

**ĐẠI HỌC QUỐC GIA HÀ NỘI  
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ**



**Nguyễn Quang Minh**

**NGHIÊN CỨU, PHÁT TRIỂN CÔNG CỤ SINH MÔ TẢ  
SẢN PHẨM CHO THƯƠNG MẠI ĐIỆN TỬ**

**LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH**

**HÀ NỘI - 2021**

**ĐẠI HỌC QUỐC GIA HÀ NỘI**  
**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ**



**Nguyễn Quang Minh**

**NGHIÊN CỨU, PHÁT TRIỂN CÔNG CỤ SINH MÔ TẢ  
SẢN PHẨM CHO THƯƠNG MẠI ĐIỆN TỬ**

**Ngành: Khoa học máy tính**

**Chuyên ngành: Khoa học máy tính**

**Mã số: 8480101.01**

**LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH**

**NGƯỜI HƯỚNG DẪN KHOA HỌC: PGS.TS. TRƯƠNG ANH HOÀNG**

**HÀ NỘI - 2021**

**VIETNAM NATIONAL UNIVERSITY, HANOI  
UNIVERSITY OF ENGINEERING AND TECHNOLOGY**

**Nguyen Quang Minh**

**BUILDING PRODUCT DESCRIPTION GENERATOR  
BASED ON GPT-2**

**THE MASTER THESIS**  
**Major: Computer Science**

**Supervisor: Assoc. Prof., Dr. Truong Anh Hoang**

**HANOI - 2021**

## **LỜI CAM ĐOAN**

Tôi là Nguyễn Quang Minh, học viên cao học lớp K26-KHMT, chuyên ngành Khoa học máy tính. Tôi xin cam đoan bài luận văn “Nghiên cứu, phát triển công cụ sinh mô tả sản phẩm cho thương mại điện tử” do tôi đề ra và thực hiện dưới sự hướng dẫn của PGS.TS Trương Anh Hoàng. Các nội dung nghiên cứu và kết quả trong luận văn là xác thực.

Tất cả các tài liệu tham khảo từ các nghiên cứu liên quan đều có nguồn gốc rõ ràng từ danh mục tài liệu tham khảo trong luận văn. Trong luận văn, không có việc sao chép tài liệu, công trình nghiên cứu của người khác mà không chỉ rõ về tài liệu tham khảo.

Hà Nội, ngày ... tháng ... năm 2021

Học viên cao học

Nguyễn Quang Minh

## LỜI CẢM ƠN

Đầu tiên, tôi xin gửi lời cảm ơn sâu sắc đến PGS.TS. Trương Anh Hoàng và TS. Nguyễn Văn Vinh đã giúp đỡ tận tình tôi trong suốt quá trình thực hiện đề tài luận văn này. Tôi cũng xin cảm ơn ThS. Nguyễn Minh Thuận cùng bạn Trần Mạnh Tùng đã hỗ trợ tôi trong suốt quá trình nghiên cứu.

Tôi cũng xin gửi lời cảm ơn chân thành đến các thầy cô, cán bộ trường Đại Học Công Nghệ - Đại Học Quốc Gia Hà Nội đã tận tình giảng dạy và hỗ trợ tôi trong suốt quá khóa học thạc sĩ.

Cuối cùng, tôi xin gửi lời cảm ơn đến gia đình và bạn bè, những người đã hỗ trợ tôi trong cuộc sống.

Xin chân thành cảm ơn!

Hà Nội, ngày ... tháng ... năm 2021

Học viên cao học

Nguyễn Quang Minh

## Tóm tắt

Trong những năm gần đây, thương mại điện tử đã đạt được những bước tiến triển mạnh mẽ và trở thành một xu hướng gần như không thể tách rời trong nhiều khía cạnh của cuộc sống. Những tiến bộ công nghệ cùng lợi ích xuất phát từ lĩnh vực này đã thu hút nhiều đối tượng tham gia, bao gồm cả các cá nhân và doanh nghiệp, tạo ra một sân chơi sôi động và mang tính cạnh tranh cao. Để tồn tại và phát triển trong một môi trường như vậy, những đối tượng này cần liên tục cải thiện chất lượng đồng thời đẩy mạnh việc quảng bá sản phẩm trên các kênh truyền thông. Bên cạnh đó, họ cũng cần quan tâm đến một yếu tố quan trọng nữa chính là nội dung truyền đạt của sản phẩm, đặc biệt là nội dung văn bản hay mô tả sản phẩm, bởi đây sẽ đóng vai trò là cầu nối, cung cấp thông tin về đặc trưng hàng hóa đến khách hàng. Một mô tả chính xác, đầy đủ thông tin và hấp dẫn không chỉ giúp người tiêu dùng đưa ra nhận định đúng đắn mà còn thúc đẩy đến quyết định mua hàng của họ. Mặc dù đóng vai trò quan trọng là vậy, việc tạo ra những mô tả chất lượng còn vướng phải nhiều vấn đề khó khăn do nhiều yếu tố, ví dụ như kinh phí và con người. Do đó, luận văn này hướng đến mục tiêu xây dựng một giải pháp sinh dữ liệu văn bản cho phép tạo ra những mô tả sản phẩm có nội dung thu hút và có thể áp dụng trong những trường hợp thực tế. Giải pháp này sẽ ứng dụng những kỹ thuật, tiên bộ mới, đặc biệt là những thành quả nghiên cứu nổi bật trong lĩnh vực học sâu để cải thiện chất lượng nội dung mô tả. Đồng thời, giải pháp cũng sẽ được chứng minh có khả năng hoạt động ổn định trong điều kiện thiếu thốn dữ liệu huấn luyện.

**Từ khóa:** Thương mại điện tử, học sâu, sinh mô tả sản phẩm

## **Abstract**

In recent years, e-commerce has made strong progress and has become an almost inseparable trend in many aspects of life. Technological advancements and benefits stemming from this sector have attracted a wide range of participants, including individuals and businesses, creating a vibrant and highly competitive playing field. In order to survive and thrive in such an environment, these entities need to continuously improve their quality and promote their products on media channels. Besides, they also need to pay attention to another important factor which is the communication content of the product, especially textual content or product description, because this will act as a bridge, providing information about product characteristics to customers. An accurate, informative and attractive description not only helps consumers make the right judgment, but also motivates their purchase decision. Despite playing such an important role, creating quality descriptions is fraught with difficulties due to many factors, such as funding or people. Therefore, this thesis aims to build a textual data generation solution that allows to create product descriptions with attractive content and can be applied in real cases. This solution will apply new techniques and advances, especially outstanding research results in the field of deep learning to improve the quality of description content. At the same time, the solution will also be proven to be able to work stably in the condition of lack of training data.

**Keywords:** E-commerce, deep learning, product description generation

## Mục lục

LỜI CAM ĐOAN.....	i
LỜI CẢM ƠN .....	ii
Tóm tắt .....	iii
Abstract .....	iv
Mục lục.....	v
Danh mục thuật ngữ và chữ viết tắt .....	vii
Danh mục hình vẽ .....	ix
Danh mục bảng biểu.....	x
Chương 1. Mở đầu .....	1
1.1. Đặt vấn đề.....	1
1.2. Các nghiên cứu liên quan.....	3
1.2.1. Sinh văn bản .....	3
1.2.2. Sinh mô tả sản phẩm .....	5
1.3. Mục tiêu.....	5
1.4. Cấu trúc luận văn.....	6
Chương 2. Cơ sở lý thuyết .....	7
2.1. Phát biểu bài toán .....	7
2.2. Transformer .....	9
2.2.1. Mã hóa vị trí .....	10
2.2.2. Bộ mã hóa.....	10
2.2.3. Bộ giải mã .....	15
2.3. GPT-2 .....	17
2.3.1. Kiến trúc .....	17
2.3.2. Mã hóa đầu vào .....	19
2.3.3. Ứng dụng.....	21



2.4. BART .....	21
2.4.1. Kiến trúc .....	21
2.4.2. Tiền huấn luyện .....	23
2.4.3. Ứng dụng .....	24
2.6. RoBERTa .....	24
2.7. Độ đa dạng từ vựng .....	27
Chương 3. Giải pháp .....	29
3.1. Sinh mô tả sản phẩm với GPT-2 .....	29
3.2. Tăng cường dữ liệu .....	31
3.2.1. Viết lại văn bản bằng mô hình dịch .....	31
3.2.2. Thay thế từ đồng nghĩa .....	32
3.2.3. Thêm từ theo ngữ cảnh .....	33
3.3. Task-Adaptive Pretraining .....	34
Chương 4. Thực nghiệm và đánh giá kết quả .....	35
4.1. Dữ liệu .....	35
4.2. Mô hình .....	36
4.3. Phương pháp đánh giá .....	37
4.3.1. Đánh giá tự động .....	37
4.3.2. Đánh giá thông qua con người .....	37
4.4. Kết quả và phân tích .....	38
Chương 5. Kết luận và định hướng phát triển .....	42
Tài liệu tham khảo .....	43
Tiếng Anh .....	43

## Danh mục thuật ngữ và chữ viết tắt

Thuật ngữ	Ý nghĩa
TMĐT	Thương mại điện tử
Token	Một dãy ký tự mang ý nghĩa cụ thể, là một đơn vị ngữ nghĩa trong xử lý ngôn ngữ, trong nhiều trường hợp token là các từ.
Word embedding	Vectơ từ nhúng
Task-Adaptive Pretraining	Tiền huấn luyện thích ứng với nhiệm vụ
Feedforward Neural Networks	Mạng truyền thẳng
Multi-Head Attention	Chú ý đa đầu
Self Attention	Kỹ thuật tự chú ý
NSP	Next Sentence Prediction - nhiệm vụ dự đoán từ tiếp theo trong quá trình huấn luyện mô hình xử lý ngôn ngữ tự nhiên
Byte-Pair Encoding	Thuật toán nén dữ liệu sử dụng các thành phần từ (Subword)
Unicode	Bộ mã chuẩn quốc tế được thiết kế để trở thành bộ mã duy nhất cho tất cả các ngôn ngữ

Downstream task	Có thể được dịch là nhiệm vụ phía sau, nhiệm vụ cụ thể - những tác vụ học giám sát được cải thiện dựa trên những mô hình tiền huấn luyện.
Pipeline	Một chuỗi chu trình, chuỗi thuật toán
Kernel	Hàm ánh xạ dữ liệu từ không gian ít chiều sang không gian nhiều chiều hơn
Seq2Seq	Sequence-to-Sequence hay mô hình Chuỗi sang chuỗi

## Danh mục hình vẽ

Hình 2.1. Ví dụ các dữ liệu của sản phẩm trên trang Amazon .....	8
Hình 2.2. Sơ đồ kiến trúc tổng thể của Transformer.....	9
Hình 2.3. Kiến trúc lớp của bộ mã hóa và bộ giải mã trong Transformer .....	11
Hình 2.4. Kiến trúc chi tiết một lớp của bộ mã hóa trong Transformer .....	12
Hình 2.5. Mô tả kiến trúc Multi-Head Attention .....	14
Hình 2.6. Kiến trúc của một lớp trong khối giải mã .....	15
Hình 2.7. Tổng quan quá trình biến đổi vector đầu ra của các lớp giải mã thành từ vựng tương ứng .....	17
Hình 2.8. Thống kê số lớp của từng phiên bản GPT-2 .....	18
Hình 2.9. Mô tả kiến trúc của GPT-2.....	19
Hình 2.10. Cơ chế mã hóa đầu vào của GPT-2.....	20
Hình 2.11. Sơ lược kiến trúc mô hình BART .....	22
Hình 2.12. Bảng so sánh hiệu suất của BART với các mô hình ngôn ngữ nổi tiếng [22] .....	23
Hình 2.13. Các thí nghiệm chứng minh tính hiệu quả khi loại bỏ nhiệm vụ NSP và sử dụng định dạng đầu vào Full-sentences [23].....	26
Hình 3.1. Định dạng dữ liệu huấn luyện trong phương pháp đề xuất.....	30

## **Danh mục bảng biểu**

Bảng 2.1. Các thông tin đầu vào cần thiết trong quá trình sinh mô tả.....	7
Bảng 4.1. Số lượng mô tả tổng quan và mô tả thuộc tính sản phẩm .....	35
Bảng 4.2. Tổng hợp kết quả đánh giá trên độ mất mát và độ đa dạng từ vựng .....	38
Bảng 4.3. Tổng hợp kết quả đánh giá trên hai tập dữ liệu đã thấy và chưa thấy .....	39
Bảng 4.4. Một số mô tả sản phẩm được tạo ra bởi các mô hình.....	40

# Chương 1. Mở đầu

## 1.1. Đặt vấn đề

Thương mại điện tử là một hình thức kinh doanh được định nghĩa là việc mua và bán hàng hóa trên mạng Internet. Các hoạt động này được thực hiện thông qua việc giao tiếp giữa các thiết bị điện tử như máy tính, điện thoại thông minh hoặc các loại thiết bị hỗ trợ khác. Đây là một xu hướng kinh doanh đang trở nên cực kỳ thịnh hành, song song với hình thức kinh doanh truyền thống. Với việc ứng dụng những tiện ích của TMĐT, các doanh nghiệp giờ đây có thể tối ưu hóa các phương hướng tiếp cận đối với tệp khách hàng của mình, mở rộng quy mô ảnh hưởng và tối đa hóa lợi nhuận.

Mặc dù có sức ảnh hưởng sâu rộng đối với nhiều nền kinh tế và xã hội, trên thực tế lịch sử phát triển của thương mại điện tử còn khá non trẻ. Hình thức có thể được coi là nguyên mẫu đầu tiên của xu hướng này xuất hiện vào cuối những năm 70 tại Anh Quốc, được phát minh bởi nhà sáng chế Michael Aldrich<sup>1</sup>. Ông đã tạo ra một hệ thống cho phép kết nối giữa máy tính của doanh nghiệp với vô tuyến của khách hàng thông qua đường dây điện thoại, đây trở thành nền móng cơ bản của mua sắm trực tuyến. Đến năm 1982, công ty Boston Computer Exchange<sup>2</sup> được thành lập, được xem là một trong những công ty thương mại điện tử đầu tiên, cung cấp giải pháp mua sắm trực tuyến máy tính cũ. Vào các năm tiếp theo - 1995, 1998, các tập đoàn công nghệ hàng đầu trong TMĐT hiện nay - Amazon<sup>3</sup> và Paypal<sup>4</sup> lần lượt ra đời. Năm 2006, lập trình viên Tobias Lutke<sup>5</sup> tạo ra Shopify<sup>6</sup>, nền tảng cho phép tạo website bán hàng tự động, sau này cũng trở thành một trong những công ty công nghệ nổi tiếng nhất trong lĩnh vực này. Năm 2007, Iphone<sup>7</sup> ra mắt lần đầu tiên, tạo nên một hiện tượng công nghệ trên toàn thế giới, thay đổi các phương thức tương tác của người sử dụng, mở rộng cơ hội tiếp cận của người dùng với các dịch vụ trực tuyến. Và trong những năm trở lại đây, những doanh nghiệp thương mại điện tử thế

---

<sup>1</sup> [https://en.wikipedia.org/wiki/Michael\\_Aldrich](https://en.wikipedia.org/wiki/Michael_Aldrich)

<sup>2</sup> [https://en.wikipedia.org/wiki/Boston\\_Computer\\_Exchange](https://en.wikipedia.org/wiki/Boston_Computer_Exchange)

<sup>3</sup> <https://www.amazon.com/>

<sup>4</sup> <https://www.paypal.com/>

<sup>5</sup> [https://en.wikipedia.org/wiki/Tobias\\_L%C3%BCtke](https://en.wikipedia.org/wiki/Tobias_L%C3%BCtke)

<sup>6</sup> <https://www.shopify.net>

<sup>7</sup> [https://en.wikipedia.org/wiki/IPhone\\_\(1st\\_generation\)](https://en.wikipedia.org/wiki/IPhone_(1st_generation))

hệ mới với quy mô khổng lồ cũng lần lượt xuất hiện, như Shopee<sup>8</sup>, Lazada<sup>9</sup>, ... đánh dấu một thời đại kinh tế số bùng nổ và cạnh tranh mạnh mẽ.

Các hình thức của thương mại điện tử luôn phát triển và thay đổi dựa trên những tiến bộ công nghệ cùng nhu cầu của xã hội. Như đã được đề cập ở trên, xu hướng này đã trải qua nhiều lần thay đổi về hình thái, từ ý tưởng nguyên bản đầu tiên của Michael Aldrich đến việc hình thành ra các sàn trao đổi hàng hóa khổng lồ như Amazon, Shopee, hay các nền tảng cho phép tạo trang mạng mua bán trực tuyến như Shopify. Có thể nói, những tiện ích công nghệ giờ đây đang ngày càng đa dạng và đơn giản hóa trong cách tiếp cận, cho phép mở rộng khả năng tham gia của nhiều đối tượng kinh doanh. Họ có thể lựa chọn việc sử dụng các sàn TMĐT để giới thiệu sản phẩm, sử dụng các công cụ tạo cửa hàng tự động hoặc tự xây dựng các trang web riêng. Điều này sẽ vừa mở ra những cơ hội cũng vừa đặt ra những thách thức mới, bởi vì khi cơ hội tiếp cận thị trường ngày càng trở nên dễ dàng thì cũng dẫn đến việc cạnh tranh sản phẩm trở nên khó khăn. Để thu hút khách hàng, ngoài việc đầu tư nâng cấp giá trị nội tại của sản phẩm và xây dựng chiến lược quảng bá đa dạng trên nhiều kênh khác nhau như trang mạng, mạng xã hội, ... thì các cá nhân, doanh nghiệp còn cần chú trọng đến nội dung truyền tải của những sản phẩm đó, bởi đây là những thông tin cốt yếu được người tiêu dùng tiếp nhận. Nội dung của một sản phẩm thông thường có thể bao gồm hình ảnh, âm thanh và văn bản. Trong ba nhân tố này, văn bản sẽ đóng vai trò quan trọng nhất bởi vì trong khi hình ảnh và âm thanh có thể gây ấn tượng tức thời, giúp thu hút sự chú ý trong thời điểm đầu tiên tiếp xúc thì văn bản biểu đạt lại là một yếu tố không thể thiếu giúp khách hàng hiểu được thông tin từ tổng quan đến chi tiết của sản phẩm, là yếu tố then chốt dẫn đến quyết định mua hàng của họ.

