To detect the relationship between word sentiments of reviews and game discount, we proposed a simple word-to-vector regression of words(unigram) vs. discounted price based on the first 1000 common words across all the reviews in the training dataset. After we get the top 1000 common words from the training set, we then applied a word2vec algorithm based on TF-IDF (IDF we used here was taking logarithm), and a linear regression is implemented on word vectors and game discounts. Games with no discount history are regarded as discounted with 0% off.

From our regression, it is noticeable that coefficients with large positive effect are mostly with words of positive sentiment (such as “easier”, “improved”, and even some emojis like “♥♥♥♥♥”). However, coefficients with negative effects have no obvious sentiments, and most of them are just nouns such as “team”, “guns”, etc. Therefore, we, although preliminarily, conclude that sentiment of words can only positively affect the price.

A simple draft code is here. Note that this code is still naïve that it contains some words that are too general to games (like “games”) and also contains some game-specific words (such as “difficulty” with “souls”), so the prediction is not effective enough, but it still helped us captured the general trend.