MID-TERM PRESENTATION

by:

VISHAKHA GUPTA

Mentored by:

RAGHAV BALI

Automatic Question Answering through Bio-Bert Fine Tuning

AGENDA

- BERT
- BERT ARCHITECTURE
- BIO-BERT
- FINE TUNING BIO-BERT FOR QUESTION ANSWERING
- DATASET
- FINE TUNING PROCEDURE
- RESULTS
- APPLICATIONS
- ROADBLOCKS
- NEXT STEPS
- MODEL PRESENTATION THROUGH FLASK (DEMO)

OBJECTIVE

To fine tune pre-trained Bio-Bert model, to automate the medical question answer system.

Rather than reading the entire document for the correct context of a given question, our model picks the most succinct and relevant span of text from the given context text.

MOTIVATION

- An automated biomedical text mining is the need of an hour.
- Cannot directly apply biomedical dataset to pre-trained state of art model.
- Reason being word distribution shift from general domain corpora to biomedical corpora.

So we extended our work to recently introduced, Bio-Bert model, completely trained on the biomedical dataset.

BERT:
Bidirectional Encoder Representation from Transormers

WHY BERT?

- First deeply bidirectional unsupervised language representation, pre-trained using a plain text corpus.
- BERT word vector output encodes rich linguistic structure.
- BERT encodes both syntactic and semantic features in word vectors.
- Masked language model, with masking is performed randomly in both direction.

BERT ARCHITECTURE

BERT ARCHITECTURE

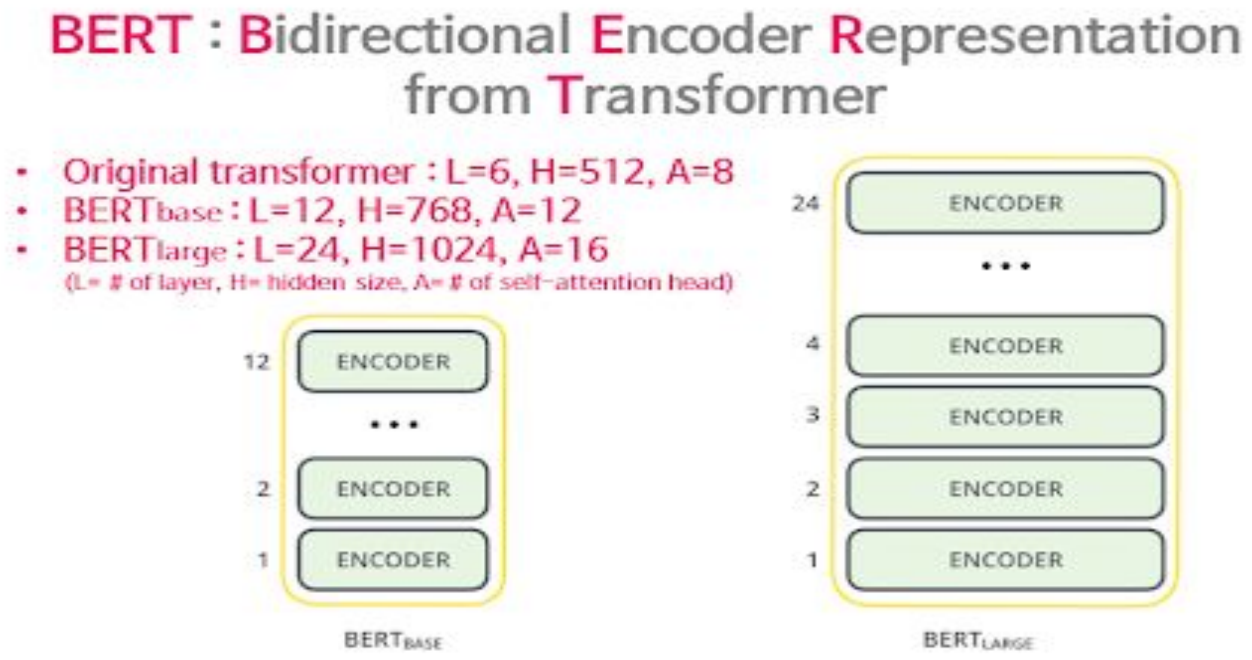


Figure 1: An overview of BERT architecture [2]

INSIDE A TRANSFORMER ENCODER

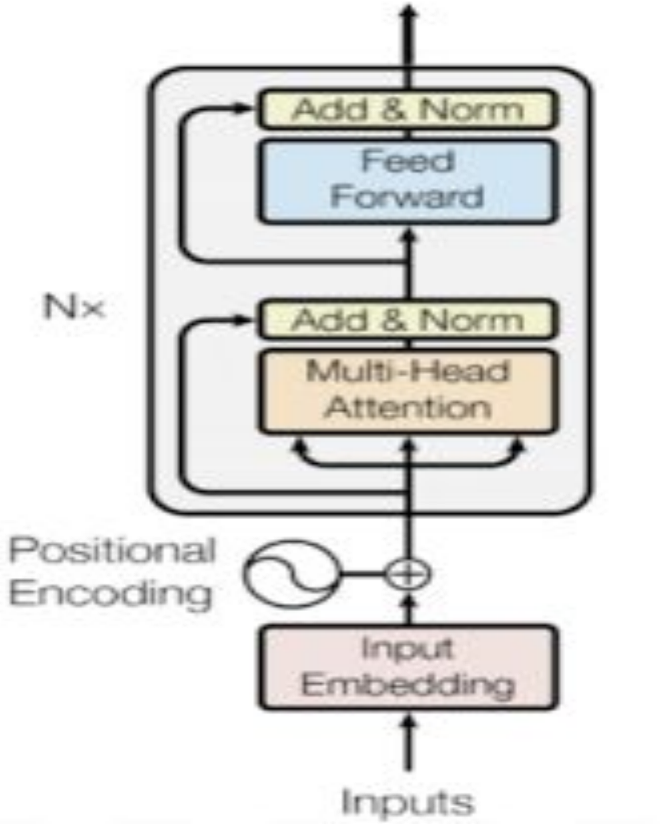
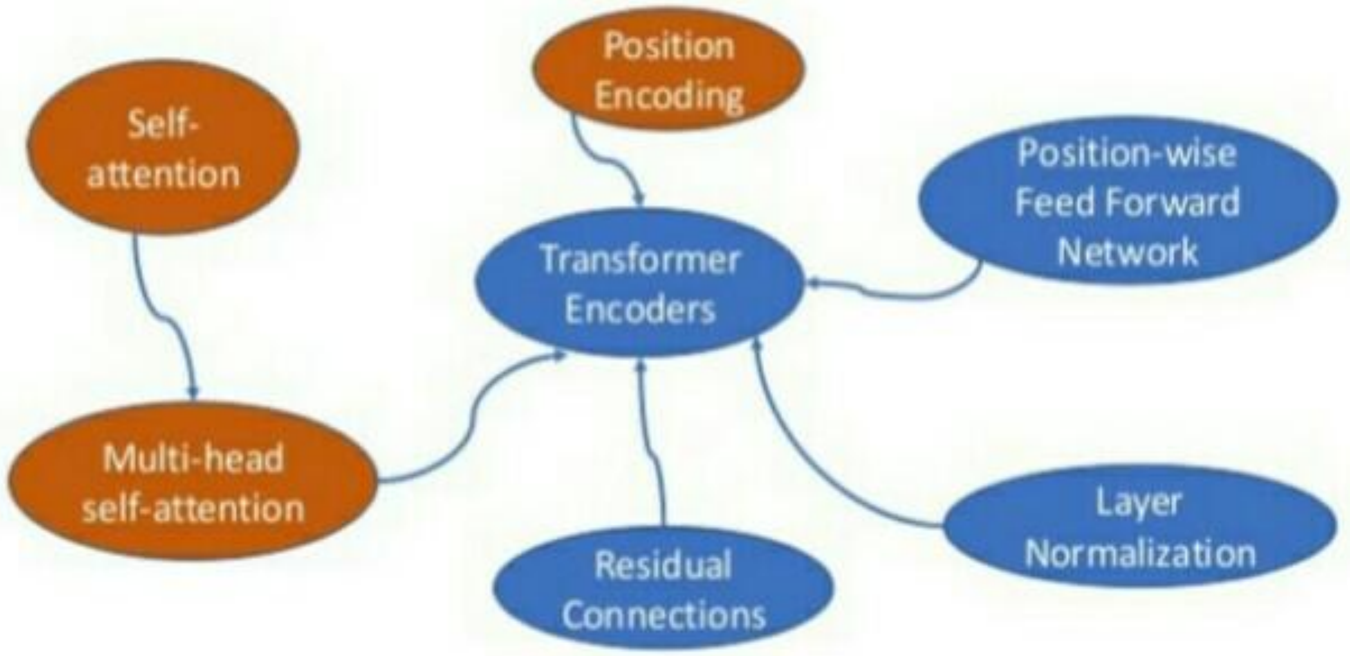


Figure 2: Detailed view of a Transformer Encoder [1]

- **BERT is a multi layer bidirectional Transformer encoder.**
- **In BERT number of such encoder blocks are chosen to be either 12(base) or 24(large).**
- **Blocks do not share weights with each other.**
- **The output from the 12th encoder block is taken as the final embedding for the given token.**

TRANSFORMER ENCODERS: KEY CONCEPTS



POSITIONAL ENCODING

- Position Encoding is used to make use of the order of the sequence
  - Since the model contains no recurrence and no convolution
- In Vawasni et al., 2017, authors used sine and cosine functions of different frequencies

$$PE_{(pos,2i)} = \sin\left(\frac{pos}{10000^{\frac{2i}{d_{model}}}}\right)$$

$$PE_{(pos,2i+1)} = \cos\left(\frac{pos}{10000^{\frac{2i}{d_{model}}}}\right)$$

  - *pos* is the position and i is the dimension

INPUT TEXT: My dog is cute. He likes playing.

