**Báo cáo đề tài phân tích cảm xúc dựa trên khía cạnh**

Giảng viên: Đặng Thị Phúc

Thành viên nhóm: Trần Hồ Phi Long

Nguyễn Đức Thịnh

Quách Trọng Nghĩa

1. **Giới thiệu đề tài**

Với sự phát triển của công nghệ và Internet, các loại hình truyền thông xã hội như mạng xã hội, diễn đàn đã cho phép mọi người không chỉ chia sẻ thông tin mà còn bày tỏ quan điểm, thái độ của mình về các sản phẩm, dịch vụ và các vấn đề xã hội khác. Internet trở thành một nguồn thông tin vô cùng quý giá và quan trọng. Mọi người ngày nay sử dụng nó như một tài liệu tham khảo để định hướng các quyết định của họ khi mua một sản phẩm hoặc sử dụng một dịch vụ. Hơn nữa, loại thông tin này cũng cho phép các nhà sản xuất và nhà cung cấp dịch vụ nhận được phản hồi về những hạn chế của sản phẩm và do đó cải thiện chúng để đáp ứng tốt hơn nhu cầu của khách hàng. Hơn nữa, nó cũng có thể giúp các cơ quan chức năng biết được thái độ và ý kiến ​​của người dân đối với các sự kiện xã hội để từ đó có những điều chỉnh phù hợp.

Kể từ đầu những năm 2000, khai thác ý kiến ​​và phân tích cảm tính đã trở thành một chủ đề nghiên cứu mới và tích cực trong Xử lý ngôn ngữ tự nhiên và Khai thác dữ liệu. Các chủ đề trong lĩnh vực này bao gồm: Spam detection, Rating, sentiment classification

Gần đây, Phân tích tình cảm dựa trên khía cạnh (ABSA), nhiệm vụ khai thác và tổng hợp ý kiến ​​từ văn bản về các thực thể cụ thể và các khía cạnh của chúng, đã và đang thu hút nhiều nghiên cứu hơn.

1. **Mô tả nhiệm vụ**

Trong mỗi bài đánh giá về quán ăn của khách hàng sẽ xác định các cặp {entity (thực thể), attribute (thuộc tính)}, tương đương với polarity (thuộc tính ) của nó

1. **Data**
2. Polarity label

* Postive
* Neutral
* Negative

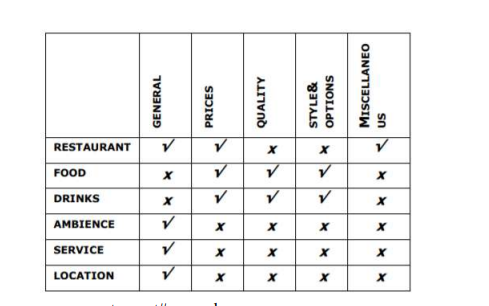
1. Aspect label làm thành cặp {entity (thực thể), attribute (thuộc tính)}

Entities:

* Restaurant
* Ambience
* Location
* Food
* Drinks
* Service

Attribute:

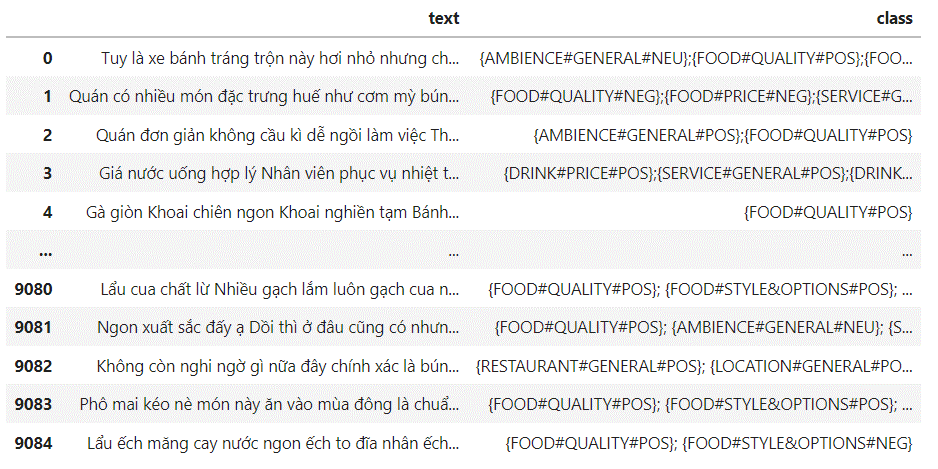
* General
* Quality
* Price
* Style\_options
* Miscellaneous



