52000782 – Nguyễn Quốc Minh

1. Mô hình Bert

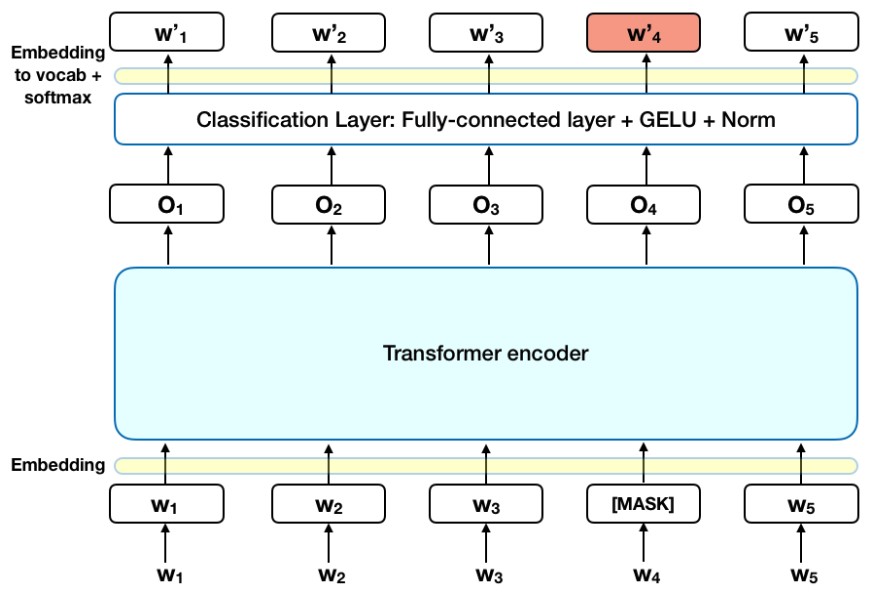
1.1 Khái niệm

Mô hình BERT (Bidirectional Encoder Representations from Transformers) là một mô hình học sâu (deep learning) trong lĩnh vực xử lý ngôn ngữ tự nhiên (NLP), được giới thiệu bởi các nhà nghiên cứu tại Google vào năm 2018. BERT đã đạt được những kết quả rất ấn tượng trong nhiều tác vụ NLP, bao gồm cả phân loại văn bản, dự đoán từ tiếp theo và sinh văn bản.

Mô hình BERT được huấn luyện trên một lượng lớn dữ liệu văn bản đa ngôn ngữ (được thu thập từ các trang web khác nhau) bằng cách sử dụng một kiến trúc mạng neural rất phổ biến là Transformer. Một điểm nổi bật của BERT đó là nó có khả năng hiểu được ngữ nghĩa của từ trong ngữ cảnh của câu, nhờ vào việc áp dụng cơ chế tự đánh giá lại (self-attention) của Transformer.

Mô hình BERT thực sự ấn tượng khi nó giúp cải thiện đáng kể các kết quả của nhiều tác vụ NLP so với các mô hình trước đó. Một trong những ứng dụng phổ biến của BERT là sử dụng nó cho việc phân loại văn bản, trong đó BERT được sử dụng để rút trích các đặc trưng của câu và sau đó được kết hợp với một mô hình phân loại để dự đoán nhãn của câu.

Mô hình BERT đã được cải tiến và phát triển thành nhiều phiên bản khác nhau để phù hợp với các nhiệm vụ NLP khác nhau. Ví dụ, mô hình RoBERTa là một phiên bản cải tiến của BERT, đạt được kết quả tốt hơn đáng kể trên nhiều tác vụ NLP.



1.2 Tại sao lại có Mô hình BERT

Mô hình BERT được phát triển để giải quyết các vấn đề trong lĩnh vực xử lý ngôn ngữ tự nhiên (NLP) như hiểu và sinh văn bản tự động, phân loại văn bản, dịch máy, và nhiều ứng dụng khác. Trong quá khứ, các mô hình NLP dựa trên mô hình ngôn ngữ (language model) như Word2Vec hay GloVe đã đạt được nhiều thành công. Tuy nhiên, các mô hình này vẫn có những hạn chế như không hiểu được ngữ cảnh của câu hay không đủ thông minh để thực hiện các tác vụ phức tạp hơn.

BERT đã đưa ra một tiếp cận mới trong việc sử dụng mô hình ngôn ngữ để giải quyết các vấn đề NLP bằng cách sử dụng kiến trúc Transformer và huấn luyện trên một lượng lớn dữ liệu. Các mô hình trước đó chỉ huấn luyện trên một khoảng thời gian ngắn và trên một lượng dữ liệu nhỏ hơn. BERT đã cho thấy rằng việc huấn luyện trên lượng dữ liệu lớn và thêm kiến trúc Transformer có thể đưa ra kết quả tốt hơn cho nhiều tác vụ NLP.

1.3 Ngữ cảnh (Contextual) và vai trò trong NLP

Ngữ cảnh (Context) là khái niệm chỉ thông tin bổ sung liên quan đến một đối tượng, sự việc hoặc hành động trong một tình huống cụ thể. Trong ngôn ngữ tự nhiên, một từ, cụm từ hay câu có ý nghĩa khác nhau tùy vào ngữ cảnh mà nó được sử dụng. Vì vậy, việc hiểu được ngữ cảnh là vô cùng quan trọng trong NLP.

Trong NLP, ngữ cảnh đóng vai trò quan trọng trong các nhiệm vụ như:

* Xác định ý nghĩa của một từ hay câu: Việc xác định ý nghĩa của một từ hay câu đôi khi phụ thuộc vào ngữ cảnh sử dụng của nó. Ví dụ: từ "bàn" có thể mang nhiều ý nghĩa khác nhau tùy vào ngữ cảnh như "bàn ăn", "bàn làm việc", "bàn đạp xe",...
* Phân tích cú pháp: Ngữ cảnh có thể giúp phân tích cú pháp, đặc biệt là trong tiếng Anh, nơi thứ tự từ có thể thay đổi nhưng vẫn giữ được ý nghĩa của câu.
* Dịch máy: Việc dịch một câu từ một ngôn ngữ sang ngôn ngữ khác cũng đòi hỏi phải hiểu được ngữ cảnh của câu đó để dịch đúng nghĩa.

Trong những nhiệm vụ NLP phức tạp, việc xử lý ngữ cảnh đòi hỏi mô hình phải có khả năng hiểu và xử lý thông tin phức tạp. Mô hình BERT đã ra đời với khả năng đưa ra các kết quả dự đoán chính xác trong nhiều nhiệm vụ NLP nhờ khả năng hiểu và xử lý ngữ cảnh.

1.4 Cách tiếp cận

Trong NLP, để hiểu và phân tích ngôn ngữ, chúng ta cần tiếp cận theo hướng xử lý ngôn ngữ tự nhiên (Natural Language Processing - NLP). Tiếp cận này bao gồm việc sử dụng các kỹ thuật và công cụ để xử lý và hiểu các dạng ngôn ngữ như văn bản, âm thanh và hình ảnh.

Một cách tiếp cận phổ biến trong NLP là tiếp cận dựa trên ngữ cảnh (contextual approach). Theo cách tiếp cận này, mỗi từ trong câu được hiểu và đánh giá dựa trên ngữ cảnh của nó trong câu, chứ không đơn thuần chỉ dựa trên nghĩa cơ bản của từ đó. Ví dụ, trong câu "Anh ta đã ăn bánh trái cây tối qua", từ "trái cây" được hiểu là một loại thực phẩm, trong khi trong câu "Anh ta đã chơi bóng trái cây tối qua", từ "trái cây" được hiểu là một loại bóng.

Các mô hình ngôn ngữ dựa trên ngữ cảnh như BERT (Bidirectional Encoder Representations from Transformers) và GPT (Generative Pre-trained Transformer) đã được phát triển để giải quyết các vấn đề trong NLP dựa trên tiếp cận ngữ cảnh. Các mô hình này được đào tạo trên các dữ liệu lớn và cho phép đưa ra dự đoán chính xác hơn về ý nghĩa và ngữ cảnh của từ trong câu.

