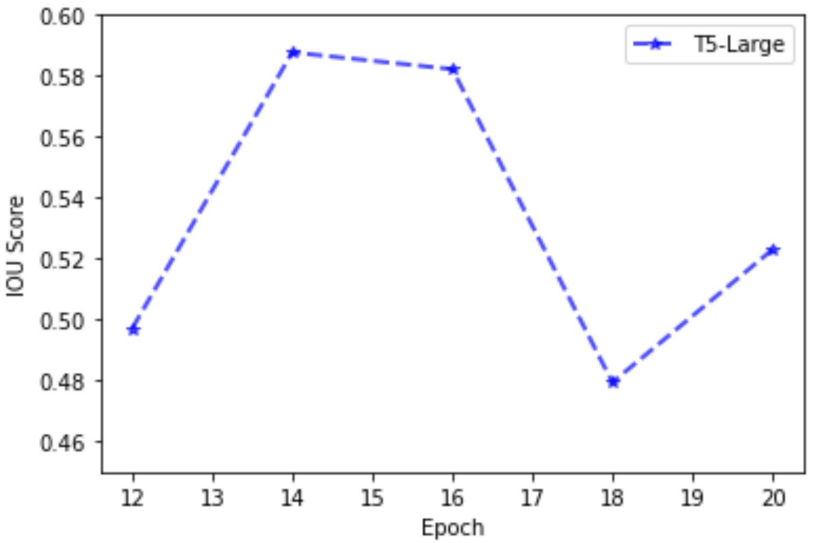


Figure 4:IOU Score on validation dataset with respect to training epoch

T5 vs Bert model

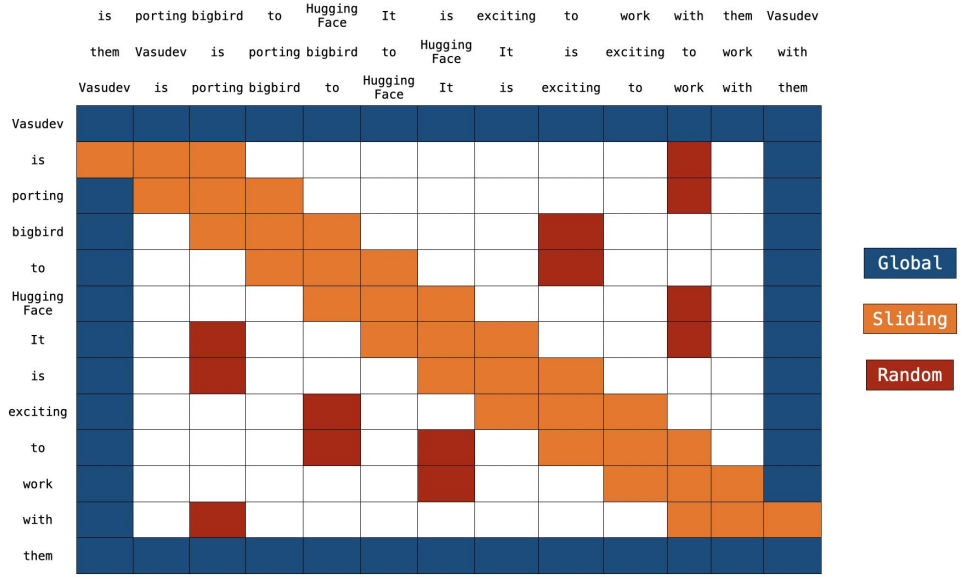
```
Loss:
 0.16677483916282654
```

```
Some weights of the model checkpoint at bert-base-uncased were not used when initializing BertModel: ['cls.predict
ons.transform.LayerNorm.weight', 'cls.predictions.transform.dense.bias', 'cls.seq_relationship.weight', 'cls.predi
tions.transform.dense.weight', 'cls.predictions.bias', 'cls.predictions.decoder.weight', 'cls.predictions.transfor
m.LayerNorm.bias', 'cls.seq_relationship.bias']
- This IS expected if you are initializing BertModel from the checkpoint of a model trained on another task or wit
another architecture (e.g. initializing a BertForSequenceClassification model from a BertForPreTraining model).
- This IS NOT expected if you are initializing BertModel from the checkpoint of a model that you expect to be exac
ly identical (initializing a BertForSequenceClassification model from a BertForSequenceClassification model).
```

```
Stamp prediction: 84.26 - 107.0

Stamp gold: 29 - 104
```

