

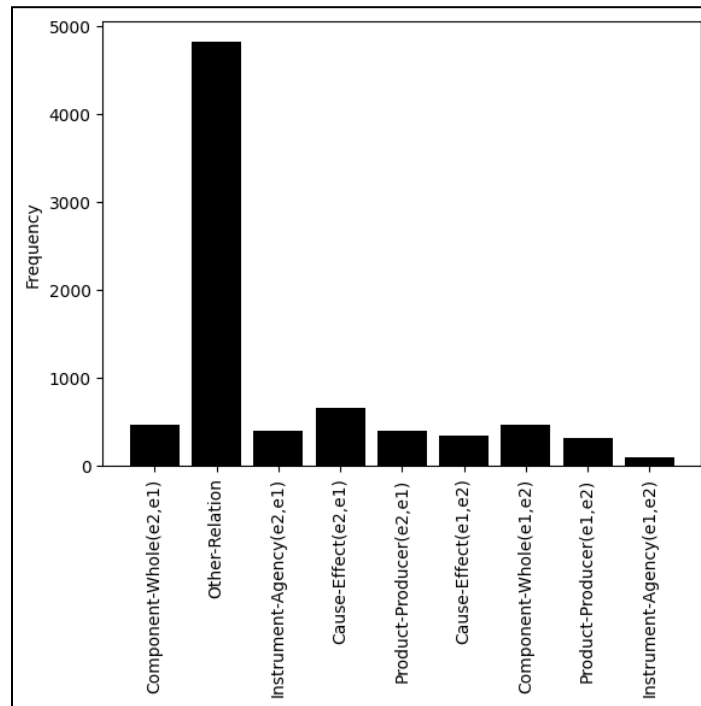
CS 6320
Natural Language Processing

Assignment 3
Semantic Relation Classification

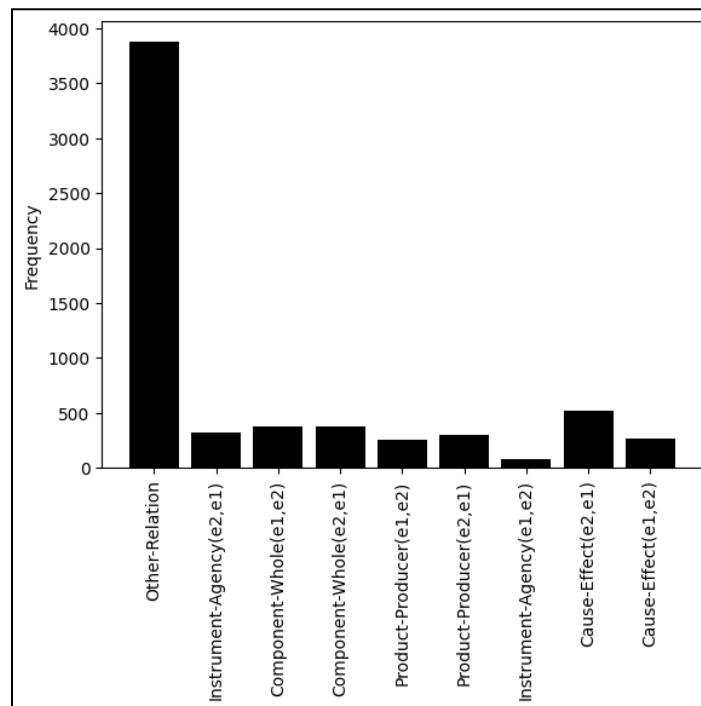
Prof. Dan Moldovan

Akash Biswal (axb200166)