1.5 Các phương pháp trong mô hình BERT

Mô hình BERT sử dụng phương pháp pre-training trên một lượng lớn văn bản để học các biểu diễn ngữ nghĩa của từ và câu. Sau đó, các biểu diễn này được fine-tuning trên các tác vụ NLP cụ thể.

Các phương pháp chính trong mô hình BERT bao gồm:

* 1. Masked Language Modeling (MLM): Đây là phương pháp chính để học biểu diễn từ của BERT. Trong MLM, một số từ được chọn ngẫu nhiên trong câu và được thay thế bằng ký tự [MASK]. Mô hình phải dự đoán từ ban đầu của các ký tự [MASK] này từ các thông tin ngữ cảnh khác trong câu.
  2. Next Sentence Prediction (NSP): Phương pháp này giúp BERT học được khả năng dự đoán sự liên kết giữa hai câu trong một văn bản. Ví dụ, nếu câu B đến ngay sau câu A trong văn bản, mô hình sẽ được huấn luyện để dự đoán xác suất của sự kết nối giữa hai câu.
  3. Fine-tuning: Sau khi đã được huấn luyện trước, mô hình BERT có thể được fine-tuning trên các tác vụ NLP cụ thể như phân loại văn bản hoặc phát hiện thực thể trong văn bản. Trong quá trình fine-tuning, các trọng số của mô hình được điều chỉnh để tối ưu hóa cho tác vụ cụ thể đó.

Tất cả các phương pháp này đều giúp BERT học được các biểu diễn ngữ nghĩa của từ và câu, giúp mô hình đạt được kết quả tốt trong nhiều tác vụ NLP khác nhau.

1.5.1 Masked Language Modeling (MLM)

Masked Language Modeling (MLM) là một trong những phương pháp trong quá trình huấn luyện mô hình BERT. MLM giúp mô hình BERT hiểu được cách mà các từ trong câu được liên kết với nhau. Cụ thể, MLM yêu cầu mô hình BERT dự đoán từ bị mask (che đi) trong một câu cho trước.

Để thực hiện MLM, đầu tiên, một phần của từ vựng đầu vào được chọn ngẫu nhiên để bị mask (ví dụ: chuyển từ "I like eating ice cream" thành "I like eating [MASK] cream"). Sau đó, mô hình BERT được huấn luyện để dự đoán từ bị mask bằng cách sử dụng các từ còn lại trong câu.

Việc huấn luyện MLM cho phép mô hình BERT học cách biểu diễn từ mà không phụ thuộc vào vị trí của từ đó trong câu. Điều này có nghĩa là BERT có thể hiểu được nghĩa của một từ dựa trên ngữ cảnh của nó, thay vì chỉ dựa vào từ đó đơn lẻ.

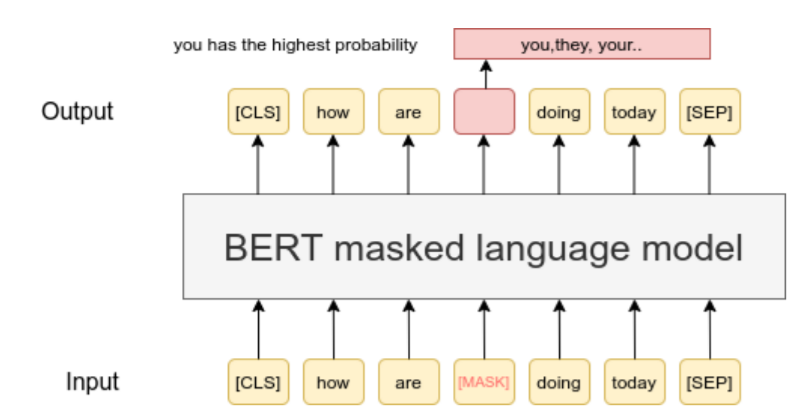
MLM được coi là một phương pháp hiệu quả để huấn luyện mô hình ngôn ngữ BERT, đặc biệt là trong việc giải quyết các nhiệm vụ liên quan đến xử lý ngôn ngữ tự nhiên.

Tuy nhiên, việc đặt ngẫu nhiên một số từ để mô hình dự đoán sẽ làm cho nhiều từ có thể được dự đoán từ các từ còn lại trong câu. Để khắc phục vấn đề này, mô hình BERT sử dụng hai loại mask:

* 1. Mask đơn giản (simple mask): Một từ được đặt mask và thay thế bằng một ký tự đặc biệt [MASK]. Mô hình sẽ cố gắng dự đoán từ bị mask dựa trên các từ còn lại trong câu.
  2. Mask ngẫu nhiên (random mask): Một số từ trong câu được chọn ngẫu nhiên để đặt mask. Điều này làm giảm khả năng mô hình chỉ dựa trên các từ xung quanh để dự đoán từ bị mask.

Sau khi tạo ra câu đã bị mask, BERT sử dụng kiến trúc mã hóa Transformer để học các biểu diễn cho từng từ trong câu. Mô hình được huấn luyện trên một tập hợp các tác vụ khác nhau, bao gồm cả MLM và Next Sentence Prediction (NSP).

Để sử dụng mô hình BERT cho một tác vụ NLP cụ thể, chúng ta chỉ cần thêm một lớp đầu ra cho mô hình và huấn luyện nó trên dữ liệu của tác vụ đó. Ví dụ, để sử dụng BERT cho tác vụ phân loại văn bản, chúng ta có thể sử dụng lớp đầu ra là một mạng nơ-ron đơn giản hoặc một lớp tuyến tính để phân loại văn bản.



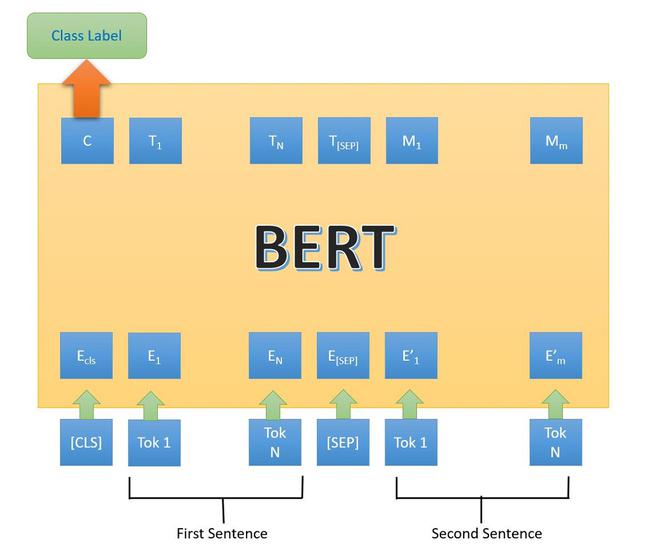
1.5.2 Next Sentence Prediction (NSP)

Next Sentence Prediction (NSP) là một phương pháp huấn luyện trong mô hình BERT để giải quyết vấn đề hiểu ngữ cảnh bằng cách dự đoán xem hai câu tiếp theo nhau có liên quan đến nhau hay không. Mục tiêu của phương pháp này là giúp mô hình học được khả năng hiểu được quan hệ giữa các câu trong văn bản, thay vì chỉ tập trung vào từng câu một.

NSP được thực hiện bằng cách sử dụng đầu vào là hai câu văn bản A và B. Trong đó, câu A là câu đầu vào và câu B có thể là câu đứng ngay sau câu A trong văn bản hoặc là một câu ngẫu nhiên khác trong văn bản. Mô hình sẽ được huấn luyện để dự đoán xem câu B có phải là câu tiếp theo của câu A hay không.

Cụ thể, để thực hiện NSP, đầu vào của mô hình BERT sẽ được cung cấp thêm một token đặc biệt [CLS] ở đầu câu A và một token [SEP] để phân biệt giữa câu A và câu B. Sau đó, mô hình sẽ đưa đầu vào qua một lớp phân loại nhị phân để dự đoán xem câu B có phải là câu tiếp theo của câu A hay không.

