L05 AWS MLU Lab Reflection Journal

Ryan Yauch

ITAI 2376

Lab Module 2

2/24/2025

Ryan Yauch
ITAI 2376

Lab 01: Processing Text

Learning Insights

This lab focused on some of the fundamentals of Natural Language Processing (NLP), which I think is one of the most important aspects of Machine Learning. The main concepts it covered were word clouds, part-of-speech tagging, stemming/lemmatization, and named entity recognition.

With some of my previous classes in Natural Language Processing, I already knew some of the basics of text processing, specifically I remembered stemming/lemmatization and part-of-speech tagging. However the other two concepts I wasn't fully aware of, with word clouds I've seen and used them before but wasn't fully aware what they were.

 I think word clouds help display how some models understand documents as a whole, specifically as a representation of weights within words. Words that are huge and take up a lot of space hold a lot of weight, and are going to influence models the most, just like how they catch our attention.

POS tagging I don't remember encountering before, but I think helped me better understand how models understand sentences as a whole. If a model is able to gather what word fits under what tag, such as a noun describing a person, place, or thing, they're able to properly understand the influence that word should have on the words around it. Since an adjective changes a noun, like "small dog" it understands the adjective "small" modifies the noun "dog".

Challenges and Struggles

This lab didn't give me any challenges or struggles, mainly since none of the cells had me write my own code or covered a concept I struggled to understand. I think lots of my previous understanding of NLP helped prevent any major challenges for this first lab.

Personal Growth

The main growth and understanding I gained from this lab came from understanding how machine learning models try to understand text documents or sentences as a whole. Word Clouds helping visualize the concept of "weight" or the importance of words, which although not directly related, it surprised me how much I think the two concepts connect.

Same with Part-Of-Speech Tagging, I still struggle sometimes with how these models understand "context" just from deciphering words, but piece by piece, such as now learning how they identify nouns and adjectives, I'm beginning to understand how they put together full context.

Ryan Yauch
ITAI 2376

Critical Reflection

Considering I already had experience with much of this lab, I don't feel like there's much new information to reflect on besides what I've covered in the other sections. There wasn't much experimenting or change to do in this lab, or many new concepts that made me rethink machine learning as a whole, only understanding a few concepts I didn't before. I think as a whole though, the idea of pre-processing text helps connect the idea that machines are very sensitive to the data you feed them, so reducing clutter and having a clear understanding of what you're giving a model is important.

Lab 02: Using the BoW Method

Learning Insights

This lab expanded on text processing in NLP models, mainly going over how to use sklearn, Bag-of-Word methods like Binary Classification, Word Counts, Term Frequency, and Term Frequency-Inverse Document Frequency. Like the last lab, I was familiar with some of these concepts as I've worked with sklearn, BoW, TF, and TF-IDF so I learned specifics of these concepts rather than a new concept entirely.

One of the first new concepts I thought was interesting was connecting Binary Classification and Sentiment Analysis. Since Binary Classification is able to pick words out of a document, it can be trained off specific negative words like "bad" or "poor" and analyze their context to see whether or not it contributes to the sentiment score.

Another new thing I learned was the existence of Inverse Document Frequency or IDF, since I was aware of TF, and TF-IDF but not IDF by itself. I thought about how it might work differently from how a Word Cloud is formed, since repetitive words get a lower score rather than getting bigger like a word cloud represents.

Challenges and Struggles

Like Lab 1, this lab didn't offer much of a challenge since it was all concepts I was familiar with, although there was one challenge section that had me write my own code. This was easy to solve though by going through the Binary Classification section and referencing the portion of code I needed to complete the segment.

Personal Growth

I wouldn't say I grew very much from this lab, I've used and deployed these concepts directly in classes before so I was even more familiar than the first lab. The only concept that I'd say sort of surprised me was the existence of IDF, and also how familiar these concepts still are to me.

Ryan Yauch
ITAI 2376

Critical Reflection

I still don't feel much to reflect on for this lab in particular, but I feel like I can connect this one with machine learning since it teaches about how text data is handled. Machines still aren't able to understand text as is, without being changed into a format it can read like a numeric format, which I think shows one of the main differences between human and machine learning.

Lab 03: Using GloVe Embeddings

Learning Insights

This lab focused on Word Embeddings, specifically going into GloVe embeddings which like the other two labs I've worked with before. Since I've used GloVe previously in other assignments, I didn't learn or gain many insights from how to load and use GloVe to make word vectors. Although a term I'm not too familiar with in the lab was cosine similarity.

I can only really find or think of deeper uses for cosine similarity beyond a search or recommendation system within an NLP model. I think it displays how machine learning models can have multiple functions, but beyond that I don't think I gained any insights.

Challenges and Struggles

I didn't have many challenges understanding or working with GloVe since I used it before, there were only two minor challenges with coding where I could just reference the code above to figure out what was asked.

Personal Growth

This lab I feel impacted me the least out of the three I've done so far, the concept of word embeddings as a whole fits in more with what I thought from the last lab, with how numerical data needs to be used for machines since they cant understand raw text data.

Critical Reflection

Since I don't feel like I learned much from this lab, I don't have much reflection or additional experiences and connections to add on.

Ryan Yauch
ITAI 2376

Lab 04: Introducing RNNs

Learning Insights

This last lab focused on Recurrent Neural Networks (RNNs), which was the neural network I focused on for my zoo assignment so I have a bit more knowledge from research. Most of the information in this lab was previously learned or gone over, but using all I learned to put together an RNN did feel like a learning experience. Importing the data/index, processing and transforming the text into numerical data, creating the vocabulary with tokenization, then using word embeddings takes a part from the previous labs so far.

Challenges and Struggles

There were no struggles with this lab since the challenge just had me messing with the number of epochs. The previous lab assignments pretty much filled in any missing context or misunderstanding I felt like I might have.

Personal Growth

Although this lab in particular might not have led to any growth, the combination of all the labs before makes me feel like I have a better understanding of this neural network than I had before.

Critical Reflection

This lab probably gave me the most to reflect over since it had a piece from the past three labs before it. Most of my understanding and growth came from the other labs, combined with my knowledge I already had I now feel like I have a better understanding of not just recurrent but other neural networks as well. With how data needs to be sourced, optimized, and transformed for a model then tagged and labeled, I understand more of how these machine learning models think and work.