

# Tinh chỉnh mô hình BERT cho nhiệm vụ ABSA Tiếng Việt

Sinh viên: Nguyễn Thị Quý 20185396  
GVHD: TS. Lê Chí Ngọc

Toán Tin 01-k63  
Viện Toán ứng dụng và Tin học  
Trường đại học Bách Khoa Hà Nội

Hà nội, 8/2022



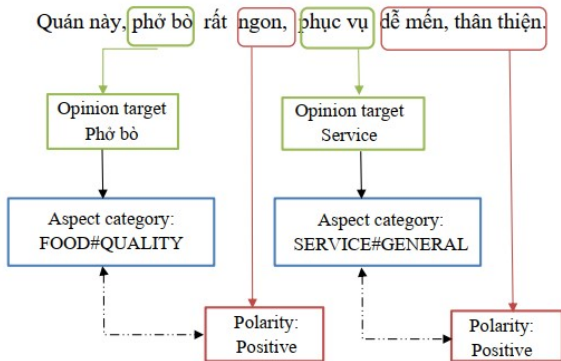
# Nội dung

- 1 Giới thiệu - nhiệm vụ ABSA
- 2 Mô hình BERT
- 3 Mô hình thực nghiệm
- 4 Kết quả thực nghiệm

## Giới thiệu - nhiệm vụ ABSA

# Giới thiệu

## Nhiệm vụ ABSA - Aspect-based sentiment analysis



# Các mô hình nhiệm vụ

- Xây dựng mô hình nhận diện {thực thể#khía cạnh} - AD, mô hình nhận diện {thực thể#khía cạnh#nhấn cực} - AP riêng biệt.
- Mô hình phân cực cảm xúc cho các khía cạnh BERT POLARITY, và mô hình kết hợp đầu cuối ABSA.

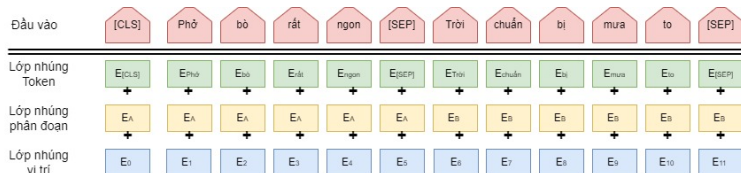
# Mô hình BERT

# Giới thiệu

## BERT [1] - Bidirectional Encoder Representations from Transformers

- Là một mô hình ngôn ngữ tạo bởi Google AI.
- GLUE score - chỉ số tổng quát mức độ hiểu ngôn ngữ
- Bert sử dụng transformers, huấn luyện dữ liệu 2 chiều, học ngữ cảnh của từ.

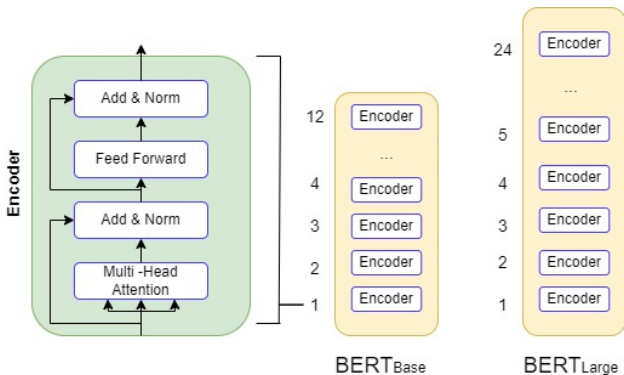
# Đại diện đầu vào mô hình BERT



Hình 1: Đại diện đầu vào của mô hình BERT



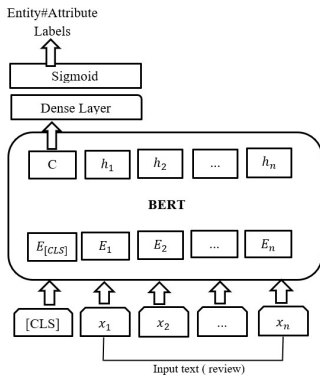
# Kiến trúc mô hình BERT



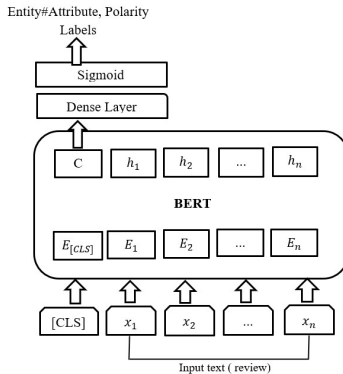
Hình 2: Kiến trúc mô hình BERT BASE và BERT LARGE