Việc huấn luyện mô hình BERT bằng cách kết hợp cả phương pháp MLM và NSP giúp cho mô hình có khả năng hiểu ngữ cảnh một cách toàn diện hơn.



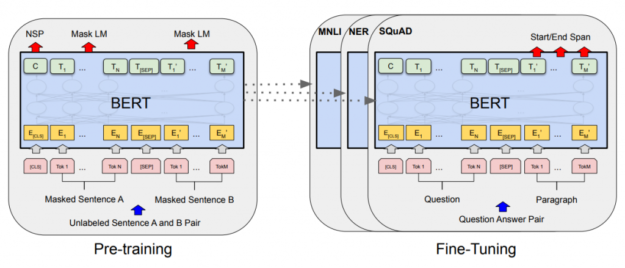
1.5.3 Fine-tuning

Fine-tuning là quá trình tinh chỉnh mô hình ngôn ngữ tổng quát như BERT để phù hợp với một nhiệm vụ cụ thể, chẳng hạn như phân loại văn bản, dự đoán thực thể, tóm tắt văn bản, vv.

Trong quá trình huấn luyện mô hình ngôn ngữ BERT, mạng nơ-ron đã được đào tạo trên hàng triệu đoạn văn bản từ nhiều nguồn khác nhau. Sau đó, khi thực hiện Fine-tuning, ta chỉ cần đưa ra một tập dữ liệu có nhãn mới (ví dụ: tập dữ liệu phân loại tin tức) và huấn luyện lại một số phần của mô hình (thông qua backpropagation) để thích nghi với tập dữ liệu mới này.

Để thực hiện Fine-tuning, ta sẽ thêm một lớp phân loại đầy đủ vào cuối của mô hình BERT, sau đó huấn luyện mô hình để điều chỉnh các trọng số của lớp phân loại và lớp trước đó (lớp tổng hợp) đồng thời thông qua backpropagation. Sau khi huấn luyện xong, mô hình BERT đã được tinh chỉnh để phù hợp với tập dữ liệu cụ thể và có thể được sử dụng để dự đoán kết quả trên tập dữ liệu mới.

Fine-tuning là một kỹ thuật rất hiệu quả trong xử lý ngôn ngữ tự nhiên, và đã được áp dụng rộng rãi trong nhiều nhiệm vụ khác nhau, bao gồm phân loại văn bản, dự đoán thực thể, tóm tắt văn bản, vv.



1.6 Các kiến trúc model BERT

Hiện tại, có ba kiến trúc mô hình BERT được công bố: BERT-base, BERT-large và BERT-multi-lingual.

1. BERT-base: Đây là mô hình cơ bản nhất và nhỏ nhất trong số các mô hình BERT. Nó có 12 lớp encoder, tổng cộng 110 triệu tham số. BERT-base được huấn luyện trên 16GB dữ liệu Wikipedia và BookCorpus.
2. BERT-large: BERT-large là mô hình lớn hơn với 24 lớp encoder và 340 triệu tham số. Nó được huấn luyện trên cùng một bộ dữ liệu như BERT-base.
3. BERT-multi-lingual: BERT-multi-lingual được huấn luyện trên 104 ngôn ngữ khác nhau và có khả năng xử lý nhiều ngôn ngữ. Nó có 12 lớp encoder và 170 triệu tham số.

Ngoài ra, có thêm các biến thể khác của BERT như RoBERTa, DistilBERT, ALBERT,... mỗi loại đều có các điểm mạnh và yếu khác nhau tùy vào mục đích sử dụng.

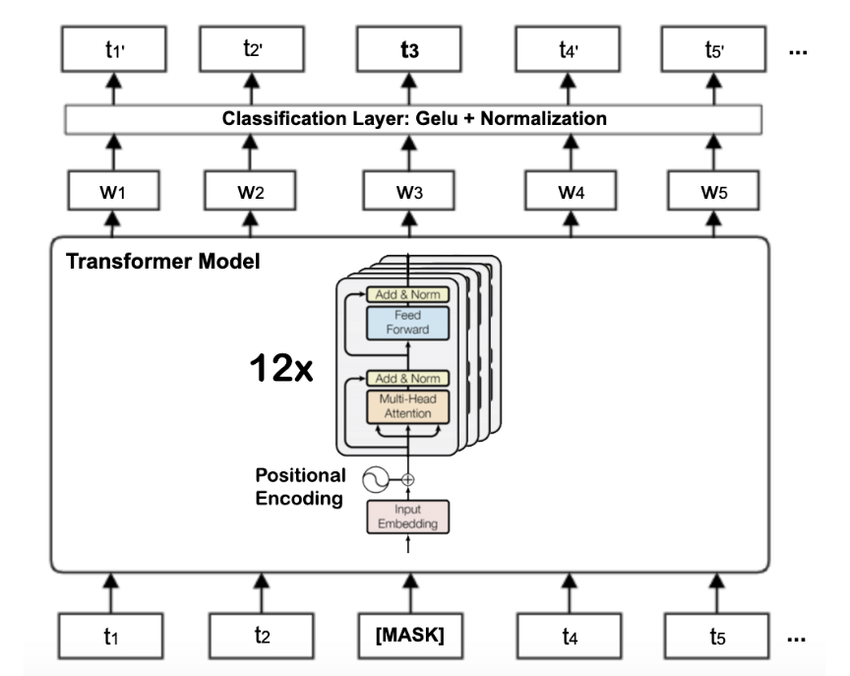
1.6.1 BERT-base

BERT-base là một kiến trúc mô hình BERT được đào tạo trên tập dữ liệu lớn, gồm 12 lớp Transformer và khoảng 110 triệu tham số. Đây là phiên bản BERT phổ biến nhất và thường được sử dụng như là một kiến trúc cơ bản cho các ứng dụng NLP.

Mỗi lớp Transformer trong BERT-base bao gồm một lớp self-attention và hai lớp fully-connected (hay còn gọi là lớp feed-forward). Trong quá trình đào tạo, mô hình được huấn luyện với cả hai phương pháp MLM và NSP.

BERT-base được đào tạo trên tập dữ liệu lớn, gồm khoảng 3,3 tỷ từ. Tập dữ liệu này bao gồm các văn bản từ nhiều nguồn khác nhau, chẳng hạn như các trang web, tài liệu khoa học, báo chí, vv. Vì vậy, BERT-base có khả năng hiểu và xử lý rất nhiều ngữ cảnh và định dạng dữ liệu khác nhau.

BERT-base đã đạt được nhiều kết quả ấn tượng trong nhiều nhiệm vụ NLP khác nhau, chẳng hạn như phân loại văn bản, dự đoán từ tiếp theo, phân tích ý kiến, tóm tắt văn bản, vv.



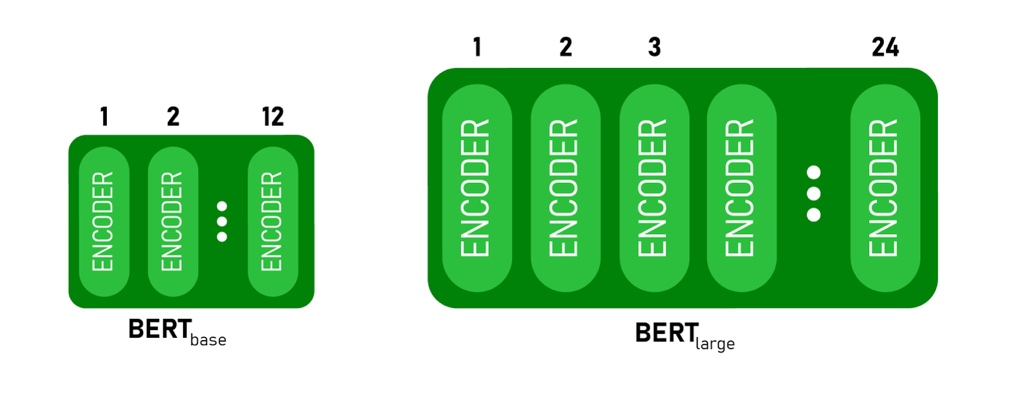
1.6.2 BERT-large

BERT-large là một phiên bản lớn hơn của BERT-base, có tổng cộng 24 lớp transformer và 340 triệu tham số.