1. **Ví dụ**

mình ăn ở đây mấy lần rồi **ngon lắm luôn** mình nghĩ **đồ ăn ở đây ngon** **và chất lượng** nhất sài gòn đấy **ăn phần lớn 25k là no cả buổi** luôn, {FOOD#QUALITY#POS};{FOOD#PRICE#POS}

**Không gian nhỏ nhưng yên tĩnh và sạch sẽ** Chị chủ dễ thương **Đồ ăn ngon và giá cả khá rẻ trang trí siêu dễ thương** luôn Rất thích hợp cho những ai muốn có không gian yên tĩnh để đọc sách hay làm việc,{AMBIENCE#GENERAL#POS};{SERVICE#GENERAL#POS};{FOOD#QUALITY#POS};{FOOD#PRICE#POS}



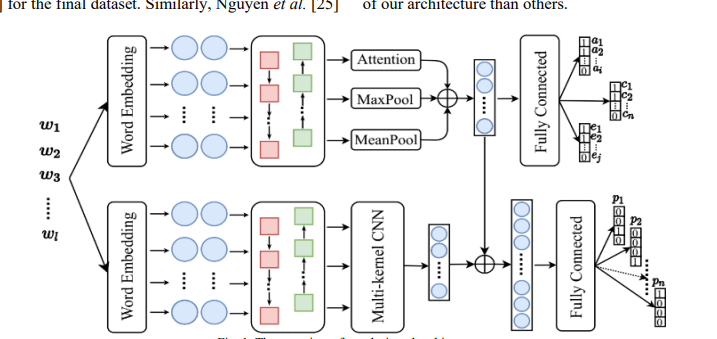
D **Xử lý**

* Xóa stopword
* Sủa lỗi chính tả
* Xóa những câu không liên quan
* Xóa những từ như 20k , tên riêng , xóa icon, xóa url
* Xóa chấm câu, xóa số
* Chuyển về chữ thường

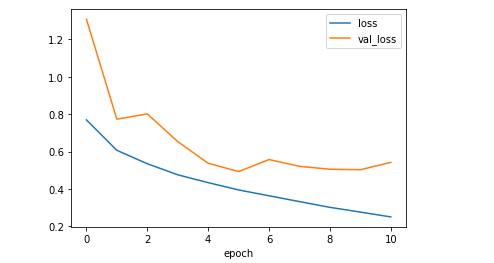
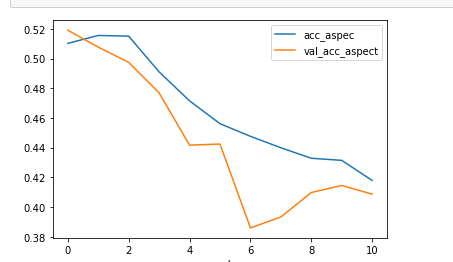
1. **Mô hình**

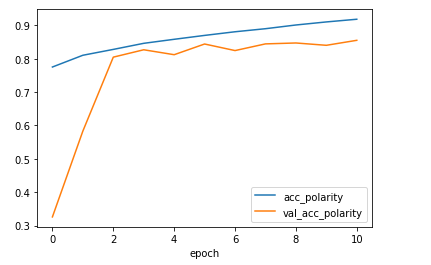
Mô hình được huấn luyện với hai mô hình cho hai sub-task : phân loại khía cạnh và phân loại cảm xúc tương đương với khía cạnh.

**Sơ đồ mô hình 1**

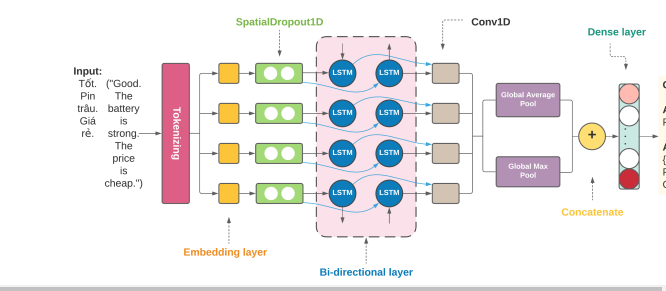


Biểu đồ đánh giá mô hình 1

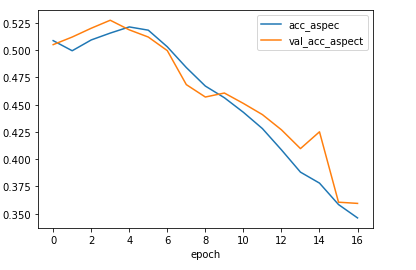
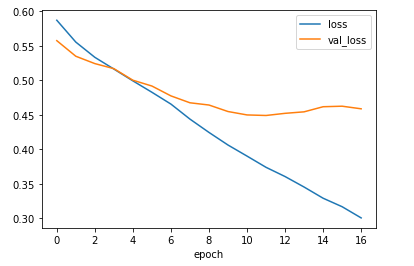
 

