CSE 587: Deep Learning for NLP

Midterm Project

Spam Detection Using CNN and GloVe Embeddings

Jeremy Huang

Date: 2/14/2025

A. Problem Definition and Dataset Curation

Spam message is a big issue nowadays, especially in emails and SMS. Many spam messages contain unwanted advertisements, scams, or harmful links. Spam detection is important because it helps filter out these messages and protect users. The goal of this project is to build a model that can automatically classify messages as either spam or non-spam (ham).

The dataset used for this project is the SMS Spam Collection Dataset, which contains 5,574 messages labeled as either spam or ham. The dataset was downloaded from Kaggle. Before using it for training, I processed the text by removing unnecessary characters, converting everything to lowercase, and tokenizing the words. This could help improve how the model understands the text.

B. Word Embeddings, Algorithm, and Training Process

B.1 Word Embeddings

In order to represent text so that a neural network can understand, I used pre-trained GloVe embeddings (100-dimensional). GloVe embeddings help capture the meaning of words by analyzing how frequently they appear together in large datasets. Instead of treating words as separate entities, embeddings allow words with similar meanings to have similar numerical representations.

B.2 Model Architecture

For this project, I used a Convolutional Neural Network (CNN) to classify messages. CNNs are usually used for image recognition, but they also work well for text classification because they can identify important word patterns. The model consists of several layers:

- Embedding Layer: Converts words into vector representations using pre-trained GloVe embeddings.
- 1D Convolutional Layer: Detects important patterns and word sequences in messages.
- Global Max Pooling Layer: Reduces the size of the data while keeping the most important features.
- Dense Layer: A fully connected layer that helps the model learn complex patterns.
- Dropout Layer: Prevents overfitting by randomly turning off some neurons during training.
- Output Layer: Uses the sigmoid function to classify messages as spam or ham.

B.3 Training Details

- Loss Function: Binary Cross-Entropy
- Optimizer: Adam
- Epochs: 5
- Batch Size: 32
- Validation Split: 20% of the dataset was used for validation

C. Results and Presentation

After training for 5 epochs, the model achieved the following results:

	Training Set	Validation Set
Accuracy	0.9984	0.9874
Loss	0.0103	0.0448

Figure 1. Model Training and Validation Performance

The accuracy of 98.74% on the validation set shows that the model is highly effective in classifying spam and ham messages.

D. In-Depth Analysis and Experiments

D.1 Checking for Overfitting

Overfitting happens when a model performs well on training data but poorly on new data. In this case, the training accuracy is 99.84%, while the validation accuracy is 98.74%. Since the difference is small, it suggests that the model is not overfitting.

D.2 Testing on New Messages

To test the model further, I provided some new messages and checked if the predictions were correct. The following table shows the results:

Input	Predicted Probability	Interpretation
Congratulations! You won a free iPhone. Click here!	9.6652335e-01	Spam
Hey, are we meeting tomorrow?	8.1521016e-08	Ham

Figure 2. Predictions on New Messages

The model correctly identified spam messages and ham messages with very low probability, meaning it should work well for real-world examples.

E. Lessons&Experience

- GloVe embeddings help the model understand words better by assigning similar meanings to similar words.
- CNN is good for short text classification, as it efficiently captured spam-related patterns.
- Even with fewer training epochs, the model performed well, showing that early stopping could be useful.
- Testing on new messages confirmed the model's accuracy, proving that it generalizes well to unseen text.
- We could compare CNN with LSTM to see if sequential relationships in text improve spam detection for future improvements.