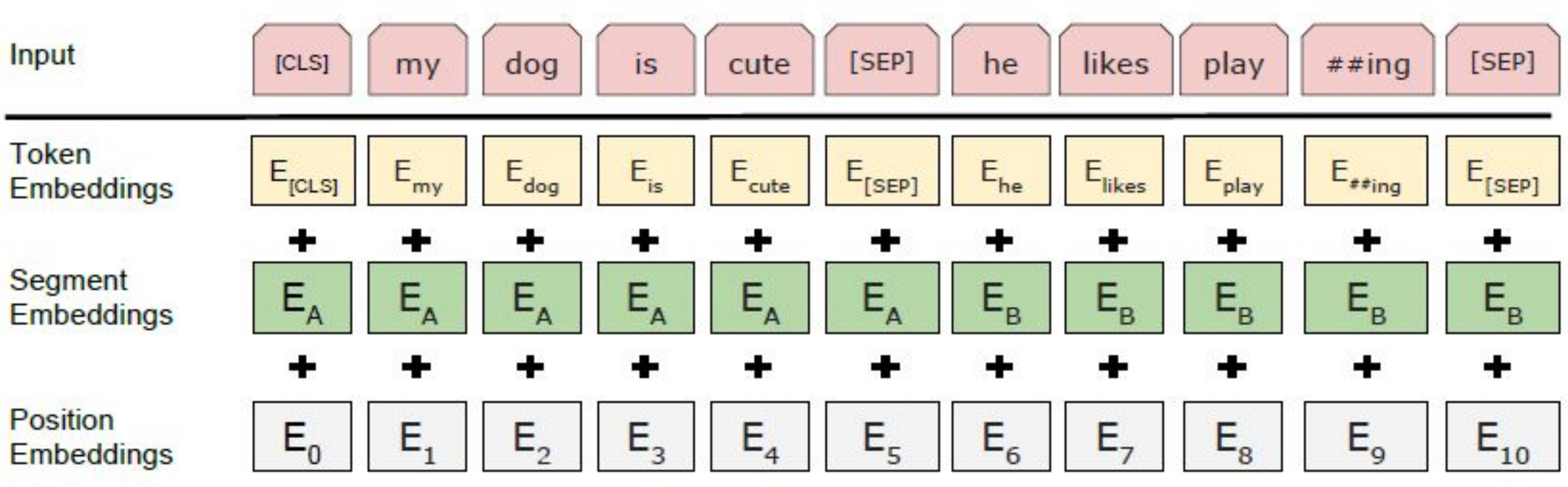


Figure 6: Input Representation [2]

INPUT REPRESENTATION
(In detail)

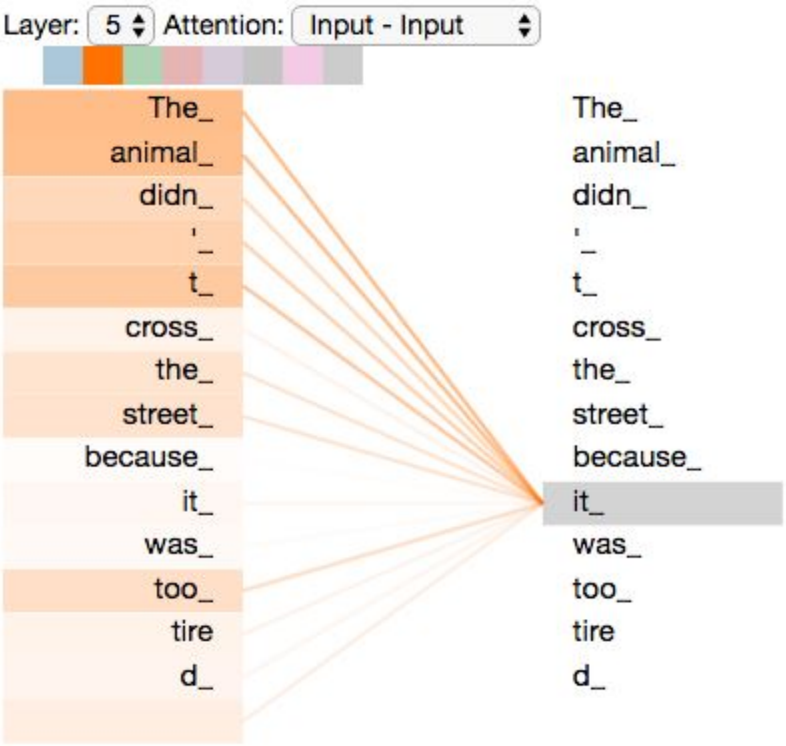
- **TOKEN EMBEDDINGS:**
  Pre-trained WordPiece Embeddings is used

- **SEGMENT EMBEDDINGS**
  0 and 1 are marked depending on the need.

- **POSITION EMBEDDINGS**
  Learned position embeddings

[CLS] is used for classification task
[SEP] is used to separate sentences by using a special token.

SELF- ATTENTION (An Overview)



Self Attention layer, takes as input a position injected naive form of embeddings and outputs more context aware embeddings.

For every input word a score is calculated with respect to every word in the sentence, by "relevance"

The model processes each word (each position in the input sequence), self attention allows it to look at other positions in the input sequence for clues that can help lead to a better encoding for this word.

Figure 3: Self Attention (the word "it" is encoded iin encoder #5 (the top encoder in the stack), part of the attention mechanism was focusing on "The Animal", and baked a part of its representation into the encoding of "it".) [1]

MULTI HEAD ATTENTION

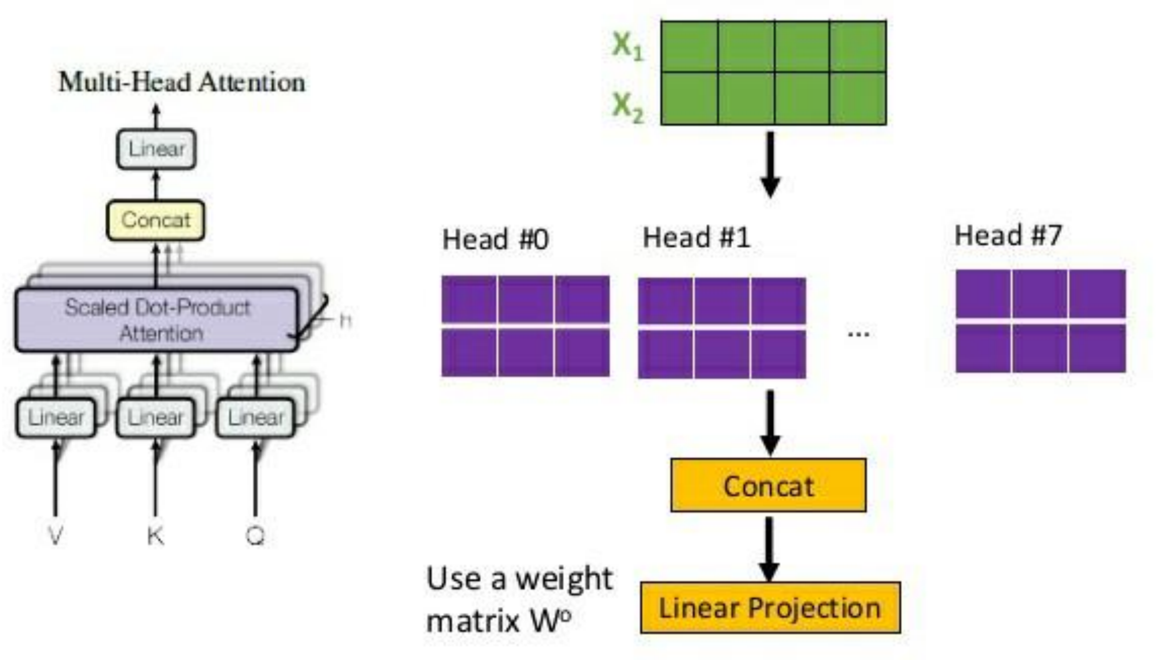


Figure 5: Multihead attention detailed view [1]

TASK #1: MASKED LANGUAGE MODEL

- 15% of the words are masked at random
  - and the task is to predict the masked words based on its left and right context
- Not all tokens were masked in the same way (example sentence "My dog is hairy")
  - 80% were replaced by the <MASK> token: "My dog is <MASK>"
  - 10% were replaced by a random token: "My dog is apple"
  - 10% were left intact: "My dog is hairy"

BIO-BERT:

a pre-trained biomedical language representation model for biomedical text mining

1. BioBERT is initialized with BERT pre-trained model trained on Wikipedia 2.5 billion words & Books Corpus 0.8 billion words. Rather than using random initialization of weights, pre-trained weights from BERT model are taken.

2. The next step is to pre-training on the domain data, BioBERT is pre-trained on PubMed Abstracts 4.5 billion words & PMC Full-text articles 13.5 billion words.

The pre-trained model can be used to fine-tune on various biomedical text mining tasks like NER, question & answer, relation extraction.

FINE TUNING BIO-BERT FOR QUESTION ANSWERING

QUESTION ANSWERING TASK

Document

Question

Answer

Or

impossible

DATASET

DATASET

Bio-ASQ question answer dataset is taken from large-scale biomedical semantic indexing and question answering competition.

The dataset consists of 3266 questions in training set and 935 questions in development dataset.

BioASQ QA dataset is pre-processed publicly available dataset built in similar format of that of SQUAD dataset.

The dataset consists of question and a context answer which contains answer in form of span of text.

GLIMPSE AT BIO-ASQ DATASET

"data": [{
"paragraphs": [
{
"qas": [
{
"id": "52bf208003868f1b06000019_002",
"question": "What is the inheritance pattern of Li\u2013Fraumeni syndrome?",
"answers": [
{
"text": "autosomal dominant",
"answer_start": 213
}] }],
"context": "... breast cancer patient from a Li-Fraumeni syndrome family. Li-Fraumeni Syndrome (LFS) is characterized by early-onset carcinogenesis involving multiple tumor types and shows autosomal dominant inheritance. Approximately 70% of LFS cases are due to germline mutations in the TP53 gene on chromosome 17p13.1. Mutations have also been found in the CHEK2 gene on chromosome 22q11, and others have been mapped to chromosome 11q23. ....
},

Question:

"What is the inheritance pattern of Li\u2013Fraumeni syndrome?"