Cụ thể, kiến trúc của BERT-large bao gồm:

* Lớp nhúng từ (Embedding layer): Tương tự như trong BERT-base, lớp này sẽ chuyển đổi các từ trong câu thành các vector trong không gian embedding.
* Lớp mã hóa (Encoder layer): BERT-large có tổng cộng 24 lớp transformer, mỗi lớp có 16 self-attention heads và 1,024 đơn vị đầu ra trong lớp fully connected. Vì vậy, tổng cộng BERT-large sử dụng 384 self-attention heads và 24,576 đơn vị đầu ra trong lớp fully connected.
* Pooling layer: Tương tự như BERT-base, BERT-large cũng sử dụng lớp Pooling để lấy vector đại diện của toàn bộ câu từ đầu ra của lớp mã hóa cuối cùng.
* Fully connected layer: Cuối cùng, lớp fully connected có 2 đơn vị đầu ra được sử dụng để dự đoán xác suất của 2 lớp (positive/negative hoặc entailment/non-entailment) cho tác vụ cụ thể.

Với số lượng lớp và tham số lớn hơn, BERT-large cho phép mô hình học được mối quan hệ phức tạp hơn giữa các từ trong câu, đồng thời có khả năng khai thác được nhiều thông tin hơn từ dữ liệu huấn luyện để cải thiện độ chính xác của mô hình. Tuy nhiên, điều này cũng đồng nghĩa với việc BERT-large có kích thước lớn hơn và cần nhiều tài nguyên tính toán hơn để huấn luyện và sử dụng.



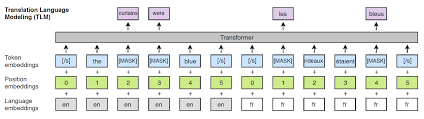
1.6.3 BERT-multi-lingual

BERT-multilingual là một phiên bản của BERT được đào tạo trên một tập dữ liệu lớn hơn, gồm nhiều ngôn ngữ khác nhau, bao gồm cả tiếng Anh và các ngôn ngữ khác như tiếng Tây Ban Nha, tiếng Bồ Đào Nha, tiếng Pháp, tiếng Đức, tiếng Nga, tiếng Trung và nhiều ngôn ngữ khác. Vì vậy, BERT-multilingual được coi là một mô hình đa ngôn ngữ.

Một điểm đáng chú ý về BERT-multilingual là nó chia sẻ một bộ mã hóa từ vựng chung cho tất cả các ngôn ngữ. Điều này có nghĩa là mỗi từ trong tất cả các ngôn ngữ được biểu diễn bằng cùng một token và được đưa vào cùng một bộ mã hóa từ vựng.

Với việc chia sẻ một bộ mã hóa từ vựng chung và đào tạo trên nhiều ngôn ngữ, BERT-multilingual có thể xử lý nhiều ngôn ngữ khác nhau trong cùng một mô hình, giúp cho việc phân loại và dịch thuật trở nên hiệu quả hơn.

Tuy nhiên, vì mỗi ngôn ngữ có những đặc thù riêng của nó, nên đôi khi sử dụng BERT-multilingual không thể đạt được độ chính xác cao nhất. Vì vậy, khi xử lý một ngôn ngữ cụ thể, nên sử dụng các phiên bản BERT được đào tạo riêng cho ngôn ngữ đó để đạt được độ chính xác tốt nhất.



1.7 Ưu điểm và nhược điểm của mô hình BERT

1.7.1 Ưu điểm

Mô hình BERT (Bidirectional Encoder Representations from Transformers) là một trong những mô hình học sâu tiên tiến nhất trong lĩnh vực xử lý ngôn ngữ tự nhiên (NLP). Dưới đây là một số ưu điểm của mô hình BERT:

1. Hiệu suất tốt: Mô hình BERT đã đạt được nhiều kết quả tốt nhất trên nhiều tác vụ NLP khác nhau, bao gồm phân loại văn bản, dịch máy, tóm tắt văn bản, và nhiều tác vụ khác.
2. Học không giám sát: Mô hình BERT được huấn luyện bằng cách sử dụng dữ liệu không được gắn nhãn, điều này có nghĩa là mô hình có khả năng học và hiểu được ngôn ngữ tự nhiên một cách tự động, mà không cần phải giám sát hoặc hướng dẫn từ con người.
3. Khả năng biểu diễn ngôn ngữ: Mô hình BERT có khả năng biểu diễn ngôn ngữ tốt hơn các mô hình trước đó. Điều này có nghĩa là mô hình có khả năng hiểu được ngữ cảnh và ý nghĩa của từng từ và cụm từ trong văn bản.
4. Tính tương thích cao: Mô hình BERT có khả năng tích hợp với các mô hình và hệ thống khác để tăng cường khả năng xử lý ngôn ngữ tự nhiên của chúng.
5. Mở rộng và tùy chỉnh: Mô hình BERT được xây dựng dựa trên kiến trúc Transformer, cho phép các nhà nghiên cứu tùy chỉnh mô hình để phù hợp với các tác vụ và mục đích cụ thể.

1.7.2 Nhược điểm

Mặc dù mô hình BERT có nhiều ưu điểm, nhưng nó cũng có một số nhược điểm như sau:

1. Khả năng tính toán cao: Mô hình BERT có số lượng tham số lớn và khối lượng tính toán cao hơn so với các mô hình trước đó. Điều này làm cho việc huấn luyện và triển khai mô hình khó khăn hơn.
2. Tốn kém thời gian và tài nguyên: Do mô hình BERT được huấn luyện trên nhiều dữ liệu, nên việc huấn luyện mô hình yêu cầu thời gian và tài nguyên tính toán lớn.
3. Không phù hợp cho các văn bản ngắn: Mô hình BERT được huấn luyện trên các văn bản dài, do đó nó không phù hợp cho các tác vụ xử lý ngôn ngữ tự nhiên trên các văn bản ngắn.
4. Khó hiểu: Do mô hình BERT được huấn luyện không giám sát, nên việc giải thích cách mà mô hình đưa ra các dự đoán của nó có thể khó hiểu và phức tạp.

Tóm lại, mô hình BERT có một số nhược điểm, đặc biệt là khả năng tính toán cao và tốn kém thời gian và tài nguyên. Ngoài ra, mô hình cũng không phù hợp cho các văn bản ngắn và có khả năng giải thích khó hiểu. Tuy nhiên, những nhược điểm này có thể được giải quyết bằng cách tối ưu hóa mô hình và cải thiện phương pháp huấn luyện.

1.8 Tính ứng dụng của mô hình BERT

Mô hình BERT đã có nhiều ứng dụng trong lĩnh vực xử lý ngôn ngữ tự nhiên (NLP), bao gồm:

* Xử lý ngôn ngữ tự nhiên: BERT đã đạt được kết quả tốt trong nhiều tác vụ NLP như phân loại văn bản, phân đoạn văn bản, dự đoán từ khóa, truy vấn hỏi đáp, phân tích tâm trạng, và tổng hợp văn bản.
* Dịch máy: BERT được sử dụng để cải thiện chất lượng dịch máy bằng cách sử dụng nó như một phương pháp đánh giá chất lượng dịch văn bản.
* Tìm kiếm: BERT có thể được sử dụng để cải thiện kết quả tìm kiếm văn bản bằng cách đưa ra các kết quả liên quan đến các từ khóa hoặc câu hỏi được tìm kiếm.
* Tổng hợp thông tin: BERT có thể được sử dụng để tổng hợp thông tin từ các nguồn khác nhau và đưa ra các kết quả tóm tắt văn bản.
* Trích xuất thông tin: BERT cũng có thể được sử dụng để trích xuất thông tin quan trọng từ các văn bản, như tên, địa chỉ, điện thoại, ngày tháng, giá cả, v.v.
* Phân tích ngôn ngữ: BERT có thể được sử dụng để phân tích các trường hợp sử dụng ngôn ngữ và từ vựng khác nhau trong các văn bản, nhằm đưa ra kết quả đánh giá và phân tích ngôn ngữ.

