**MBTI MODEL EXPLANATION**

1. Utilize word embeddings to forecast the co-occurrence of words for the primary characters in the Harry Potter movies, as well as for Reddit users who have been attributed an MBTI personality.

Example with the word 'ever':

|  |  |
| --- | --- |
| **The probability that word2 occurs within the context window of word1** | |
| Harry Potter | Reddit users with the MBTI personality type INTJ |
|  |  |

1. Compare the probabilities of *each* word co-occurrence in the Harry Potter corpus with those of Reddit users, grouped by their MBTI type. The closer the probabilities are, the stronger the association between the Harry Potter character and that particular MBTI personality type.

Example with Harry and four types of MBTI personalities:

|  |  |
| --- | --- |
| **Word co-ocurrence with the word 'think'** | |
| *Harry* | *MBTI types* |
|  | INTJ |
| ESFP  *distance = |0.1|*  *distance = |0.3|* |
| ISTP  *distance = |0.1|* |
| ESFJ  *distance = |0.7|* |

1. After calculating the distances for each word in Harry's corpus, calculate the average. The MBTI type with the *lowest* average distance will be the personality associated with Harry.

Example with Harry and four types of MBTI personalities:



In the example, Harry is assigned to the MBTI personality type **ISTP**.