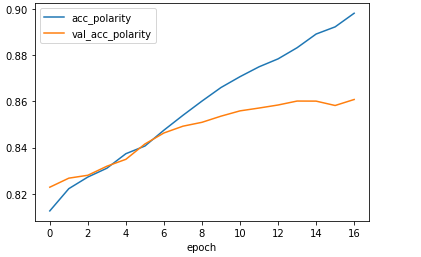


**Sơ đồ mô hình 2:**

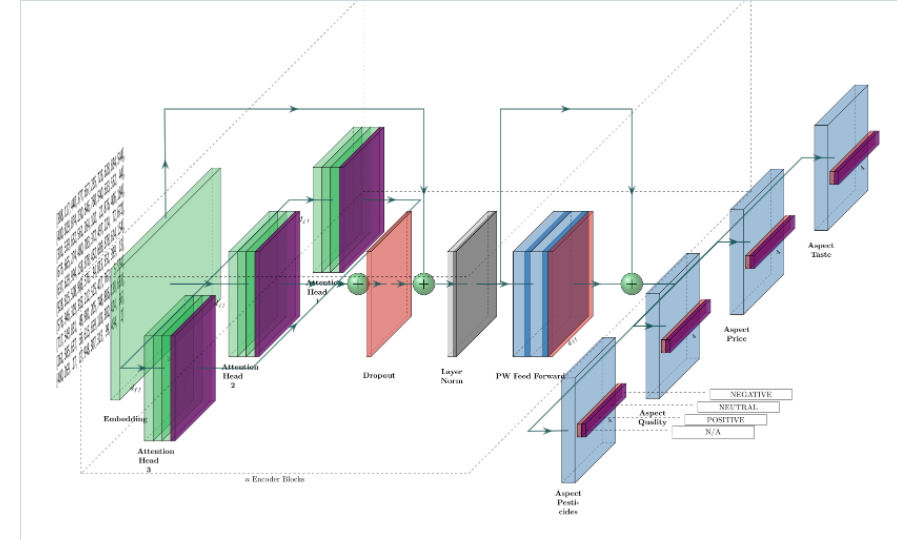
****

Biểu đồ đánh giá mô hình 2

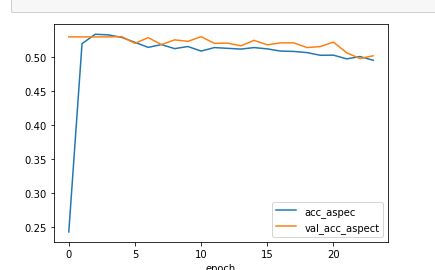
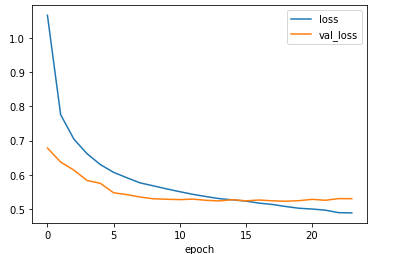


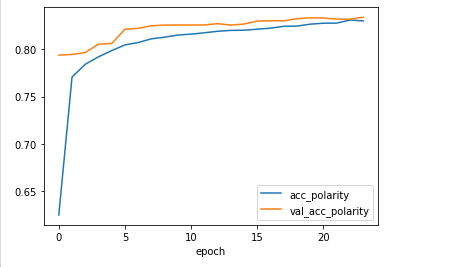


**Sơ đồ mô hình 3:**

****

Biểu đồ đánh giá mô hình 3





|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Mô hình | Các layer chính | Total params | Loss | Val\_loss | Accuracy\_aspect | Accuracy\_polarity | val Accuracy\_aspect | Val\_accuracy\_polarity |
| Mô hình 1 | dùng cnn, gru, attention | 5,701,336 | 0.24 | 0.46 | 0.51 | 0.91 | 0.51 | 0.85 |
| Mô hình 2 | Lstm, conv1d | 1,415,100 | 0.33 | 0.44 | 0.51 | 0.52 | 0.86 | 0.89 |
| Mô hình 3 | transformer | 154,080 | 0.48 | 0.51 | 0.53 | 0.83 | 0.53 | 0.83 |

1. **Triển khai model trên local dùng fastapi**