Tóm lại, mô hình BERT đã đóng góp quan trọng vào việc nâng cao chất lượng và hiệu suất của các ứng dụng NLP, giúp cho việc xử lý ngôn ngữ tự nhiên trở nên đơn giản và hiệu quả hơn.

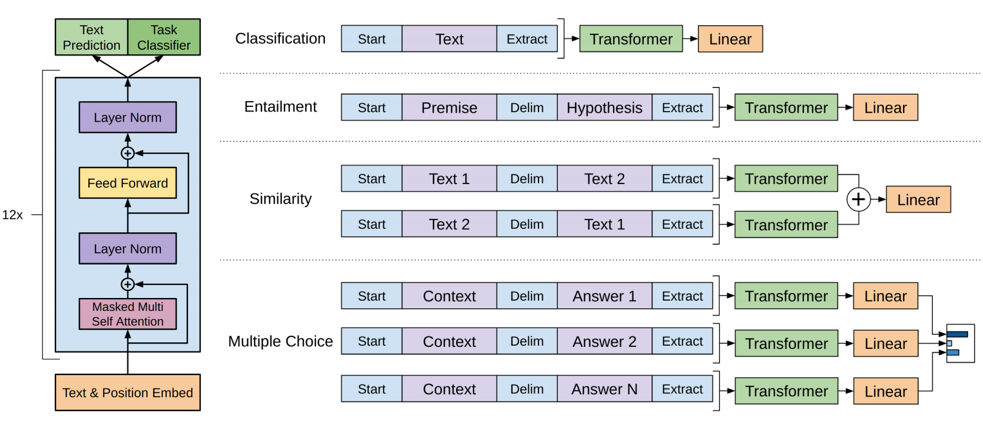
2. Mô hình GPT

2.1 Khái niệm

GPT (Generative Pre-trained Transformer) là một mô hình học sâu cho NLP (Natural Language Processing) được phát triển bởi OpenAI. GPT được đào tạo trên các nhiệm vụ dự đoán từ tiếp theo (next word prediction) trong văn bản tự động sinh ra. Mô hình này sử dụng kiến trúc Transformer, cùng với việc tiền đào tạo (pre-training) trên một tập dữ liệu lớn, để học cách hiểu và sản xuất các chuỗi văn bản tự nhiên.

GPT là một mô hình ngôn ngữ tự nhiên (NLP) mạnh mẽ và đa dạng vì nó có thể được sử dụng cho nhiều loại tác vụ, bao gồm hoàn thành đoạn văn bản, dịch ngôn ngữ, phân loại văn bản và tự động sinh văn bản.

GPT được phát triển bởi OpenAI và có thể được sử dụng thông qua gói phần mềm mã nguồn mở của họ, có tên là "GPT-2". GPT-2 có kích thước lớn và được đào tạo trên tập dữ liệu rất lớn (hơn 40GB), do đó nó có khả năng sinh ra văn bản tự nhiên đầy đủ và đáng ngạc nhiên.



2.2 Tại sao lại có Mô hình GPT

Mô hình GPT (Generative Pre-trained Transformer) là một mô hình học sâu sử dụng kiến trúc transformer được huấn luyện trước trên dữ liệu lớn. Nó được tạo ra với mục đích tạo ra các đoạn văn bản tự động và tự nhiên hơn.

Một trong những thách thức lớn nhất trong lĩnh vực xử lý ngôn ngữ tự nhiên là tạo ra các đoạn văn bản tự động và tự nhiên hơn. Để tạo ra các đoạn văn bản như vậy, cần phải hiểu sâu sắc về ngôn ngữ và có thể "dự đoán" các từ tiếp theo của câu.

Mô hình GPT giải quyết vấn đề này bằng cách sử dụng một mô hình học sâu được huấn luyện trên một tập dữ liệu lớn. Việc huấn luyện này cho phép mô hình "học" các đặc trưng của ngôn ngữ và sử dụng chúng để tạo ra các đoạn văn bản mới.

Mô hình GPT đã được sử dụng rộng rãi trong nhiều ứng dụng như tạo ra các mô tả tự động cho ảnh, tạo ra các đoạn văn bản cho bài viết tự động và tạo ra các trả lời tự động cho các cuộc trò chuyện.

2.3 Cách tiếp cận

Mô hình GPT (Generative Pre-trained Transformer) là một mô hình NLP được huấn luyện trên cơ sở của kiến trúc Transformer và một lượng lớn dữ liệu văn bản. Mục đích của mô hình là sinh ra các đoạn văn bản liên tục, tự nhiên và chính xác dựa trên đầu vào ban đầu.

Cách tiếp cận của GPT giống với BERT, đó là pre-training trên một lượng lớn dữ liệu văn bản không có nhãn để học cách hiểu và trích xuất thông tin từ dữ liệu đó. Sau đó, mô hình được fine-tuning trên các tác vụ cụ thể để đạt được hiệu suất tốt trên các tác vụ đó.

Tuy nhiên, điểm khác biệt lớn nhất giữa GPT và BERT là mục đích của mô hình. Trong khi BERT được huấn luyện để phân loại, GPT được huấn luyện để sinh ra văn bản. GPT được coi là một mô hình "đơn ngữ", có nghĩa là nó được huấn luyện để đưa ra các dự đoán tiếp theo trong một chuỗi văn bản duy nhất, trong khi BERT là một mô hình "đa ngôn ngữ", có khả năng xử lý nhiều ngôn ngữ.

Một cách tiếp cận khác của GPT là sử dụng mô hình để tạo ra văn bản tự động, được gọi là mô hình sinh văn bản. GPT-2 là một phiên bản tiếp theo của GPT, được huấn luyện trên một lượng lớn dữ liệu văn bản và cho phép tạo ra các đoạn văn bản có chất lượng rất cao và rất giống với những gì được viết bởi con người.

2.4 Các phương pháp trong mô hình GPT

Mô hình GPT (Generative Pre-trained Transformer) là một mô hình học sâu dựa trên kiến trúc Transformer và được đào tạo dựa trên bộ dữ liệu lớn để học cách dự đoán từ tiếp theo trong một câu. Sau đó, mô hình được điều chỉnh để tạo ra các đoạn văn bản mới dựa trên ngữ cảnh đầu vào.

Các phương pháp trong mô hình GPT bao gồm:

1. Pre-training: Tương tự như mô hình BERT, mô hình GPT cũng được đào tạo trước trên một lượng lớn dữ liệu không giám sát. Trong giai đoạn này, mô hình học cách dự đoán từ tiếp theo trong một câu. Việc đào tạo được thực hiện thông qua một nhiệm vụ dự đoán từ tiếp theo (next-word prediction) trong một câu đã bị mất mát từ (masked language model). Để tăng độ khó của bài toán, một số từ trong câu được mất mát từ theo tỉ lệ nhất định.
2. Fine-tuning: Sau khi đào tạo, mô hình GPT được điều chỉnh trên một tập dữ liệu nhỏ hơn để tạo ra các đoạn văn bản mới dựa trên ngữ cảnh đầu vào. Việc điều chỉnh được thực hiện thông qua một nhiệm vụ giám sát (supervised learning) như dự đoán từ tiếp theo trong một đoạn văn bản hoặc phân loại văn bản.
3. Decoding: Khi được sử dụng để tạo ra các đoạn văn bản mới, mô hình GPT sử dụng phương pháp tạo ra chuỗi (autoregressive generation) để dự đoán từ tiếp theo dựa trên từ trước đó. Mô hình sử dụng một chiến lược giải mã (decoding strategy) để tạo ra các đoạn văn bản mới dựa trên ngữ cảnh đầu vào. Các chiến lược giải mã phổ biến bao gồm sampling, beam search và top-k sampling.