Để tạo ra những nội dung văn bản chất lượng, thường sẽ có hai cách chính. Một là sao chép các nội dung sản phẩm tương tự trên các trang web thương mại điện tử lớn, nội dung trên các trang này thường đã được đầu tư sản xuất một cách chuyên nghiệp. Nhược điểm của phương pháp này là các nội dung trùng lặp thường bị đánh giá rất thấp bởi các công cụ tìm kiếm, từ đó khiến cho trang web của bên bán hàng khó tiếp cận với khách hàng hơn. Cách làm thứ hai là thuê người viết mô tả sản phẩm, đây là một phương án cho phép tạo ra những mô tả sản phẩm có nội dung phong phú

---

<sup>8</sup> <https://shopee.com/>

<sup>9</sup> <https://www.lazada.com/>

nhưng lại ẩn chứa hai bất cập chính là: Yêu cầu sự đầu tư đáng kể về tài chính và bị phụ thuộc hoàn toàn vào yếu tố con người. Trong đó, yếu tố về con người là một rủi ro đáng cân nhắc bởi con người thường có năng suất giới hạn, khó có thể đáp ứng yêu cầu về tần suất công việc cao và còn có thể bị ảnh hưởng bởi các vấn đề như sức khỏe, tâm lý, ...

Có thể nói, những vấn đề đề cập ở trên đây vô hình chung sẽ gây ra những khó khăn cho các cá nhân, doanh nghiệp có nhu cầu tham gia thị trường số, đặc biệt là những đối tượng có khả năng đầu tư hạn chế. Do đó, cần thiết một công cụ có thể tự động hóa việc sinh nội dung sản phẩm để hỗ trợ các đối tượng kinh doanh thương mại điện tử vừa có thể gia tăng khả năng cạnh tranh, vừa có thể tiết kiệm tối đa chi phí. Để xây dựng được công cụ này, cần xem xét bài toán sinh văn bản (text generation) trong lĩnh vực Xử lý ngôn ngữ tự nhiên. Bài toán này được định nghĩa là quá trình sinh văn bản đầu ra dựa trên điều kiện đầu vào với mục tiêu bắt chước càng giống càng tốt với văn phong do con người viết ra.

## **1.2. Các nghiên cứu liên quan**

Trong những năm gần đây, sinh văn bản nói chung và sinh mô tả sản phẩm nói riêng đã và đang nhận được nhiều sự quan tâm và đầu tư nghiên cứu. Đã có nhiều giải pháp được công bố, một vài trong số đó đã cho thấy những kết quả hứa hẹn, cho phép mở ra những hướng đi mới trong lớp bài toán này. Tuy vậy, vẫn tồn tại một số vấn đề còn chưa được giải quyết, cần được tiếp tục cải thiện và phát triển trong thời gian tương lai. Trong chương này, luận văn sẽ đi qua một số công trình tiêu biểu về sinh văn bản nói chung và sinh mô tả sản phẩm nói riêng, từ đó phân tích những ưu nhược điểm của chúng cùng những liên hệ với phương pháp mới.

### **1.2.1. Sinh văn bản**

Sinh văn bản tự động vẫn luôn là một bài toán quan trọng trong xử lý ngôn ngữ tự nhiên và đã thu hút được những sự đầu tư nghiên cứu từ rất sớm. Một trong số những hướng tiếp cận đầu tiên có thể kể đến là tập hợp những phương pháp mô hình hóa cấu trúc diễn ngôn và học các biểu diễn quan hệ giữa các đơn vị văn bản. Những công trình nổi bật trong lớp giải pháp này có thể kể đến như Lý thuyết cấu trúc tu từ (Rhetorical Structure Theory) của Mann và Thompson [1] hay Lý thuyết thể hiện diễn ngôn được phân đoạn (Segmented Discourse Representation Theory) của Asher và Lascarides [2]. Ngoài ra, còn có một số hướng nghiên cứu khác tập trung vào việc



xây dựng các mô hình sinh dựa trên khuôn mẫu kết hợp cùng những kỹ thuật dựa trên lý thuyết thống kê như nén câu, diễn giải từ vựng, chuyển đổi cú pháp, ... để cải thiện chất lượng sinh văn bản như công trình của Sporleder [3], hay của Clarke và M. Lapata [4]. Đặc điểm chung của những nghiên cứu trong thời gian này là đều chưa tạo ra được những kết quả thật sự đột phá. Tuy nhiên, đây có thể coi là những bước tiến quan trọng đầu tiên, góp phần tạo nên nền móng vững chắc để phát triển những ý tưởng mới về sau này.

Trong những thập kỷ tiếp theo, cùng sự nổi lên của học máy và sự sẵn có của những nguồn dữ liệu mở, sinh văn bản đã đạt được những bước phát triển vượt bậc, những nghiên cứu trong thời kỳ này có xu hướng dịch chuyển từ việc áp dụng những mô hình đơn giản như sử dụng luật, khuôn mẫu sang ứng dụng những kỹ thuật tiên bộ của học máy. Sự thay đổi này bắt đầu diễn ra trong khoảng từ những năm đầu của thập niên 2010, với những nghiên cứu chủ yếu tập trung vào việc ứng dụng những mạng nơ-ron hồi quy (Recurrent Neural Network) như công trình của Graves (2013) [5] đề xuất việc ứng dụng mô hình Bộ nhớ dài ngắn hạn (Long short-term memory [6], một biến thể của mạng nơ-ron hồi quy) để tạo ra những văn bản phức tạp hay giải pháp của Cho và các cộng sự [7] với việc xây dựng mạng Nút hồi tiếp có cổng (Gated Recurrent Unit - GRU) cho việc học các biểu diễn ngôn ngữ và đặc biệt là sự xuất hiện của mô hình chuỗi sang chuỗi (Sequence-to-Sequence - Seq2Seq) [8] - một mô hình đột phá với kiến trúc mã hóa-giải mã kết hợp cùng mạng tuần tự. Mô hình hoàn toàn mới này đã chứng minh được tính hiệu quả trong rất nhiều tác vụ khác nhau như sinh văn bản, dịch máy, sinh hội thoại, ... Tuy nhiên, các giải pháp dựa trên Seq2Seq trong thời điểm này cũng đã gặp phải một vấn đề nan giải là không thể nắm bắt được sự phức thuộc ngữ nghĩa xa trong văn bản dài. Điều này đã thúc đẩy những nỗ lực nghiên cứu mới như công trình xây dựng các mạng “attention” (Bahdanau cùng các cộng sự [9]) hay các mạng “pointer” (Vinyals cùng các cộng sự [10]). Đến năm 2017, Google đã công bố kiến trúc Transformer với cơ chế mã hóa kết hợp kỹ thuật attention mềm dẻo (soft attention) trên bộ giải mã đã thực sự giải quyết được phần nào vấn đề trên. Mô hình này cũng đã trở thành một tiêu chuẩn mới trong sự phát triển của những mô hình ngôn ngữ thế hệ tiếp theo, tiêu biểu là BERT, GPT-2, GPT-3, XLNet, ...

### 1.2.2. Sinh mô tả sản phẩm

Giống như Sinh văn bản, Sinh mô tả sản phẩm cũng đã trải qua một thời gian dài nghiên cứu và phát triển. Một trong những nỗ lực đầu tiên để giải quyết bài toán này có thể kể đến như phương pháp sử dụng khuôn mẫu kết hợp với các kỹ thuật thống kê của Wang cùng các cộng sự [11]. Trong nội dung giải pháp, các tác giả đã đưa ra một số kỹ thuật mới bao gồm: Lựa chọn giá trị thuộc tính của sản phẩm khi điền vào mẫu (Value Preference); Trích trợn xác suất xuất hiện của thuộc tính theo phương pháp thống kê (Attribute Prior); Tính toán điểm cấu trúc của văn bản dựa trên các câu của nó cùng xác suất có điều kiện giữa các thuộc tính trong từng câu; Đánh giá, xếp hạng các mô tả dựa trên các thông tin đã được trích xuất cùng phương pháp SVM với hạt nhân tuyến tính (linear kernel) [12]. Kết quả báo cáo của nghiên cứu cũng đã cho thấy giải pháp có thể tạo ra tính liên kết giữa các giá trị thuộc tính của sản phẩm đầu vào với nội dung của những khuôn mẫu có sẵn. Tuy nhiên, phương pháp này vẫn còn gặp nhiều hạn chế, tiêu biểu là việc không thể tạo ra những văn bản có cấu trúc phong phú và chứa hàm lượng thông tin cao. Để giải quyết vấn đề này, một vài công trình nghiên cứu khác chuyển hướng sang áp dụng những kỹ thuật học sâu như công trình của Chen cùng các cộng sự [13] đã sử dụng mô hình Transformer để cải thiện chất lượng ngữ nghĩa của mô tả sản phẩm. Mặc dù nghiên cứu này cho kết quả đầu ra tương đối khả quan nhưng vẫn đối mặt một nhược điểm quan trọng là mô hình đòi hỏi rất nhiều dữ liệu trong quá trình huấn luyện. Đây là một vấn đề cần được cân nhắc kỹ lưỡng trong nhiều tình huống áp dụng bởi vì dữ liệu có thể thu thập trong thực tế sẽ rất có thể bị giới hạn trong một số điều kiện nhất định.

### 1.3. Mục tiêu

Từ những phân tích trên, luận văn này sẽ hướng đến mục tiêu xây dựng một giải pháp sinh dữ liệu văn bản cho phép tạo ra những mô tả sản phẩm có nội dung thu hút và có thể áp dụng được trong những trường hợp thực tế. Giải pháp này sẽ ứng dụng những kỹ thuật, tiến bộ mới, đặc biệt là những thành quả nghiên cứu nổi bật trong lĩnh vực học sâu để cải thiện chất lượng nội dung mô tả. Đồng thời, giải pháp cũng sẽ được chứng minh khả năng hoạt động ổn định trong điều kiện thiếu thốn dữ liệu huấn luyện. Ngoài ra, luận văn cũng sẽ đưa ra chi tiết những thử nghiệm, đánh giá trên những mô hình / phương pháp có sử dụng trong quá trình nghiên cứu để phân tích những ưu, nhược điểm của chúng khi áp dụng vào bài toán này.

#### **1.4. Cấu trúc luận văn**

Luận văn này được chia thành 5 phần, với Chương 1 giới thiệu bài toán cùng việc phân tích những nghiên cứu liên quan đã được trình bày ở trên. Tiếp theo, Chương 2 sẽ trình bày cách mô hình hóa cùng một số kiến thức về nền tảng học sâu có liên quan đến phương pháp mới. Chi tiết triển khai, các kỹ thuật xử lý sẽ được nêu ở Chương 3. Chương 4 sẽ trình bày báo cáo, đánh giá kết quả thực nghiệm của giải pháp và cuối cùng Chương 5 sẽ nêu lên kết luận của luận văn cùng những định hướng phát triển của nghiên cứu này trong thời gian sắp tới.

## Chương 2. Cơ sở lý thuyết

Trong phần này, luận văn sẽ mô tả cách thức mô hình hóa bài toán, chuyển từ bài toán sinh dữ liệu mô tả sản phẩm sang bài toán Seq2Seq có điều kiện. Một số kiến thức phục vụ trong quá trình triển khai phương pháp đề xuất cũng sẽ được giới thiệu và trình bày tổng quan, như mô hình Transformer, GPT-2, Độ đa dạng từ vựng, ...

### 2.1. Phát biểu bài toán

Đầu tiên, luận văn sẽ mô hình hóa bài toán ở dạng cơ bản. Ở bước này, có thể phát biểu nội dung bài toán như sau: “Đối với một đầu vào văn bản là tên sản phẩm, xây dựng một hệ thống cho phép tự động tạo ra một mô tả tương ứng, giả định rằng tiêu đề sản phẩm và mô tả của chúng được định nghĩa là  $x = \{x_1, x_2, \dots, x_{T_x}\}$  và  $y = \{y_1, y_2, \dots, y_{T_y}\}$ .”

Mục tiêu ở bước này là phân tích xác suất có điều kiện giữa kết quả đầu ra đối với chuỗi đầu vào  $P_{\Theta}(y|x)$  thành tích các xác suất có điều kiện của các token trong chuỗi đối với các token được tạo ra trước đó của chúng cùng chuỗi đầu vào:

$$P_{\Theta}(y|x) = \prod_{t=1}^{T_y} P_{\Theta}(y_t|y_{<t}, x) \quad (1)$$

Tuy nhiên, có thể dễ dàng nhận thấy rằng đầu vào văn bản được đề xuất trong phát biểu trên chưa thật sự cung cấp những thông tin bao quát về sản phẩm. Điều này là cần thiết bởi vì để đi đến quyết định mua hàng, người tiêu dùng cần nắm được nhiều thông tin về sản phẩm hơn là chỉ tiêu đề của chúng. Do đó, đầu vào cần được bổ sung một số thông tin quan trọng của sản phẩm như cấu trúc danh mục, thương hiệu và các thuộc tính chi tiết. Mô tả ngắn gọn của những dữ liệu này sẽ được tóm tắt như trong Bảng 2.1:

Bảng 2.1. Các thông tin đầu vào cần thiết trong quá trình sinh mô tả

Tên	Ý nghĩa
Category structure	Cấu trúc danh mục sản phẩm
title	Tiêu đề của sản phẩm

brand	Nhãn hiệu của sản phẩm
attribute	Thuộc tính của sản phẩm

Bảng 2.1 đã liệt kê những thông tin đầu vào cần thiết trong quá trình sinh mô tả. Lý do những thông tin này được chọn là bởi cấu trúc danh mục có thể giúp người tiêu dùng phân loại được sản phẩm, thương hiệu giúp định vị được độ tin cậy lẫn nguồn gốc xuất xứ và cuối cùng các thuộc tính sẽ cung cấp cái nhìn chi tiết về những đặc trưng của sản phẩm.

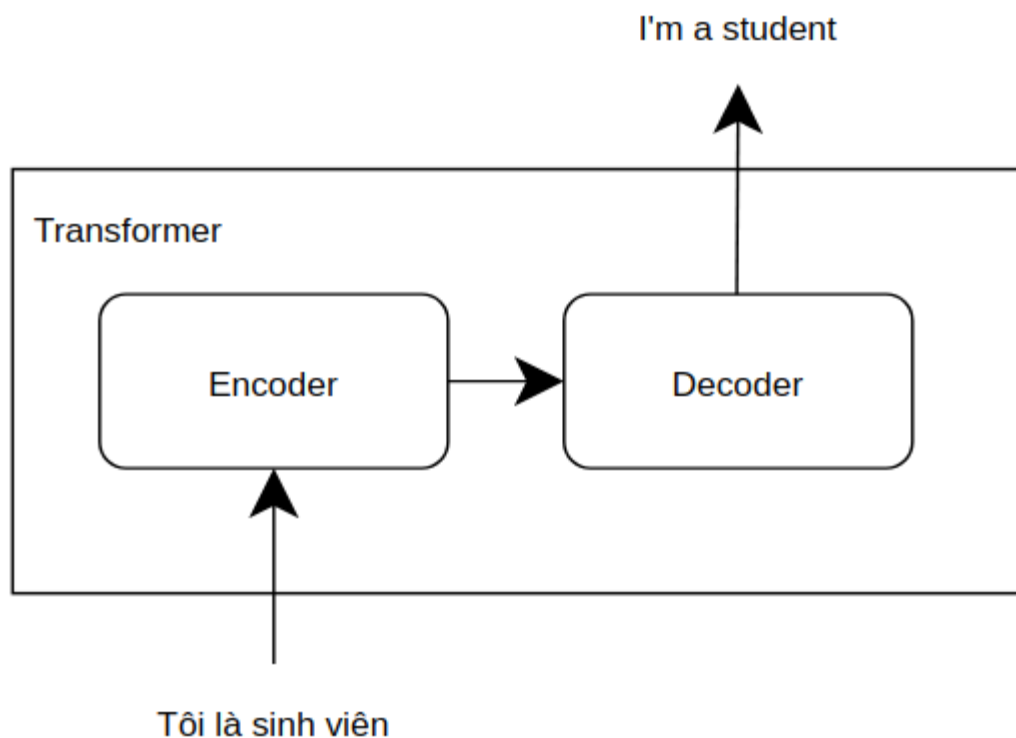
**Bose QuietComfort Noise Cancelling Earbuds - Bluetooth** ← **Tiêu đề**  
**Wireless Earphones, Triple Black, the World's Most Effective Noise Cancelling Earbuds**  
 Visit the Bose Store  
 ★★★★★ 13,453 ratings | 507 answered questions  
**Amazon's Choice** In Earbud & In-Ear Headphones by Bose ← **Cấu trúc thư mục**  
 Only 1 left in stock - order soon.  
 Color: **Triple Black**  
 \$178.95 2 options from \$199.00 \$198.49  
 Roll over image to zoom in  
 Brand: Bose  
 Color: Triple Black  
 Connectivity Technology: Wireless ← **Thuộc tính**  
 Model Name: QuietComfort Earbuds  
 Form Factor: In Ear  
**About this item**  
 • Noise cancelling earbuds – Engineered with the world's most effective noise cancelling. Bose controllable noise cancellation and full Transparency Mode allow you to eliminate distractions, let your surroundings in, or a little bit of each.  
 • High-fidelity audio – An exclusive acoustic architecture brings your music, podcasts, and videos to life, while Volume-optimized Active EQ keeps the sound crisp and balanced at any level.  
 • Comfortable wireless earbuds – Three sizes of StayHear Max eartips are included for a secure fit. Crafted from soft silicone, they provide passive noise blocking with total comfort. ← **Mô tả**  
 • Simple touch controls – It's as easy as tap, swipe, or press. Control your noise cancellation settings, pause, play, and volume right from the side of the earbud itself. Enable it all in the Bose Music app.  
 • Noise-rejecting microphones – Tiny microphones work together to focus on your voice, filtering out surrounding noises for fewer distractions during phone calls.  
 • Long battery life – Listen for up to 6 hours on a single charge. The charging case provides an additional 12 hours and charges via the included USB-C cable or wirelessly with any Qi-certified charging mat (not included).  
 • Weather and sweat resistant – These IPX4-rated wireless earbuds can withstand sweat, water, and some inclement weather.  
 Show more

Hình 2.1. Ví dụ các dữ liệu của sản phẩm trên trang Amazon

Như trong ví dụ của Hình 2.1, một sản phẩm tai nghe trên trang TMĐT Amazon có thể có những dữ liệu quan trọng như tiêu đề, cấu trúc thư mục, thuộc tính và các mô tả chi tiết về sản phẩm.