## Mô hình thực nghiệm

# Tinh chỉnh mô hình BERT cho nhiệm vụ AD, AP [2]



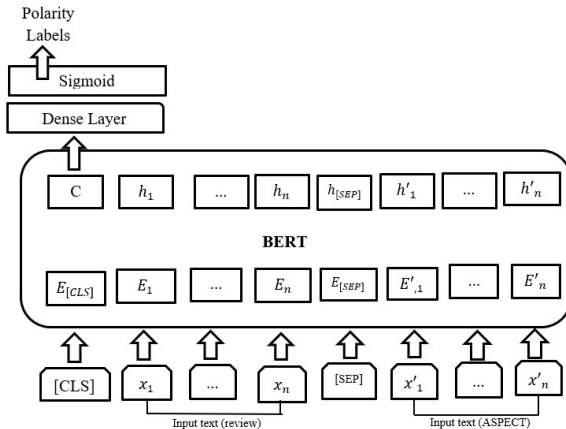
BERT - AD



BERT - AP

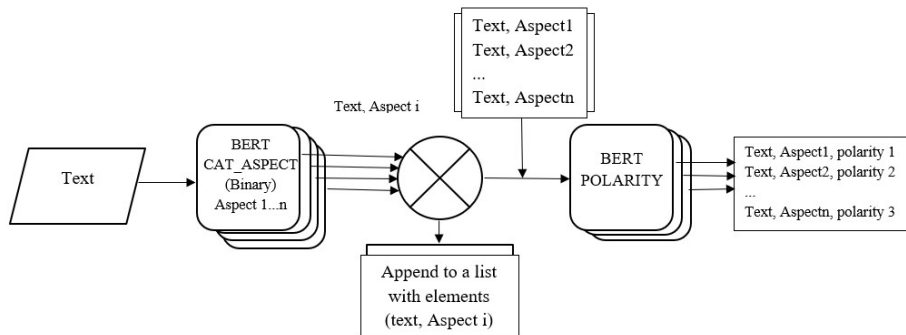
Hình 3: Tinh chỉnh BERT cho nhiệm vụ AD, AP

# Tinh chỉnh mô hình BERT cho nhiệm vụ BERT POLARITY



Hình 4: Mô hình tinh chỉnh BERT POLARITY

# Mô hình kết hợp



Hình 5: Flowchart nhiệm vụ đầu cuối ABSA

# Dữ liệu [3]

Miền	Tập dữ liệu	Số đánh giá	Tổng số khía cạnh
Nhà hàng	Train	2961	9297
	Development	1290	3443
	Test	500	2419
Khách sạn	Train	3000	13949
	Development	2000	7111
	Test	600	2584

Bảng 2.1: Thống kê tập dữ liệu VLSP

## Example 1: Restaurant domain.

#lozi #lozisaigon #chaovit #anchinh ngon - bổ - rẻ.

Khuyết là hơi xa trung tâm

(delicious - good for health - cheap, far from center)

{FOOD#PRICE, positive},

{FOOD#QUALITY, positive},

{LOCATION#GENERAL, negative}

## Example 2. Hotel domain.

Phòng ốc sạch, giường thoải mái, nhân viên thân thiện.

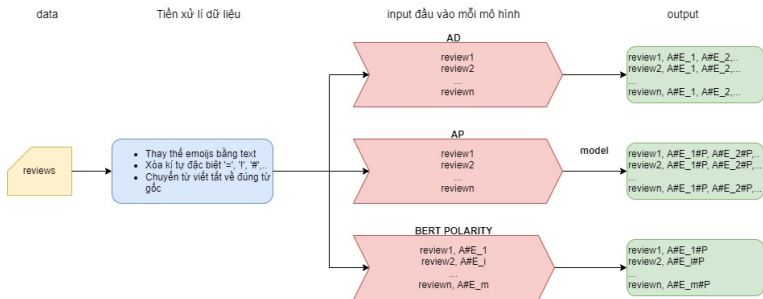
(clean rooms, comfortable beds, friendly staffs)

{ROOMS#CLEANLINESS, positive},

{ROOMS#COMFORT, positive},

{SERVICE#GENERAL, positive}

# Quá trình tiền xử lí dữ liệu



Hình 6: Luồng dữ liệu cho thực nghiệm

# Quá trình thực thi và siêu tham số

- pre-trained BERT: "bertbase-multilingual-cased", PhoBERT
- sử dụng BERTBASE cho việc tinh chỉnh mô hình với 12 khối transformer, 768 lớp ẩn, và 12 self-attention head và tổng tham số là 110M
- thực thi đào tạo trên google colab, hỗ trợ GPU, trung bình 1p40s cho 40 mẫu.
- Siêu tham số sử dụng:
  - MAX\_LENGTH: 256
  - epochs: 10
  - Learning rate: 1e-5
  - batch-size: 16
  - valid batch size: 16
  - dropout: 0.1
- Ngưỡng phân loại:

Miền	Subtask1	Subtask2	BERT POLARITY
Nhà hàng	0.5	0.6	0.5
Khách sạn	0.45	0.45	0.5



## Kết quả thực nghiệm

# Độ đo đánh giá

$$\mathbf{Accuracy(T)} = \frac{TP + TN}{P + N} = \frac{TP + TN}{TP + FP + TN + FN}(\%) \quad (1)$$

$$\mathbf{Precision} = \frac{TP}{TP + FP}(\%) \quad (2)$$

$$\mathbf{Recall} = \frac{TP}{TP + FN}(\%) \quad (3)$$

$$\mathbf{F1\ Score} = \frac{2precision \times recall}{precision + recall}(\%) \quad (4)$$

- Sử dụng micro-averaged trên các độ đo precision, recall, f1

# Kết quả chạy mô hình AD, AP, so sánh

Miền	Tác giả	Precision	recall	F1
Nhà hàng	Đội về 3			
	Đội về nhì[9]	0.78	0.65	0.71
	Đội về nhất	0.75	0.85	0.79
	Dang[18]			
	Lam	0.77	0.88	0.82
	<b>BERT_MULTI</b>	<b>0.77</b>	<b>0.85</b>	<b>0.81</b>
Khách sạn	<b>BERT_Phobert</b>	<b>0.8</b>	<b>0.84</b>	<b>0.82</b>
	Đội về 3			
	Đội về nhì[9]	0.83	0.51	0.63
	Đội về nhất	0.75	0.64	0.69
	Dang[18]			
	Lam[10]	0.79	0.78	0.69
Khách sạn	<b>BERT_MULTI</b>	<b>0.86</b>	<b>0.49</b>	<b>0.63</b>
	<b>BERT_Phobert</b>	<b>0.84</b>	<b>0.56</b>	<b>0.67</b>

## AD - dev

Miền	Tác giả	Precision	recall	F1
Nhà hàng	Đội về 3	0.88	0.38	0.54
	Đội về nhì[9]	0.62	0.62	0.62
	Đội về nhất	0.79	0.76	0.77
	Dang[11]	0.85	0.76	0.80
	Lam	0.79	0.83	0.81
	<b>BERT_MULTI</b>	<b>0.79</b>	<b>0.77</b>	<b>0.78</b>
Khách sạn	<b>BERT_Phobert</b>	<b>0.85</b>	<b>0.74</b>	<b>0.79</b>
	Đội về 3	0.85	0.42	0.56
	Đội về nhì	0.83	0.58	0.68
	Đội về nhất	0.76	0.66	0.7
	Dang[18]	0.82	0.6	0.7
	Lam[10]	0.79	0.79	0.79
Khách sạn	<b>BERT_MULTI</b>	<b>0.88</b>	<b>0.50</b>	<b>0.63</b>
	<b>BERT_Phobert</b>	<b>0.84</b>	<b>0.63</b>	<b>0.72</b>

## AD - test

Miền	Tác giả	Precision	recall	F1
Nhà hàng	Đội về 3			0.59
	Đội về nhì[9]	0.71	0.59	0.64
	Đội về nhất	0.63	0.71	0.67
	Dang[18]			
	Lam[10]	0.66	0.71	0.68
	<b>BERT_MULTI</b>	<b>0.83</b>	<b>0.52</b>	<b>0.64</b>
Khách sạn	<b>BERT_Phobert</b>	<b>0.69</b>	<b>0.68</b>	<b>0.68</b>
	Đội về 3			
	Đội về nhì[9]	0.78	0.48	0.6
	Đội về nhất	0.67	0.58	0.62
	Dang[11]			
	Lam[10]	0.79	0.59	0.67
Khách sạn	<b>BERT_MULTI</b>	<b>0.84</b>	<b>0.57</b>	<b>0.68</b>
	<b>BERT_Phobert</b>	<b>0.76</b>	<b>0.58</b>	<b>0.66</b>

