**So sánh chất lượng của các bộ Words Embedding**

1. **So sánh trên các mô hình phân loại:**

Đánh giá dựa trên độ tương thích và chính xác trên các 4 mô hình phân loại: Ở đây nhóm sử dụng 4 mô hình phổ biến thường hay được dùng trong phân loại báo.

* ***Multi layer perceptron*** *(MLPs)*
* LSTM
* Sequential\_30 (custom model)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Fast Text** | **GloVe** | **ELMo** | **Word2Vec** |
| **MLPs** *(custom)* | 96.55 | 95.47 | 95.04 | **97.37** |
| **LSTM** | **96.09** | 95.67 | 94.62 | 94.58 |
| **CNN + LSTM** | **97.15** | 96.15 | 95.88 | 95.05 |
| **Bi - LSTM** | **96.89** | 95.85 | 95.1 | 95.63 |

* LSTM + CNN

*Bảng 1.1: đánh giá độ chính xác trong phân loại báo trên tập kiểm thử*

Bên cạnh đó, ta cũng có đồ thị chi tiết về quá trình đánh giá trên

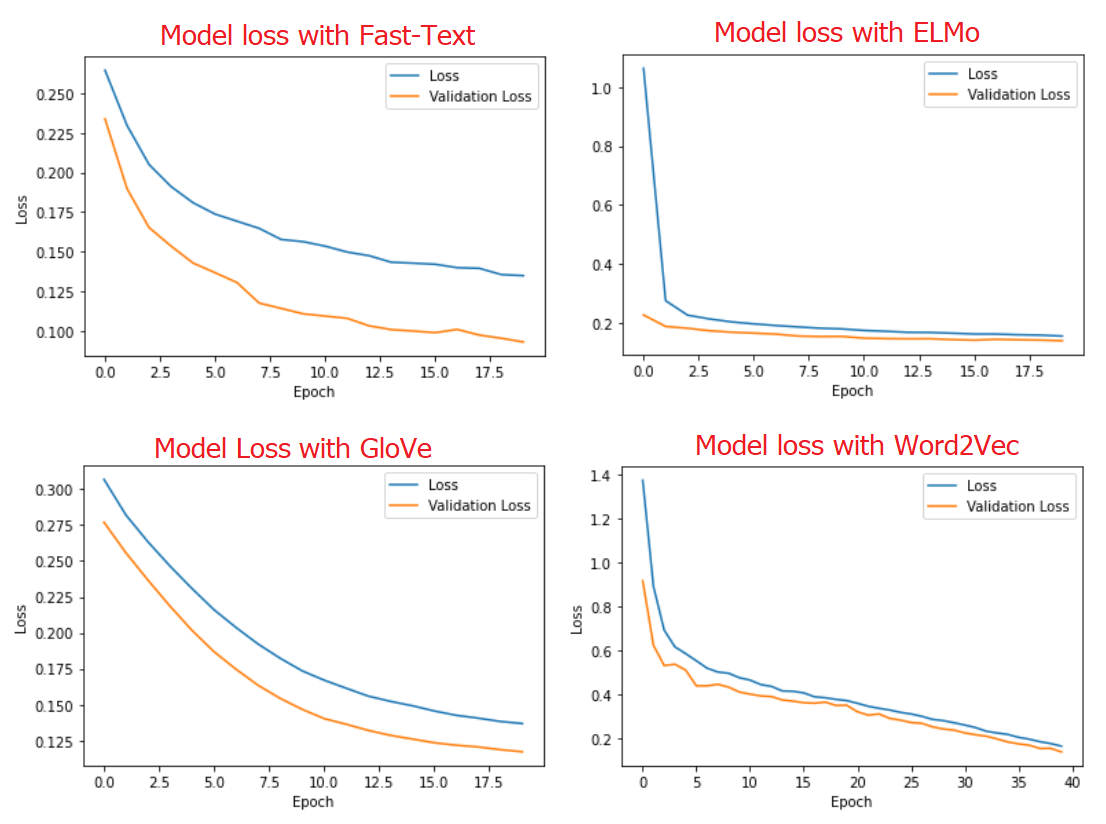
*Hình 1.1: Biểu đồ miêu tả chi tiết độ chính xác đo được trên tập test của các bộ embedding trên từng mô hình phân lớp*

1. **Chi tiết quá trình đánh giá trên dữ liệu:**
   1. **. Đánh giá trên mô hình mạng neural đa tầng (custom):**

2.1.1. Cấu trúc mạng neural được sử dụng:

*Hình 2.1.1: Mô hình mạng neural đa tầng MLPs (custom)*

2.1.2 Đồ thị hàm mất mát của các word embedings trên bộ phân lớp:

****

*Hình 2.1.2: Đồ thị hàm loss của các mô hình word-embeding so với mô hình thứ nhất MLPs (custom)*

Như ta thấy được rằng Mô Hình ELMo có điểm hội ban tốt hơn so với các mô hình khác. Tiếp theo là GloVe có đồ thị loss với đường cong khá tốt. Fast-Text tuy thời gian xử lý 1 câu rất nhanh tuy nhiên như ta thấy đồ thị hàm loss của nó khá rồ rề có một số thời điểm loss sẽ tăng nên khó dự đoán số epoch cần thiết cho mẫu này hơn.