## 2.2. Transformer

Transformer là một mô hình ngôn ngữ được Google phát triển và công bố vào năm 2017 trong bài báo “Attention is All You Need” [14]. Kiến trúc của mô hình này bao gồm hai thành phần chính là Bộ mã hóa và Bộ giải mã. Nhiệm vụ của chúng là biến đổi nội dung đầu vào thành một nội dung đầu ra tương ứng. Cụ thể hơn, bộ mã hóa sẽ ánh xạ chuỗi token đầu vào  $x = (x_1, x_2, \dots, x_n)$  thành một chuỗi liên tục tương ứng  $z = (z_1, z_2, \dots, z_n)$  và bộ giải mã sẽ sử dụng chuỗi này để tạo ra một chuỗi đầu ra  $y = (y_1, y_2, \dots, y_n)$ . Dưới đây là sơ đồ kiến trúc tổng thể của mô hình.



Hình 2.2. Sơ đồ kiến trúc tổng thể của Transformer

Như trong Hình 2.2, câu “Tôi là sinh viên” là một văn bản nguồn đã được bộ mã hóa xử lý thông qua các phép biến đổi để trở thành đầu vào của bộ giải mã. Sau đó, bộ giải mã lại biến đổi tiếp đầu vào này thành kết quả cuối cùng là “I’m a student”. Để làm rõ hơn quy trình cũng như tính hiệu quả của kiến trúc này, trong những phần nội dung tiếp theo, luận văn sẽ đi vào chi tiết hơn về những thành phần cấu tạo cùng những kỹ thuật, lý thuyết liên quan đến mô hình.

### 2.2.1. Mã hóa vị trí

Như chúng ta đã biết, trước khi các từ trong văn bản thực sự được xử lý bởi mô hình, chúng phải trải qua quá trình mã hóa để chuyển sang một hình thức biểu diễn thông tin khác mà chương trình có thể hiểu được. Hình thức biểu diễn thông tin này được gọi là các từ nhúng, vector từ nhúng hoặc word embedding. Các vector này sau đó được nối với nhau trở thành một ma trận hai chiều và được xử lý bởi các tiến trình tiếp theo của mô hình. Quá trình này rất phổ biến và không có vấn đề gì khi được áp dụng ở các mô hình tuần tự. Nhưng trong Transformer, có một vấn đề khác nảy sinh là mô hình không thể nhận biết được vị trí của các từ thông qua word embedding do cơ chế xử lý các từ song song của mô hình. Để giải quyết vấn đề này, các tác giả đã giới thiệu cơ chế mã hóa mới có tên gọi là mã hóa vị trí (position encoding), với mục tiêu mã hóa vị trí của các từ bằng một vector có kích thước bằng word embedding và sẽ được cộng trực tiếp vào vector embedding tương ứng. Dưới đây là công thức mã hóa vị trí được nêu trong bài báo:

$$p_t^i = f(t)^i = \begin{cases} \sin(w_k * t) & \text{if } i = 2k \\ \cos(w_k * t) & \text{if } i = 2k + 1 \end{cases} \quad (2)$$

Trong đó:

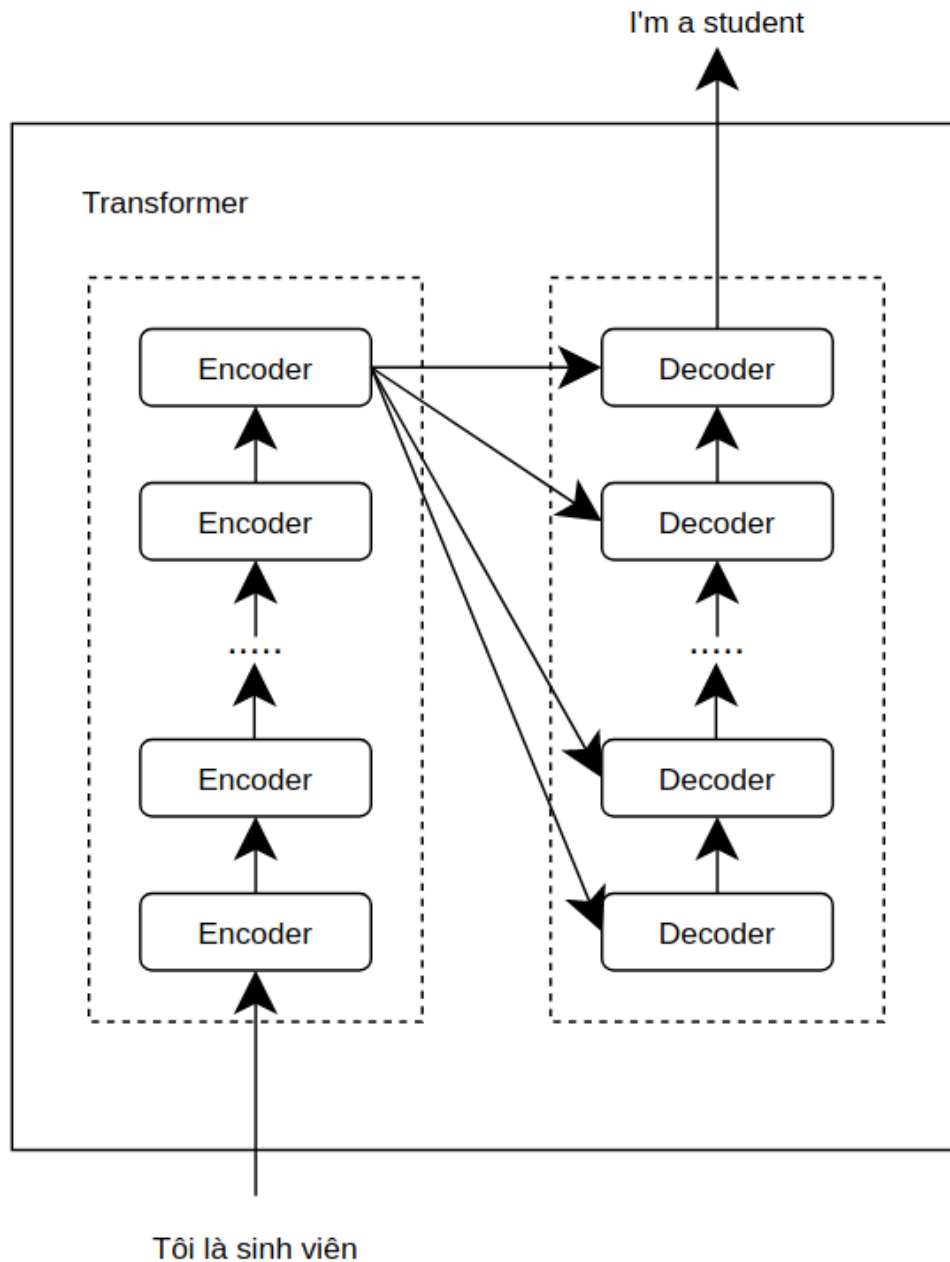
$$w_k = \frac{1}{10000^{2k/d}} \quad (3)$$

Ở công thức (2),  $t$  là vị trí của từ trong văn bản,  $i$  là vị trí của chiều trong không gian embedding, các tác giả dùng hàm sin để tính giá trị tương ứng với vị trí chiều chẵn và cos với vị trí chiều lẻ.

### 2.2.2. Bộ mã hóa

#### 2.2.2.1. Kiến trúc

Như đã được mô tả trong Hình 2.2, Bộ mã hóa là một trong hai thành phần quan trọng của mô hình Transformer, đóng vai trò mã hóa đầu vào thành một biểu diễn mới giàu thông tin hơn gọi là “vector ngữ cảnh”. Kiến trúc của bộ mã hóa được tạo nên bằng cách xếp chồng các lớp mã hóa với nhau (thường là 6 lớp), tạo thành một mạng truyền thẳng (Feedforward Neural Networks). Kiến trúc này cho phép các lớp xử lý đồng thời các từ, thay vì xử lý tuần tự giống các mô hình khác như LSTM.

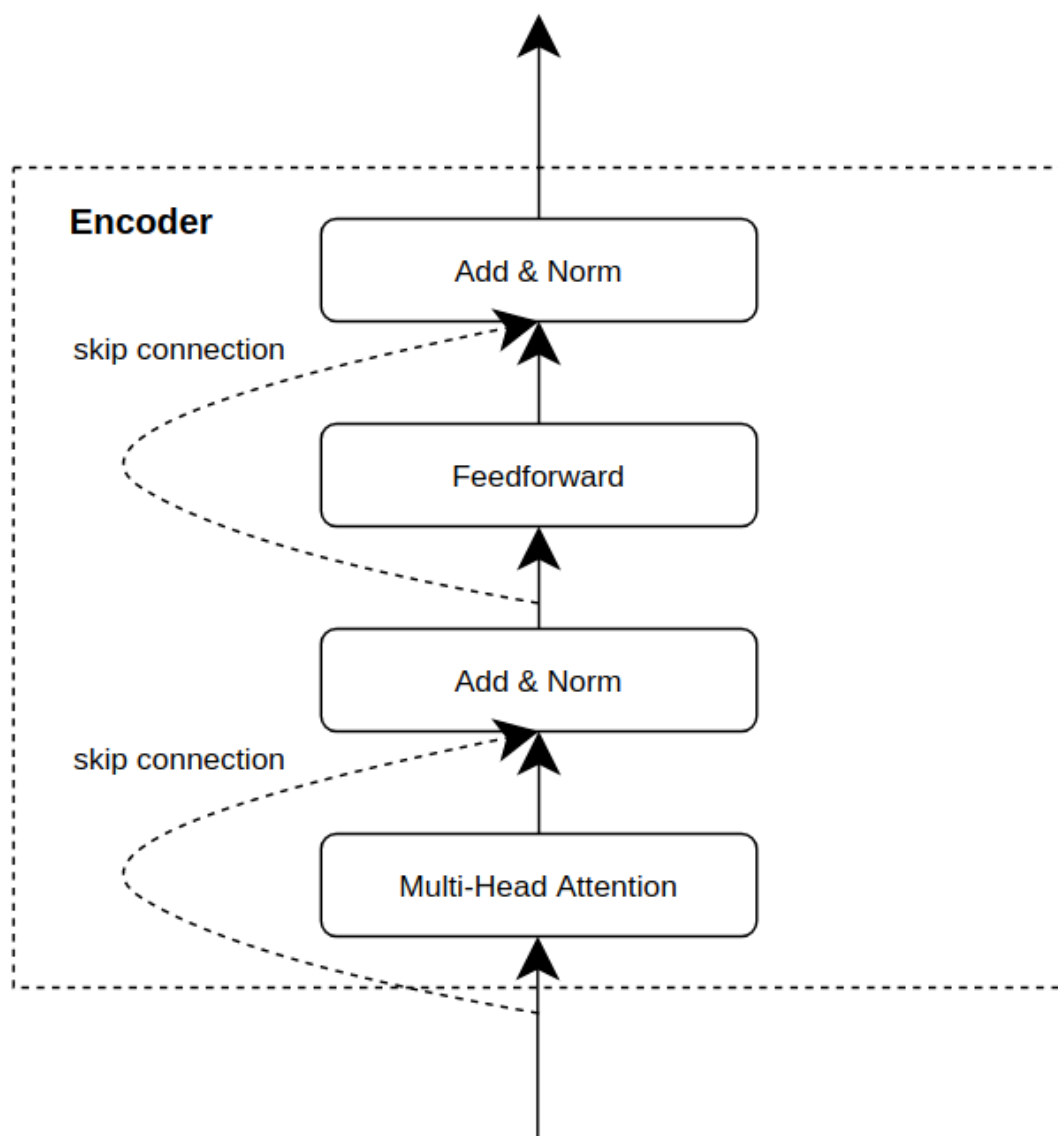


Hình 2.3. Kiến trúc lớp của bộ mã hóa và bộ giải mã trong Transformer

Như trong Hình 2.3, khối bộ mã hóa bản chất là một chuỗi các lớp xếp chồng lên nhau, có nhiệm vụ xử lý dữ liệu đầu vào để cung cấp thông tin cho các lớp trong khối Decoder.

Tiếp theo, trong mỗi bộ mã hóa lại có hai thành phần chính là tầng tập trung đa đầu (multi-head attention) và mạng truyền thẳng (feedforward network), ngoài ra còn có bỏ kết nối (skip connection) và lớp chuẩn hóa (normalization). Sơ đồ kiến trúc chi tiết của một lớp mã hóa sẽ được trình bày như dưới đây:





Hình 2.4. Kiến trúc chi tiết một lớp của bộ mã hóa trong Transformer

Như trong hình, sau khi nhận được đầu vào là ma trận biểu diễn của các từ đã được thêm thông tin ở bước mã hóa vị trí (Positional Encoding), mô đun chú ý đa đầu sẽ là nơi xử lý đầu tiên của dữ liệu. Mô đun này là một phần trong cơ chế chú ý, bao gồm nhiều lớp tự chú ý (self-attention) để có thể xử lý nhiều mẫu khác nhau trong dữ liệu.

#### 2.2.2.2. Tự chú ý

Tự chú ý hay Self Attention là một cơ chế cho phép mô hình xem xét những thông tin có liên quan tới một từ trong ngữ cảnh văn bản (thường là những từ khác trong

chính văn bản đó) để tìm ra manh mối có thể giúp dẫn đến mã hóa tốt hơn cho từ này. Ví dụ trong một câu “Tôi đã mua một quyển sách vào chiều thứ bảy và nó có giá 100 nghìn đồng”, khi “nó” được mã hóa, mô hình sẽ đặc biệt chú ý vào các từ có liên quan như “quyển sách”, điều này cũng rất hợp lý khi trong thực tế không phải bất kỳ cặp từ nào cũng có mối tương quan giống nhau. Việc áp dụng kỹ thuật này sẽ giúp mô hình cải thiện khả năng học phụ thuộc xa, giảm thiểu sự phức tạp trong mỗi lớp đồng thời song song hóa khối lượng tính toán, qua đó tăng tốc các tác vụ của mô hình. Tự chú ý cũng đã được ứng dụng rất thành công trong nhiều bài toán khác nhau như đọc hiểu, tóm tắt trừu tượng văn bản, ...