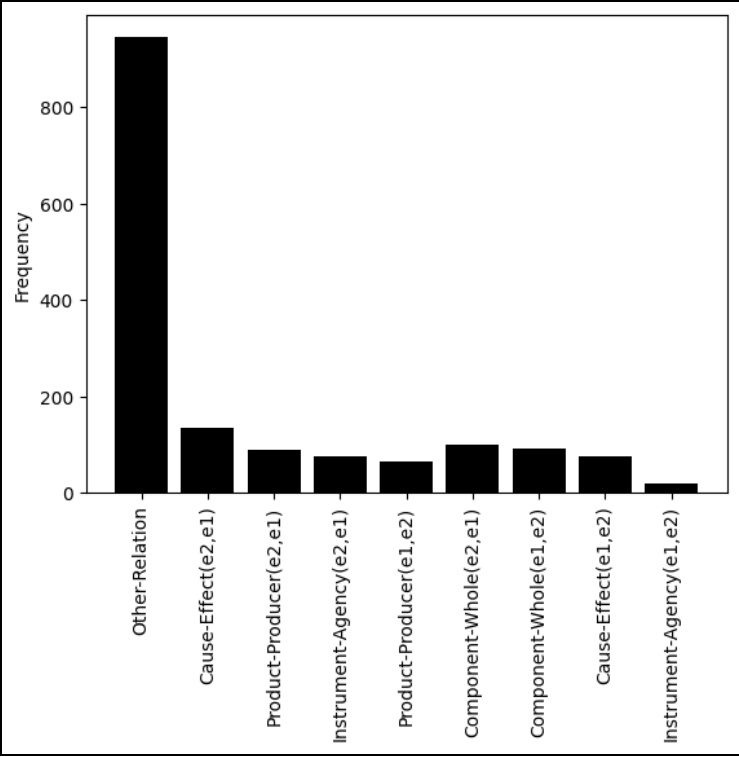
Results and Observations:



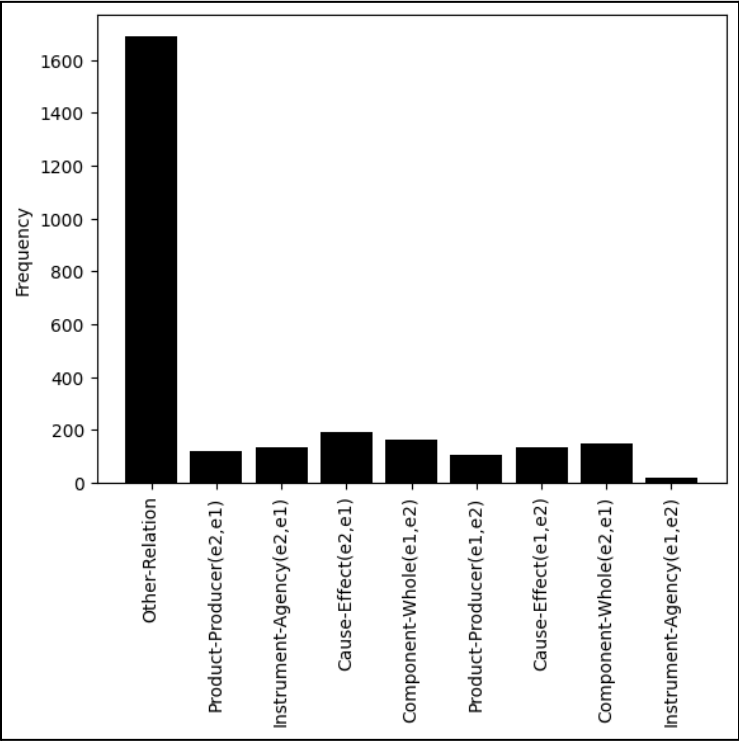
Frequency plot for the entire training dataset