Context:

…..breast cancer patient from a Li-Fraumeni syndrome family. Li-Fraumeni Syndrome (LFS) is characterized by early-onset carcinogenesis involving multiple tumor types and shows **autosomal dominant** inheritance. Approximately 70% of LFS cases are due to germline mutations in the TP53 gene on chromosome 17p13.1. Mutations have also been found in the CHEK2 gene on chromosome 22q11, and others have been mapped to chromosome 11q23. While characterizing an LFS family with a documented defect in TP53, we found one family member who developed bilateral breast cancer at age 37 yet was homozygous for wild-type TP53………………… These data may implicate the region at breakpoint 11q23 and/or 15q15 as playing a significant role in predisposition to breast cancer development.

Answer:

Autosomal Dominant

FINE TUNING BIO-BERT PROCEDURE

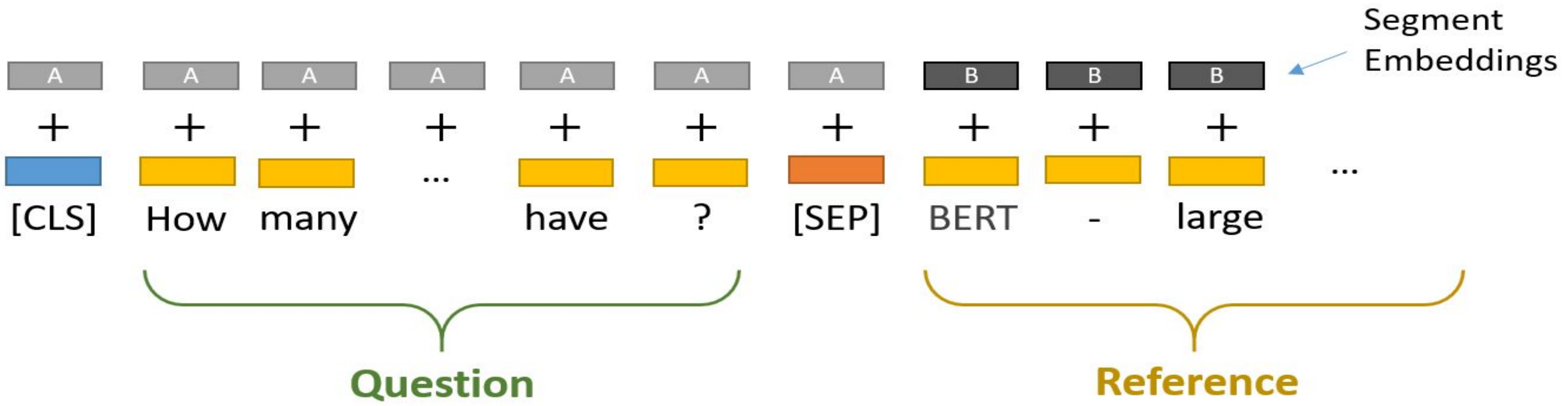
SPAN LEVEL TASK: (BIO-ASQ)

- Represent the input question and paragraph as a single packed sequence
  - The question uses the A embedding and the paragraph uses the B embedding
- New parameters to be learned in fine-tuning are start vector $S \in \mathbb{R}^H$ and end vector $E \in \mathbb{R}^H$
- Calculate the probability of word i being the start of the answer span

$$P_i = \frac{e^{S \cdot T_i}}{\sum_j e^{S \cdot T_j}}$$

- The training objective is the log-likelihood the correct and end positions

BIO-BERT
INPUT FORMAT



Question: What is the inheritance pattern of Li\u2013Fraumeni syndrome?

Reference Text: … breast cancer patient from a Li-Fraumeni syndrome family. Li-Fraumeni Syndrome (LFS) is characterized by early-onset carcinogenesis involving multiple tumor types and shows autosomal dominant inheritance. Approximately 70% of LFS cases are due to germline mutations in the TP53 gene on chromosome 17p13.1. Mutations have also been found in the CHEK2 gene on chromosome 22q11, and others have been mapped to chromosome 11q23. ……………………

Figure 7: Formation of input embeddings [5]

START TOKEN CLASSIFIER

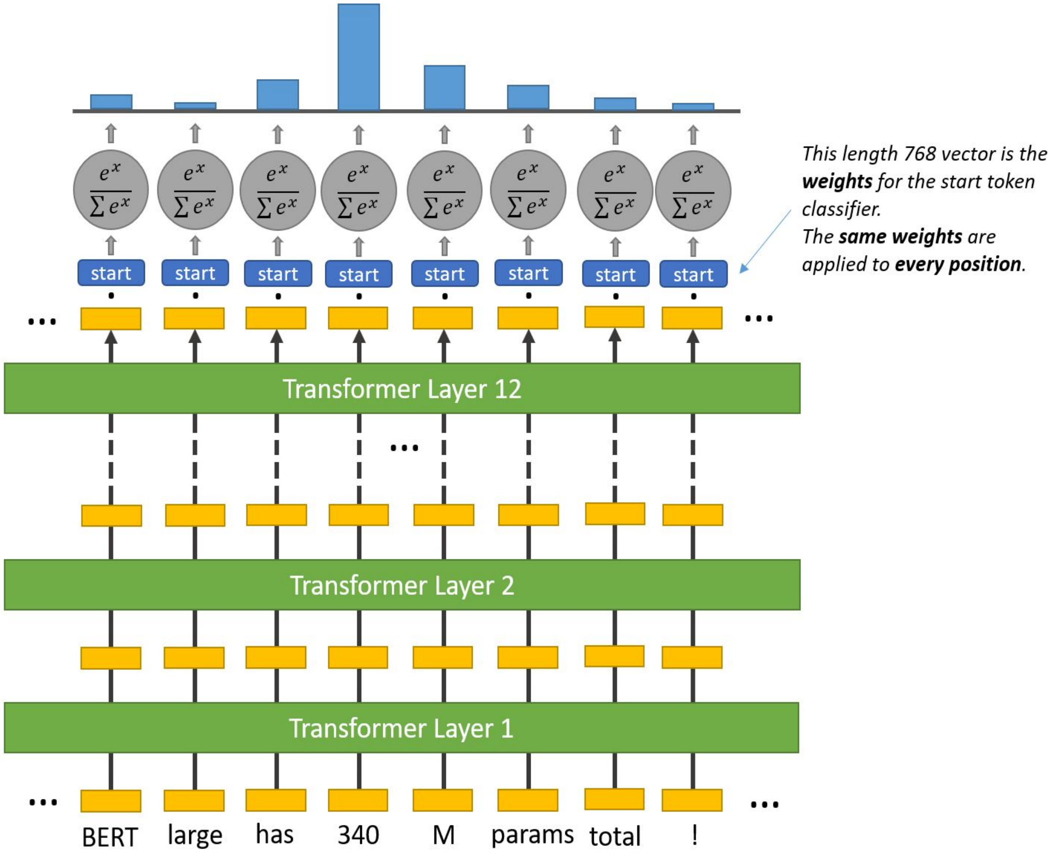


Figure 8: pictorial representation of start token classification[5]

BERT highlights a "span" of text containing the answer.

For every token it's final embeddings is taken, and dot product is performed with 'start' weights.

Softwax activation, used to produce a probability distribution over all of the words.

Highest probability word is chosen as start token.