Để tính toán giá trị attention cho toàn bộ các từ trong một câu, mô hình Transformer đã sử dụng ba ma trận là Query, Keys và Values (viết tắt là Q, K, V tương ứng). Trong đó, Query dùng để chứa thông tin của từ được tìm kiếm, mỗi dòng của Q sẽ là một vector đại diện cho các từ đầu vào, Key dùng để biểu diễn thông tin của các từ được so sánh với từ cần tìm kiếm đó (mỗi dòng cũng là một vector của từ đầu vào) và Values dùng để biểu diễn nội dung, ý nghĩa của các từ. Hai ma trận Q, K sẽ được sử dụng để tính giá trị attention của các từ trong câu đối với một từ xác định. Tiếp theo, các giá trị này sẽ được sử dụng để tính ra các vector attention dựa trên việc trung bình hóa có trọng số với ma trận V. Phương trình tính toán Attention được mô tả như sau:

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \quad (4)$$

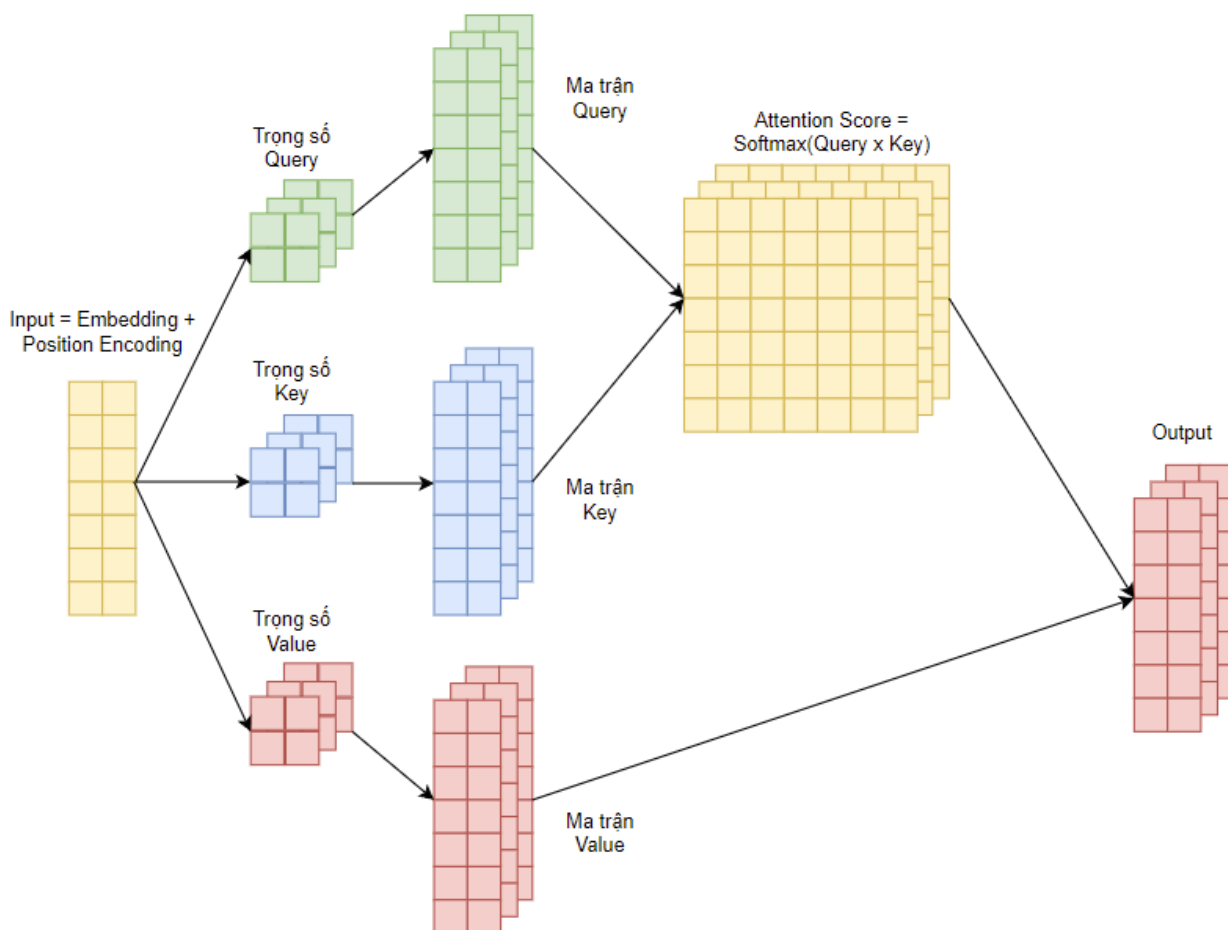
Trong đó:

- $d_k$  là số chiều của vector Key. Việc chia cho  $\sqrt{d_k}$  là nhằm mục đích tránh tràn luồng nếu số mũ trở nên quá lớn.

### 2.2.2.3. Tầng chú ý đa đầu

Áp dụng kỹ thuật tự chú ý, Transformer có thể học được những mối quan hệ giữa các từ với nhau trong một văn bản. Tuy nhiên, trong thực tế những mối quan hệ này cũng rất đa dạng và phong phú, không thể chỉ gói gọn trong một hình thức thể hiện nhất định. Do đó, để mở rộng cũng như cải thiện hiệu suất của mô hình, các tác giả đã đề xuất sử dụng nhiều lớp tự chú ý nhằm mục đích nắm bắt nhiều nhất có thể những mối quan hệ trong văn bản. Và để phân biệt các lớp tự chú ý với nhau, các ma

trận trọng số Query, Key, Value sẽ có thêm một chiều “depth” chứa định danh của lớp.



Hình 2.5. Mô tả kiến trúc Multi-Head Attention

Trên hình có thể thấy, mô-đun chú ý đa đầu (Multi-Head Attention) bao gồm nhiều lớp Attention, cùng xử lý dữ liệu đầu vào là Embedding sau khi đã được thêm dữ liệu mã hóa vị trí và cho ra các kết quả đầu ra tương ứng.

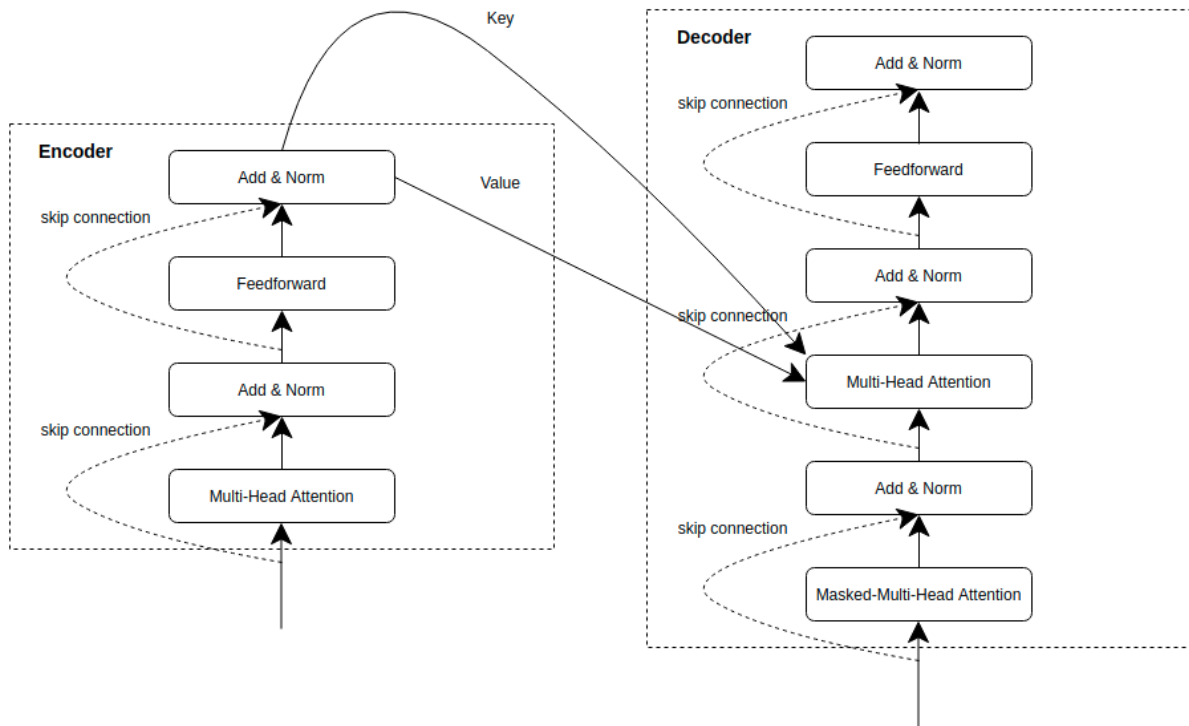
#### 2.2.2.4. Kết nối phần dư và Lớp chuẩn hóa

Trong mỗi lớp encoder, các kết nối phần dư (residual connection) và lớp chuẩn hóa đóng vai trò giúp mô hình có thể hội tụ nhanh hơn và tránh mất mát thông tin trong quá trình huấn luyện. Chi tiết minh họa của hai thành phần này đã được đề cập đến trong Hình 3.3.

## 2.2.3. Bộ giải mã

### 2.2.3.1. Kiến trúc

Giống như bộ mã hóa, bộ giải mã cũng có kiến trúc xếp chồng với 6 lớp tạo thành mạng truyền thẳng. Mỗi lớp sẽ nhận thông tin đầu vào từ bộ mã hóa để thực hiện tác vụ giải mã vector của câu nguồn thành đầu ra tương ứng. Kiến trúc của những lớp này cơ bản giống với những lớp trong bộ mã hóa, ngoại trừ có thêm một lớp chú ý đa đầu nằm ở giữa đóng vai trò học mối quan hệ tương quan giữa các từ đang được dịch với các từ trong văn bản gốc. Dưới đây là mô tả kiến trúc chi tiết của từng lớp giải mã.



Hình 2.6. Kiến trúc của một lớp trong khối giải mã

Như trong Hình 2.6, mỗi lớp giải mã sẽ bao gồm 3 thành phần chính là tầng tập trung đa đầu có mặt nạ (Masked Multi Head Attention), chú ý đa đầu (Multi-Head Attention) và mạng truyền thẳng (Feedforward). Trong đó ngoài lớp truyền thẳng đã biết thì tập trung đa đầu có mặt nạ là lớp có chức năng mã hóa các token trong văn bản đích trong quá trình dịch nhưng có ẩn đi những từ chưa được mô hình dịch đến. Đầu ra của lớp này và đầu ra của khối mã hóa được lớp chú ý đa đầu sử

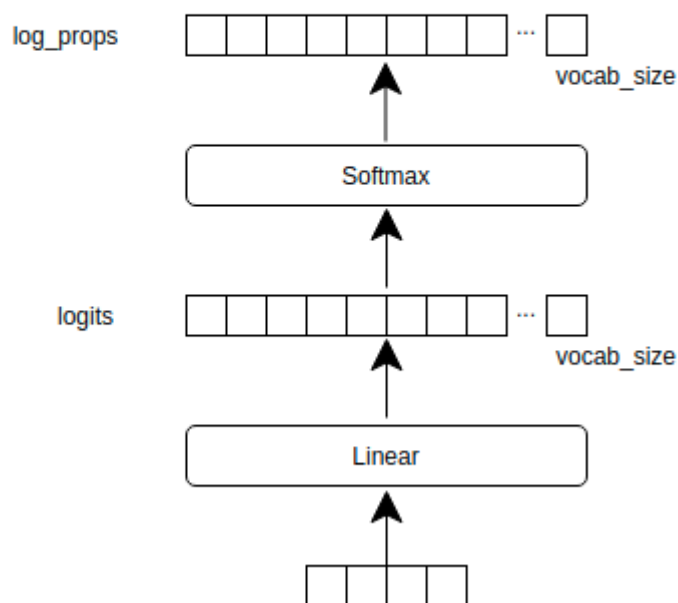
dụng để học ra mối tương quan giữa từ nguồn và từ đang được dịch như đã đề cập ở phần trên.

### **2.2.3.2. Lớp tuyến tính cuối cùng và lớp Softmax**

Sau tất cả quá trình xử lý, các lớp giải mã sẽ cho ra đầu ra cuối cùng là một vector bao gồm những số thực. Để chuyển hóa kết quả này thành một từ có ý nghĩa, mô hình sẽ sử dụng lớp tuyến tính cuối cùng (Final Linear) cùng lớp Softmax. Trong đó, lớp tuyến tính cuối cùng bản chất là một mạng nơ-ron kết nối đầy đủ có nhiệm vụ biến đổi vector đầu vào thành một vector lớn hơn rất nhiều được gọi là vector “logits”. Vector này có số chiều bằng với số từ trong tập từ vựng của mô hình, mỗi ô trong vector chứa giá trị tương đương với một điểm của một từ duy nhất. Những điểm này sẽ được chuyển hóa thành giá trị xác suất thông qua lớp Softmax. Cuối cùng mô hình sẽ chọn ra từ tương ứng với ô chứa giá trị xác suất cao nhất trong vector trên.

Từ trong tập từ vựng tương ứng am

Chỉ số của ô có giá trị lớn nhất 5



Hình 2.7. Tổng quan quá trình biến đổi vector đầu ra của các lớp giải mã thành từ vựng tương ứng

Như trong Hình 2.7, vector sẽ được biến đổi lần đầu tiên thông qua lớp tuyến tính và sau đó được xử lý thêm một lần nữa qua lớp Softmax. Cuối cùng chỉ số của ô có giá trị xác suất lớn nhất sẽ được xác định và ánh xạ ra nội dung của từ tương ứng.

## 2.3. GPT-2

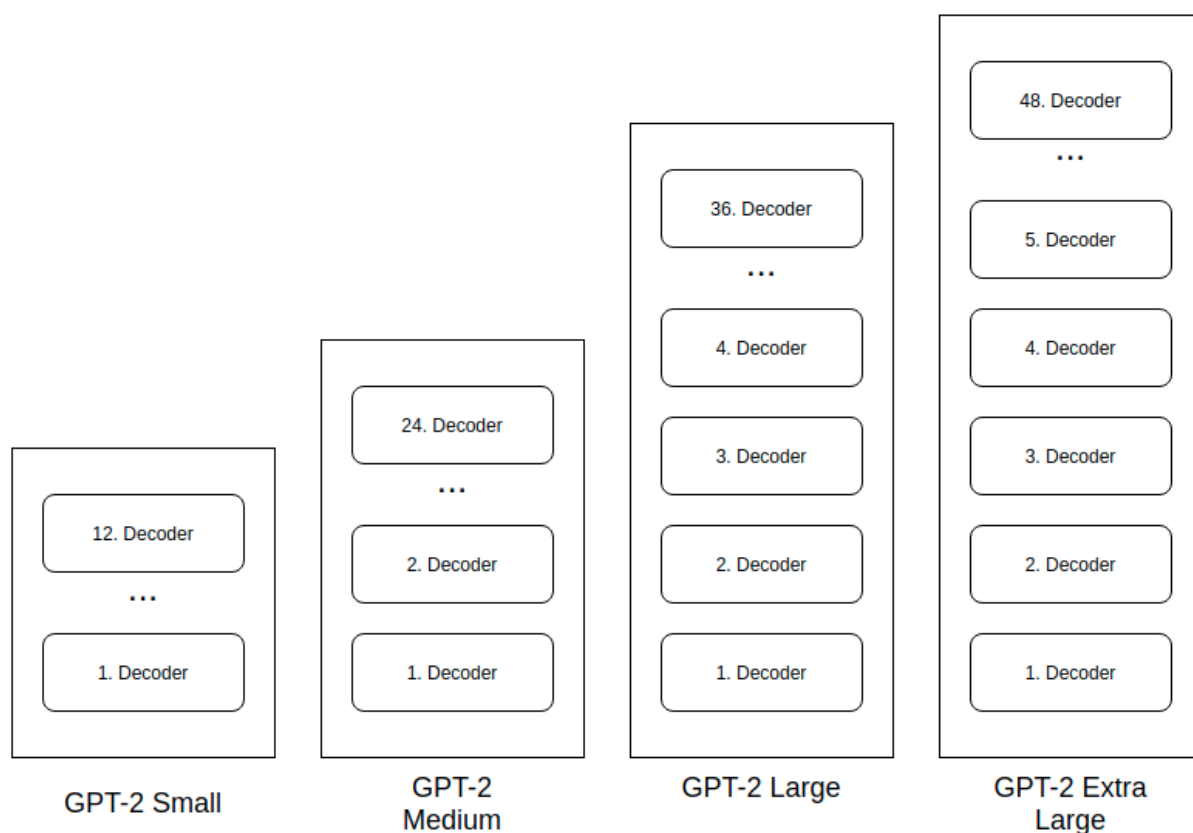
### 2.3.1. Kiến trúc

GPT-2 [15] hay Generative Pre-Trained Transformer 2 là một phiên bản nâng cấp của GPT [16]. So với tiền nhiệm của mình, GPT-2 đã có sự cải thiện đáng kể khi mô hình được nâng số lượng tham số lên gấp 10 lần cũng như đạt được nhiều kết quả vượt trội hơn trong nhiều tác vụ. Hiện tại, GPT-2 có 4 phiên bản được phân loại theo kích thước như sau:

- **GPT-2 Small:** Mô hình có 117 triệu tham số
- **GPT-2 Medium:** Mô hình có 345 triệu tham số

- **GPT-2 Large:** Mô hình có 762 triệu tham số
- **GPT-2 Extra Large:** Mô hình có 1.542 tỉ tham số

Kiến trúc của GPT-2 được thiết kế dựa trên mô hình Transformer, tuy nhiên đã lược bỏ đi bộ mã hóa và được gia tăng số lớp trong bộ giải mã. Số lượng lớp giải mã trong GPT-2 giao động tùy thuộc vào từng phiên bản, với 12 lớp trong GPT-2 Small cho đến 48 lớp trong GPT-2 Extra Large. Đây là một xu hướng thiết kế đã được áp dụng ở nhiều mô hình nổi tiếng khác như BERT (chỉ có bộ mã hóa), Transformer XL [17] (Chỉ có bộ giải mã), ... Dưới đây là một thống kê ngắn số lượng lớp trong từng mô hình của GPT-2.

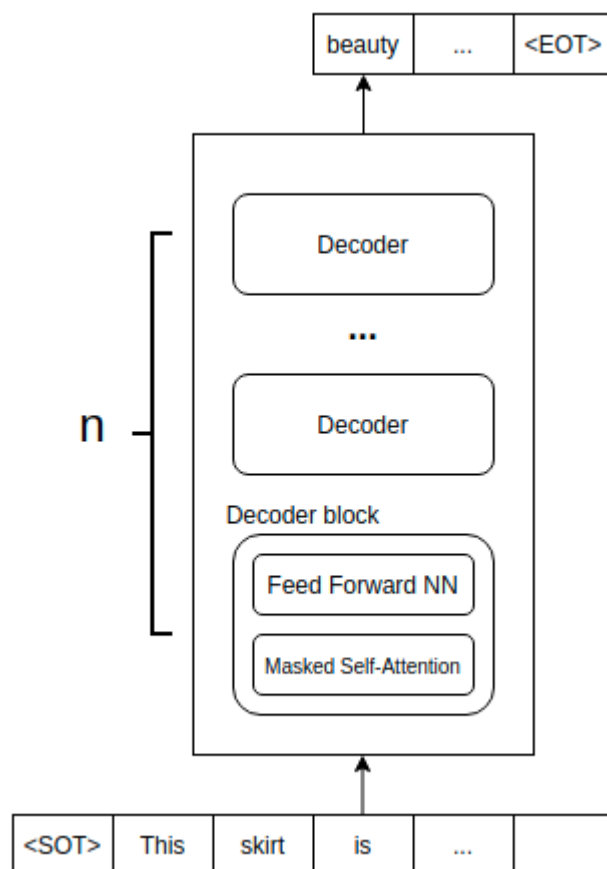


Hình 2.8. Thống kê số lớp của từng phiên bản GPT-2

Như trong Hình 2.8, số lớp giải mã của GPT-2 Small là 12, bằng với mô hình GPT trước đó. Các phiên bản còn lại có số lớp tăng dần, với GPT-2 Medium là 24, GPT-2 Large là 36 và GPT-2 Extra Large là 48.