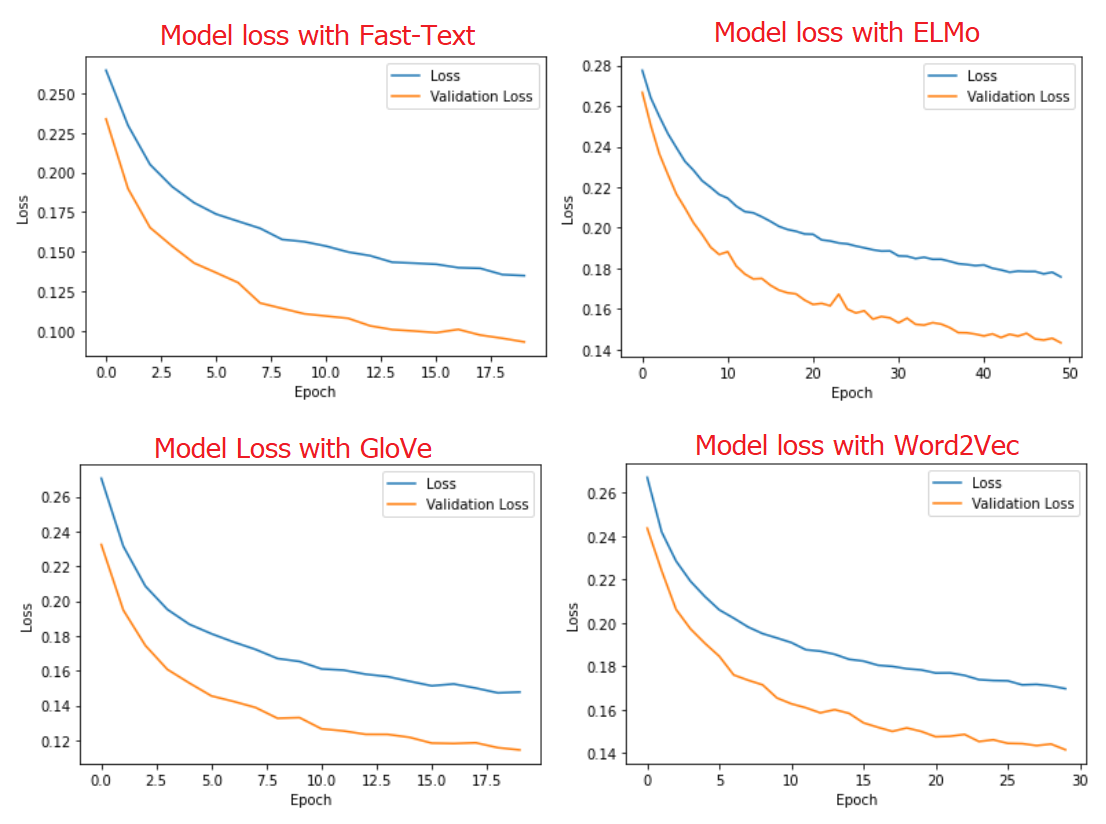
* 1. **. Đánh giá trên mô hình mạng LSTM:**

2.1.1. Cấu trúc mạng LSTM được sử dụng:

Đây là một mạng LSTM thuần với 1 layer LSTM 100 unit node và 1 layer neural 50 unit, mục đích của model này là để trích xuất và lưu lại các đặt trung cơ bản cũng như tăng tốc độ xử lý với dữ liệu nhiều loại câu. Có thể được xem là một mạng thuần LSTM.