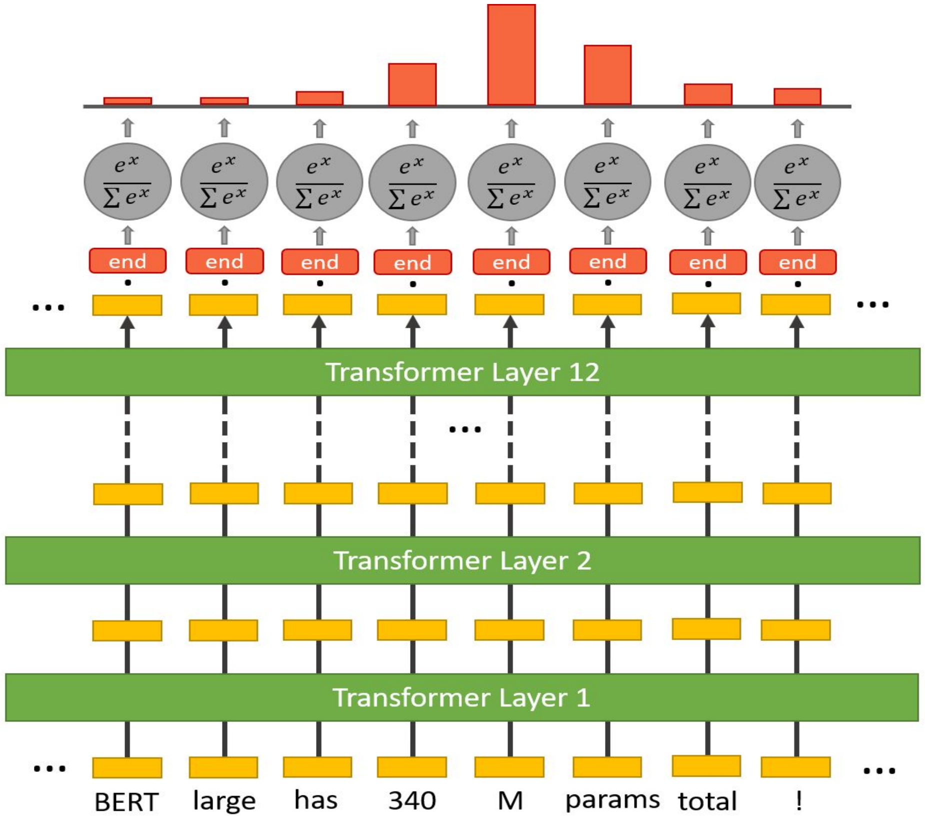


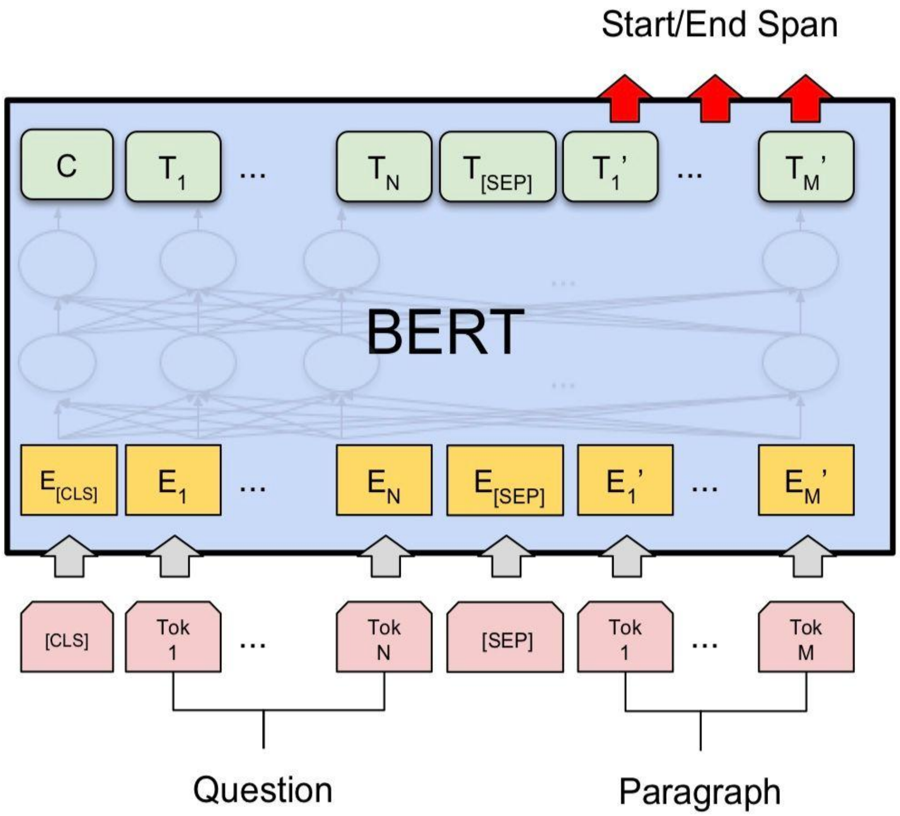
Figure 9: pictorial representation of start token classification[5]

Same process is repeated to pick the end token.

For end token we take a seperate end token weight

(represented by the red "end" rectangle in the aside illustration)

Fine Tuning FOR QA



Possible to answer if:

max(start) + max(end)

> threshold + (start(CLS) + end(CLS))

Where threshold is selected on development set to maximize F1.

Figure 10: pictorial representation for fine tuning Bio-Bert for QA[2]

RESULT

FINE- TUNING BERT ON BIO-ASQ DATASET

```
Iteration:  95% 239/251 [05:55<00:17,  1.49s/it]
Iteration:  96% 240/251 [05:56<00:16,  1.49s/it]
Iteration:  96% 241/251 [05:58<00:14,  1.49s/it]
Iteration:  96% 242/251 [05:59<00:13,  1.49s/it]
Iteration:  97% 243/251 [06:01<00:11,  1.49s/it]
Iteration:  97% 244/251 [06:02<00:10,  1.49s/it]
Iteration:  98% 245/251 [06:04<00:08,  1.49s/it]
Iteration:  98% 246/251 [06:05<00:07,  1.49s/it]
Iteration:  98% 247/251 [06:07<00:05,  1.49s/it]
Iteration:  99% 248/251 [06:08<00:04,  1.49s/it]
Iteration:  99% 249/251 [06:09<00:02,  1.49s/it]
Iteration: 100% 250/251 [06:11<00:01,  1.49s/it]
Iteration: 100% 251/251 [06:12<00:00,  1.48s/it]
Epoch: 100% 5/5 [30:47<00:00, 369.47s/it]
05/22/2020 08:49:17 - INFO - __main__ -   ***** Running predictions *****
05/22/2020 08:49:17 - INFO - __main__ -     Num orig examples = 935
05/22/2020 08:49:17 - INFO - __main__ -     Num split examples = 1139
05/22/2020 08:49:17 - INFO - __main__ -     Batch size = 8
05/22/2020 08:49:17 - INFO - __main__ -   Start evaluating
Evaluating:   0% 0/143 [00:00<?, ?it/s]05/22/2020 08:49:17 - INFO - __main__ -   Processing example: 0
Evaluating:  87% 125/143 [00:40<00:05,  3.08it/s]05/22/2020 08:49:58 - INFO - __main__ -   Processing examp
Evaluating: 100% 143/143 [00:46<00:00,  3.11it/s]
05/22/2020 08:50:03 - INFO - __main__ -   Writing predictions to: output/predictions.json
05/22/2020 08:50:03 - INFO - __main__ -   Writing nbest to: output/nbest_predictions.json
05/22/2020 08:50:28 - INFO - __main__ -   F1 score : 62.712041733484156
05/22/2020 08:50:28 - INFO - __main__ -   Accuracy : 52.72727272727273
05/22/2020 08:50:28 - INFO - __main__ -   Saving model with dev score : 62.712041733484156
```

FINE- TUNING BIO-BERT ON BIO-ASQ DATASET

```
Iteration:  96% 244/253 [05:43<00:12,  1.41s/it]
Iteration:  97% 245/253 [05:44<00:11,  1.40s/it]
Iteration:  97% 246/253 [05:45<00:09,  1.41s/it]
Iteration:  98% 247/253 [05:47<00:08,  1.40s/it]
Iteration:  98% 248/253 [05:48<00:07,  1.41s/it]
Iteration:  98% 249/253 [05:50<00:05,  1.41s/it]
Iteration:  99% 250/253 [05:51<00:04,  1.41s/it]
Iteration:  99% 251/253 [05:53<00:02,  1.41s/it]
Iteration: 100% 252/253 [05:54<00:01,  1.41s/it]
Iteration: 100% 253/253 [05:54<00:00,  1.40s/it]
Epoch: 100% 5/5 [29:34<00:00, 354.83s/it]
05/19/2020 21:27:07 - INFO - __main__ -   ***** Running predictions *****
05/19/2020 21:27:07 - INFO - __main__ -     Num orig examples = 935
05/19/2020 21:27:07 - INFO - __main__ -     Num split examples = 1206
05/19/2020 21:27:07 - INFO - __main__ -     Batch size = 8
05/19/2020 21:27:07 - INFO - __main__ -   Start evaluating
Evaluating:   0% 0/151 [00:00<?, ?it/s]05/19/2020 21:27:07 - INFO - __main__ -   Processing example: 0
Evaluating:  83% 125/151 [00:37<00:07,  3.29it/s]05/19/2020 21:27:45 - INFO - __main__ -   Processing example: 16
Evaluating: 100% 151/151 [00:45<00:00,  3.31it/s]
05/19/2020 21:27:52 - INFO - __main__ -   Writing predictions to: output/predictions.json
05/19/2020 21:27:52 - INFO - __main__ -   Writing nbest to: output/nbest_predictions.json
05/19/2020 21:28:09 - INFO - __main__ -   F1 score : 87.92892883058533
05/19/2020 21:28:09 - INFO - __main__ -   Accuracy : 76.2566844919786
05/19/2020 21:28:09 - INFO - __main__ -   Saving model with dev score         : 87.92892883058533
```

COMPARISON IN RESULT OF BERT AND BIO-BERT

	F1 SCORE	ACCURACY
BERT	62.712	52.727
BIO-BERT	87.92	76.25

TRAINING DATASET EXAMPLE

"qas": [{
"id": "**5325fdf0600967d132000001_002"**,
"question": "**What is the gold standard treatment for Iatrogenic male incontinence?**",
"answers": [{
"text": "AUS",
"answer_start": 371
}]}],
"context": "Slings in iatrogenic male incontinence: Current status. OBJECTIVES: The increasing number of prostatectomies entails an increasing number of patients suffering from iatrogenic incontinence despite improved surgical techniques. The severity of this problem often requires invasive treatments such as periurethral injection of bulking agents, artificial urinary sphincter **(AUS) implantation, and sub-urethral sling positioning. The artificial urethral** sphincter has represented, until today, the gold standard but, in the recent years, sling systems have been investigated as minimally invasive alternative options. Today, three different sling procedures are commonly performed: bone-anchored, readjustable, and trans-obturator slings systems. The aim of this review is to critically report the current status of sling systems in the treatment of iatrogenic male incontinence. MATERIALS AND METHODS: MEDLINE and PubMed databases were searched and all articles between 1974 and 2009 were evaluated. RESULTS: With regard to bone-anchored, readjustable, and trans-obturator slings systems, cure rates ranged between 58.0% and 86.0%, 55.5% and 73.0%, and 40.0% and 63.0%, respectively, while major complication rates ranged between 0 and 14.5%, 10.0 and 22.2%, and 0 and 10.0%, respectively. CONCLUSIONS: Suburethral slings are the only alternative techniques which can be favorably compared with the AUS, showing more advantages with respect to AUS implantations which are mainly represented by a quick and less invasive approach, low morbidity, and low costs. In spite of the difficulty in identifying the most effective sling procedure, overall, sling systems can be recommended for patients with persistent mild or moderate incontinence. ....”
},

N-BEST PREDICTIONS

BERT

"5325fdf0600967d132000001_002": [
    {
        "text": "aus) implantation , and sub - urethral sling positioning . the artificial urethral sphincter",
        "probability": 0.392093349439298,
        "start_logit": 7.307885646820068,
        "end_logit": 6.496368885040283
    },
    {
        "text": "artificial urethral sphincter",
        "probability": 0.23292811224625753,
        "start_logit": 6.787115573883057,
        "end_logit": 6.496368885040283
    },
    {
        "text": "aus",
        "probability": 0.20102315384979216,
        "start_logit": 7.307885646820068,
        "end_logit": 5.828289031982422
    },