Frequency plot for the train-set split



Frequency plot for the validation-set split

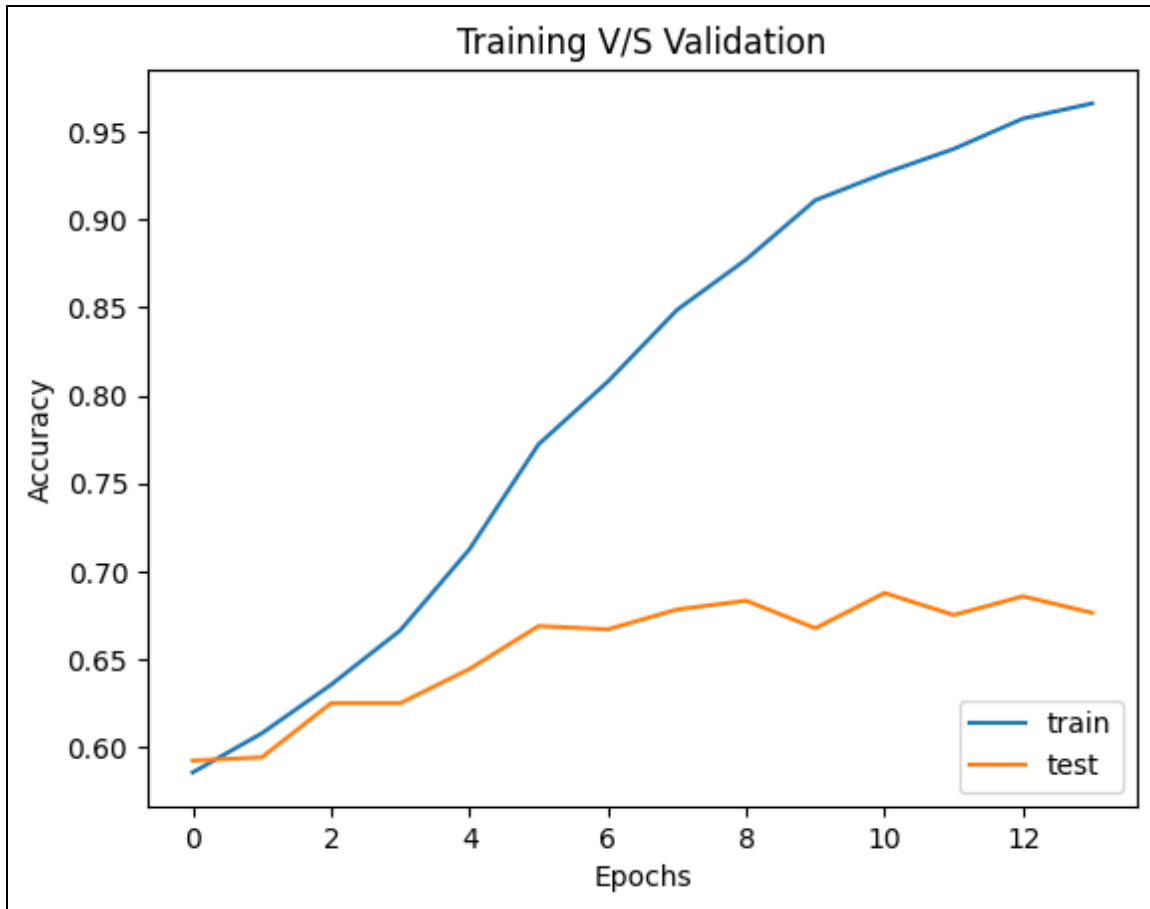


Frequency plot for the testing data

- As observed from the frequency plots of the data above, the data post-preprocessing is heavily skewed and unbalanced as there is a majority of “Other-Relation” labels.
- To tackle this issue the model is trained with balanced and unbalanced data.
- The data is balanced by assigning a higher weight to the minority class and a lower weight to the majority class to ensure that the model pays more attention to the minority classes when training.
- Below are the training v/s validation plots for both cases



Training v/s validation - balanced weights data



Training v/s validation - unbalanced data

- During training the training accuracy reaches about 96% and the validation accuracy caps at the high 60s, therefore the ceiling in terms of epochs was 14.
- Tuning other parameters and model dimensions did not produce different results.
- From the above plots, it's clear that the model performs better on validation data when the training data is balanced with weights
- We can observe this in the results and the model's accuracy on unseen test data when trained on balanced data is 56% compared to 55% on normal unbalanced data.
- The accuracy reached on testing sentences is 56% for my model.
- The manual error analysis is performed based on this model

Manual Error Analysis:

The model trained on balanced data is tested on fifty randomly picked sentences, below are the incorrectly predicted ones, and a manual error analysis on them.