Second Model - BigBird

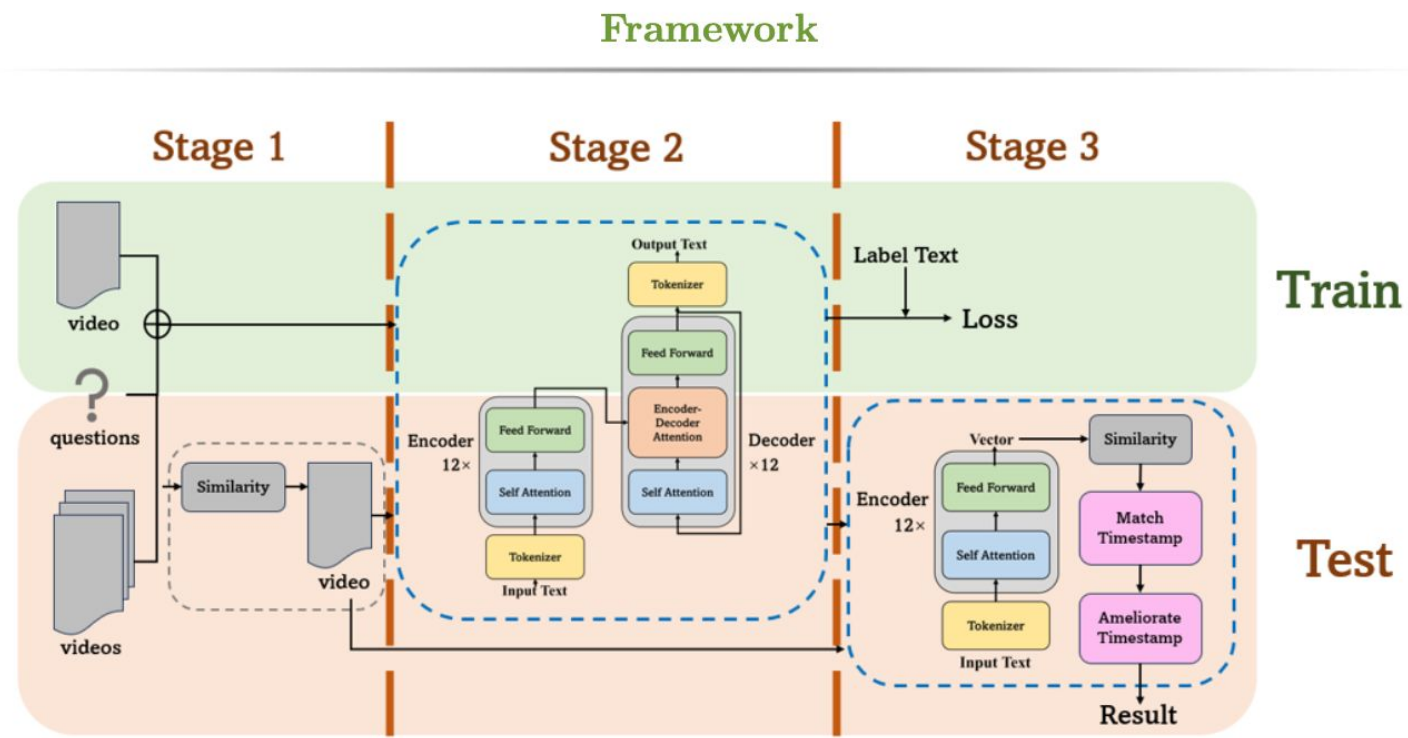
- BigBird is a transformer-based model designed to handle long sequences of data.
- BigBird implements a sparse attention mechanism to balance computational resources.
- Despite this, BigBird can still be very computationally expensive.



Second Model - BigBird

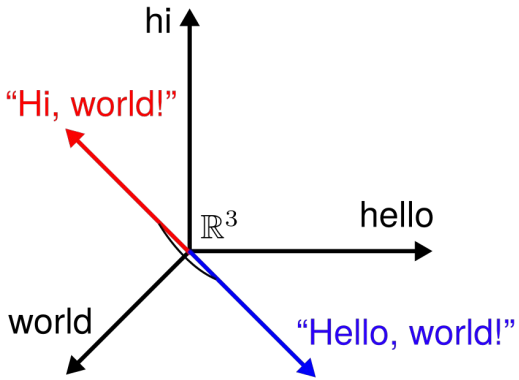
```
print("Begin Training, last modified at 2023-7-15-15:08")

device = torch.device("cuda" if torch.cuda.is_available() else "cpu")
epochs = 100
big_bird_learning_rate = 1e-5
learning_rate = 3e-5
adam_epsilon = 1e-8
crf_learning_rate = 0.01
num_labels = 3  # B-Answer, I-Answer, Other


dataset = MyTensorDataset()
train_dataloader = DataLoader(dataset,batch_size=4,shuffle=True)


# training loop
model = BigBirdCRF(num_labels=3).to(device)
#model = torch.load('CheckPoints/checkpoints_5.pt')
```

General Framework



Stage 1

- We faced challenges with ranking the videos.
- A simple approach was to try and look at the cosine similarity between question-subtitles. We used a similar “bins” approach as the Dossier team from MedVidQA 2022.
- Our main idea was to find which videos did not have any relevance to the questions.
- Once the non-relevant videos were determined, we put the rest through our model and then ranked the output.
- Identified “top-20” videos.