BIO-BERT

"5325fdf0600967d132000001_002": [
    {
        "text": "aus",
        "probability": 0.4380434270051029,
        "start_logit": 3.1561317443847656,
        "end_logit": 5.1763505935668945
    },
    {
        "text": "aus) implantation , and sub - urethral sling positioning . the artificial urethral sphincter",
        "probability": 0.26625265832018236,
        "start_logit": 3.1561317443847656,
        "end_logit": 4.678478240966797
    },
    {
        "text": "artificial urethral sphincter",
        "probability": 0.11724444087628094,
        "start_logit": 2.335947036743164,
        "end_logit": 4.678478240966797
    },

DEV DATASET EXAMPLE


```
"qas": [
        {
          "id": "56b1f4300a360a5e4500001b_020",
          "question": "Which peptide plays a pivotal role in human cystatin C fibrillization?",
          "answers": [
            {
              "text": "LQVVR",
              "answer_start": 767
            }
          ]
        }
      ],
      "context": "The pentapeptide LQVVR plays a pivotal role in human cystatin C fibrillization. Human cystatin C (HCC) is a low molecular weight member of the cystatin family (type2). HCC consists of 120 amino acids. Normally it is an inhibitor of cysteine proteases, but in pathological conditions it forms amyloid fibrils in brain arteries of young adults. An 'aggregation-prone' pentapeptide ((47)LQVVR(51)) was located within the HCC sequence using AmylPred, an 'aggregation-prone' peptide prediction algorithm developed in our lab. This peptide was synthesized and self-assembled into amyloid-like fibrils in vitro, as electron microscopy, X-ray fiber diffraction, Attenuated Total Reflectance Fourier-Transform Spectroscopy and Congo red staining studies reveal. Thus, the (47)LQVVR(51) peptide seems to have an important role in HCC fibrillization."
```

N-BEST PREDICTIONS

BERT

"56b1f4300a360a5e4500001b_020": [
{
"text": "120",
"probability": 0.3069954292715951,
"start_logit": 0.7720412611961365,
"end_logit": -1.608727216720581
},
{
"text": "lqvvr",
"probability": 0.23284531520186857,
"start_logit": -0.10071486979722977,
"end_logit": -1.0124295949935913
},
{
"text": "lqvvr plays a pivotal role in human cystatin c fibrillization . human cystatin c (hcc) is a low molecular weight member of the cystatin family (type2) . hcc consists of 120",
"probability": 0.12826221563318593,
"start_logit": -0.10071486979722977,
"end_logit": -1.608727216720581
},

BIO-BERT

"56b1f4300a360a5e4500001b_020": [
{
"text": "lqvvr",
"probability": 0.8182389457019558,
"start_logit": 1.7428301572799683,
"end_logit": 1.013503909111023
},
{
"text": "the pentapeptide lqvvr",
"probability": 0.13748406421406154,
"start_logit": -0.040816232562065125,
"end_logit": 1.013503909111023
},
{
"text": "lqvvr plays a pivotal role in human cystatin c fibrillization . human cystatin c (hcc) is a low molecular weight member of the cystatin family (type2) . hcc consists of 120 amino acids . normally it is an inhibitor of cysteine proteases , but in pathological conditions it forms amyloid fibrils in brain arteries of young adults . an ' aggregation - prone ' pentapeptide ((47) lqvvr (51)) was located within the hcc sequence using amylpred",
"probability": 0.009112227093689043,
"start_logit": 1.7428301572799683,
"end_logit": -3.4840333461761475
},

APPLICATION & BUSINESS VALUE

APPLICATION

- Information Retrieval
- Entity Recognition
- Medical / clinical Chat bots
- Suggest advise on a scanned reports(findings of the report needs to be highlighted.)
- Text Summarization

"

ROADBLOCKS / LIMITATIONS

LIMITATIONS OF OUR PROPOSED MODEL

- **Fails to answer such questions were reasoning is involved.**

- **As bert highlights the span of text in the given context, so fails to concatenate the answer randomly spread in the document.**

NEXT STEPS

- **Dockerize the implementation**
- **Extension of bio-bert on other downstream task such as sentiment analysis, entity relations.**
- **Exploration of latest state of art models. (Roberta, XLNet)**

REFERENCES

1. Jay Alammar. 2018. The Illustrated Transformer [Blog post]. (2018). https://jalammar.github.io/illustrated-transformer/

2. Jay Alammar. 2018. The Illustrated BERT, ELMo, and co. (How NLP Cracked Transfer Learning) [Blog post]. (2018)http://jalammar.github.io/illustrated-bert/

3. Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, Jaewoo Kang, BioBERT: a pre-trained biomedical language representation model for biomedical text mining, Bioinformatics, Volume 36, Issue 4, 15 February 2020, Pages 1234–1240, https://doi.org/10.1093/bioinformatics/btz682

4. Devlin, Jacob, Ming-Wei Chang, Kenton Lee and Kristina Toutanova. “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.” ArXiv abs/1810.04805 (2019): n. pag.

REFERENCES

5. Chris McCormick. 2020. Question Answering with a Fine-Tuned BERT[Blog post]. (2020). https://mccormickml.com/2020/03/10/question-answering-with-a-fine-tuned-BERT/

The complete code repository is present on my github account.

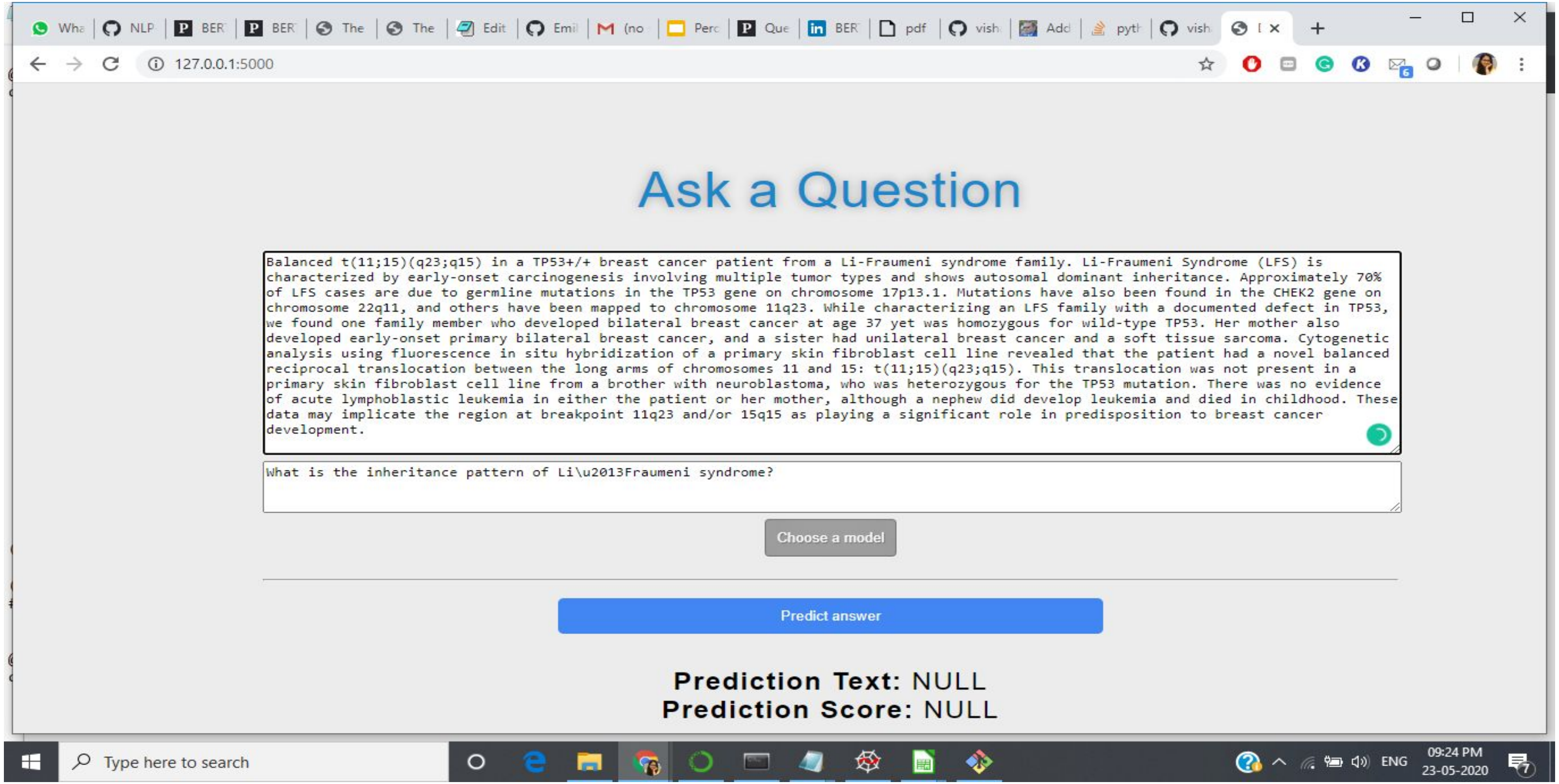
Link for the same:

https://github.com/vishakhagupta10/mid-work

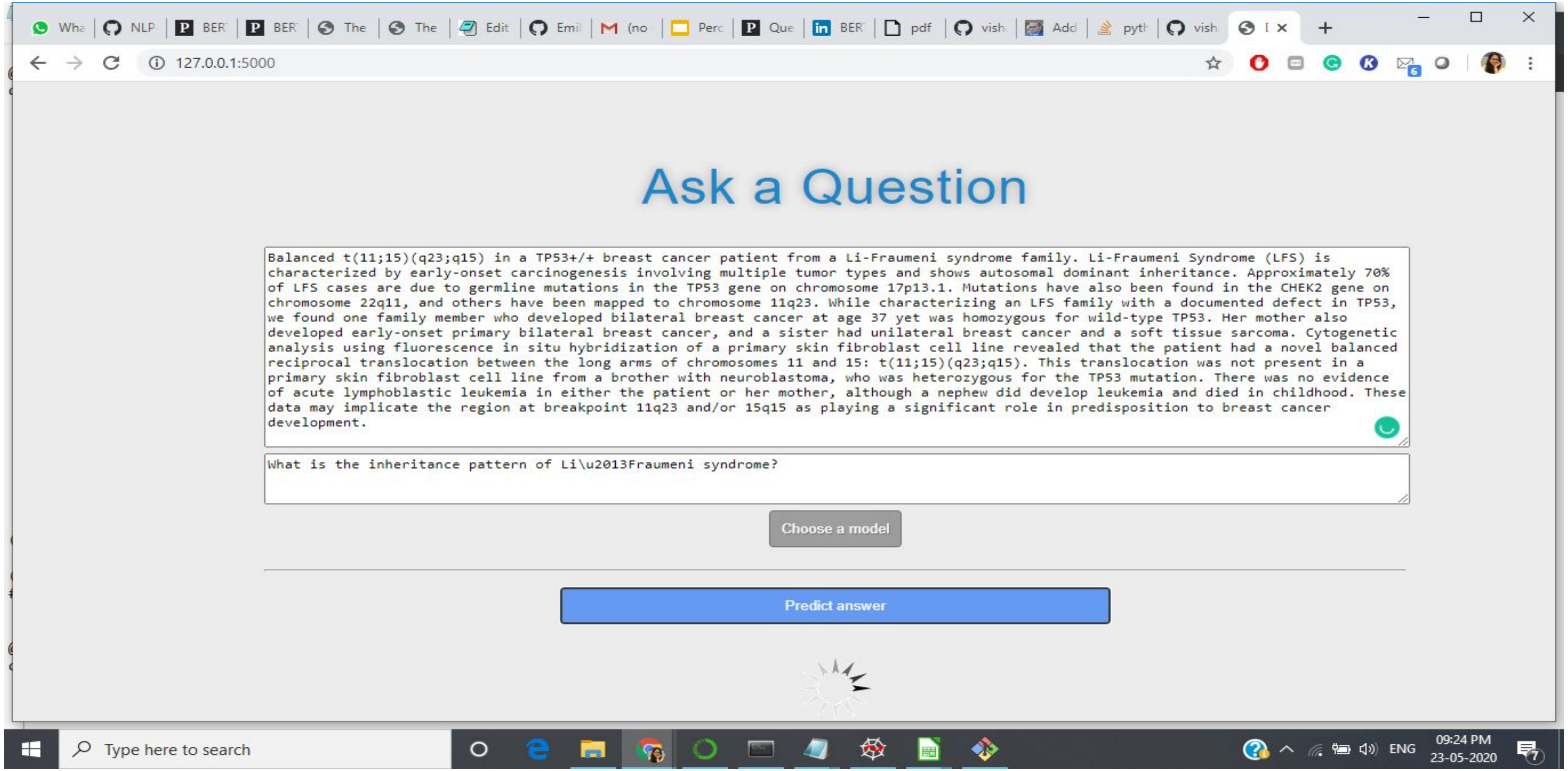
THANK YOU!

