**Approach and Reasoning**

1. First, we make sure all the packages that we need: virtual environment and prodigy

* Check python version
* Install the virtual environment package
* Create a virtual environment
* Install prodigy with the license that we were given.
* Check if the prodigy is ready to go (then start the next step)

1. We started the annotating job (see the annotation guideline) with the unlabeled data. 20 annotations were done in the first round.
2. Then we looked at the annotated data, transferred the annotated data from db to jsonl. As jsonl would be the file that could be read through VScode for us.
3. And then we trained the model with first round of annotation and get the result in figure 1. The low score (0.14) showed that this model needs refining. Therefore, we referred to the two existing dataset that we had: Pydata and ChatGPT3.5.
4. We then combined the dataset (with the two existing dataset) and trained data again, create dataset combine01(Pydata) and combine02 (ChatGPT3.5).

We first started from combing the Pydata with our labeled data, and trained the data for the second round, and then got the results in Figure 2. With 3000 examples, we reached the score of 0.51. The model did perform better.

Then we combined our labeled data with the ChatGPT3.5 data, and trained the data for the third time, and then got the results in Figure 3. With 4400 examples, we reached the score of 0.53. The model did perform better.

When we combined the data, there’s no preference of the order of the two existing dataset. We started with the Pydata one, and then continued another combination with the ChatGPT3.5.

1. Then we corrected some annotations in the first round, based on the annotation guideline, to improve the precision of the model. Then we combined the corrected model and the Pydata, and ChatGPT3.5 data separately, creating combine06 (ChatGPT3.5) and combine07(Pydata).For combine 06, the score reached 0.55 (Figure 4); for combine07, the score reached 0.54 (Figure 5).
2. Based on the foundation of the sixth step, we combined all the data groups together and retrained. However, the results didn't turn out as expected; the scores didn't improve, and in fact, they even decreased (Figure 6). We analyzed that this was due to overfitting, leading to a decrease in scores. Therefore, we decided to consider the combine06 the best model, and we submitted it)

**Annotation Guidelines**

|  |  |  |  |
| --- | --- | --- | --- |
| Label | Definition | Examples (common) | Examples (edge) |
| Dish | Known food dishes, usually a prepared food item; unseparated. | Meatball, salad, fish soup, etc. | * Braising meat * Tuna salad sandwich: it should include all words for this item. |
| InGredient | Any food or substances that are combined to make a dish/recipe; individual parts of a dish. For some common dishes, like dumplings and noodles, if they are treated as part of the dish, then they are also ingredients. | Chicken, beef, veggie, etc. | * Cioppino; san marzano: unfamiliar ingredients for us. * ground mustard: includes ground as it is the form of the mustard. |
| equipment | Tools/apparatus/machinery/ utensils needed for cooking. | Kuhn Rikon peeler, microwave, measure cup, bamboo steamer, etc. | * Le Creuset, casserole: unfamiliar items for us. * cast iron skillet: includes every part of the equipment word, not only cast iron. * Towel: uncommon in the kitchen context, but if it’s for baking, then include it as a tool. |

**Appendix**

**Appendix\_1: Codes**

python --version #check the version of python

pip3 install virtualenv #install the virtual environment package

python3 -m virtualenv venv #create a virtual environment

python -m pip install prodigy -f [https://323A-43DE-964B-2B40@download.prodi.gy](mailto:https://323A-43DE-964B-2B40@download.prodi.gy)

python -m prodigy stats

python -m prodigy ner.manual hmwk-1-manual blank:en unlabeled.jsonl --label DISH,INGREDIENT,EQUIPMENT

python -m prodigy db-out hmwk-1-manual > labeled-class.jsonl

python -m prodigy train --ner hmwk-1-manual

python -m prodigy db-merge hmwk-1-manual,pydata-nyc combined01

python -m prodigy train --ner combined01 ./output

python -m prodigy db-in gpt3-5 .\gpt3-5-zeroshot.jsonl

python -m prodigy db-merge hmwk-1-manual,gpt3-5 combine02

python -m prodigy train --ner combined02 ./output

python -m prodigy ner.correct --help

python -m prodigy ner.correct dataset-corrected output/model-best unlabeled.jsonl --label DISH,INGREDIENT,EQUIPMENT

