



**CS 769**  
**Advanced NLP**  
**Assignment 1 Report**

**Name: Sujay Chandra Shekara Sharma**

**Wisc ID: 9085871714**

**Wisc Email: [schandrashe5@wisc.edu](mailto:schandrashe5@wisc.edu)**

**Model Architecture:**

**Embedding Layer:**

- The model uses an embedding layer to convert input tokens into vectors of size 300.
- Pre-trained embeddings from FastText are loaded into this embedding layer.
- The embeddings are regularized for robustness using word dropout which randomly drops word embeddings during training as suggested in the deep averaging network paper.

**Classifier:**

- The classifier comprises a feedforward neural network with 3 linear layers and ReLU activations.
  - **First layer:** Takes the embedding size as input and outputs to a hidden layer.
  - **Second layer:** Hidden layer.
  - **Third layer:** Outputs to the number of classes or tags.
- The final output is a set of unnormalized scores for each class.

**Initialization:**

- Embeddings and classifier weights are initialized using a uniform distribution within the range  $[-0.08, 0.08]$ .

**Other Approaches:**

- The primary improvements in the model's accuracy over the baseline came from using pre-trained embeddings from FastText instead of randomly initialized embeddings.
- In addition to this, the addition of random embedding dropout prior to computing the average of the embeddings for an input also led to an improvement in accuracy.
- After experimenting with various different activations like ReLU, LeakyReLU, ELU, GELU, and Tanh, it was observed that there didn't seem to be any significant difference in accuracy regardless of the activation function being used.
- The implementation of batch normalization also didn't result in any accuracy gains.
- Finally, some text pre-processing to remove certain tags like `<br>` in the reviews also didn't help improve the accuracy of the model.