MODEL REPRESENTATION THROUGH FLASK

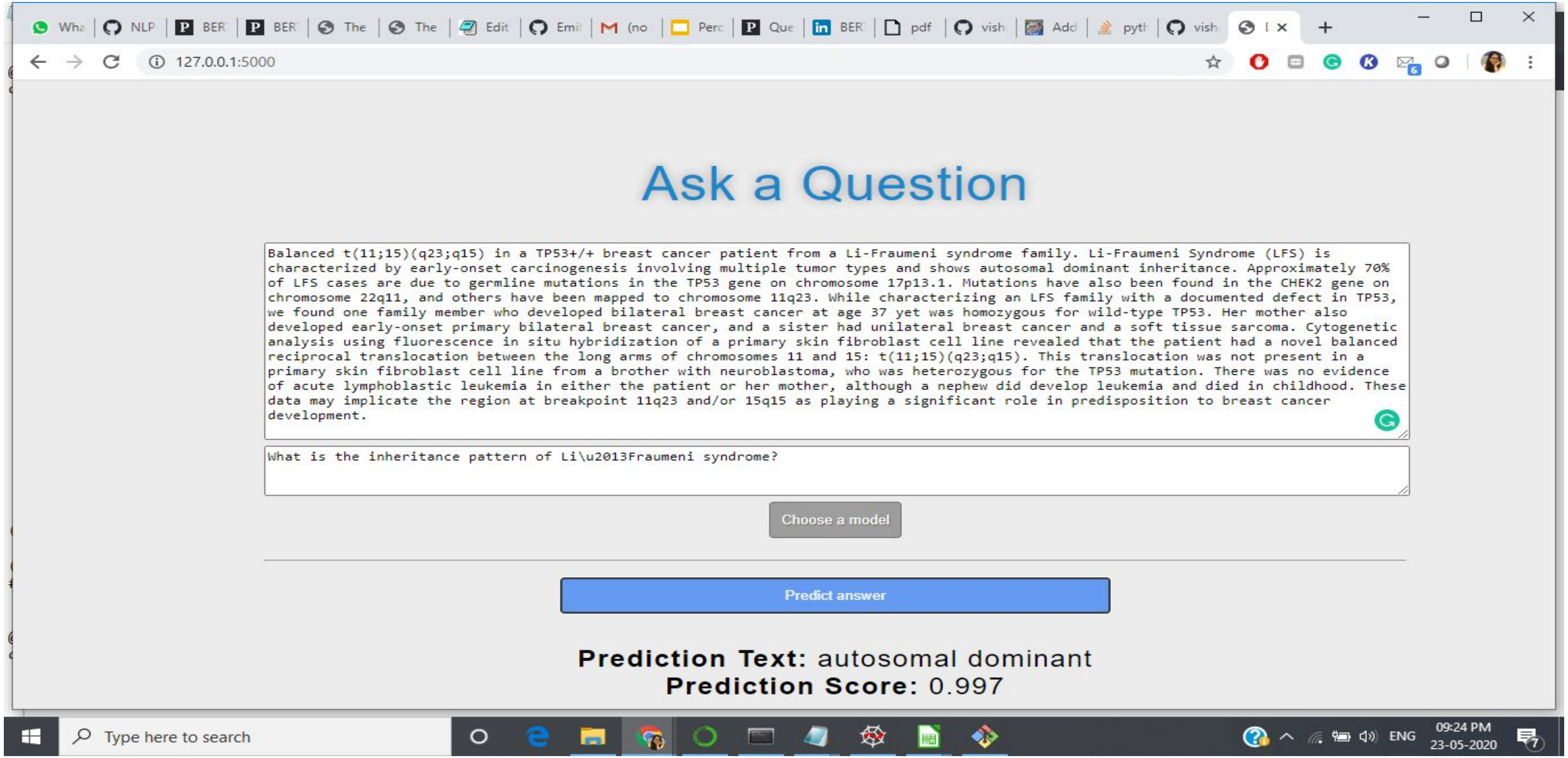
GLIMPSES

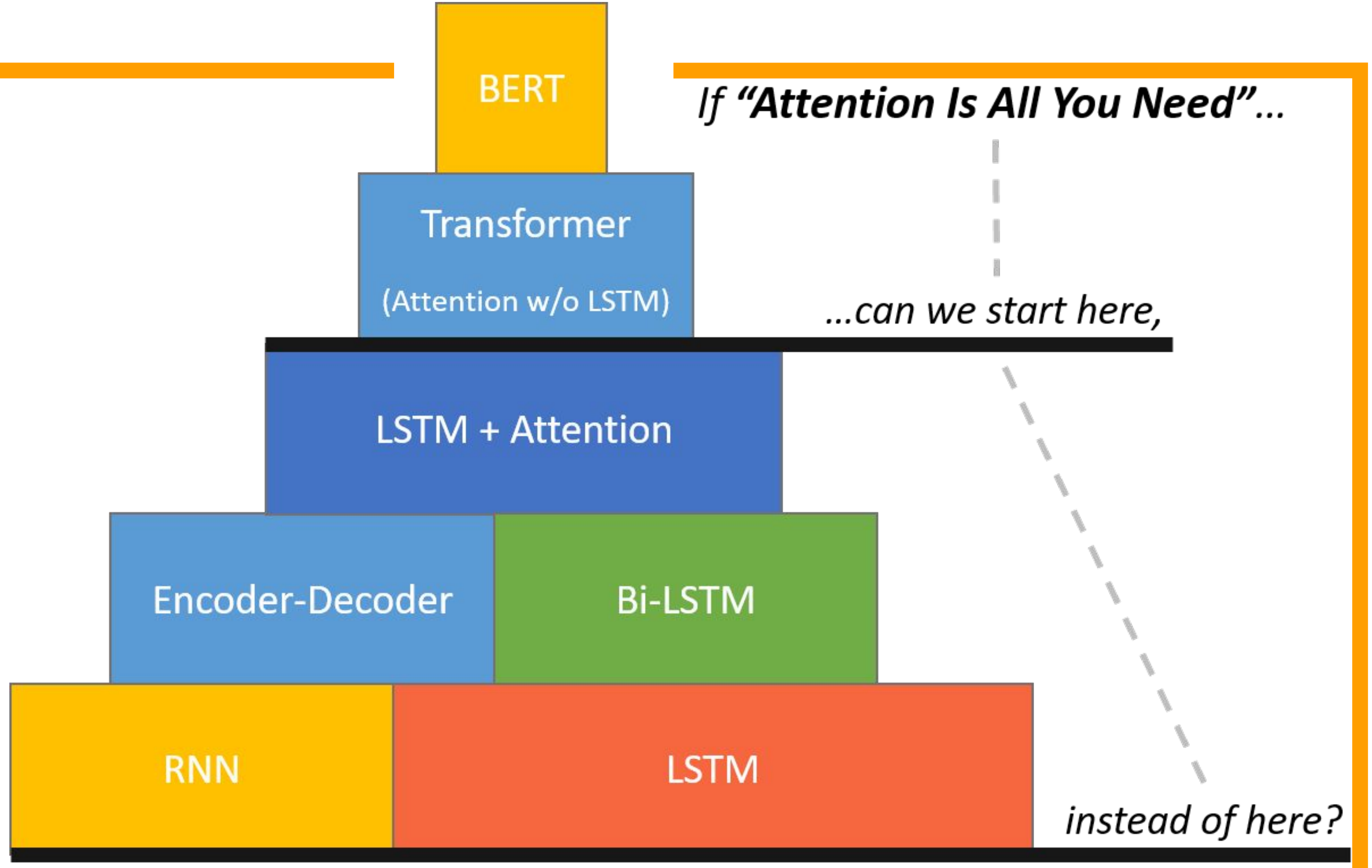

127.0.0.1:5000
Ask a Question
Balanced t(11;15)(q23;q15) in a TP53+/+ breast cancer patient from a Li-Fraumeni syndrome family. Li-Fraumeni Syndrome (LFS) is characterized by early-onset carcinogenesis involving multiple tumor types and shows autosomal dominant inheritance. Approximately 70% of LFS cases are due to germline mutations in the TP53 gene on chromosome 17p13.1. Mutations have also been found in the CHEK2 gene on chromosome 22q11, and others have been mapped to chromosome 11q23. While characterizing an LFS family with a documented defect in TP53, we found one family member who developed bilateral breast cancer at age 37 yet was homozygous for wild-type TP53. Her mother also developed early-onset primary bilateral breast cancer, and a sister had unilateral breast cancer and a soft tissue sarcoma. Cytogenetic analysis using fluorescence in situ hybridization of a primary skin fibroblast cell line revealed that the patient had a novel balanced reciprocal translocation between the long arms of chromosomes 11 and 15: t(11;15)(q23;q15). This translocation was not present in a primary skin fibroblast cell line from a brother with neuroblastoma, who was heterozygous for the TP53 mutation. There was no evidence of acute lymphoblastic leukemia in either the patient or her mother, although a nephew did develop leukemia and died in childhood. These data may implicate the region at breakpoint 11q23 and/or 15q15 as playing a significant role in predisposition to breast cancer development.
What is the inheritance pattern of Li\u2013Fraumeni syndrome?
Choose a model
Predict answer
Prediction Text: NULL
Prediction Score: NULL
Type here to search
ENG
09:24 PM
23-05-2020

GLIMPSES

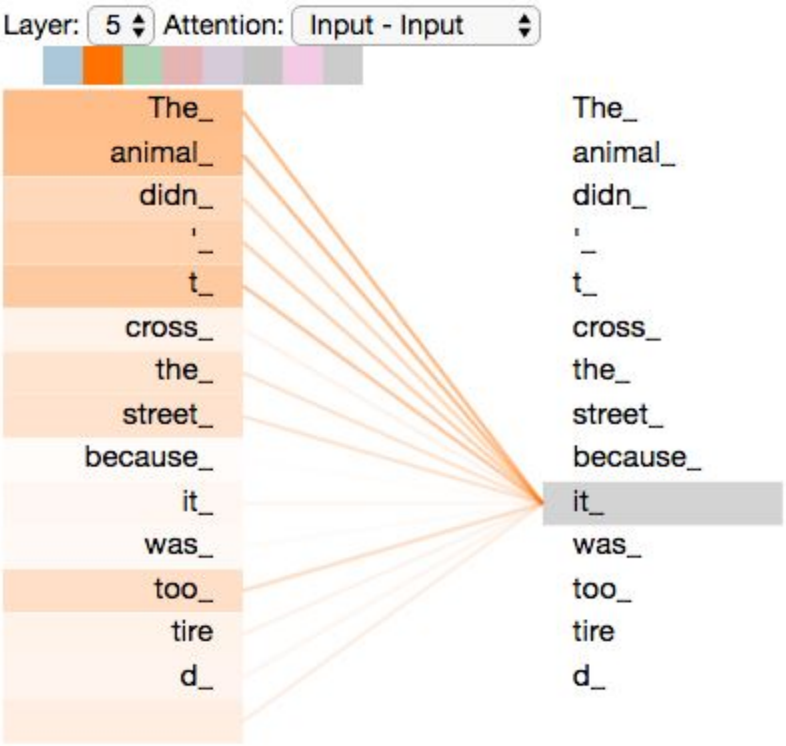

127.0.0.1:5000
Ask a Question
Balanced t(11;15)(q23;q15) in a TP53+/+ breast cancer patient from a Li-Fraumeni syndrome family. Li-Fraumeni Syndrome (LFS) is characterized by early-onset carcinogenesis involving multiple tumor types and shows autosomal dominant inheritance. Approximately 70% of LFS cases are due to germline mutations in the TP53 gene on chromosome 17p13.1. Mutations have also been found in the CHEK2 gene on chromosome 22q11, and others have been mapped to chromosome 11q23. While characterizing an LFS family with a documented defect in TP53, we found one family member who developed bilateral breast cancer at age 37 yet was homozygous for wild-type TP53. Her mother also developed early-onset primary bilateral breast cancer, and a sister had unilateral breast cancer and a soft tissue sarcoma. Cytogenetic analysis using fluorescence in situ hybridization of a primary skin fibroblast cell line revealed that the patient had a novel balanced reciprocal translocation between the long arms of chromosomes 11 and 15: t(11;15)(q23;q15). This translocation was not present in a primary skin fibroblast cell line from a brother with neuroblastoma, who was heterozygous for the TP53 mutation. There was no evidence of acute lymphoblastic leukemia in either the patient or her mother, although a nephew did develop leukemia and died in childhood. These data may implicate the region at breakpoint 11q23 and/or 15q15 as playing a significant role in predisposition to breast cancer development.
What is the inheritance pattern of Li\u2013Fraumeni syndrome?
Choose a model
Predict answer
Type here to search
ENG
09:24 PM
23-05-2020

GLIMPSES


127.0.0.1:5000
Ask a Question
Balanced t(11;15)(q23;q15) in a TP53+/+ breast cancer patient from a Li-Fraumeni syndrome family. Li-Fraumeni Syndrome (LFS) is characterized by early-onset carcinogenesis involving multiple tumor types and shows autosomal dominant inheritance. Approximately 70% of LFS cases are due to germline mutations in the TP53 gene on chromosome 17p13.1. Mutations have also been found in the CHEK2 gene on chromosome 22q11, and others have been mapped to chromosome 11q23. While characterizing an LFS family with a documented defect in TP53, we found one family member who developed bilateral breast cancer at age 37 yet was homozygous for wild-type TP53. Her mother also developed early-onset primary bilateral breast cancer, and a sister had unilateral breast cancer and a soft tissue sarcoma. Cytogenetic analysis using fluorescence in situ hybridization of a primary skin fibroblast cell line revealed that the patient had a novel balanced reciprocal translocation between the long arms of chromosomes 11 and 15: t(11;15)(q23;q15). This translocation was not present in a primary skin fibroblast cell line from a brother with neuroblastoma, who was heterozygous for the TP53 mutation. There was no evidence of acute lymphoblastic leukemia in either the patient or her mother, although a nephew did develop leukemia and died in childhood. These data may implicate the region at breakpoint 11q23 and/or 15q15 as playing a significant role in predisposition to breast cancer development.
What is the inheritance pattern of Li\u2013Fraumeni syndrome?
Choose a model
Predict answer
Prediction Text: autosomal dominant
Prediction Score: 0.997
Type here to search
ENG
09:24 PM
23-05-2020


BERT
Transformer
(Attention w/o LSTM)
LSTM + Attention
Encoder-Decoder
Bi-LSTM
RNN
LSTM
If "Attention Is All You Need"...
...can we start here,
instead of here?

SELF- ATTENTION
(An Overview)



Self Attention layer, takes as input a position injected naive form of embeddings and outputs more context aware embeddings.

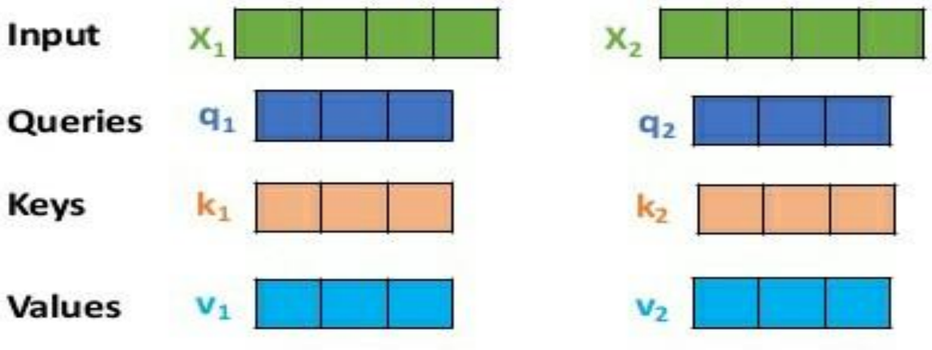
For every input word a score is calculated with respect to every word in the sentence, by "relevance"

The model processes each word (each position in the input sequence), self attention allows it to look at other positions in the input sequence for clues that can help lead to a better encoding for this word.