#corrected dataset + hmwk-1-manual (gpt3-5)

python -m prodigy db-out dataset-corrected > labeled-class.jsonl

python -m prodigy db-merge dataset-corrected,combined02 combined06

python -m prodigy train --ner combined06 ./output

#corrected dataset + hmwk-1-manual (pydata)

python -m prodigy db-out dataset-corrected > labeled-class.jsonl

python -m prodigy db-merge dataset-corrected,combined01 combined07

python -m prodigy train --ner combined07 ./output

#corrected dataset + pydata + gpt3-5

python -m prodigy db-merge hmwk-1-manual,pydata-nyc,gpt3-5,dataset-corrected combined11

python -m prodigy train --ner combined11 ./output

**Appendix\_2: Scores**

1. The scores after step 4 (training with the first round of annotations).

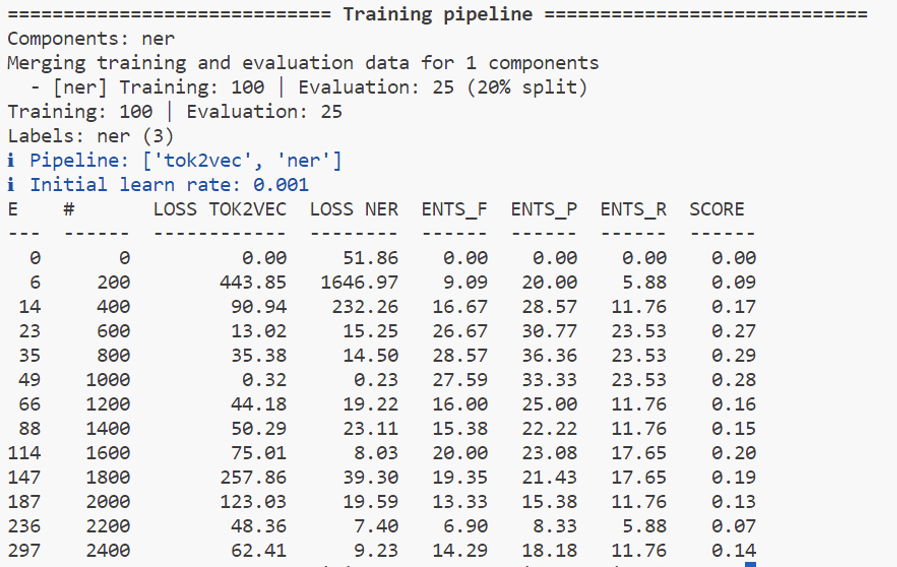


Figure 1

1. The scores after step 5 (training after referring to the pydata).

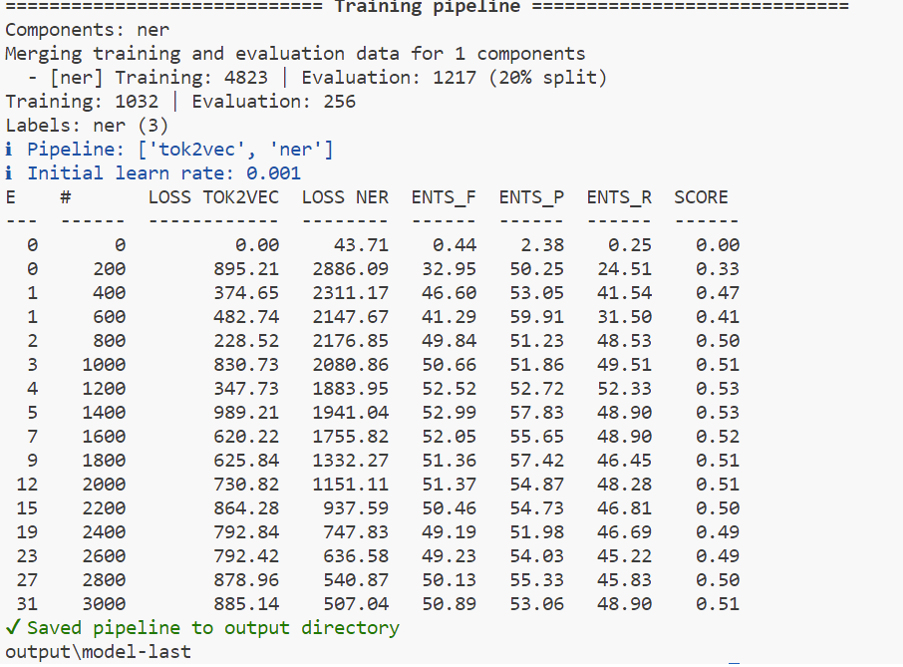


Figure 2

1. The scores after step 5 (training after referring to the ChatGPT3.5 data)