2.4.1 Pre-training

Pre-training là quá trình huấn luyện mô hình ngôn ngữ trên một lượng lớn dữ liệu không được gán nhãn trước khi fine-tuning trên tác vụ cụ thể. Mục đích của pre-training là học ra các trọng số của mô hình sao cho chúng có khả năng tổng quát hóa cao trên các tác vụ liên quan đến ngôn ngữ tự nhiên.

Trong mô hình GPT, pre-training được thực hiện bằng cách sử dụng thuật toán Transformer và độ lớn của mô hình được điều chỉnh bằng cách thêm lượng lớn các lớp Transformer.

Cụ thể, quá trình pre-training trong mô hình GPT bao gồm hai bước chính:

* Masked Language Model (MLM): Tương tự như trong mô hình BERT, bước này sẽ thực hiện việc mô hình hóa từ bị che đi trong văn bản. Tuy nhiên, khác với BERT, ở đây việc mask sẽ được thực hiện ngẫu nhiên trên mỗi từ trong văn bản, không chỉ riêng trên các từ đầu vào như trong BERT.
* Next Sentence Prediction (NSP): Bước này được sử dụng để đánh giá khả năng hiểu và kết nối giữa các câu. Mô hình sẽ được đưa vào hai câu liên tiếp và được huấn luyện để đoán xem chúng có liên kết với nhau hay không.

Sau khi pre-training được hoàn thành, mô hình sẽ được fine-tuning trên các tác vụ liên quan đến ngôn ngữ tự nhiên như phân loại văn bản, dịch máy, tự động tóm tắt văn bản, và các tác vụ khác.

Sau khi có dữ liệu tiền huấn luyện, mô hình GPT sẽ được tiền huấn luyện trên tập dữ liệu đó với mục đích học cách tạo ra một mô hình ngôn ngữ tự nhiên thông qua việc dự đoán từ tiếp theo trong một câu. Để làm được điều này, mô hình sẽ sử dụng kiến trúc Transformer, cho phép nó học được những thông tin về cấu trúc ngữ pháp và ngữ nghĩa của ngôn ngữ.

Cụ thể, mô hình GPT sử dụng một phương pháp tiền huấn luyện gọi là Masked Language Modeling (MLM). Trong phương pháp này, một số từ trong câu sẽ bị che đi và mô hình cần phải dự đoán từ đó để hoàn thành câu. Việc che các từ này sẽ giúp mô hình học được những thông tin liên quan đến ngữ pháp và ngữ nghĩa của câu, cũng như tạo ra một khả năng "giả đoán" cho mô hình về các từ không được che.

Một phương pháp khác được sử dụng trong tiền huấn luyện của GPT là Next Sentence Prediction (NSP). Trong phương pháp này, mô hình sẽ được huấn luyện để dự đoán xem hai câu có liên quan đến nhau hay không. Việc này giúp mô hình hiểu được cách các câu liên kết với nhau trong một văn bản và hỗ trợ cho quá trình tiền huấn luyện về ngữ pháp và ngữ nghĩa của ngôn ngữ.

Quá trình tiền huấn luyện của mô hình GPT có thể mất nhiều ngày hoặc thậm chí nhiều tuần để hoàn tất trên một bộ xử lý đồng thời. Tuy nhiên, một khi đã hoàn tất, mô hình có thể được fine-tuning trên các tác vụ cụ thể trong lĩnh vực xử lý ngôn ngữ tự nhiên để đạt được kết quả tốt.

2.4.2 Fine-tuning

Fine-tuning trong mô hình GPT là quá trình sử dụng mô hình đã được huấn luyện trước đó trên dữ liệu mới để thực hiện một tác vụ cụ thể, chẳng hạn như phân loại văn bản, tạo tiêu đề, tạo câu hỏi và trả lời, v.v. Fine-tuning là một kỹ thuật rất hiệu quả trong học sâu, cho phép tái sử dụng các mô hình đã huấn luyện trước đó để giải quyết các vấn đề khác.

Trong quá trình fine-tuning, chúng ta sử dụng một mô hình GPT được huấn luyện trước đó và tiếp tục huấn luyện trên một tập dữ liệu mới cho tác vụ cụ thể. Các tham số của mô hình được điều chỉnh để phù hợp với tập dữ liệu mới và tối ưu hóa độ chính xác của mô hình.

Fine-tuning cho mô hình GPT thường bao gồm hai bước chính:

* 1. Đầu tiên, ta đưa vào một tập dữ liệu mới và tiền xử lý tập dữ liệu này để đưa về định dạng phù hợp với đầu vào của mô hình. Sau đó, ta sử dụng mô hình GPT đã được huấn luyện trước đó để sinh ra dự đoán cho từng ví dụ trong tập dữ liệu mới.
  2. Sau đó, ta sử dụng hàm mất mát và thuật toán tối ưu hóa để điều chỉnh các tham số của mô hình GPT để tối ưu hóa độ chính xác trên tập dữ liệu mới. Thông thường, ta sẽ huấn luyện mô hình GPT trên tập dữ liệu mới trong nhiều epochs để tối ưu hóa độ chính xác của mô hình. Cuối cùng, sau khi quá trình fine-tuning kết thúc, ta sẽ sử dụng mô hình GPT mới được huấn luyện để thực hiện các tác vụ cụ thể trên dữ liệu mới.

Với quá trình fine-tuning, mô hình GPT đã được huấn luyện trước đó trên tập dữ liệu lớn, đã học được cấu trúc ngôn ngữ và kiến thức về thế giới tự nhiên, và vì vậy nó có thể tự động học được các quy luật và thông tin mới từ tập dữ liệu mới và cải thiện độ chính xác cho tác vụ cụ thể.

2.4.3 decoding

Trong mô hình GPT, decoding là quá trình tạo ra chuỗi đầu ra từ chuỗi đầu vào. Decoding được sử dụng để sinh ra các dự đoán tiếp theo trong các tác vụ thế hệ văn bản, chẳng hạn như hoàn thành câu, dự đoán từ tiếp theo trong câu hoặc tạo ra câu hỏi từ đoạn văn bản.

GPT sử dụng mô hình Transformer để thực hiện quá trình decoding. Trong quá trình decoding, đầu vào được mã hóa dưới dạng vector embedding và truyền qua một loạt các lớp self-attention và feed-forward network để tạo ra đầu ra. Quá trình này được thực hiện tuần tự từng từ một, với mỗi từ được tạo ra dựa trên tất cả các từ trước đó.

Trong quá trình decoding, các từ được tạo ra bằng cách lấy trọng số softmax của các giá trị logits được tính toán từ mô hình Transformer. Các giá trị logits này đại diện cho xác suất của mỗi từ trong bộ từ vựng được sử dụng để tạo ra đầu ra. Để đảm bảo tính coerency của văn bản, các giá trị logits cho các từ không phù hợp với ngữ cảnh đang xét được đặt về một giá trị rất nhỏ gần bằng 0.

Tuy nhiên, quá trình decoding trong GPT không được thực hiện hoàn toàn tự động mà được điều chỉnh bằng cách sử dụng beam search hoặc top-k sampling. Beam search là phương pháp tìm kiếm được sử dụng để giảm thiểu sự mất mát (loss) của mô hình trong quá trình sinh ra chuỗi đầu ra. Top-k sampling là phương pháp lấy ra k từ có xác suất cao nhất từ bộ từ vựng và sử dụng chúng để tạo ra đầu ra tiếp theo.

Tóm lại, quá trình decoding trong mô hình GPT là quá trình tạo ra chuỗi đầu ra từ chuỗi đầu vào bằng cách sử dụng mô hình Transformer và các kỹ thuật tìm kiếm, nhằm đảm bảo tính coerency và độ chính xác của văn bản được sinh ra.