Từng lớp giải mã của GPT-2 được thiết kế rất giống với các lớp giải mã trong Transformer, ngoại trừ việc bỏ qua lớp tự chú ý thứ hai. Ngoài ra, lớp chuẩn hóa

cũng được chuyển đến đầu vào của mỗi lớp, tương tự với đề xuất của He và các cộng sự [18] và một lớp chuẩn hóa bổ sung cũng được thêm vào sau khối tự chú ý cuối cùng.



Hình 2.9. Mô tả kiến trúc của GPT-2

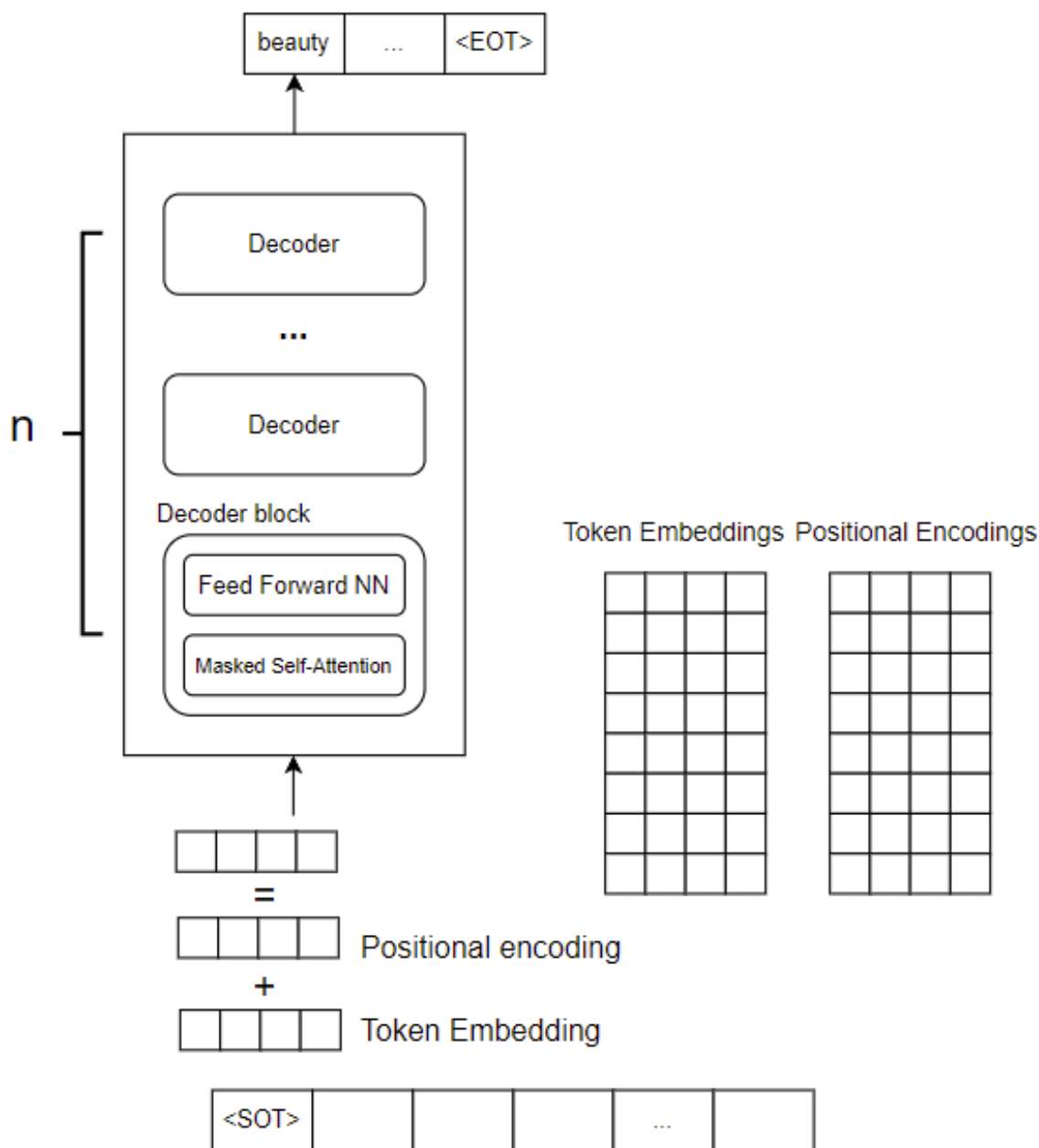
Mô hình GPT-2 được huấn luyện với tác vụ dự đoán từ tiếp theo khi đã biết tập hợp những từ trước đó. Như trong hình trên, định dạng của dữ liệu đầu vào yêu cầu một token đặc biệt đánh dấu vị trí bắt đầu của một chuỗi - *< SOT >*. Sau mỗi lần mô hình sinh ra một token mới, token này sẽ được thêm vào chuỗi đầu vào và chuỗi mới này sẽ trở thành đầu vào của mô hình trong bước lặp tiếp theo. Quá trình này sẽ được lặp lại liên tục cho đến khi token *< EOT >* được sinh ra, đánh dấu sự kết thúc của chuỗi văn bản.

### 2.3.2. Mã hóa đầu vào

Mỗi từ trước khi được mô hình xử lý đều được mã hóa dưới hai dạng chính là vector nhúng và vector mã hóa vị trí (mã hóa vị trí của từ), hai vector này sẽ được gộp



vào thành đầu vào của mô hình. Dưới đây là một minh họa ngắn gọn của quá trình này.



Hình 2.10. Cơ chế mã hóa đầu vào của GPT-2

Khối giải mã đầu tiên sẽ xử lý các token bằng cách truyền nó vào quá trình tự chú ý, sau đó truyền tiếp vào lớp mạng nơ-ron của nó. Sau khi khối đầu tiên xử lý xong token, nó sẽ chuyển vector kết quả đến khối xử lý tiếp theo. Quá trình xử lý của

các khối là độc lập và mỗi khối sẽ duy trì các trọng số ở lớp tự chú ý và mạng nơ-ron của chúng một cách độc lập.

### 2.3.3. Ứng dụng

Như trong bài báo gốc, GPT-2 được huấn luyện trên một nguồn dữ liệu khổng lồ (WebText), với 40GB dữ liệu được thu thập bằng cách cào (crawl) từ các trang web được liên kết với các bài đăng trên Reddit<sup>10</sup> có ít nhất ba phiếu tán thành trước tháng 12 năm 2017. Dữ liệu này được các tác giả đánh giá là tốt hơn so với Common Crawl<sup>11</sup>, một tập dữ liệu khác được sử dụng khá thường xuyên trong quá trình huấn luyện mô hình xử lý ngôn ngữ.

Do thừa hưởng những ưu điểm của Transformer và được huấn luyện trên một tập dữ liệu khổng lồ và phong phú, GPT-2 đã cho thấy khả năng thực hiện tốt nhiều tác vụ khác nhau ngoài sinh văn bản như: Trả lời câu hỏi, Tóm tắt văn bản hay thậm chí là dịch máy, ... Trong bài báo, các tác giả cũng đã chứng minh hiệu suất đáng kinh ngạc của mô hình này trên một số tập dữ liệu phổ biến như WMT-14-Fr-En [19] cho dịch máy, CoQA [20] cho đọc hiểu văn bản, CNN and Daily Mail [21] cho tóm tắt văn bản, ... Đây là một kết quả ngoài mong đợi khi GPT-2 chỉ được thiết kế cho bài toán dự đoán từ tiếp theo trong chuỗi.

## 2.4. BART

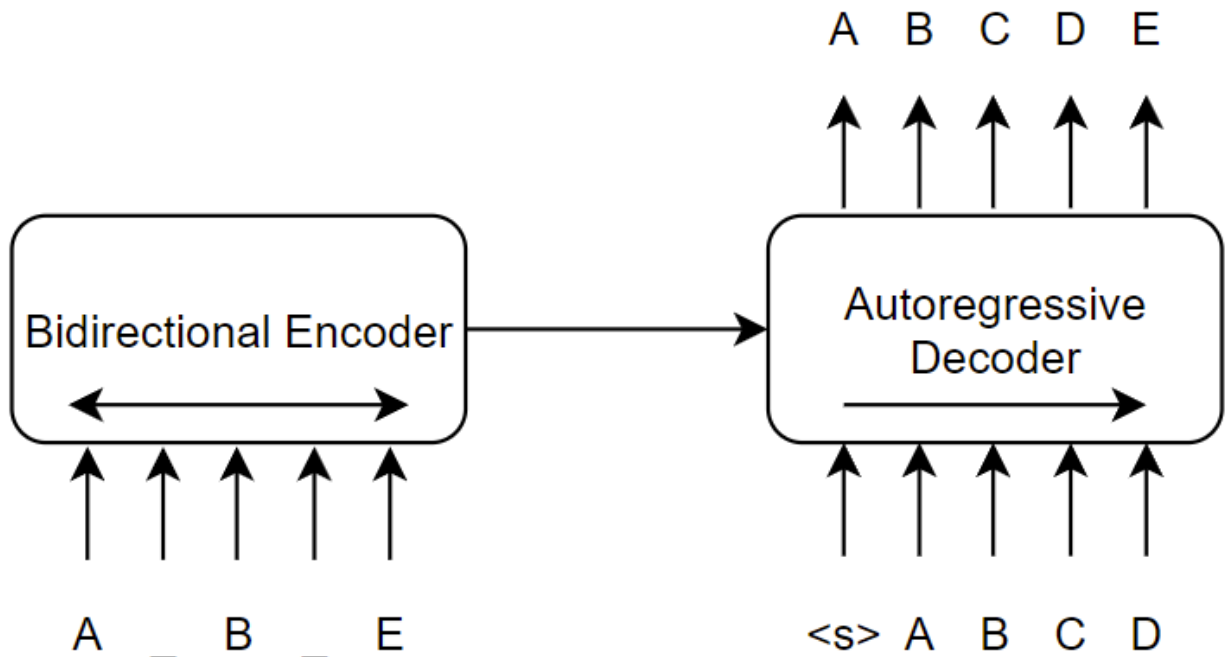
### 2.4.1. Kiến trúc

BART [22] là một bộ mã hóa tự động khử nhiễu có nhiệm vụ ánh xạ một văn bản bị biến đổi có chủ ý sang trạng thái gốc của nó. BART được xây dựng dựa trên kiến trúc seq2seq, với một bộ mã hóa hai chiều (Bidirectional Decoder) để xử lý văn bản bị biến đổi và một bộ giải mã tự động có chiều từ trái sang phải (Left-to-right Decoder) để dự đoán nội dung gốc của văn bản đó. Dưới đây là kiến trúc sơ lược của mô hình BART:

---

<sup>10</sup> <https://www.reddit.com/>

<sup>11</sup> <https://commoncrawl.org/connect/blog/>



Hình 2.11. Sơ lược kiến trúc mô hình BART

Trong BART, đầu vào sẽ được bộ mã hóa xử lý mà không cần phải được căn chỉnh với đầu ra của bộ giải mã, điều này cho phép những biến đổi nhiều tùy ý, bao gồm cả việc thay đổi độ dài của dữ liệu. Ưu điểm chính của phương pháp này là nâng cao khả năng xử lý nhiễu của mô hình. Ở Hình 2.11, một văn bản đã được tiền xử lý bằng cách che dấu đi một số phần bằng các ký hiệu đặc biệt (trái) sẽ được mã hóa với một mô hình 2 chiều (Bidirectional Encoder), và sau đó văn bản khôi phục tương ứng sẽ được tính toán bằng một bộ giải mã hồi quy tự động (phải).

Kiến trúc của BART cũng được xây dựng dựa trên Transformer giống như các mô hình nổi tiếng khác như GPT, GPT2, ... ngoại trừ việc sử dụng các hàm kích hoạt GeLUs thay vì ReLU và khởi tạo các tham số theo cơ chế khác. BART cũng có hai phiên bản chính, phiên bản cơ sở với 6 lớp trong mỗi khối mã hóa và giải mã và phiên bản lớn hơn với 12 lớp mỗi khối. Nhìn chung, kiến trúc này cũng khá tương đồng khi so sánh với BERT, ngoại trừ một số điểm khác biệt bao gồm: Mỗi lớp trong khối giải mã sẽ thực hiện thêm cơ chế chú ý chéo (cross-attention) trên lớp ẩn cuối cùng của khối mã hóa và không sử dụng một mạng truyền thẳng trước khi dự đoán các từ.

## 2.4.2. Tiền huấn luyện

Được huấn luyện với mục tiêu khôi phục lại văn bản, BART cho phép xử lý bất kỳ dạng mất mát thông tin nào của văn bản. Trong trường hợp đặc biệt khi toàn bộ thông tin bị lược bỏ, có thể coi BART tương đương với một mô hình ngôn ngữ. Một số phương thức biến đổi văn bản mà BART áp dụng trong quá trình huấn luyện bao gồm:

- **Mặt nạ:** Thay đổi một số token trong văn bản bằng token [MASK] một cách ngẫu nhiên.
- **Xóa Token:** Xóa ngẫu nhiên một số token trong văn bản gốc.
- **Điền văn bản:** Một số đoạn văn bản được thay thế bằng token [MASK] trong văn bản gốc, độ dài văn bản được lấy mẫu tuân theo phân phối Poisson với  $\lambda=3$ . Các đoạn văn bản có độ dài bằng 0 sẽ tương đương với việc thêm token [MASK].
- **Hoán vị câu:** Văn bản sẽ được tách thành tập hợp các câu và những câu này sẽ được hoán vị theo một thứ tự mới.
- **Xoay văn bản:** Để mô hình học được đâu là điểm bắt đầu của văn bản, một token sẽ được chọn ngẫu nhiên và văn bản gốc sẽ chọn đây là điểm bắt đầu của văn bản mới.

Khi được so sánh với các mô hình nổi tiếng khác như BERT, RoBERTa, ... BART cho thấy những kết quả rất tốt trên nhiều tập dữ liệu / bài toán khác nhau như tóm tắt văn bản, sinh hội thoại, ... Dưới đây là bảng so sánh tóm tắt hiệu suất của BART với những mô hình phổ biến.

	SQuAD 1.1 EM/F1	SQuAD 2.0 EM/F1	MNLI m/mm	SST Acc	QQP Acc	QNLI Acc	STS-B Acc	RTE Acc	MRPC Acc	CoLA Mcc
BERT	84.1/90.9	79.0/81.8	86.6/-	93.2	91.3	92.3	90.0	70.4	88.0	60.6
UniLM	-/-	80.5/83.4	87.0/85.9	94.5	-	92.7	-	70.9	-	61.1
XLNet	<b>89.0</b> /94.5	86.1/88.8	89.8/-	95.6	91.8	93.9	91.8	83.8	89.2	63.6
RoBERTa	88.9/ <b>94.6</b>	<b>86.5/89.4</b>	<b>90.2/90.2</b>	96.4	92.2	94.7	<b>92.4</b>	86.6	<b>90.9</b>	<b>68.0</b>
BART	88.8/ <b>94.6</b>	86.1/89.2	89.9/90.1	<b>96.6</b>	<b>92.5</b>	<b>94.9</b>	91.2	<b>87.0</b>	90.4	62.8

Hình 2.12. Bảng so sánh hiệu suất của BART với các mô hình ngôn ngữ nổi tiếng [22]

Có thể thấy trong nhiều tác vụ phân loại, BART cho ra hiệu quả tương đương với RoBERTa và XLNet, cho thấy rằng các lớp trong bộ giải mã một hướng của

BART không làm giảm hiệu suất đối với các tác vụ loại này. Ngoài ra, BART còn cho thấy khả năng xử lý vượt trội khi được so sánh với BERT, UniLM trong suốt quá trình thử nghiệm.

