

(۱)

در این مدل، یک ترکیب از مدل های Google، Mozilla، Nemo و Whisper داریم که به ترتیب به خروجی هر کدام یک وزن می دهد یا خروجی پرتکرارترین را انتخاب می کند. این روش یک ماتریس از کلمات می سازد. برای توسعه این مدل هم علاوه بر تأمین داده ها، وزن بیشتری به مدل اصلی ما اختصاص داده می شود.

(۲)

می تواند مشابه سوال ۱ عمل کند یا بر اساس احتمال درستی متن های خروجی، مدل ها را مرتب کند و سپس 16 خروجی با بیشترین احتمال را انتخاب کند.

همچنین در یک روش دیگر ابتدا خروجی های هر مدل بر اساس اولویت هایی که مدل خود تعیین کرده است مرتب می شوند، بعد از این ترتیب اولویت ها برای انتخاب 16 خروجی نهایی استفاده می شود.

(۳)

• **Normalizer:**

- عملیات نرمال سازی متن ورودی را به شکل استاندارد تبدیل می کند. این عملیات شامل حذف punctuation، تبدیل حروف به حروف کوچک، حذف فاصله های اضافی و موارد مشابه است.

• **Formalizer:**

- این ابزار برعکس قبلی عمل می کند و متن را به یک قالب استانداردتر تبدیل می کند. برای مثال، اگر یک متن به زبان عامیانه باشد، Formalizer می تواند آن را به زبان رسمی تغییر دهد و می تواند برای text generation یا sentiment analysis استفاده شود.

• **Lemmatizer:**

- این عملیات شامل تبدیل کلمات به شکل اصلی یا ریشه ی آن ها است. به عنوان مثال، "بهترین" به "خوب" تبدیل می شود و یا فعل ها به بن مضارع و ماضی تبدیل می شوند.

• **Stemmer:**

- این فرایند ساده تر از Lemmatizer است و به دنبال حذف پسوندها یا پیشوندهای یک کلمه است تا به ریشه آن یا stem دست پیدا کند.

• **Chunker:**

- chunker کلمات یا نشانه ها را بر اساس برجسب های part of speech یا سایر الگوهای زبانی آن ها در تکه های نحوی (مانند اسم ها، صفات، افعال، و...) شناسایی و گروه بندی می کند. این ابزار معمولاً برای شناسایی named entity ها یا استخراج اطلاعات استفاده می شود.

• **Tagger:**

- Tagger به اجزای مختلف یک جمله برجسب می زند. این برجسب ها معمولاً نشان دهنده نوع و نقش کلمات در جمله هستند.

• **POSTagger:**

- این اصطلاح دیگری برای برجسب گذاری است که به طور خاص به ابزار یا سیستمی اشاره دارد که part-of-speech tagging را روی داده های متنی انجام می دهد.

• **Embedder:**

- مدلی است که برای تبدیل کلمات، عبارات یا جملات به نمایش های عددی متراکم یا بردارهایی چند بعدی به نام embeddings استفاده می شود. این نمایش اطلاعات معنایی و متنی را در مورد متن دریافت می کنند و برای مثال با استفاده از شباهت کسینوسی بین بردار دو کلمه می توان به شباهت معنای آنها پی برد.

• Word Embedder :

- نوعی از Embedder است که برای تک تک کلمات در یک vocabulary جاسازی می کند. این word embedding ها بردارهایی در فضای چند بعدی هستند و معمولاً بر اساس معنای کلمات در جمله و رابطه آن ها با کلمات دیگر ساخته می شوند. Word Embedder ها اغلب با استفاده از تکنیک هایی مانند Word2Vec یا GloVe بر روی مجموعه های متنی بزرگ از قبل آموزش داده می شوند.

• Parser :

- ابزار یا الگوریتمی است که ساختار دستوری جملات را برای تعیین روابط نحوی آنها تجزیه و تحلیل می کند. این برای کارهایی مانند تجزیه جملات به درختان تجزیه یا استخراج اطلاعات ساختاریافته از متن بسیار مهم است.