Cosine Similarity

Stage 2

- Text answer is generated by T5 model / BigBird.
- Models have been fine-tuned on medical domain knowledge from training datasets.
- We noticed that all of the output answers were summaries of the correct answer. Models could not find direct excerpts.
- If the question context was broad we noticed issues with answer generation.

Stage 3

- Similarity between generated text and subtitle is calculated, match between subtitle and time stamp are used to generate final result.

- Challenges included: lengthy answers were hard to map back to a subtitle timestamp, ensuring the matching reflects the context of the answer.

- String matching was also used to find the timestamps from the generated text.
- If answers spanned multiple timestamps then we needed to adjust to the exact moment but faced challenges.

Official Results

Visual Answer Localization

Team	RunID	IoU=0.3	IoU=0.5	IoU=0.7	mIoU
UNCWAI	run-2.json	42.5	32.5	22.5	31.37
UNCWAI	run-1.json	10	7.5	0	9.32
UNCWAI	run-3.json	25	10	5	15.78
Min		10	7.5	0	9.32
Mean		40.5	29.5	20.5	30.336
Max		67.5	62.5	50	55.24

Video Retrieval

Team	RunID	MAP	R@5	R@10	P@5	P@10	nDCG
UNCWAI	run-2.json	0.1839	0.1903	0.1903	0.29	0.145	0.2858
UNCWAI	run-1.json	0.3669	0.2221	0.3654	0.395	0.3575	0.5094
UNCWAI	run-3.json	0.3669	0.2221	0.3654	0.395	0.3575	0.5094
Min		0.1839	0.1903	0.1903	0.29	0.145	0.2858
Mean		0.31288	0.24766	0.31664	0.387	0.2755	0.44596
Max		0.404	0.3549	0.4132	0.545	0.3625	0.5448

Acknowledgements

We would like to thank NSF project 1053575 - XSEDE: eXtreme Science and Engineering Discovery Environment, for the computational resources used in our project. We would also like to thank NC State University for hosting the GEARS summer program.

Questions?

Thank you all for listening!

References

[1] G. Awad, K. Curtis, A. A. Butt, J. Fiscus, A. Godil, Y. Lee, A. Delgado, E. Godard, L. Diduch, D. Gupta, D. D. Fushman, Y. Graham, , and G. Que´not, “Trecvid 2023 - a series of evaluation tracks in video understanding,” in *Proceedings of TRECVID 2023*, NIST, USA, 2023.

[2] C.Raffel,N.Shazeer,A.Roberts,K.Lee,S.Narang,M.Matena,Y.Zhou, W. Li, and P. J. Liu, “Exploring the limits of transfer learning with a unified text-to-text transformer,” *Journal of Machine Learning Research*, vol. 21, no. 140, pp. 1–67, 2020.

[3] W. Kusa, G. Peikos, O. Espitia, A. Hanbury, and G. Pasi, “Dossier at medvidqa 2022: Text-based approaches to medical video answer localization problem,” in *Proceedings of the 21st Workshop on Biomedical Language Processing*, pp. 432–440, 2022.

[4] D. Gupta and D. Demner-Fushman, “Overview of the medvidqa 2022 shared task on medical video question-answering,” in *Proceedings of the 21st Workshop on Biomedical Language Processing*, pp. 264–274, 2022.

[5] M. Zaheer, G. Guruganesh, A. Dubey, J. Ainslie, C. Alberti, S. Ontanon, P. Pham, A. Ravula, Q. Wang, L. Yang, and A. Ahmed, “Big bird: Transformers for longer sequences,” 2021.

[6] B. Li, Y. Weng, F. Xia, B. Sun, and S. Li, “Vpai lab at medvidqa 2022: a two-stage cross-modal fusion method for medical instructional video classification,” in *Proceedings of the 21st Workshop on Biomedical Language Processing*, pp. 212–219, 2022.

[7] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, “Automatic differentiation in pytorch,” 2017.