به نام خدا

آزمونک شماره سه و چهار - پردازش زبان طبیعی

نام و نام خانوادگی: ..... ۲۹ / ۲/ ۱۴۰۳

سوال ۱) در الگوریتم Beam-Search اگر سایز beam را افزایش بدهیم، کدام موارد اتفاق می‌افتد؟(جواب می‌تواند چند گزینه باشد.)(۴ نمره-پاسخ نادرست منفی یک نمره) گزینه a,b,c

a. این الگوریتم کندتر اجرا می‌شود.

b. این الگوریتم به حافظه بیشتری نیاز خواهد داشت.

c. این الگوریتم بطور کلی جواب بهتری را پیدا خواهد کرد.

پاسخ

سوال ۲) برای کدام کاربردهای زیر معماری Encoder-Decoder مناسب است؟(جواب می‌تواند چند گزینه باشد.)(۴ نمره-پاسخ نادرست منفی یک نمره) گزینه a,d

a. Summarization

b. Image captioning

c. Sentiment Analysis

d. Machine translation

سوال ۳) تصویر زیر معماری یک بلوک از بخش انکودر ترنسفورمر را نشان می‌دهد. جاهای خالی را پر کرده و کاربرد آن‌ها را بطور مختصر توضیح دهید. (۲۰ نمره)

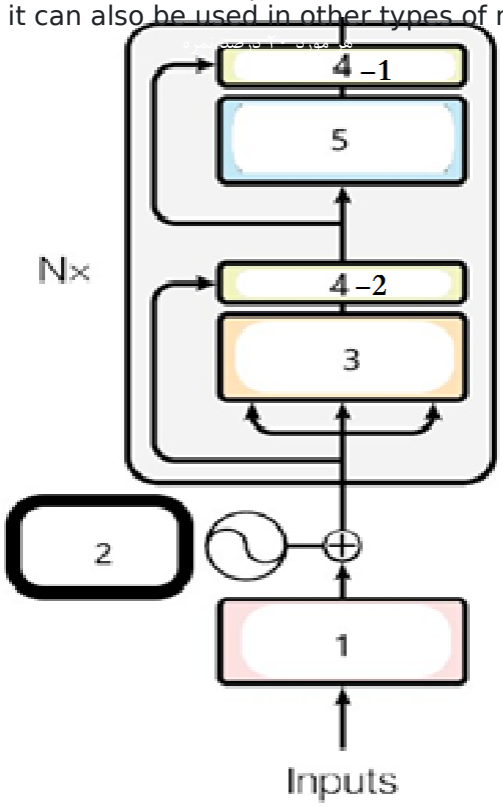
2- **Positional encoding/embedding**: it provides the transformer model with information about where the words are in the input sequence.

3- **Multi-headed Self-attention:** where the representation of a sequence (or sentence) is computed by relating different words in the same sequence.

4-2- **Residual Connections:** Residual connections alleviate unstable gradients and aids in preserving information across different layers of the network.

4-1- **Layer Normalization:** layer norm just normalizes each layer with the mean and variance of each activation. Layer Normalization improves both the training time and the generalization performance

5- **Feed Forwards:** used to transform the output of the self-attention mechanism into the final output of the model. However, it can also be used in other types of models and tasks.



مثال:

۱- امبدینگ توکن‌ها: این بخش متن ورودی را به یک بازنمایی عددی نگاشت می‌کند تا شبکه بتواند با اعداد کار کند.

سوال ۴) کدام ماژول یا ماژول‌ها از معماری ترنسفرمر مشابه مکانیسم gating (مشابه آنچه در lstm و gru دیدیم) عمل می‌کند؟ چرا؟ (۹ نمره)

سوال ۵) چگونه می‌توان دو تسک NLI و NER را به وسیله‌ی مدل برت انجام داد؟ ترجیحا با رسم شکل پاسخ دهید. (۱۵ نمره)

پاسخ

In the Transformer model, residual connections act as a gating mechanism, similar to those in GRUs and LSTMs. These connections allow the model to pass a portion of the input directly to deeper layers, effectively bypassing certain transformations. This mechanism helps in mitigating the vanishing gradient problem and aids in preserving information across different layers of the network.

(NLI یا Natural Language Inference تسکی است که در آن دو جمله ارائه می‌گردد و می‌بایست نسبت جمله‌ی دوم با جمله‌ی اول تشخیص داده شود. مثلا تشخیص داده شود که آیا از جمله‌ی اول می‌توان جمله‌ی دوم را نتیجه گرفت یا خیر).

سوال ۶) به سوالات زیر پاسخ کوتاه بدهید. هر کدام ۶ نمره

a. دو منبع مهم برای آموزش مترجم ماشینی را نام ببرید.

پاسخ

NLI:

.اعمال شود نیز پاسخ درست لحاظ می‌گردد (C یا بدون C ها با T) روی خروجی توکن‌ها feed forward اگر لایه

NER:

۱. Parallel text, 2. Bitext

b. Context vector در معماری Encoder-Decoder چیست؟

چکیده ورودی را به دیکودر منتقل می‌کند. به عبارت دیگر contextualized representation از کل ورودی است که در آخرین وضعیت مخفی قرار دارد.

c. چرا نمونه برداری از توزیع softmax برای تولید خروجی در Decoder مناسب نیست؟
دو دلیل میتوان متصور شد: ۱. این روش حریصانه است و شاید خروجی آخر بهترین نباشد. ۲. ممکن است جواب خیلی بد هم تولید شود. چون به هر حال شانس انتخاب بد، هرچند کم وجود دارد.

d. در الگوریتم Beam-Search وقتی یکی از خروجی‌ها به توکن </S> می‌رسد الگوریتم در ادامه چی کار می‌کند؟
در اینجا، این دنباله از frontier حذف می‌شود و سایز beam یکی کاهش پیدا می‌کند. و این دنباله به عنوان یک کاندید نهایی انتخاب می‌شود. در آخر بهترین کاندید نهایی به عنوان جواب انتخاب خواهد شد.

e. در ترجمه ماشینی، اگر از Beam-Search بدون استفاده از Sentence normalization بهره ببریم چه اتفاقی خواهد افتاد؟
در اینصورت ترجمه‌های کوتاه‌تر ترجیح داده خواهند شد.

f) تفاوت self attention و cross attention در چیست و در ترنسفرمر ها از کدام نوع attention استفاده می‌گردد؟

g) دلیل استفاده از Muti head self attention چیست؟

h) دلیل استفاده از ماسک در مکانیسم attenion، هنگام آموزش دیکدر در معماری ترنسفورمر چیست؟ (

پاسخ

e- Self-Attention: Focuses on relations within the same sequence. Cross-Attention: Consider the relations between different sequences – Both
g- Each head can learn to focus on different aspects, e.g., subject and object of a sentence.
h- The attention mask in the Transformer decoder makes sure that during the self-attention part, each token can only attend on tokens that come before it in the sequence.