

MAIS 202 Deliverable 2

Problem statement

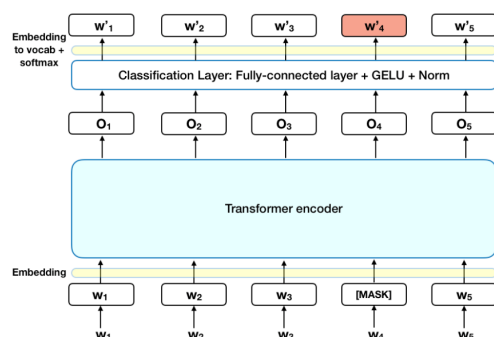
In our project, we aim to develop a user-friendly web application that offers a personalized recipe recommendation system. Users will input their cuisine preferences, dietary restrictions, type of meal, or preferred ingredients, which our system will use to generate tailored recipe suggestions that users can then rate and review to improve future recommendations. This recommendation system will employ a hybrid approach, combining collaborative and content filtering methods.

Data Preprocessing

We are using the Food Ingredients and Recipes Dataset with Images from Kaggle. The dataset includes 13,582 samples, each including the dish title, ingredients, instructions, and the dish image. Due to the large size of the dataset, we had to reduce the size in order to process it, starting with the first 8 recipes to test our model. The feature for our model will be the user's input of one ingredient. More preferences such as spice levels, cuisine types, and nutritional information can be added later on for user inputs. For our first try we will be using one feature.

Machine learning model

We used a BERT based model from the Hugging Face's Transformers library. The architecture of the model is shown in the figure below.



Regularization

We are planning to implement grid search in regards to hyperparameters. As for learning rate, we will use a rate of 0.01 and adjust as we go.

Preliminary results

Here is our work so far: [MAIS202 Deliverable 2 Model](#) however we struggled to determine the following things:

- We were unable to remove all the numbers and weird symbols from the original dataset, as well as make it into a string (for some reason the iteration was not working) so when we want to calculate cosine similarity and use a for loop for sentence2, we are unable to call on that due to iteration issues previously found.

- The highest result from the cosine similarity of the 8 recipes was 0.44526622 and the recipe did not include the inputted ingredient. The performance of this model requires significant improvement.

Next steps

We aim to work through cleaning up the data appropriately for further calculations and deal with issues in calculating cosine similarity as indicated above. We also hope to train our model and incorporate learning into it once the preliminary issues are resolved.