*Hình 2.2.1: Mô hình mạng LSTM Vanila*

2.1.2 Đồ thị hàm mất mát của các word embedings trên bộ phân lớp LSTM:

****

*Hình 2.2.2: Đồ thị hàm loss của các mô hình word-embeding so với mô hình thứ hai LSTM.*

Rõ ràng ta có thể thấy 4 đồ thị trên, Fastext hội tụ nhanh nhất đối với model LSTM trong khoảng 18-20 epoch. Tiếp theo ta có GloVe có khoảng hội tụ khá là nhanh tương tự và như model thứ nhất đã so sánh lần này GloVe cho đồ thị loss rất đẹp. Mặc khác ELMo không còn cho kết quả đáng mong đợi như mô hình thứ nhất nữa mà đồ thị có phần nhấp nhô ở một vài epoch. Và cuối cùng là Word2Vec với mức hội tụ tốt nhưng lượng epoch tiêu hao khá cao gần như gấp đôi các phương pháp khác ~ 25-30 epoch

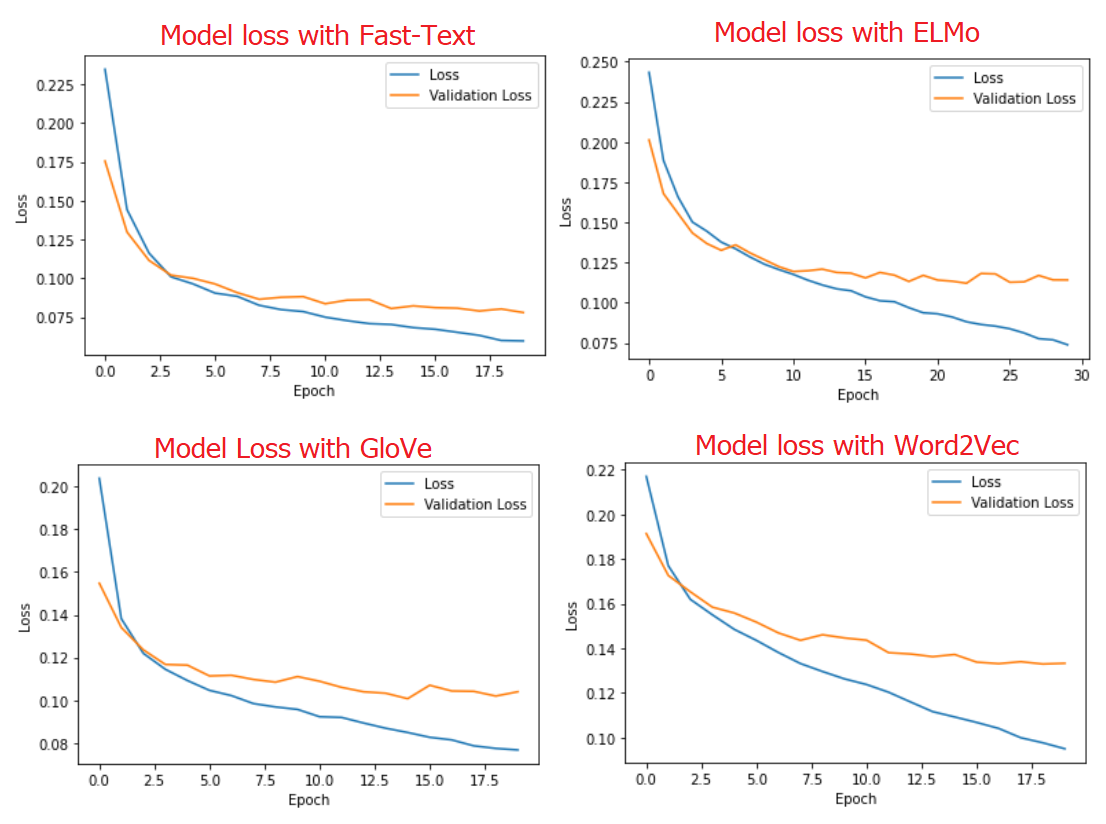
* 1. **. Đánh giá trên mô hình mạng kết hợp CNN + LSTM:**

2.1.1. Cấu trúc mạng LSTM được sử dụng:

Đây là một mô hình kết hợp giữa các tầng đầu là mang tính chất CNN bao gồm 1 tầng tích chập với filter là 32 và một tần max pooling giúp trích xuất đặc trưng trước khi đưa vào train ở 1 tầng LSTM. Tầng này dùng để lưu lại các đặc trưng cần thiết và loại bỏ các đặc trưng không mong muốn. Và sau đây là sơ đồ biểu diễn các tầng trong mạng trên.

