**Analysis For TF-IDF Homework**

Analysis 1: What answers get confused with each other most easily? What kinds of mistakes does this guesser make?

For question that answer is confused with guesses, its answer has a higher probability to be a specific name, e.g. name for a human or a place, that has no or few synonyms.

* Answer Examples: Mass\_(music), Byzantine\_Empire, Russian\_Empire, Sandman, Los\_Angeles, Lord\_Byron, Matthew\_Arnold, S-waves, Density, Georgia\_(country), Luminosity, Apollo

By counting the number of answers’ synonyms of three classes of results (close, hit and miss), it can be found that confused answer is more concentrate around 0, while answers in the other two classes tend to have more synonyms.

A graph with different colored lines

Description automatically generated

Figure Count of Synonyms for Close, Hit and Miss Results

According to the question, answer and guesses in miss category, it is found that the guesser is overfitting the training dataset. The f1 score on training is 0.97, which is much higher than F1 score on testing dataset (0.51).

From cases study, it is found that the guesser prefers to select guesses which words appear in the original question (see Figure 2). Except this, the guesser also fails to find answers based on multiple given information. Except for constraint of psychologist, there are other constraints for the person. But the guesser wrongly found the other psychologist (see Figure 3).

|  |  |
| --- | --- |
| A screenshot of a computer  Description automatically generated  Figure 2 Preference in Repeat of the Words in Question | A screenshot of a computer  Description automatically generated  Figure 3 Unable to Handle Multiple Constraints |

Analysis 2: Where does your buzzer make mistakes? How might you further improve the buzzer?

According to the coefficient of logistic regression classifier, the buzzer mainly relies on the confidence of tf-idf guesser. This means when guesser make a wrong guess but with a high confidence, the buzzer cannot figure it out.

For aggressive category of the evaluation result, while facing the guess that has the same words with the question, the buzzer is more likely to give a higher buzzing probability.

To further improve the buzzer, a different tf-idf guesser could be trained to generate a difference confidence score for this question & guesser. For example, using the Wikipedia page and doc as the data source to train an additional guesser.

Besides, a feature presents whether the word from guess is in the question, especially when guess is name and other specific definition could be added.