Figure 3: Self Attention (the word "it" is encoded iin encoder #5 (the top encoder in the stack), part of the attention mechanism was focusing on "The Animal", and baked a part of its representation into the encoding of "it".) [1]

SELF- ATTENTION (In Detail)

- Attention maps a query and a set of key-value pairs to an output
  - query, keys, and output are all vectors



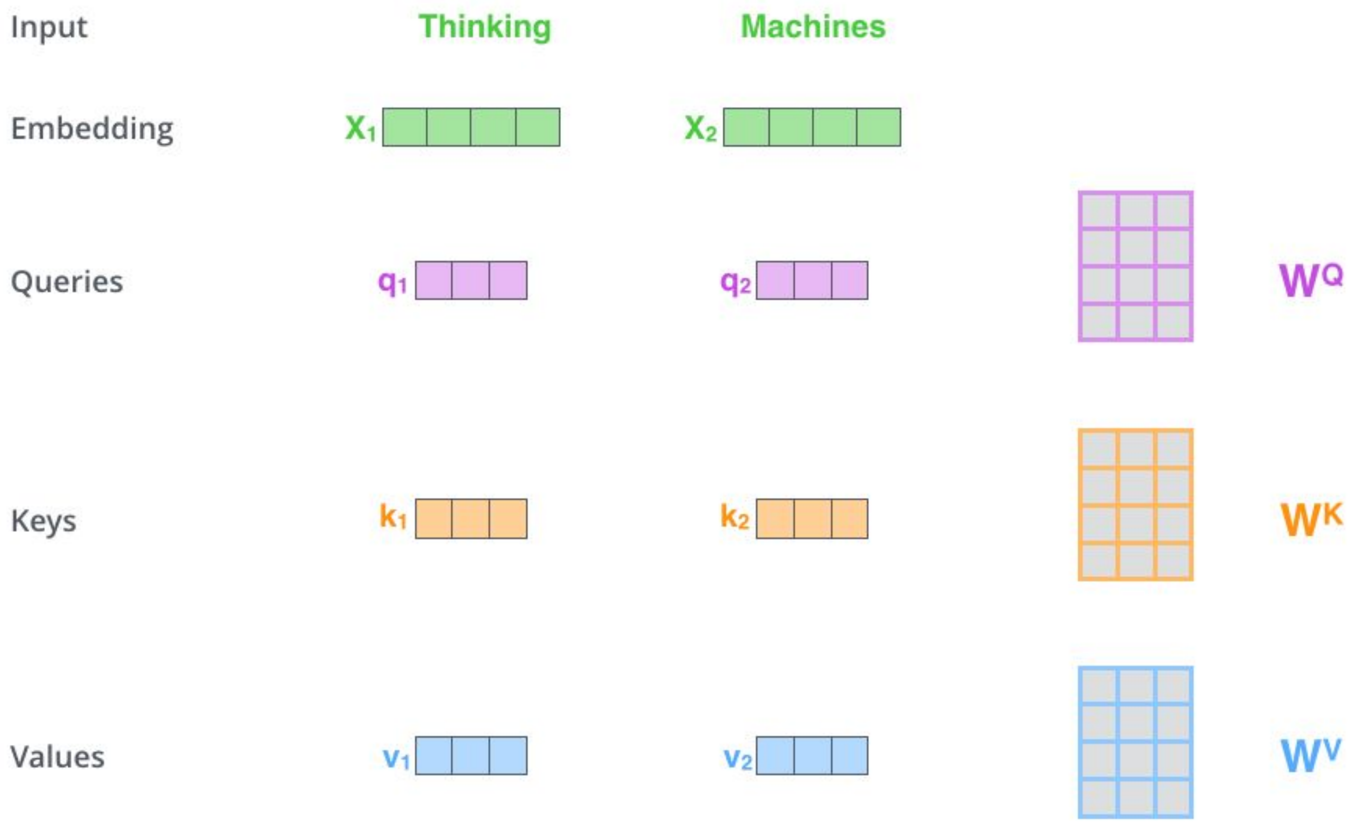
Use matrices W^Q, W^K and W^V to project input into query, key and value vectors

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

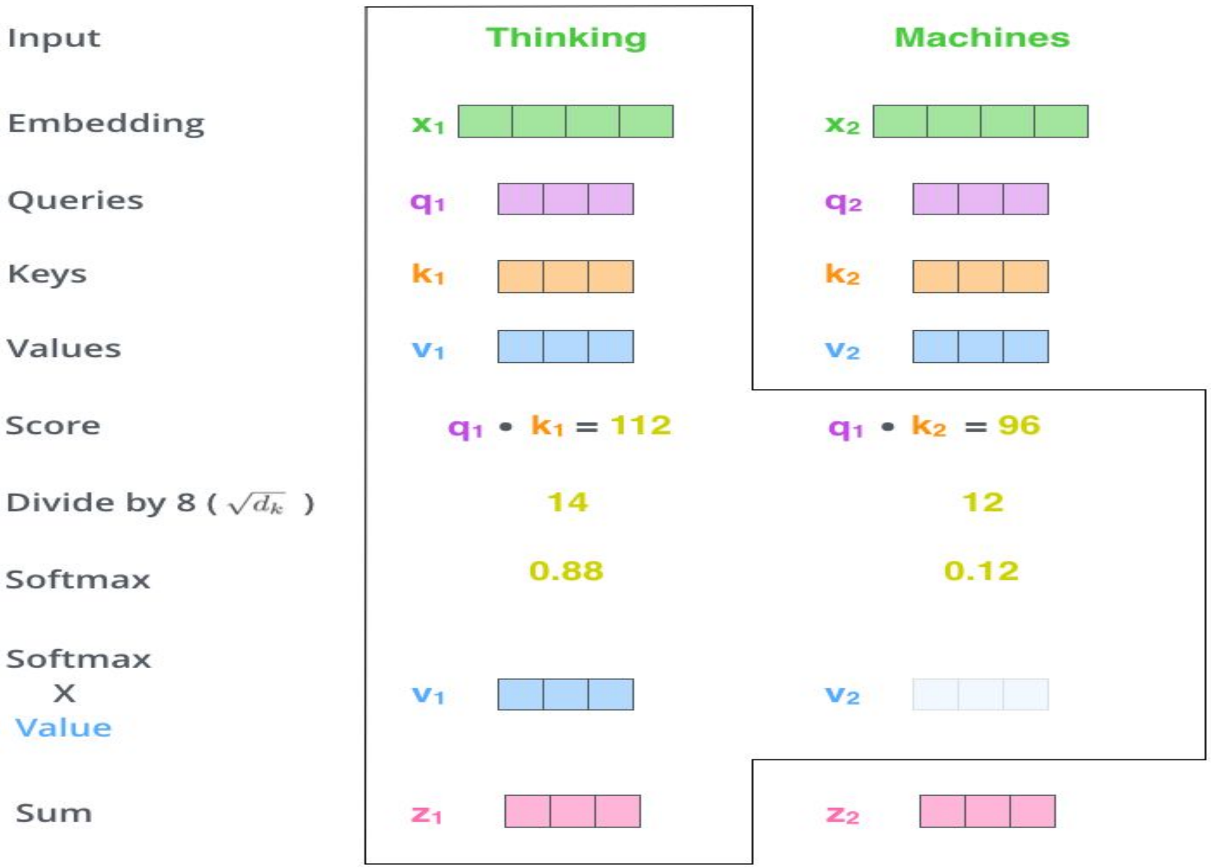
d_k is the dimension of key vectors

13

SELF- ATTENTION (In Detail)



SELF- ATTENTION (In Detail)



MULTI HEAD ATTENTION

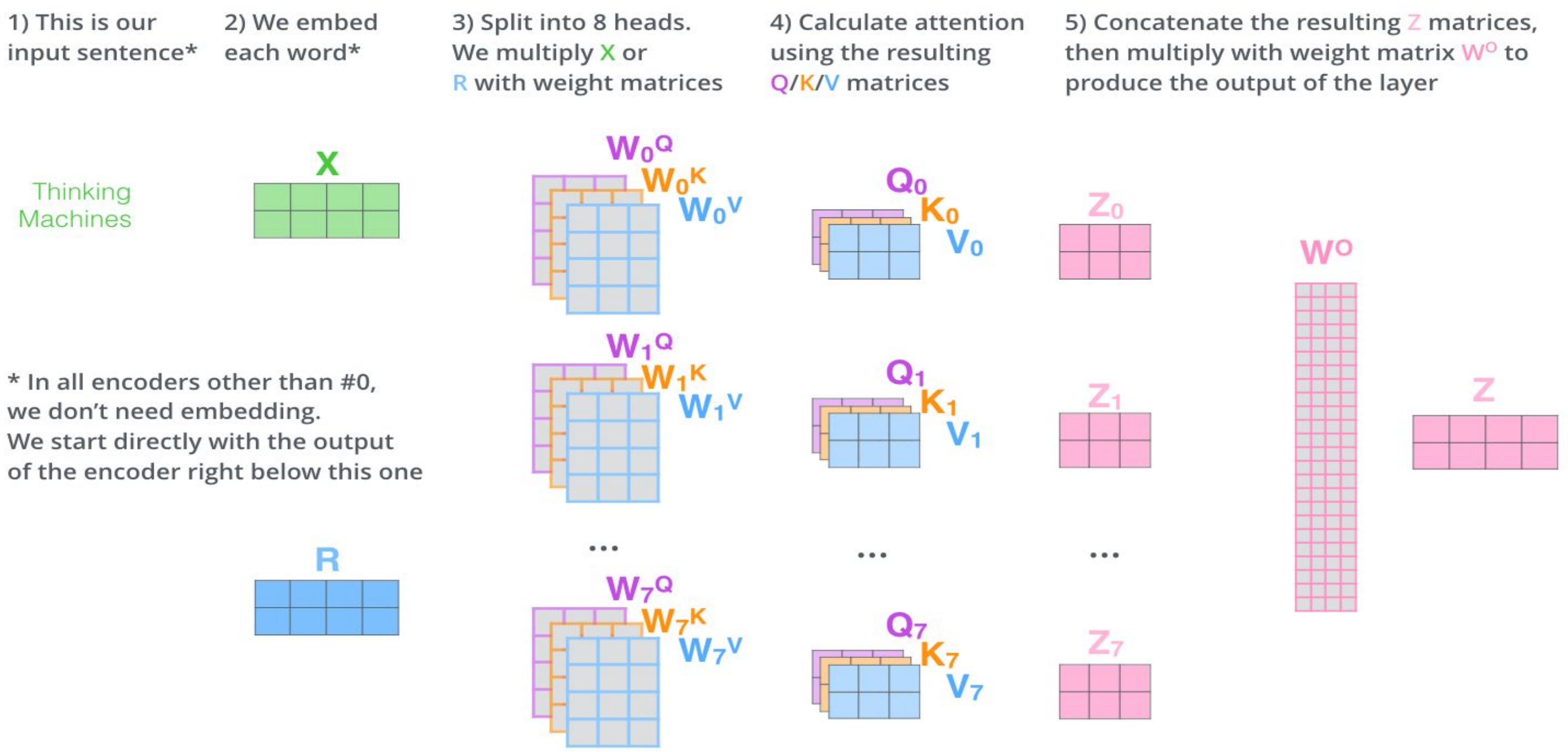


Figure 4: Multihead attention detailed view [1]