2.5 Các kiến trúc model GPT

Hiện tại, đã có ba phiên bản của mô hình GPT:

* 1. GPT-1: GPT-1 là mô hình đầu tiên trong loạt mô hình GPT, được đào tạo trên khoảng 40 GB dữ liệu văn bản và có tổng cộng 117 triệu tham số. Mô hình này sử dụng kiến trúc Transformer với 12 block Transformer và 768 chiều ở mỗi block, đầu vào được token hoá bằng Byte Pair Encoding (BPE).
  2. GPT-2: GPT-2 là phiên bản tiếp theo của GPT-1, với số lượng tham số lớn hơn gấp 10 lần. Nó được đào tạo trên 1,5 TB dữ liệu văn bản và có tổng cộng 1,5 tỷ tham số. Kiến trúc của GPT-2 tương tự như GPT-1, tuy nhiên có 24 block Transformer và 1.024 chiều ở mỗi block. GPT-2 được coi là một trong những mô hình NLP tốt nhất hiện nay.
  3. GPT-3: GPT-3 là phiên bản mới nhất và lớn nhất của mô hình GPT, với tổng cộng 175 tỷ tham số. Nó được đào tạo trên 45 TB dữ liệu văn bản, là một trong những mô hình NLP lớn nhất và mạnh nhất hiện nay. GPT-3 sử dụng kiến trúc Transformer với số block Transformer và số chiều trong mỗi block khác nhau tùy vào cỡ của mô hình. Điều đó có nghĩa là các mô hình nhỏ hơn có số block Transformer ít hơn, số chiều nhỏ hơn, trong khi các mô hình lớn hơn có số block Transformer và số chiều lớn hơn. GPT-3 được coi là một trong những mô hình NLP tiên tiến nhất hiện nay và đã đạt được nhiều thành tựu đáng chú ý trong các tác vụ ngôn ngữ tự nhiên như dịch máy, tóm tắt văn bản, và thậm chí viết bài học toàn bộ cho một bài kiểm tra.

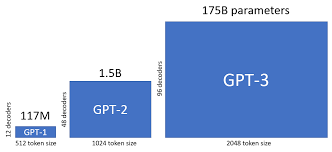
2.5.1 GPT-1

GPT-1 (Generative Pre-trained Transformer 1) là một mô hình ngôn ngữ sinh tự động sử dụng kiến trúc transformer và được pre-training trên một lượng lớn văn bản tiếng Anh để học cấu trúc ngôn ngữ tự nhiên.

Mô hình GPT-1 có hơn 117 triệu tham số, được pre-training trên bộ dữ liệu trên 40GB các tài liệu từ các nguồn khác nhau như Wikipedia và các trang tin tức. Trong quá trình pre-training, mô hình học cách dự đoán từ tiếp theo trong một đoạn văn bản và sử dụng kiến trúc transformer để học được mối quan hệ giữa các từ trong câu.

Với mỗi đầu vào, mô hình GPT-1 sẽ tạo ra các đầu ra định hướng bởi các từ trước đó để tạo ra một văn bản hoàn chỉnh và có ý nghĩa. Với kiến trúc transformer, GPT-1 có khả năng xử lý dữ liệu lớn và tạo ra các đầu ra dựa trên ngữ cảnh của đầu vào.

Một ưu điểm của GPT-1 là nó có thể sinh ra các đoạn văn bản khá dài và tự nhiên, không chỉ là các câu hoặc đoạn ngắn. Tuy nhiên, mô hình này vẫn có một số hạn chế như không thể hiểu được ý nghĩa của câu hoặc văn bản mà chỉ có khả năng sinh ra các đoạn văn bản dựa trên ngữ cảnh đã học được.



2.5.2 GPT-2

GPT-2 là một mô hình ngôn ngữ tự động tiên tiến, được phát triển bởi OpenAI vào năm 2019. GPT-2 được xây dựng dựa trên kiến trúc Transformer và đào tạo trên tập dữ liệu lớn, bao gồm khoảng 45 triệu bài viết từ trang web. Điều này giúp GPT-2 hiểu được các quy tắc và cấu trúc của ngôn ngữ tự nhiên, đồng thời tạo ra những kết quả tốt trong nhiều nhiệm vụ liên quan đến xử lý ngôn ngữ tự nhiên.

Kiến trúc của GPT-2 tương tự như GPT-1, bao gồm một kiến trúc Transformer Encoder được sử dụng để mã hóa thông tin đầu vào và một kiến trúc Transformer Decoder để dự đoán các từ tiếp theo trong chuỗi đầu ra. Tuy nhiên, GPT-2 có nhiều cải tiến vượt trội so với GPT-1, bao gồm:

1. Số lượng tham số lớn hơn: GPT-2 có 1,5 tỉ tham số, gấp khoảng 10 lần so với GPT-1. Điều này giúp GPT-2 hiểu được các mối quan hệ phức tạp hơn trong dữ liệu đầu vào.
2. Quy mô dữ liệu đào tạo lớn hơn: GPT-2 được đào tạo trên một tập dữ liệu lớn gấp nhiều lần so với GPT-1, bao gồm khoảng 45 triệu bài viết từ trang web. Điều này giúp GPT-2 hiểu được các quy tắc và cấu trúc của ngôn ngữ tự nhiên.
3. Khả năng sinh ra đầu ra dài hơn: GPT-2 có thể sinh ra đầu ra dài hơn so với GPT-1, cho phép mô hình tạo ra các bài viết, truyện ngắn và các đoạn văn bản tự nhiên có chủ đề cụ thể.

GPT-2 đã đạt được nhiều kết quả ấn tượng trong nhiều nhiệm vụ NLP, bao gồm sinh văn bản tự động, dịch thuật tự động, tóm tắt văn bản tự động và phân loại văn bản tự động.

2.5.3 GPT-3

GPT-3 (Generative Pre-trained Transformer 3) là một mô hình ngôn ngữ tự động phát triển bởi OpenAI, ra mắt vào năm 2020. GPT-3 được xây dựng dựa trên mô hình transformer, tương tự như GPT-2 và BERT. GPT-3 được coi là một bước tiến đáng kể so với các mô hình trước đó vì kích thước của nó rất lớn (175 tỷ tham số), là một trong những mô hình lớn nhất và phức tạp nhất hiện nay.

GPT-3 được huấn luyện trên một tập dữ liệu lớn hơn bao gồm các ngôn ngữ khác nhau, từ tiếng Anh đến tiếng Trung Quốc và tiếng Tây Ban Nha, cho phép nó hiểu và sử dụng các từ vựng và cấu trúc câu của các ngôn ngữ khác nhau. Nó cũng có khả năng tạo ra văn bản tự nhiên với chất lượng rất cao, từ những câu đơn giản đến các bài viết dài và đầy đủ ý nghĩa.

Một trong những tính năng đáng chú ý của GPT-3 là khả năng thực hiện nhiều tác vụ khác nhau trong lĩnh vực NLP, bao gồm dịch máy, tóm tắt văn bản, phân tích cảm xúc, trả lời câu hỏi, tạo ra các bài viết và văn bản truyện ngắn, và nhiều hơn nữa. Điều này cho thấy khả năng tiềm năng của GPT-3 trong việc giải quyết các vấn đề phức tạp trong thế giới thực và cải thiện trải nghiệm người dùng của các ứng dụng AI.