## AP - dev

Miền	Tác giả	Precision	recall	F1
Nhà hàng	Đội về 3	0.79	0.35	0.48
	Đội về nhì	0.52	0.52	0.52
	Đội về nhất	0.62	0.6	0.61
	Dang[11]			
	Lam[10]	0.63	0.65	0.64
	<b>BERT_MULTI</b>	<b>0.67</b>	<b>0.69</b>	<b>0.68</b>
Khách sạn	<b>BERT_Phobert</b>	<b>0.7</b>	<b>0.63</b>	<b>0.66</b>
	Đội về 3	0.8	0.39	0.53
	Đội về nhì	0.71	0.49	0.58
	Đội về nhất	0.66	0.57	0.61
	Dang[11]			
	Lam[10]	0.7983	0.5882	0.6774
Khách sạn	<b>BERT_MULTI</b>	<b>0.83</b>	<b>0.55</b>	<b>0.66</b>
	<b>BERT_Phobert</b>	<b>0.86</b>	<b>0.54</b>	<b>0.66</b>

## AP - test

# Kết quả đánh giá mô hình combine

Mô hình	Miền	Precision	Recall	F1
BERT ASPECT	Nhà hàng	0.77	0.85	0.81
	Khách sạn	0.84	0.57	0.68
BERT POLARITY	Nhà hàng	0.81	0.80	0.80
	Khách sạn	0.88	0.89	0.89
Combine	Nhà hàng	0.66	0.7	0.68
	Khách sạn	0.61	0.53	0.6

Hình 7: Kết quả đánh giá nhiệm vụ ABSA trên tập dev

Mô hình	Miền	Precision	Recall	F1
BERT ASPECT	Nhà hàng	0.79	0.77	0.78
	Khách sạn	0.88	0.5	0.63
BERT POLARITY	Nhà hàng	0.75	0.73	0.74
	Khách sạn	0.87	0.88	0.88
Combine	Nhà hàng	0.65	0.55	0.6
	Khách sạn	0.6	0.51	0.55

Hình 8: Kết quả đánh giá nhiệm vụ ABSA trên tập test

# Thử nghiệm trên dữ liệu thực tế

Nhà hàng Nihonkai Sushi - Japanese Cuisine

**#2. Đồ ăn tươi ngon, lên nhanh chóng. Nhân viên phục vụ rất tốt**

*subtask1+Polarity:FOOD#QUALITY#positive, SERVICE#GENERAL#positive*

*subtask2:FOOD#STYLE&OPTIONS#positive, FOOD#QUALITY#positive, AMBIENCE#GENERAL#positive, SERVICE#GENERAL#positive*

**#3. Lần gần đây nhất mình qua, thì nhà hàng đóng cửa để phòng chống covid**

*subtask1+Polarity:FOOD#QUALITY#positive, SERVICE#GENERAL#positive*

*subtask2:FOOD#QUALITY#neutral, FOOD#QUALITY#positive*

**#4. Đồ ăn ngon, nhà hàng phục vụ tốt, mình sẽ quay lại nhiều lần**

*subtask1+Polarity:FOOD#QUALITY#positive,*

*RESTAURANT#GENERAL#positive, SERVICE#GENERAL#positive*

*subtask2:FOOD#QUALITY#positive, AMBIENCE#GENERAL#positive, SERVICE#GENERAL#positive*

# Thử nghiệm trên dữ liệu thực tế

Khách sạn Marina Hotel Hanoi

**#4. Dịch vụ mà nhân viên chào đón với một nụ cười rạng rỡ rất tốt và phòng sạch sẽ nên tôi sẽ có một khoảng thời gian vui vẻ haha!**

*subtask1+Polarity:*HOTEL#GENERAL#positive,

SERVICE#GENERAL#positive,

HOTEL#COMFORT#positive, ROOMS#CLEANLINESS#positive,

ROOMS#DESIGN&FEATURES#positive

*subtask2:*HOTEL#COMFORT#positive HOTEL#GENERAL#positive

ROOMS#CLEANLINESS#positive SERVICE#GENERAL#positive

**#5. Nó đối diện với Marriott nên rất dễ nhớ vị trí**

*subtask1+Polarity:*HOTEL#GENERAL#positive,

LOCATION#GENERAL#positive

*subtask2:*LOCATION#GENERAL#positive

**#6. Không có wifi, vì vậy tôi đã bỏ cuộc.**

*subtask1+Polarity:*HOTEL#GENERAL#negative

*subtask2:*

# Thank you!

# Tài liệu tham khảo

- [1] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “Bert: Pre-training of deep bidirectional transformers for language understanding,” *arXiv preprint arXiv:1810.04805*, 2018.
- [2] N. C. Lê, N. The Lam, S. H. Nguyen, and D. Thanh Nguyen, “On vietnamese sentiment analysis: A transfer learning method,” in *2020 RIVF International Conference on Computing and Communication Technologies (RIVF)*, 2020, pp. 1–5. DOI: 10.1109/RIVF48685.2020.9140757.
- [3] “Vlsp 2018 dataset,”. [Online]. Available: <https://vlsp.org.vn/resources-vlsp2018>.