- 1943: "The conference started with the general debate at which heads of delegations are delivering their national statements."
 - True tag: Component-Whole(e2,e1) Predicted Tag: Other-Relation
 - The golden tag is Component-Whole(e2,e1) but the predicted tag is Other-Relation, this could be because of insufficient information for the model to determine the relationship of Component-Whole(e2,e1). It could also be because the model was not complex enough to learn the relationship.
- 1122: "I first became acquainted with the book through an audio book I was given by a dear friend of mine a few years ago."
 - True tag: Other-Relation Predicted Tag: Product-Producer(e1,e2)
 - In our data the true tag is Other-Relation but the golden tag is Entity-Origin(e1,e2). The predicted tag is Product-Producer(e1,e2), this could be because the model may be associating some words with the Product-Producer(e1,e2) relation, and seeing "book" and "friend", it could have assumed a product and producer relationship.
- 6: "A child is told a lie for several years by their parents before he/she realizes that a Santa Claus does not exist."
 - True tag: Product-Producer(e1,e2) Predicted Tag: Other-Relation
 - The golden tag is Product-Producer(e1,e2) but the predicted tag is Other-Relation, this could be because of the same reason as id "1943", insufficient information for the model to determine the relationship of Product-Producer(e1,e2). It could also be because the model was not complex enough to learn the relationship.
- 595: "Cancer is the abnormal growth of cells that causes illness in the body."
 - True tag: Cause-Effect(e1,e2) Predicted Tag: Other-Relation
 - The reason for this mistake is also the same as id "1122". Insufficient information on Cause-Effect(e1,e2) relations for the model to learn the relationship.
- 1802: "For wader enthusiasts each spring provides an opportunity to discover a trip of dotterel."
 - True tag: Other-Relation Predicted Tag: Instrument-Agency(e2,e1)
 - The model could have "trip"(e1) as an instrument used by wader enthusiasts to discover the "dotterel"(e2) which leads to Instrument-Agency(e2,e1) being predicted.

- 240: "Bist du bei mir (en: "Be thou with me") (BWV 508) is an aria by a German composer named Gottfried Heinrich Stölzel."
 - True tag: Product-Producer(e1,e2) Predicted Tag: Other-Relation
 - The reason for this misprediction could be that the model was not trained on enough Product-Producer(e1,e2) examples to accurately learn the relationship, this is similar to the previous IDs for ex "595".
- 817: "Complementing the unique arsenal is a vibrant world that employs a cel-shaded animation reminiscent of a comic book."
 - True tag: Other-Relation Predicted Tag: Cause-Effect(e1,e2)
 - The predicted tag could be Cause-Effect(e1,e2) because the model may have seen "vibrant"(e1) as the reason/cause for the "animation"(e2). Since there could be relationships the True tag is appropriate.
- 305: "I read the report from Somalia on the agreement reached by faction leaders on the form of a future government that has been warmly welcomed."
 - True tag: Other-Relation Predicted Tag: Product-Producer(e1,e2)
 - The model could have associated a Product-Producer(e1,e2) relationship between "report"(e1) and "agreement"(e2), leading to misprediction. Since the association is general the True tag is correct.
- 2103: "The fortress has four towers corresponding to the cardinal points."
 - True tag: Component-Whole(e2,e1) Predicted Tag: Other-Relation
 - The model may have predicted Other-Relation as it is generic and there are also more training examples in the data when compared to examples for Component-Whole(e2,e1). The model could have seen a Model-Collection relationship and therefore predicted Other-Relation as well. The reason is again similar to previously seen IDs.
- 1657: "Unemployment and poverty cause abuse of children and serious situations such as the breakdown of families where the rights of the child are easily infringe."
 - True tag: Cause-Effect(e1,e2) Predicted Tag: Other-Relation
 - Other-Relation could be a predicted tag because it is generic and due to the possible lack of training examples for Cause-Effect(e1,e2). This sentence also displays the Content-Container relationship as "abuse" could be contained within "unemployment". Therefore the prediction was Other-Relation.
- 357: "After they finish wiring, electricians use ohmmeters, voltmeters, and oscilloscopes to measure the amount of electricity running through the system."
 - True tag: Instrument-Agency(e2,e1) Predicted Tag: Component-Whole(e1,e2)
 - The model could have failed to understand "electricians"(e1) and "ohmmeters"(e2) in the Instrument and Agency context. It instead saw a part-whole relationship between "electricians"(e1) and "ohmmeters"(e2) since the two words are of the same domain, i.e electronics