2.6 Ưu điểm và nhược điểm của mô hình GPT

2.6.1 Ưu điểm

Mô hình GPT (Generative Pre-trained Transformer) là một mô hình mạng nơ-ron đa tác vụ được huấn luyện trước trên dữ liệu lớn với kiến trúc transformer. Dưới đây là một số ưu điểm của mô hình GPT:

1. Khả năng tạo ra văn bản tự nhiên: Mô hình GPT có khả năng tạo ra các đoạn văn bản tự nhiên có ý nghĩa và trôi chảy. Nó có thể được sử dụng để tạo ra các câu chuyện, bài viết blog, trả lời câu hỏi tự động và nhiều ứng dụng khác trong lĩnh vực xử lý ngôn ngữ tự nhiên.
2. Tính đa nhiệm: Mô hình GPT có thể được sử dụng cho nhiều tác vụ khác nhau trong xử lý ngôn ngữ tự nhiên, bao gồm cả tạo ra văn bản, dịch máy, phân loại văn bản và trích xuất thông tin.
3. Hiệu suất cao: Mô hình GPT đã đạt được nhiều kết quả tốt trên nhiều bộ dữ liệu xử lý ngôn ngữ tự nhiên, đặc biệt là trong các tác vụ liên quan đến việc tạo ra văn bản tự nhiên.
4. Cải thiện khả năng hiểu ngữ cảnh: Mô hình GPT được huấn luyện trên dữ liệu lớn, do đó nó có khả năng hiểu ngữ cảnh tốt hơn và cung cấp các dự đoán chính xác hơn.
5. Học tập không giám sát: Mô hình GPT được huấn luyện không giám sát, do đó nó có khả năng tự học và cải thiện hiệu suất theo thời gian.

2.6.2 Nhược điểm

Mặc dù mô hình GPT (Generative Pre-trained Transformer) có nhiều ưu điểm, nhưng nó cũng có một số nhược điểm:

1. Độ dài quá lớn: Mô hình GPT được huấn luyện trên rất nhiều dữ liệu, điều này dẫn đến độ phức tạp tính toán cao, do đó kích thước mô hình cũng rất lớn. Điều này khiến cho việc triển khai và sử dụng mô hình trên các thiết bị có tài nguyên hạn chế gặp khó khăn.
2. Không thể xử lý được một số tác vụ phức tạp: Mặc dù GPT có khả năng sinh ra các đoạn văn bản tự nhiên rất tốt, nhưng nó không phải là giải pháp tối ưu cho tất cả các tác vụ ngôn ngữ tự nhiên. Các tác vụ có tính tương tác cao hoặc yêu cầu kiến thức chuyên môn cần thiết có thể không được xử lý tốt bởi mô hình GPT.
3. Dễ bị ảnh hưởng bởi dữ liệu huấn luyện: Mô hình GPT được huấn luyện trên một lượng lớn dữ liệu, do đó nếu dữ liệu huấn luyện không đại diện cho đủ các trường hợp thực tế, mô hình có thể cho kết quả không chính xác hoặc kém chất lượng.
4. Khả năng sinh ra thông tin sai lệch: Mô hình GPT có thể tự động sinh ra các đoạn văn bản tự nhiên, tuy nhiên điều này cũng có thể dẫn đến việc sinh ra thông tin sai lệch hoặc kém chính xác nếu mô hình không được kiểm soát đầy đủ.
5. Thiếu khả năng giải thích: Mô hình GPT là một mô hình học sâu (deep learning model), có nghĩa là nó hoạt động dựa trên các phép toán trên ma trận. Tuy nhiên, điều này khiến cho việc giải thích kết quả của mô hình trở nên khó khăn và không rõ ràng, dẫn đến khả năng khó hiểu các quyết định của mô hình.

2.7 Tính ứng dụng của mô hình GPT

Mô hình GPT đã được ứng dụng rộng rãi trong các lĩnh vực như:

* Tạo văn bản: GPT có thể tạo ra các đoạn văn bản mới dựa trên các đầu vào từ khóa hoặc các đoạn văn bản có sẵn. Điều này đã được sử dụng để tạo nội dung cho các trang web, các bài báo và các ứng dụng phần mềm.
* Tăng cường ngôn ngữ: GPT có thể học cách nói, viết và hiểu ngôn ngữ tự nhiên. Các ứng dụng tăng cường ngôn ngữ bao gồm hỗ trợ ngôn ngữ và các chatbot thông minh.
* Phân tích ngôn ngữ tự nhiên: GPT có thể phân tích các đoạn văn bản, nhận dạng thực thể và xác định mối quan hệ giữa chúng. Điều này đã được sử dụng trong các công cụ phân tích cảm xúc và các công cụ dự đoán yếu tố thị trường.
* Xử lý ngôn ngữ tự nhiên: GPT cũng có thể được sử dụng để giải quyết các vấn đề liên quan đến xử lý ngôn ngữ tự nhiên, chẳng hạn như dịch thuật, trích xuất thông tin và trả lời câu hỏi.
* Các ứng dụng thông minh nhân tạo khác: GPT cũng đã được sử dụng để tạo ra các ứng dụng thông minh nhân tạo khác, chẳng hạn như phát hiện gian lận và giúp đỡ trong các quyết định kinh doanh.

3 Sự khác nhau giữa mô hình BERT và GPT

Mô hình BERT và GPT là hai mô hình NLP tiên tiến và được sử dụng rộng rãi hiện nay. Tuy nhiên, chúng có một số điểm khác nhau quan trọng:

* 1. Mục tiêu của Pre-training: BERT được huấn luyện với mục đích học đại diện cho ngôn ngữ trong cả hai hướng (được gọi là pre-training liên kết) thông qua việc huấn luyện trên các nhiệm vụ mask language modeling và next sentence prediction. Trong khi đó, GPT được huấn luyện trên nhiệm vụ dự đoán từ tiếp theo (next word prediction) thông qua mô hình Transformer đơn hướng.
  2. Kiến trúc mạng: BERT sử dụng kiến trúc Transformer với cả mã hóa (encoder) và giải mã (decoder) để học mối quan hệ giữa các từ trong một văn bản. Trong khi đó, GPT chỉ sử dụng kiến trúc mã hóa (encoder) trong việc huấn luyện và dự đoán các từ tiếp theo.
  3. Quy mô mô hình: GPT-3 là một trong những mô hình lớn nhất đã được xây dựng, với hơn 175 tỷ tham số, trong khi BERT-large chỉ có 340 triệu tham số. Kích thước lớn của GPT-3 cho phép nó đạt được hiệu suất tốt hơn trong nhiều tác vụ NLP phức tạp.
  4. Ứng dụng: BERT thường được sử dụng cho các tác vụ như phân loại văn bản, dò tìm thực thể và trích xuất thông tin từ văn bản. Trong khi đó, GPT thường được sử dụng cho các ứng dụng tạo văn bản tự động, như tự động viết tin tức, tạo đoạn văn mô tả hoặc đối thoại tự động.

Tóm lại, mặc dù cả BERT và GPT đều là những mô hình NLP tiên tiến, chúng có mục đích và ứng dụng khác nhau và được xây dựng dựa trên các phương pháp huấn luyện và kiến trúc mạng khác nhau.

4 Sự giống nhau giữa mô hình BERT và GPT

Mô hình BERT và GPT đều là những mô hình ngôn ngữ mạnh trong lĩnh vực xử lý ngôn ngữ tự nhiên (NLP) và được huấn luyện trên lượng dữ liệu lớn, có khả năng tạo ra các đại diện phân loại tốt cho các nhiệm vụ NLP khác nhau.

Cả BERT và GPT đều sử dụng kiến trúc mạng transformer, cho phép mô hình học được các đặc trưng phức tạp của ngôn ngữ trong không gian vector đa chiều. Cả hai mô hình đều được huấn luyện trên lượng dữ liệu lớn và có khả năng đưa ra các đại diện phân loại tốt cho các nhiệm vụ NLP như phân loại văn bản, dịch máy, sinh văn bản, v.v.

Tuy nhiên, sự khác biệt chính giữa hai mô hình nằm ở cách chúng được huấn luyện và nhiệm vụ mà chúng được tối ưu hóa. BERT được huấn luyện trên hai nhiệm vụ, gồm Masked Language Modeling (MLM) và Next Sentence Prediction (NSP), trong khi đó, GPT được huấn luyện trên một nhiệm vụ đơn lẻ, đó là tiếp tục văn bản (language modeling). Do đó, GPT thường được sử dụng cho các nhiệm vụ liên quan đến sinh văn bản và tạo ra những đoạn văn bản liền mạch, trong khi BERT thường được sử dụng cho các nhiệm vụ phân loại văn bản và phân tích ngôn ngữ tự nhiên.