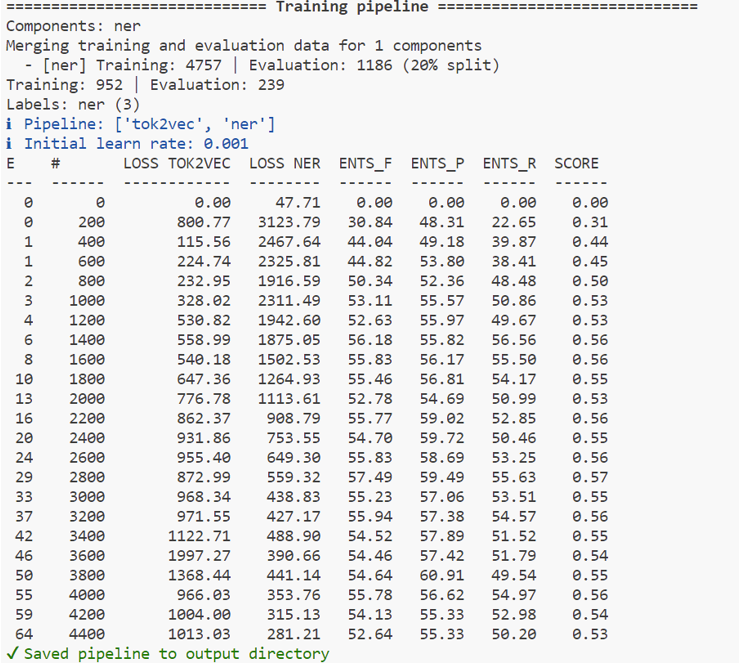


Figure 3

1. The scores after correcting some annotations and retraining it with combine06 (ChatGPT3.5), after step 6.



Figure 4

1. The scores after correcting some annotations and retraining it with combine07 (Pydata), after step 6.

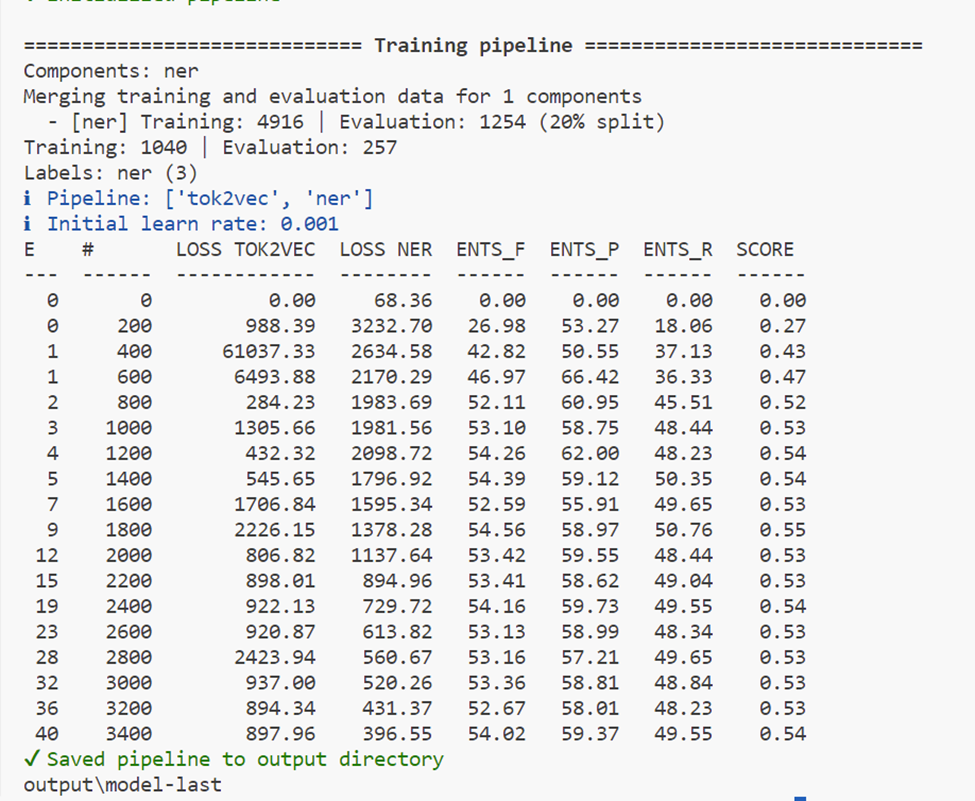


Figure 5

1. The scores after correcting some annotations and retraining it with Pydata and Gpt3-5, after step 7.

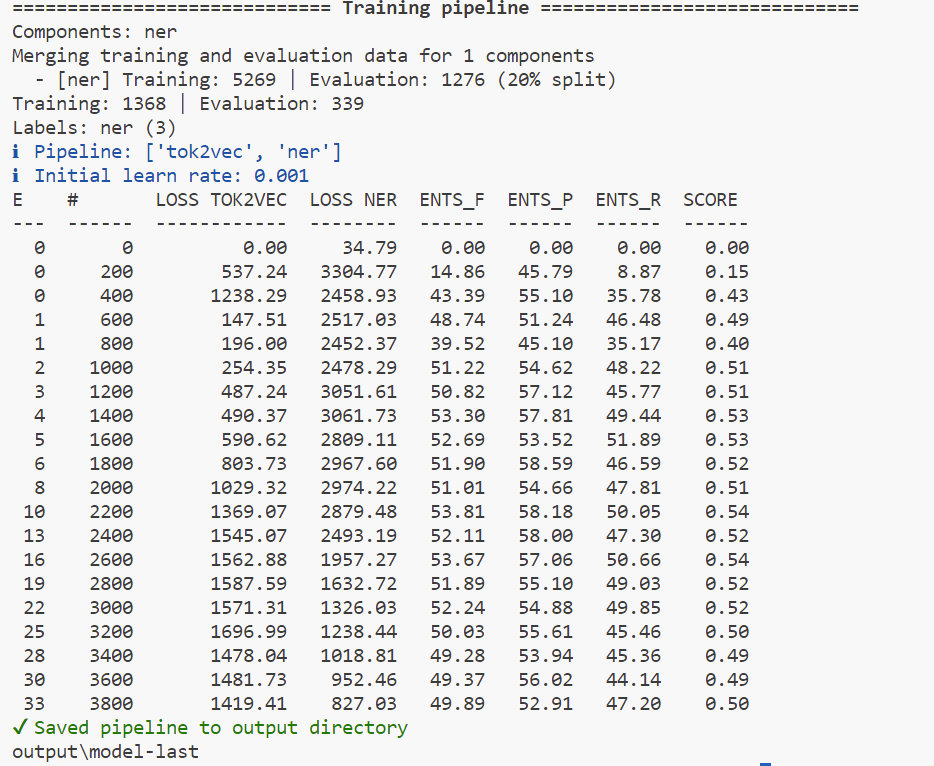


Figure 6