- 2711: "The umbrella frame is provided with a movable yoke at the upper end and a stationary yoke at an intermediate point."
 - True tag: Component-Whole(e2,e1) Predicted Tag: Other-Relation
 - One of the reasons could be the same as previous IDs, i.e. lack of training examples for the Component-Whole(e2,e1) relationship. The Model could also have seen a Member-Collection relationship between "umbrella"(e1) and "frame"(e2) which could have led to Other-Relation as the prediction.

- 914: "The wind lifted the umbrella, tipped the table, smashed some of the umbrella ribs, and ripped about half the umbrella seams."
 - True tag: Component-Whole(e2,e1) Predicted Tag: Other-Relation
 - The reason for this misprediction is the same as the previous sentence. Additionally, the model could have seen a Part-Whole relationship between "umbrella"(e1) and "ribs"(e2). Therefore Other-Relation is predicted.

- 1735: "The artwork and artifacts that make-up this exhibit give students the opportunity to physically and personally experience some of the things (and weight of the things) that "grunts" carried in Vietnam."
 - True tag: Component-Whole(e1,e2) Predicted Tag: Other-Relation
 - In the data, there could be fewer training examples for the Component-Whole(e1,e2) relationship, due to which the model fails to recognize "artifacts"(e1) as part of the "exhibit"(e2). It may have instead seen it as an Entity-Origin or Member-Collection relationship, therefore predicting Other-Relation.

- 2209: "This is the eye of the bull in the constellation Taurus."
 - True tag: Component-Whole(e1,e2) Predicted Tag: Instrument-Agency(e1,e2)
 - The model could have determined that the "eye"(e1) is being used as an instrument by the "bull"(e2). This could also be due to the nature of the words. Therefore an Instrument-Agency(e1,e2) relation is predicted.

- 2545: "All of the products grown by the defendant are grown either under glass or in special plots of ground."
 - True tag: Product-Producer(e1,e2) Predicted Tag: Product-Producer(e2,e1)
 - Here the model was unable to identify the directionality of the relationship. It could have been seen as the "defendant"(e2) producing the "products"(e1) when in fact the "products"(e1) are produced by the "defendant"(e2). Therefore the prediction is Product-Producer(e2,e1).

- 210: "Musicians used their real nails of the right hand to pluck the strings."
 - True tag: Component-Whole(e1,e2) Predicted Tag: Other-Relation
 - As seen in previous sentences the reason for this misprediction could be similar, i.e. lack of training examples. Another common reason could be a misinterpretation. The model could have misinterpreted the relation as

Part-Whole as “nails”(e1) are part of the “hand”(e2), which is also implied in the sentence. There we see a misprediction.

Erroneous predictions: 17

Summary:

- From the above results, we saw that the majority of the Erroneous predictions were the label “Other-Relation”
- The reason for this could be as mentioned in the analysis above, limited training examples for other labels and the majority of the labels being the generic class “Other-Relation”.
- Since in our data for simplicity we defaulted many relation classes to the “Other Relation” type, the model could have experienced semantic ambiguity in the sentences. This is possible by categorizing multiple types of sentences with different structures and meanings would create a chunk of ambiguous data.
- This semantic ambiguity along with the skewed data and lack of adequate training examples for all the labels could have caused the mispredictions to be “Other-Relation” as the model defaulted more to the general option.
- There were also other errors such as directionality errors and misinterpretations in general. These could again be attributed to less data. This is because, in the errors of this type, we observed that the misprediction was not very far-off from the true class and the reasons for the mispredictions were tiny nuances.
- Therefore, based on the understanding, observation, and manual error analysis, to improve the performance of the model, more data is required with more training examples such that the label distribution is balanced.

Link to the Colab Notebook:

<https://drive.google.com/file/d/1Qgydo7nOOc6-JqDMkA8LXAC34GYima-S/view?usp=sharing>