### 2.4.3. Ứng dụng

Là một mô hình mạnh mẽ, BART có thể được áp dụng trong rất nhiều nhiệm vụ, bài toán khác nhau, ví dụ như:

- **Phân loại chuỗi:** Đối với nhiệm vụ này, một đầu vào sẽ cùng được đưa vào bộ mã hóa và bộ giải mã, và trạng thái ẩn cuối cùng của token cuối cùng của bộ giải mã sẽ được đưa vào bộ phân loại tuyến tính đa lớp mới. Bộ phân loại tuyến tính này sẽ đóng vai trò phân loại kết quả.
- **Phân loại token:** Trong nhiệm vụ này, các tài liệu văn bản cũng được đưa vào bộ mã hóa, giải mã một cách hoàn toàn. Sau đó, trạng thái ẩn trên cùng của bộ giải mã sẽ được sử dụng làm vector đại diện cho mỗi từ. Vector này sẽ tiếp tục được sử dụng để phân loại token.
- **Sinh chuỗi:** Bởi vì BART có bộ giải mã tự động, nó có thể được tinh chỉnh trực tiếp cho các nhiệm vụ tạo chuỗi như trả lời và tóm tắt câu hỏi trừu tượng. Trong cả hai nhiệm vụ này, thông tin được sao chép từ đầu vào nhưng bị thao túng, liên quan chặt chẽ đến mục tiêu huấn luyện khử nhiễu. Ở đây, đầu vào của bộ mã hóa là chuỗi đầu vào và bộ giải mã tạo ra đầu ra tự động.
- **Dịch máy:** BART cũng có thể được sử dụng để cải thiện bộ giải mã dịch máy trong tác vụ dịch văn bản sang tiếng anh. Toàn bộ mô hình của BART (bao gồm bộ mã hóa và bộ giải mã) có thể được sử dụng để làm một bộ giải mã được huấn luyện trước cho bài toán dịch máy, bằng việc thêm một tập tham số mã hóa mới được học từ bitext.

## 2.6. RoBERTa

Năm 2019, Liu và các cộng sự [23] đã công bố RoBERTa - một mô hình ngôn ngữ được thiết kế dựa trên mô hình BERT [24] của Google và được tối ưu theo bài toán che giấu ngôn ngữ chủ động (Masked-Language Modeling). Về tổng quan, RoBERTa giữ nguyên kiến trúc và có một số tinh chỉnh trên các siêu tham số quan trọng của BERT, loại bỏ nhiệm vụ dự đoán câu tiếp theo đồng thời huấn luyện với cơ chế sinh mặt nạ mới cùng các thông số mini-batch và learning-rate lớn hơn nhiều

so với mô hình cũ. Để thấy được rõ lợi ích của những thay đổi này, luận văn sẽ trình bày và phân tích chúng chi tiết hơn ở những phần tiếp theo.

Đầu tiên, trong quá trình tiền huấn luyện, các tác giả đã thay thế phương pháp sinh mặt nạ tĩnh (Static Masking) bằng sinh mặt nạ động (Dynamic Masking). Theo đó, thay vì thực hiện việc thay thế token trong văn bản bằng token MASK chỉ một lần trong quá trình tiền xử lý dữ liệu và phải nhân bản dữ liệu huấn luyện lên gấp 10 lần để đa dạng hóa đầu vào, kỹ thuật mặt nạ động cho phép tạo ra các mẫu mặt nạ cho dữ liệu tại bất kỳ thời điểm nào mô hình cần sử dụng chúng trong quá trình huấn luyện. Phương pháp này đã được các tác giả chứng minh là không chỉ có thể giữ nguyên hiệu suất của mô hình mà còn giúp tăng tính hiệu quả về mặt thời gian chạy trong trường hợp cần huấn luyện mô hình trên những tập dữ liệu có kích thước lớn.

Thứ hai, bài báo chỉ ra rằng khi loại bỏ tác vụ dự đoán câu tiếp theo (Next Sentence Prediction) và thay đổi định dạng đầu vào huấn luyện thành Full-Sentences, mô hình BERT sẽ có những cải thiện hiệu suất đáng kể. Ở đây, cần làm rõ các khái niệm Full-sentences và Next Sentence Prediction. Full-sentences là một kiểu định dạng của đầu vào huấn luyện trong đó mỗi nội dung được đóng gói với với các câu đầy đủ được lấy mẫu liên nhau từ một hoặc nhiều tài liệu, sao cho tổng độ dài tối đa là 512 token. Khi đến phần cuối của mỗi tài liệu, mẫu các câu từ tài liệu tiếp theo sẽ được trích xuất và thêm một mã phân tách bổ sung giữa các tài liệu. Next Sentence Prediction là một tác vụ có nội dung dự đoán xem hai câu bất kỳ có phải là hai câu liên tiếp trong cùng một văn bản hay không. Mục tiêu thiết kế của tác vụ này là cải thiện hiệu suất của mô hình trên các nhiệm vụ cụ thể ví dụ như Suy luận ngôn ngữ (Natural Language Inference) và được cho rằng sẽ ảnh hưởng đến tính hiệu quả của mô hình nếu bị bỏ qua [24]. Tuy nhiên những thí nghiệm trong bài báo đã chứng minh kết quả ngược lại, khi loại bỏ NSP thì mô hình không những không bị ảnh hưởng mà còn cho kết quả tốt hơn. Dưới đây là bảng mô tả tổng kết thí nghiệm của Liu và các cộng sự:

Model	SQuAD 1.1/2.0	MNLI-m	SST-2	RACE
<i>Our reimplementation (with NSP loss):</i>				
SEGMENT-PAIR	90.4/78.7	84.0	92.9	64.2
SENTENCE-PAIR	88.7/76.2	82.9	92.1	63.0
<i>Our reimplementation (without NSP loss):</i>				
FULL-SENTENCES	90.4/79.1	84.7	92.5	64.8
DOC-SENTENCES	90.6/79.7	84.7	92.7	65.6
BERT <sub>BASE</sub>	88.5/76.3	84.3	92.8	64.3
XLNet <sub>BASE</sub> (K = 7)	-/81.3	85.8	92.7	66.1
XLNet <sub>BASE</sub> (K = 6)	-/81.0	85.6	93.4	66.7

Hình 2.13. Các thí nghiệm chứng minh tính hiệu quả khi loại bỏ nhiệm vụ NSP và sử dụng định dạng đầu vào Full-sentences [23]

Trên Hình 2.13 có thể thấy rằng, khi thực hiện việc đánh giá trên các tập dữ liệu phổ biến như SQuAD 1.1 / 2.0, MNLI-m, SST-2, RACE với các cách triển khai mô hình khác nhau thì việc không sử dụng NSP cùng sự thay đổi định dạng đầu vào thành Full-sentences và Doc-sentences cho ra kết quả tốt nhất. Lý do các tác giả chọn Full-sentences thay vì Doc-sentences là bởi việc sử dụng Doc-sentences có thể dẫn đến kích thước lô thay đổi.

Một số thay đổi còn lại của RoBERTa liên quan đến việc chỉnh sửa hai tham số mini-batches, learning-rate và thay đổi bộ từ vựng BPE. Cụ thể, các tác giả đã thực hiện một vài thử nghiệm và quan sát thấy rằng khi sử dụng giá trị mini-batch đủ lớn cùng learning-rate phù hợp, quá trình huấn luyện có thể tăng tốc đồng thời cũng cải thiện được hiệu suất của một số nhiệm vụ cụ thể của mô hình. Cụ thể trong bài báo, kích cỡ lô được khuyến nghị là 8000 và learning-rate là  $10^{-3}$ , so với những giá trị mặc định tương ứng của BERT là 256 và  $10^{-4}$  thì có sự chênh lệch khá lớn. Ngoài ra, các tác giả cũng đề xuất thay bộ từ vựng BPE cấp độ ký tự với kích cỡ 30K trong triển khai gốc của BERT với bộ từ vựng BPE cấp độ bytes kích cỡ 50K được giới thiệu bởi Radford và các cộng sự [15]. Lý do là bởi bộ từ vựng BPE cấp độ bytes có thể mã hóa bất kỳ dữ liệu đầu vào nào mà không cần sử dụng các token “unknown”, điều này là không thể khi sử dụng bộ từ vựng BPE cũ với các ký tự unicode. Tuy việc thay đổi này có thể gây ra kết quả kém đi một chút trong một vài tác vụ nhưng

các tác giả cũng nhấn mạnh rằng ưu điểm khi sử dụng cơ chế mã hóa này là không thể chối cãi.

Bằng các cải tiến nêu trên, RoBERTa đã cho thấy sự hiệu quả trong nhiệm vụ dự đoán từ ngữ được che giấu tốt hơn so với mô hình gốc và chứng minh được kết quả vượt trội hơn trên các nhiệm vụ cụ thể (downstream task).

## 2.7. Độ đa dạng từ vựng

Độ đa dạng từ vựng (Lexical Diversity) [25] là một phép đo cho phép đánh giá số lượng từ vựng khác nhau có trong một văn bản. Những từ vựng này có thể là danh từ, tính từ, động từ và trạng từ có chức năng truyền tải ý nghĩa. Đây được xem như là một chỉ số quan trọng phản ánh mức độ phức tạp của văn bản và có thể được sử dụng ở nhiều lĩnh vực khác nhau như ngôn ngữ học, thống kê, tâm lý học, ...

Lexical Diversity có thể được tính bằng nhiều cách thức khác nhau, cách phổ biến nhất là dựa trên giá trị tỷ số giữa số lượng từ xuất hiện duy nhất (type) đối với tổng số từ (token) trong một văn bản (type-token ratio, TTR; Chotlos, 1944 [26]; Templin, 1957 [27]). Phương pháp này có thể áp dụng tốt trong một số trường hợp nhất định như so sánh các văn bản có cùng kích thước hoặc rất giống nhau nhưng không thích hợp trong các trường hợp tổng quát hơn. Lý do là bởi khi độ dài tăng lên, sẽ xảy ra khả năng những văn bản ngắn có thể đạt được chỉ số TTR cao hơn so với những văn bản dài, đây là một kết quả không hợp lý. Để khắc phục nhược điểm này, một số phương pháp khác đã được đề xuất, tiêu biểu như MTLT [28] và VoC-D [29][30]. Những phương pháp này cho phép giảm thiểu sự ảnh hưởng của yếu tố chiều dài văn bản trong quá trình tính toán. VoC-D bao gồm việc lấy ngẫu nhiên một số lượng từ nhất định (thường giao động từ 35 đến 50) từ dữ liệu, sau đó tính toán chỉ số TTR trung bình cho mỗi độ dài này và tìm đường cong phù hợp nhất với đường cong được tạo bởi TTR (giữa một họ đường cong được tạo bởi các biểu thức chỉ khác nhau giá trị của một tham số duy nhất). Giá trị tham số tương ứng với đường cong phù hợp nhất được kết luận như là kết quả của phép đo độ đa dạng. Toàn bộ quy trình này có thể được lặp lại nhiều lần và được tính trung bình. Còn MTLT hay với tên đầy đủ là Measure of Textual Lexical Diversity là một phương pháp sử dụng sự phân tích tính tuần tự của một mẫu để ước tính điểm LD. Về cơ bản, MTLT sẽ tính toán độ dài trung bình của các đoạn có chỉ số TTR nhất định. Phép tính được thực hiện hai lần, một lần từ trái sang phải và ngược lại. Các kết quả sau đó sẽ được tổng hợp và tính trung bình. Hiện nay, TTR, VoC và MTLT đã và đang là những chỉ số được



sử dụng phổ biến trong những bài toán yêu xác định tính đa dạng của văn bản. Bằng cách sử dụng chúng cùng nhau, chúng ta có thể có được góc nhìn rõ ràng hơn về toàn bộ văn bản và tránh đưa ra những kết luận sai lầm.

## Chương 3. Giải pháp

GPT-2 đã cho thấy khả năng ứng dụng tốt trên nhiều bài toán xử lý ngôn ngữ khác nhau như tóm tắt văn bản, dịch văn bản, ... Tuy rằng chất lượng những kết quả của mô hình không phải là tối ưu nhất khi so sánh với các phương pháp khác hiện nay nhưng chúng cũng chứng minh được rằng nếu cung cấp đủ dữ liệu chưa được gán nhãn cho mô hình, các tác vụ có thể được hưởng lợi từ các phương pháp học không giám sát.

Trong chương này, luận văn sẽ mô tả cách tiếp cận sử dụng mô hình GPT-2 để tạo ra những mô tả sản phẩm có chất lượng cao dựa trên thông tin đầu vào của chúng. Ngoài ra, một số kỹ thuật khác như tăng cường dữ liệu hay task-adaptive cũng sẽ được áp dụng để cải thiện chất lượng đầu ra.

### 3.1. Sinh mô tả sản phẩm với GPT-2

Như đã được đề cập trong phần 3.1, bài toán sinh mô tả sản phẩm có thể được mô hình hóa thành nhiệm vụ sinh ra văn bản đầu ra dựa trên văn bản đầu vào có chứa một số thông tin ban đầu về sản phẩm như tiêu đề, nhãn hiệu, danh mục, thuộc tính. Do đó, trong nội dung ở phần này, luận văn sẽ tập trung vào việc xây dựng giải pháp đáp ứng yêu cầu trên bằng cách ứng dụng GPT-2.

Để tinh chỉnh mô hình GPT-2 (fine-tuning) đáp ứng nhiệm vụ tạo ra những mô tả có ý nghĩa dựa trên tiêu đề, luận văn sẽ điều kiện hóa dữ liệu đầu vào với định dạng như sau:

$$< SOT > title = description < EOT > \quad (5)$$

Trong đó, <SOT> và <EOT> là cặp token đặc biệt đã được đề cập ở phần 3.3, đại diện cho sự bắt đầu vào kết thúc của một văn bản. Sau khi nhận được đầu vào  $< SOT > title =$ , nhiệm vụ của mô hình GPT-2 sẽ là lấy mẫu với các tham số Top-K và Top-p (quá trình sampling) để tạo ra văn bản trôi chảy hơn và tránh thoái hóa văn bản (text degeneration [31]). Song song với việc sử dụng tiêu đề, phương pháp cũng sử dụng thêm các thông tin như cấu trúc danh mục, thương hiệu và các thuộc tính của sản phẩm như đã đề cập trong phần 3.1. Vì vậy, định dạng của dữ liệu huấn luyện sẽ được cấu trúc lại như sau:

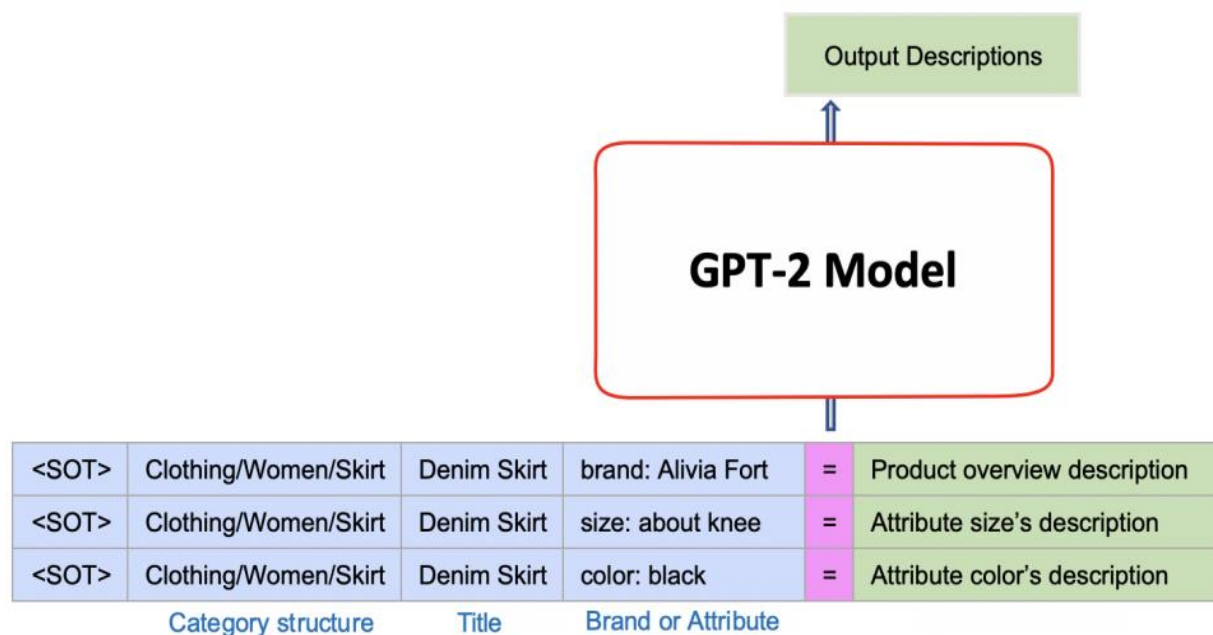
$$< SOT > [category][title][brand][attributes] = description < EOT > \quad (6)$$

Tuy nhiên, một nhược điểm nảy sinh với cách tiếp cận này là mô tả được sinh ra sẽ dài và khó bao quát được thông tin đầu vào. Bởi vì tạo ra văn bản dài là một vấn đề vẫn tồn tại nhiều khó khăn, văn bản đầu ra sẽ có xu hướng không mạch lạc hoặc nội dung bị lặp lại nhiều lần, phương pháp sẽ tiếp tục chia tác vụ tạo mô tả sản phẩm thành hai nhiệm vụ nhỏ hơn, bao gồm sinh mô tả tổng quan cho sản phẩm và sinh mô tả cho các thuộc tính riêng biệt của chúng. Sau đó, các mô tả này được kết hợp lại để có được kết quả cuối cùng. Với mô tả tổng quan, dữ liệu huấn luyện sẽ có cấu trúc:

$$< SOT > [category][title][brand] = product overview description < EOT > \quad (7)$$

Với mô tả thuộc tính, dữ liệu huấn luyện có cấu trúc:

$$< SOT > [category][title][attributes] = attribute description < EOT > \quad (8)$$



Hình 3.1. Định dạng dữ liệu huấn luyện trong phương pháp đề xuất

Hình 3.1 trên đã đưa ra một vài ví dụ trong dữ liệu huấn luyện của phương pháp mới. Bằng cách huấn luyện theo hướng này, mô hình có thể nhận được nhiều lợi ích. Đầu tiên, tốc độ suy diễn (inference) sẽ tăng lên đáng kể vì mô hình có thể tạo ra các mô tả tổng quan và thuộc tính sản phẩm một cách song song. Có thể tổ chức nhiều mô hình chạy đồng thời để gia tăng hiệu suất tổng thể của ứng dụng. Thứ hai, kết quả tổng hợp cuối cùng sẽ bao gồm những thông tin quan trọng của dữ liệu đầu vào như thương hiệu và các thuộc tính của sản phẩm. Thứ ba, giải pháp có thể

tạo ra các mô tả có số lượng thuộc tính có thể thay đổi tùy theo nhu cầu mà không làm giảm chất lượng của kết quả tổng thể. Cuối cùng, mô tả sẽ chính xác và trôi chảy hơn vì không gặp phải vấn đề phụ thuộc xa (Long range dependence) [32].