*Hình 2.3.1: Mô hình mạng CNN + LSTM*

2.1.2 Đồ thị hàm mất mát của các word embedings trên model CNN + LSTM:

****

*Hình 2.3.2: Đồ thị hàm loss của các mô hình word-embeding so với mô hình thứ ba CNN+LSTM.*

Rõ ràng ta có thể thấy 4 đồ thị trên, Fastext hội tụ nhanh nhất đối với model LSTM trong khoảng 18-20 epoch. Tiếp theo ta có GloVe có khoảng hội tụ khá là nhanh tương tự và như model thứ nhất đã so sánh lần này GloVe cho đồ thị loss rất đẹp. Mặc khác ELMo không còn cho kết quả đáng mong đợi như mô hình thứ nhất nữa mà đồ thị có phần nhấp nhô ở một vài epoch. Và cuối cùng là Word2Vec với mức hội tụ tốt nhưng lượng epoch tiêu hao khá cao gần như gấp đôi các phương pháp khác ~ 25-30 epoch

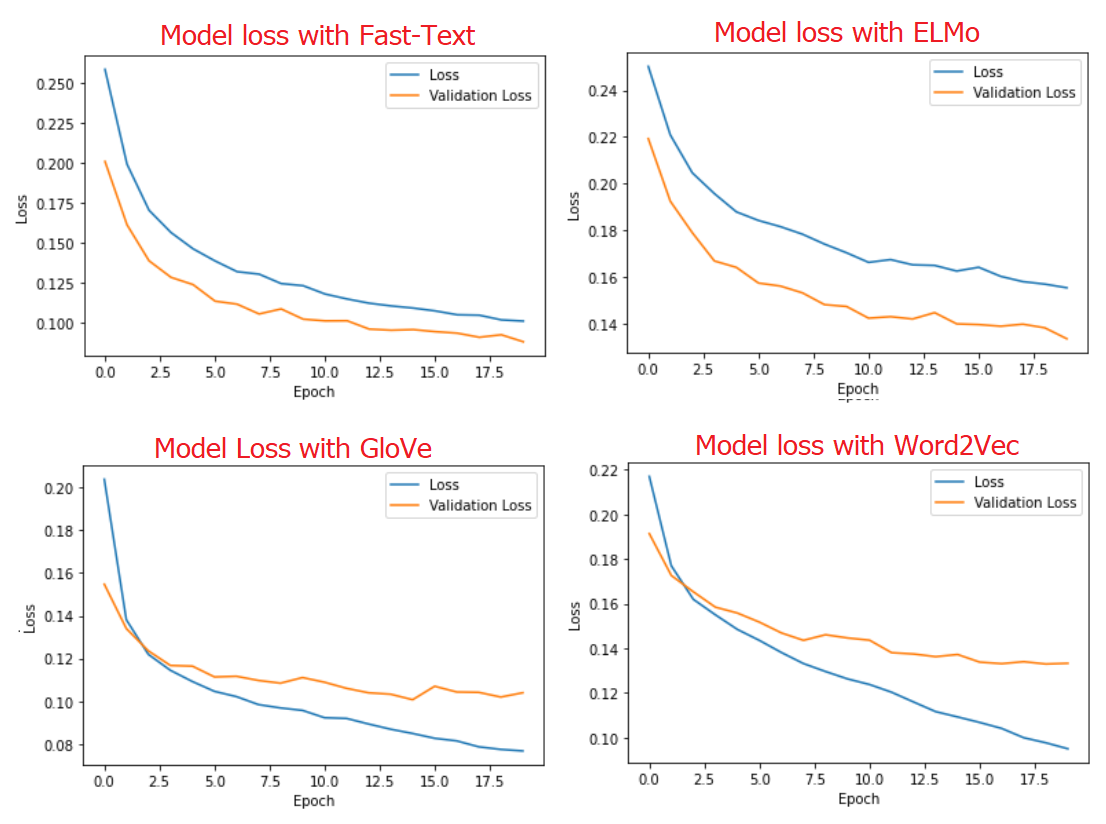
* 1. **. Đánh giá trên mô hình mạng kết hợp Bi -LSTM:**

*2.1.1. Cấu trúc mạng LSTM được sử dụng:*

Đây là mô hình được xây dựng trên mô hình mạng LSTM 2 chiều với 200 unit và 1 tầng phụ neural 100 unit và tầng output cuối sử dụng softmax.

*Hình 2.3.1: Mô hình mạng Bi-LSTM*

*2.1.2 Đồ thị hàm mất mát của các word embedings trên bộ phân lớp Bi - LSTM:*

****

*Hình 2.4.2: Đồ thị hàm loss của các mô hình word-embeding so với mô hình thứ 4 Bi-LSTM.*

Có thể thấy rõ sau hơn 4 mô hình Fast-Text vẫn giữ được tốc độ hội tụ tốt cho cả 4 model phân lớp và quá trình hội tụ ổn định từ 18-20 epoch cho số loss tốt nhất ở khoảng này *(có thể nhỏ hơn nhưng dễ gây overfitting).* Còn GloVe và W2V đã không còn hội tụ tốt đối với mô hình này nữa mà thậm chí có xu hướng bảo hòa ở các epoch gần cuối hoặc tăng lên mất ổn định. Bên cạnh đó, ELMo có sự thể hiện tốt hơn ở model này với kết quá rất sát so với Fastext.