Mặc dù tác vụ này được hưởng lợi từ tri thức được chuyển giao từ mô hình GPT-2 đã được đào tạo, chất lượng của nó vẫn được không như mong đợi vì vấn đề thiếu thôn dữ liệu và sự khác biệt giữa miền dữ liệu đào tạo của GPT2 và miền dữ liệu của bài toán sinh mô tả. Trong những phần tiếp theo của chương, luận văn sẽ trình bày một số phương pháp để giải quyết 2 vấn đề này.

### **3.2. Tăng cường dữ liệu**

Để gia tăng dữ liệu phục vụ cho quá trình huấn luyện, luận văn đã áp dụng một số kỹ thuật như là sử dụng mô hình dịch, thêm từ và thay thế một số từ loại cụ thể trong văn bản bằng từ đồng nghĩa của chúng. Chi tiết của những kỹ thuật này sẽ được trình bày như dưới đây.

#### **3.2.1. Viết lại văn bản bằng mô hình dịch**

##### **3.2.1.1. Ngôn ngữ trung gian**

Ngôn ngữ chót, trung gian hay “pivot language” là một thuật ngữ trong dịch máy, chỉ những ngôn ngữ được sử dụng làm cầu nối trung gian trong quá trình dịch. Ví dụ khi cần dịch từ Tiếng Phạn sang Tiếng Tây Ban Nha, mô hình dịch sẽ dịch nội dung Tiếng Phạn sang một ngôn ngữ trung gian và dịch tiếp từ ngôn ngữ trung gian này sang tiếng Tây Ban Nha. Điều này là cần thiết để tránh sự bùng nổ tổ hợp khi phải xây dựng từng mô hình dịch cho tất cả các cặp ngôn ngữ. Ý tưởng này được Kay [33] đề xuất lần đầu tiên vào năm 1997, khi ông đã quan sát thấy rằng sự khó khăn trong việc dịch văn bản từ một ngôn ngữ nguồn sang ngôn ngữ đích bất kỳ có thể được giải quyết trong điều kiện nếu có bản dịch chất lượng tốt nào đó của một ngôn ngữ thứ ba. Các ngôn ngữ thường được sử dụng làm ngôn ngữ trung gian bao gồm Tiếng Anh, Tiếng Pháp, Tiếng Nga, và Tiếng Ả Rập.

##### **3.2.1.2. Viết lại nội dung bằng mô hình dịch**

Áp dụng ý tưởng về ngôn ngữ chót, Mallinson và các cộng sự [34] đã giới thiệu một phương pháp cho phép viết lại nội dung văn bản bằng cách sử dụng các mô hình dịch. Trong phương pháp này, các tác giả đã sử dụng một tập hợp 3 ngôn ngữ chót

bao gồm Tiếng Pháp, Tiếng Séc, Tiếng Đức cùng 6 mô hình dịch (Tương ứng với các cặp Tiếng Anh -> Tiếng Pháp, Tiếng Anh -> Tiếng Séc, Tiếng Anh -> Tiếng Đức, Tiếng Pháp -> Tiếng Anh, Tiếng Séc -> Tiếng Anh, Tiếng Đức -> Tiếng Anh) để viết lại nội dung các văn bản Tiếng Anh. Trong quá trình xử lý, các văn bản Tiếng Anh nguồn  $E_1$  sẽ được dịch ra nhiều ngôn ngữ trung gian tạo ra các phiên bản dịch  $F_i$  của chúng. Sau đó các bản dịch này sẽ được tính toán một cách độc lập sự phân phối trên từ vựng đích  $P(y_{t'} = w | y_{<t'}, F_i)$  và được kết hợp trong một công thức để tìm ra những từ vựng hợp lý nhất trong kết quả cuối cùng. Ví dụ với hai bản dịch  $F_1$  và  $F_2$  và từ  $y_{t'}$  ở vị trí  $t'$  trong câu đích có thể xây dựng ra hai phân phối  $P(y_{t'} = w | y_{<t'}, F_1)$  và  $P(y_{t'} = w | y_{<t'}, F_2)$  cùng công thức kết hợp như sau:

$$P(y_{t'} = w | y_{<t'}, F_1, F_2) = \lambda_1 P(y_{t'} = w | y_{<t'}, F_1) + \lambda_2 P(y_{t'} = w | y_{<t'}, F_2) \quad (9)$$

Các trọng số  $\lambda$  đối với mỗi phân phối là các giá trị được tính bởi công thức  $P(F_i | E_1)$ . Do đó, công thức ở trên cũng có thể viết lại như sau:

$$P(y_{t'} = w | y_{<t'}, F_1, F_2) = P(F_1 | E_1) P(y_{t'} = w | y_{<t'}, F_1) + P(F_2 | E_1) P(y_{t'} = w | y_{<t'}, F_2) \quad (10)$$

Áp dụng một phần ý tưởng trên, luận văn sẽ sử dụng hướng tiếp cận đơn giản hơn, cũng sử dụng cặp ngôn ngữ chốt là Tiếng Pháp và Tiếng Đức nhưng thay vì tìm ra các trọng số  $\lambda$ , phương pháp sử dụng mô hình BART cho tác vụ dịch và RoBERTa cho tác vụ đánh giá khả năng giữ nguyên ý nghĩa của câu. Cách thức đánh giá là lấy câu nào có ngữ nghĩa giống với câu gốc nhất.

### 3.2.2. Thay thế từ đồng nghĩa

Một phương pháp hiệu quả khác có thể áp dụng trong tác vụ viết lại văn bản là thay thế một số từ loại bằng từ đồng nghĩa của chúng. Để triển khai phương pháp này, luận văn sử dụng cơ sở dữ liệu từ vựng PPDB 2.0<sup>12</sup> [35] với hơn 100 triệu cụm từ cùng thư viện nlpaug<sup>13</sup>. Các tham số của thư viện nlpaug được cấu hình như sau:

- **aug\_min**: Số từ thay thế tối thiểu trong văn bản được thiết lập là 1.
- **aug\_max**: Số từ thay thế tối đa trong văn bản được thiết lập là 20.
- **aug\_p**: Xác suất thay thế từ đồng nghĩa là 0.3
- **tokenizer**: thuật toán tách từ được thiết lập là thuật toán của thư viện nltk<sup>14</sup>

<sup>12</sup> <http://paraphrase.org/#/download>

<sup>13</sup> <https://github.com/makcedward/nlpaug>

<sup>14</sup> <https://www.nltk.org/>

- **stopwords**: Danh sách từ dừng được thiết lập là danh sách mặc định trong corpus của nltk.

Các bước của quá trình thay thế từ loại sử dụng cơ sở dữ liệu PPDB 2.0 bao gồm:

1. Tách từ bằng thuật toán của thư viện nltk, ra kết quả danh sách các token của văn bản đầu vào
2. Đánh dấu từ loại của các token trong danh sách, trong bước này cũng sử dụng mô hình gán nhãn từ loại của thư viện nltk
3. Loại các token là từ dừng, dấu câu hoặc các từ không có từ đồng nghĩa trong mạng từ PPDB 2.0 (có nhãn là “DT”)
4. Lấy ngẫu nhiên danh sách các từ cần tìm từ đồng nghĩa
5. Tìm các từ đồng nghĩa bằng PPDB 2.0, đối với mỗi từ có thể có nhiều từ đồng nghĩa tương ứng, do đó chỉ lấy ngẫu nhiên một từ duy nhất
6. Kết hợp văn bản gốc với các từ đồng nghĩa đã được trích xuất

### 3.2.3. Thêm từ theo ngữ cảnh

Đây là một phương pháp sử dụng mô hình ngôn ngữ để thêm những từ vựng phù hợp với ngữ cảnh của cả văn bản. Trong luận văn, phương pháp được triển khai bằng thư viện nlpaug kết hợp với mô hình RoBERTa cùng những thiết lập như sau:

- **aug\_min**: Số từ thêm tối thiểu trong văn bản được thiết lập là 1.
- **aug\_max**: Số từ thêm tối đa trong văn bản được thiết lập là 20.
- **aug\_p**: Xác suất thêm từ là 0.3
- **top\_k**: 50
- **top\_p**: 0.9
- Các tham số còn lại thiết lập theo mặc định

Quá trình thêm từ theo ngữ cảnh cơ bản bao gồm các bước như sau:

1. Tách từ sử dụng công cụ tách từ của RoBERTa, lấy ra các token từ trái qua phải cho đến hết hoặc đến khi đạt đến ngưỡng tối đa (ngưỡng số token tối đa của RoBERTa)
2. Chọn số lượng từ sẽ thêm dựa trên các thông số truyền vào như aug\_min, aug\_max, aug\_p
3. Bỏ qua các token là từ dừng, dấu câu
4. Lấy ngẫu nhiên các token bằng số lượng từ đã được khởi tạo ở bước 2

5. Thêm các Token Mask (Một token đặc biệt đại diện cho một token chưa biết, chưa rõ nội dung) tại các vị trí của các token được lấy ngẫu nhiên
6. Dùng RoBERTa dự đoán nội dung các Token Mask
7. Cập nhật nội dung của văn bản ban đầu bằng các kết quả có được

### **3.3. Task-Adaptive Pretraining**

Để cải thiện chất lượng mô tả được sinh ra bởi mô hình GPT-2, luận văn sẽ điều chỉnh miền dữ liệu của GPT-2 sang miền dữ liệu mô tả sản phẩm bằng cách sử dụng kỹ thuật task-adaptive pretraining [36]. Như đã biết, dữ liệu nhiệm vụ thường có xu hướng bao phủ chỉ một phần nhỏ trong tập dữ liệu của miền lớn hơn. Vì thế, kỹ thuật TAPT sẽ tiếp tục quá trình tiền huấn luyện với tập dữ liệu không được gán nhãn liên quan đến nhiệm vụ để cải thiện chất lượng của mô hình. Quá trình huấn luyện do đó cũng sẽ bao gồm hai bước, với bước thứ nhất là tiếp tục quá trình tiền huấn luyện GPT-2 trên kho dữ liệu (corpus) huấn luyện chưa được gán nhãn thuộc về miền mô tả sản phẩm và bước thứ hai là huấn luyện mô hình GPT-2 sau khi đã áp dụng kỹ thuật TAPT cho tác vụ sinh mô tả như trong nội dung phần 4.1.

## Chương 4. Thực nghiệm và đánh giá kết quả

Trong chương này, luận văn sẽ trình bày báo cáo kết quả thực nghiệm của những phương pháp mới. Từ đó, đưa ra những phân tích, nhận định ưu, nhược điểm của chúng để lựa chọn ra giải pháp tối ưu nhất. Tất cả thực nghiệm được thực hiện trên một máy chủ có cấu hình 2 GPU 1080Ti, 64GB RAM, CPU Intel Xeon E5-2680 v4.

### 4.1. Dữ liệu

Để phục vụ cho việc thực nghiệm và đánh giá, tôi đã xây dựng một bộ dữ liệu mới, được thu thập từ những trang thương mại điện tử lớn là Walmart.com và Amazon.com. Dữ liệu bao gồm những mô tả tổng quan và mô tả theo từng thuộc tính của sản phẩm, được phân loại thành 3 lớp như sau: “Clothing, Shoes & Accessories”, “Electronics & Office”, and “Home, Furniture & Appliances”. Bởi vì mỗi sản phẩm sẽ có thể có những thuộc tính đa dạng khác nhau, để chuẩn hóa, tôi đã chọn ra những thuộc tính được coi là đặc trưng nhất trong tất cả các loại sản phẩm, bao gồm “brand”, “size”, “material”, “color” và “style”. Dưới đây là bảng thống kê số lượng mô tả tổng quan sản phẩm và số lượng mô tả các thuộc tính.

Bảng 4.1. Số lượng mô tả tổng quan và mô tả thuộc tính sản phẩm

Phân loại	Mô tả tổng quan	Mô tả thuộc tính
Clothing, Shoes & Accessories	21207	245351
Electronics & Office	11261	173234
Home, Furniture & Appliances	13239	203248
Tổng	68707	621833

Trong khi dữ liệu mô tả tổng quan của sản phẩm có thể được thu thập một cách dễ dàng từ những website thương mại điện tử thì những dữ liệu mô tả thuộc tính của chúng cần được trích xuất bằng một số phương pháp đặc biệt như trích xuất thuộc tính sản phẩm tự động [37]. Bảng 4.1 đã liệt kê số lượng mô tả tổng quan và thuộc tính của sản phẩm trong cơ sở dữ liệu của tôi sau khi loại bỏ trùng lặp và các mô tả



ngắn dưới 10 từ. Đối với mỗi mô tả, tôi áp dụng các kỹ thuật viết lại văn bản để có được nhiều phiên bản khác nhau. Tất cả dữ liệu mô tả được chia thành 3 bộ theo tỷ lệ 8:1:1 để đào tạo (training), kiểm chứng (validation) và kiểm thử (testing) tương ứng.

## 4.2. Mô hình

Tất cả mô hình trong thử nghiệm này đều dựa trên 2 kiến trúc là Transformer và GPT-2 và được triển khai bằng các thư viện tương ứng là OpenNMT-py<sup>15</sup> và HuggingFace<sup>16</sup>. Để cấu hình các tham số Transformer, luận văn sẽ sử dụng thuật toán tối ưu Adam với giá trị learning rate là 2.0, batch type là “token”, kích cỡ batch huấn luyện là 4096, chiều dài tối đa đầu ra là 300, beam width là 10, các tham số còn lại được thiết lập mặc định. Đối với các mô hình GPT-2, bởi vì tài nguyên tính toán bị giới hạn, giải pháp sẽ sử dụng phiên bản GPT-2 cỡ nhỏ với 124 triệu tham số, 12 khối Decoder, kích thước của embedding và hidden state là 768 chiều, kích cỡ batch là 4, độ dài đầu ra tối đa là 300, top\_k là 50 và top\_p là 0.9, các tham số còn lại được thiết lập mặc định. Dưới đây là tóm tắt những mô hình được đánh giá trong thử nghiệm:

- **Baseline:** Mô hình cơ sở là Transformer, với văn bản đầu được điều kiện hóa bao gồm danh mục, thương hiệu, tiêu đề, các thuộc tính của sản phẩm và đầu ra là mô tả tương ứng.
- **GPT2-base:** Mô hình này được huấn luyện bằng việc sử dụng mô hình GPT-2 trên tập dữ liệu được thu thập trên Internet.
- **GPT-2 Plus:** Mô hình này tương tự với GPT-2 base nhưng được huấn luyện trên tập dữ liệu đã được tăng cường, bao gồm dữ liệu gốc được thu thập cùng dữ liệu đã được viết lại của chúng.
- **GPT-2 + TAPT:** Mô hình áp dụng phương pháp task-adaptive pretraining để nâng cao chất lượng của mô hình GPT-2 base.
- **GPT-2 Plus + TAPT:** Mô hình này áp dụng cả 2 phương pháp, tăng cường dữ liệu và task-adaptive để cải thiện mô hình GPT-2 base.

---

<sup>15</sup> <https://github.com/OpenNMT/OpenNMT-py>

<sup>16</sup> <https://github.com/huggingface/transformers>

### 4.3. Phương pháp đánh giá

Để đánh giá giải pháp, luận văn sử dụng 2 phương pháp chính là đánh giá tự động và đánh giá thông qua con người, chi tiết 2 phương pháp này sẽ được trình bày như dưới đây.

#### 4.3.1. Đánh giá tự động

Để kiểm tra hướng tiếp cận hoạt động hiệu quả trên bài toán sinh mô tả, đầu tiên, tôi đánh giá các mô hình dựa trên 2 khía cạnh là độ mất mát và độ đa dạng của từ vựng (Lexical Diversity). Trong đó, độ mất mát là một chỉ số phản ánh sự thiếu chính xác trong dự đoán của mô hình trên một ví dụ độc lập. Chỉ số này càng cao đồng nghĩa với khả năng mô hình sẽ cho ra những kết quả dự đoán có sự sai lệch so với mục tiêu càng lớn và ngược lại. Do đó, mục tiêu trong bước đánh giá này là giảm thiểu độ lớn của chỉ số này đối với tất cả ví dụ. Điều này sẽ được thực hiện thông qua việc tinh chỉnh tham số mô hình trong quá trình huấn luyện. Ngoài ra, tôi cũng sử dụng thêm chỉ số đánh giá độ phong phú của từ vựng (Lexical Diversity) để đánh giá khả năng tạo ra những văn bản có nội dung đa dạng của mô hình. Điều này cũng rất hợp lý khi những văn bản của con người tạo ra thường mang những nội dung phong phú.

#### 4.3.2. Đánh giá thông qua con người

Để thực hiện phương pháp, tôi lấy mẫu ngẫu nhiên một tập hợp 200 ví dụ từ kết quả chạy của từng mô hình, sau đó cho 30 người Việt Nam có trình độ tiếng anh tối thiểu B2 đánh giá theo các tiêu chí đã đặt ra, kết quả cuối cùng sẽ được lấy trung bình và phân tích. Mỗi ví dụ trong tập dữ liệu sẽ bao gồm những thông tin về sản phẩm như ảnh, nhãn phân loại, tiêu đề, nhãn hiệu, các thuộc tính và những mô tả sinh ra từ mô hình. Những người tham gia được yêu cầu đánh giá một cách độc lập và trung thực theo các tiêu chí sau:

- **Tính lưu loát:** Mô tả có độ lưu loát ở mức nào.
- **Tính liên hệ:** Mô tả được tạo ra có liên quan đến sản phẩm ở mức nào.
- **Tính thông tin:** Mô tả có cung cấp những thông tin hữu ích về sản phẩm ở mức độ nào.
- **Chất lượng tổng thể:** Mô tả có thể áp dụng trong điều kiện thực tế ở mức độ nào.

Mỗi tiêu chí đều được áp dụng thang đo Likert với 5 mức đánh giá bao gồm: Hoàn toàn không đồng ý, Không đồng ý, Đồng ý, Trung lập, Hoàn toàn đồng ý với số điểm tăng dần từ 1 đến 5. Điểm của tiêu chí càng cao thì có nghĩa là chất lượng của mô tả xét trên tiêu chí đó càng tốt.

Bên cạnh việc sử dụng tập ví dụ trên được coi là một tập kiểm thử có thể quan sát, tôi cũng chuẩn bị thêm một tập ví dụ thử nghiệm khác chứa 200 ví dụ lấy từ các kết quả chạy của từng mô hình đối với những danh mục sản phẩm không có trong dữ liệu đào tạo, cụ thể là từ các nhãn phân loại “Toys and Games” và “Food, Household & Pets”, đây được coi là tập thử nghiệm không quan sát được.

#### 4.4. Kết quả và phân tích

Dưới đây là bảng tổng hợp so sánh kết quả giữa các mô hình đề xuất dựa trên các tiêu chí đánh giá tự động, các mô hình được so sánh bao gồm: Baseline (Transformer), GPT-2 base, GPT-2 Plus, GPT-2 + TAPT, GPT-2 Plus + TAPT.

Bảng 4.2. Tổng hợp kết quả đánh giá trên độ mất mát và độ đa dạng từ vựng

Mô hình	Loss	n-gram Lexical Diversity			
		n = 1	n = 2	n = 3	n = 4
Baseline	2.63	17.22	15.84	14.25	13.01
GPT-2 base	1.37	23.31	19.18	18.42	16.56
GPT-2 Plus	1.02	24.65	23.71	21.09	20.17
GPT-2 + TAPT	0.87	23.78	20.52	19.27	17.26
GPT-2 Plus + TAPT	0.68	24.99	24.02	22.31	21.87

Từ Bảng 4.2 có thể thấy rằng các mô hình GPT-2 có hiệu suất vượt trội so với mô hình cơ sở trong cả 2 tiêu chí đánh giá là độ mất mát và độ đa dạng của từ vựng. Lý do có thể được đưa ra để giải thích cho hiện tượng này là dữ liệu huấn luyện có kích thước không đủ để huấn luyện mô hình Transformer - một mô hình thường yêu cầu hàng triệu ví dụ huấn luyện. Do đó, các mô hình GPT-2 sẽ có lợi thế hơn khi đã được huấn luyện trên nhiều tập dữ liệu khác nhau. Đồng thời, các kết quả thí nghiệm trên các mô hình GPT-2 base và GPT-2 Plus cũng đã chứng minh rằng việc thêm dữ liệu đã được viết lại sẽ cải thiện chất lượng của các mô tả trên cả hai khía cạnh đánh giá. Ngoài ra, việc áp dụng phương pháp TAPT với mô hình GPT-2 sẽ giúp nâng cao chất lượng của nó so với GPT-2 base.

Tiếp theo, Bảng 4.3 dưới đây sẽ tổng hợp kết quả đánh giá của các mô hình trên các tập dữ liệu đã thấy và chưa thấy dựa trên phương pháp đánh giá thông qua con người.

Bảng 4.3. Tổng hợp kết quả đánh giá trên hai tập dữ liệu đã thấy và chưa thấy

<b>Kết quả trên tập dữ liệu đã thấy</b>				
<b>Mô hình</b>	<b>Độ lưu loát</b>	<b>Độ liên hệ</b>	<b>Độ thông tin</b>	<b>Tổng quan</b>
Baseline	3.87	3.32	3.02	3.25
GPT-2 base	4.23	4.15	3.57	3.81
GPT-2 Plus	4.25	4.18	3.91	4.09
GPT-2 + TAPT	4.24	4.16	4.02	4.12
GPT-2 Plus + TAPT	4.25	4.31	4.23	4.24
<b>Kết quả trên tập dữ liệu chưa thấy</b>				

Mô hình	Độ lưu loát	Độ liên hệ	Độ thông tin	Tổng quan
Baseline	2.51	2.36	2.17	2.21
GPT-2 base	4.15	3.51	3.20	3.24
GPT-2 Plus + TAPT	4.21	3.92	3.59	3.61

Từ bảng kết quả 4.3 có thể thấy, mô hình cơ sở cho ra hiệu suất kém nhất trên cả 2 tập dữ liệu và 4 tiêu chí đánh giá, lý do dẫn đến điều này đã được giải thích ở phần trên. Các mô hình GPT-2 vẫn cho hiệu suất tốt, đặc biệt mô hình GPT-2 Plus + TAPT đã chứng minh kết quả vượt trội trên cả hai tập dữ liệu. Một điểm cần chú ý nữa là các chỉ số như độ lưu loát và độ liên quan của tất cả mô hình GPT-2 đều trên ngưỡng 4, điều này chứng tỏ các mô tả được sinh ra đều có độ lưu loát khá tốt và có nội dung phản ánh sát với sản phẩm tương ứng. Tuy độ đo thông tin có kết quả thấp hơn so với các độ đo khác, nó vẫn cho thấy rằng các mô tả đã cung cấp rất nhiều thông tin hữu ích cho người đọc.

Từ cả hai thử nghiệm đánh giá trên cho thấy, mô tả sản phẩm được tạo ra bởi các mô hình GPT-2, đặc biệt là GPT-2 Plus + TAPT đã đạt được chất lượng tương đối tốt, trên tất cả tiêu chí đánh giá như: Độ lưu loát, Độ liên hệ, Độ thông tin, Độ đa dạng từ vựng, Tổng quan. Tất cả kết quả và phân tích này đã củng cố rằng cách tiếp cận được đề xuất trong luận văn là phù hợp và có thể mang lại những kết quả đầy hứa hẹn cho bài toán sinh mô tả sản phẩm. Dưới đây là một số ví dụ kết quả thực tế của các mô hình trong thực nghiệm trên.

Bảng 4.4. Một số mô tả sản phẩm được tạo ra bởi các mô hình

Thông tin sản phẩm	GPT-2 base	GPT-2 Plus + TAPT
Clothing/Women/Skirt Denim Skirt	Perfect for a chic or casual day, this Alivia Ford dress is a practical addition to any	Alivia Ford's classic denim skirt are a go-to choice for work or play. Crafted with a hint of stretch for

(brand: Alivia Ford) (size: about knee) (material: spandex) (color: black)	wardrobe. The versatile knee-length dress is a practical and stylish choice for many occasions. The soft, plush feel of spandex adds a luxurious look to any outfit. Chic black is a fun way to update any wardrobe this season.	comfort, these skirt feature an elastic waistband to set you up for all-day comfort. This about knee item is a great choice for your closet to ensure a different look every time. The best spandex material for a garment that offers a charming look and feel. It's an everyday classic and is sure to become a wardrobe staple. This shade of black is a chic shade of soft and modern, but it's also a good way to stay ahead of the trends.
Toys and Game / Action Figure Beast Morphers (brand: Bioworld) (material: plastic) (size: 1:10)	A classic Bioworld action figure is a fun way to add a touch to any collection. Plastic is tough enough to stand up to regular use and is also resistant to scratching and grime. Available in a 1:10 scale, this action figure is a great size for playing or displaying with friends and family.	A must-have addition to any collection, this Bioworld action figure is a great addition to a fan's toy collection or just a decorative addition to their own. The plastic material is durable enough to withstand regular handling and is designed to stand up to everyday wear and tear. A portable size of 1:10 that fits easily into any collection, this action figure is large enough to be displayed in a room, on the shelf or on its own in a cabinet.

## Chương 5. Kết luận và định hướng phát triển

Để thích nghi và phát triển trong thời đại kinh tế số, các cá nhân, doanh nghiệp luôn phải tự hoàn thiện, làm mới mình, áp dụng những chiến lược đúng đắn để gia tăng khả năng thành công trên thị trường. Một trong số đó có thể kể đến việc cải thiện chất lượng nội dung sản phẩm, giúp cho người tiêu dùng có thể tiếp nhận đầy đủ thông tin đồng thời tạo ra động lực thúc đẩy hành vi mua của họ. Tuy nhiên trong thực tế, để triển khai giải pháp này còn vấp phải nhiều khó khăn, nhất là về tài chính và con người. Nhận thấy sự cần thiết của một giải pháp tổng thể để giải quyết vấn đề này, tôi đã vận dụng những kiến thức, kỹ thuật tiên bộ trong lĩnh vực học sâu để xây dựng một phương pháp mới giúp tạo ra các mô tả sản phẩm một cách tự động, dễ dàng và hiệu quả. Trải qua quá trình phát triển, cụ thể nghiên cứu đã đạt được những kết quả chính như sau:

- Đề xuất giải pháp ứng dụng mô hình ngôn ngữ GPT-2 cho bài toán sinh mô tả sản phẩm.
- Áp dụng các kỹ thuật tăng cường dữ liệu và TAPT cho mô hình GPT-2 để nâng cao chất lượng của mô hình sinh.
- Xây dựng chiến thuật sinh mô tả giúp tăng cường khả năng mở rộng hiệu suất tổng thể theo chiều ngang đồng thời giảm thiểu những vấn đề mà những mô hình sinh thường gặp phải là phụ thuộc xa trong văn bản dài.
- Các kết quả thử nghiệm cho thấy rằng những mô hình đề xuất đã có hiệu suất vượt trội so với mô hình cơ sở thông qua các phương pháp đánh giá bao gồm đánh giá tự động và đánh giá thông qua con người. Đặc biệt, chúng còn đạt được kết quả hứa hẹn không chỉ trên tập dữ liệu đã thấy mà còn trên cả tập dữ liệu chưa thấy.

Có thể thấy, mô hình GPT-2 đã được chứng minh hoạt động hiệu quả khi sử dụng cùng với các đặc trưng của sản phẩm như tiêu đề, nhãn hiệu, danh mục, thuộc tính. Tuy nhiên, trong thực tế sản phẩm còn nhiều thông tin tiềm năng khác còn có thể được khai thác như hình ảnh hoặc âm thanh. Do đó, trong tương lai, hướng nghiên cứu sẽ tập trung vào việc tìm cách sử dụng những loại thông tin này để tạo ra những mô tả đa dạng, phong phú hơn. Ngoài ra, tôi cũng sẽ thử nghiệm thêm các mô hình ngôn ngữ khác như BERT, T5, ... để đánh giá chất lượng cũng như xây dựng một mô hình mới hiệu quả hơn.

## Tài liệu tham khảo

### Tiếng Anh

- [1] W. C. Mann and S. A. Thompson, “Rhetorical Structure Theory: Description and Construction of Text Structures,” *Natural Language Generation*, pp. 85–95, 1987.
- [2] N. Asher and A. Lascarides, *Logics of conversation*. Cambridge: Cambridge Univ. Press, 2010.
- [3] C. Sporleder and M. Lapata, “Discourse chunking and its application to sentence compression,” *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing - HLT 05*, 2005.
- [4] J. Clarke and M. Lapata, “Global Inference for Sentence Compression: An Integer Linear Programming Approach,” *Journal of Artificial Intelligence Research*, vol. 31, pp. 399–429, 2008.
- [5] A. Graves, “Generating Sequences With Recurrent Neural Networks”, *CoRR*, vol abs/1308.0850, 2013.
- [6] S. Hochreiter en J. Schmidhuber, “Long Short-Term Memory”, *Neural Comput.*, vol 9, no 8, bll 1735–1780, 1997.
- [7] K. Cho et al., “Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation”, in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014*, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL, 2014, bll 1724–1734.
- [8] I. Sutskever, O. Vinyals, en Q. V. Le, “Sequence to Sequence Learning with Neural Networks”, in *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014*, December 8-13 2014, Montreal, Quebec, Canada, 2014, bll 3104–3112.
- [9] D. Bahdanau, K. Cho, en Y. Bengio, “Neural Machine Translation by Jointly Learning to Align and Translate”, in *3rd International Conference on Learning Representations, ICLR 2015*, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings, 2015.
- [10] O. Vinyals, M. Fortunato, en N. Jaitly, “Pointer Networks”, in *Advances in Neural Information Processing Systems 28: Annual Conference on Neural*



- Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada, 2015, bll 2692–2700.
- [11] J. Wang, Y. Hou, J. Liu, Y. Cao, en C.-Y. Lin, “A Statistical Framework for Product Description Generation”, in Proceedings of the Eighth International Joint Conference on Natural Language Processing, IJCNLP 2017, Taipei, Taiwan, November 27 - December 1, 2017, Volume 2: Short Papers, 2017, bll 187–192.
  - [12] T. Joachims, “Optimizing search engines using clickthrough data”, in Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, July 23-26, 2002, Edmonton, Alberta, Canada, 2002, bll 133–142.
  - [13] Q. Chen, J. Lin, Y. Zhang, H. Yang, J. Zhou, en J. Tang, “Towards Knowledge-Based Personalized Product Description Generation in E-commerce”, in Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2019, Anchorage, AK, USA, August 4-8, 2019, 2019, bll 3040–3050.
  - [14] A. Vaswani et al., “Attention is All you Need”, in Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA, 2017, bll 5998–6008.
  - [15] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, en I. Sutskever, “Language Models are Unsupervised Multitask Learners”, 2019.
  - [16] A. Radford en K. Narasimhan, “Improving Language Understanding by Generative Pre-Training”, 2018.
  - [17] Z. Dai, Z. Yang, Y. Yang, J. G. Carbonell, Q. V. Le, en R. Salakhutdinov, “Transformer-XL: Attentive Language Models beyond a Fixed-Length Context”, in Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers, 2019, bll 2978–2988.
  - [18] K. He, X. Zhang, S. Ren, en J. Sun, “Identity Mappings in Deep Residual Networks”, in Computer Vision - ECCV 2016 - 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part IV, 2016, vol 9908, bll 630–645.

- [19] M. Artetxe, G. Labaka, E. Agirre, en K. Cho, “Unsupervised Neural Machine Translation”, CoRR, vol abs/1710.11041, 2017.
- [20] S. Reddy, D. Chen, en C. D. Manning, “CoQA: A Conversational Question Answering Challenge”, Trans. Assoc. Comput. Linguistics, vol 7, bll 249–266, 2019.
- [21] A. See, P. J. Liu, en C. D. Manning, “Get To The Point: Summarization with Pointer-Generator Networks”, in Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 1: Long Papers, 2017, bll 1073–1083.
- [22] M. Lewis et al., “BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension”, in Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020, 2020, bll 7871–7880.
- [23] Y. Liu et al., “RoBERTa: A Robustly Optimized BERT Pretraining Approach”, CoRR, vol abs/1907.11692, 2019.
- [24] J. Devlin, M.-W. Chang, K. Lee, en K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”, in Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers), 2019, bll 4171–4186.
- [25] G. Yu, “Lexical Diversity in Writing and Speaking Task Performances”, Applied Linguistics, vol 31, bll 236–259, 2010.
- [26] J. Chotlos, “IV. A statistical and comparative analysis of individual written language samples”, The Psychological Monographs, vol 56, bll 75–111, 1944.
- [27] M. C. Templin, “Certain language skills in children: their development and interrelationships”, 1957.
- [28] P. McCarthy en S. Jarvis, “MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment”, Behavior research methods, vol 42, bll 381–392, 05 2010.
- [29] D. Malvern, B. Richards, N. Chipere, en P. Durán, “Lexical Diversity and Language Development: Quantification and Assessment”, 2004.
- [30] P. M. McCarthy en S. Jarvis, “vocd: A theoretical and empirical evaluation”, Language Testing, vol 24, bll 459–488, 2007.

- [31] A. Holtzman, J. Buys, M. Forbes, en Y. Choi, “The Curious Case of Neural Text Degeneration”, ArXiv, vol abs/1904.09751, 2020.
- [32] G. Samorodnitsky, “Long Range Dependence”, Found. Trends Stoch. Syst., vol 1, bll 163–257, 2006.
- [33] M. Kay, “The Proper Place of Men and Machines in Language Translation”, Machine Translation, vol 12, bll 3–23, 2004.
- [34] M. Lapata, R. Sennrich, en J. Mallinson, “Paraphrasing Revisited with Neural Machine Translation”, in EACL, 2017.
- [35] E. Pavlick, P. Rastogi, J. Ganitkevitch, B. Van Durme, en C. Callison-Burch, “PPDB 2.0: Better paraphrase ranking, fine-grained entailment relations, word embeddings, and style classification”, in ACL, 2015.
- [36] S. Gururangan et al., “Don’t Stop Pretraining: Adapt Language Models to Domains and Tasks”, ArXiv, vol abs/2004.10964, 2020.
- [37] R. Ghani, K. Probst, Y. Liu, M. Krema, en A. E. Fano, “Text mining for product attribute extraction”, SIGKDD Explor., vol 8, bll 41–48.