

TF-IDF

In order to combine numerical data and textual data to get a feature that can predict the potential success and failure of products, we can use the clustering model generated previously in Model 1 to get numerical index. After clustering, the texts are analyzed to identify the characteristics of the clustered category. According to the word frequency outcome, we found it just corresponds to the classification of favorable reviews and unfavorable reviews. We marked the favorable review category as category 1 and the unfavorable category as category 2.

Then, observe the useful feature nouns, and then go back to the text to extract the reviews that mentioned this specific useful feature noun, and put forward some suggestions for improving the characteristics of commodity attributes.

To quantify the difference between the two categories of text, we introduce TF-IDF model

这里加TF-IDF 原理说明部分

...

Steps

- Combining the two types of text as a whole, a dictionary is constructed. Text vectorization is conducted according to this dictionary. Each comment corresponds to a vector, which contains the TF-IDF score calculated as above.
- At this point, separate the vectors according to the previous clustering, and the words with the largest difference in tf-idf between the two categories can be found, and the ranking is conducted from large to small. In particular, since the number of reviews in each of the category is basically the same, it's reasonable to use TF-IDF value difference to measure the differences directly. By subtracting the score of each word in one review category from the score of the word in the review of the other category, words with larger difference are the characteristic word of that category.
- According to the characteristic words extracted from the above methods, the first three characteristic words related to commodity attributes were extracted in order of the degree of specificity (that is, the degree of difference between the two reviews on this word), and we got the table.

Analysis

- Now we take the first feature of class 1, space, and revert to its original review, looking for advice on microwave space usage.
(这里放表格space)
 - Some people love its size because of some special uses, while some others are actually saying something positive while teasing a little about how the microwave is too big and taking up too much space.
- Similarly, we now extract the first feature word of the second category: keypad. Expectedly, we can find improvements on keypad
(这里放表格keypad)
 - As expected, microwave keypad section didn't do well, and many people commented negatively on it. For example, the plastic film on the keypad, shown in the chart above, bubbled after a few months. It didn't do well in typing numbers as well.

- According to these two key words, we draw a conclusion that users with more positive evaluation think the space occupied by microwave oven is too large and unreasonable, while users with more negative evaluation think the keypad is not good. In other words, the defects caused by the space problem will not strongly affect the user's evaluation, while the defects of the keypad will ruin the customer's view of the microwave. If Sunshine is going to enter the online market of microwave ovens, the first thing it should improve is the keypad. Then the improvement of space factor is considered.

Advice

(这里放表feature_words)

For the other two data sets, the same analysis can be carried out to provide advice to sunshine company in the online market of these three commodities

- Microwave: Improve the keypad. As for the size of the space, the designer can make both big and small to meet different needs.
- Hair dryer: Keep the hair dryer high power, but improve the design of the nozzle as well as the cord material. Ensure high power and fast blow-drying while also considering how to avoid damage to the skin.
- Pacifier: The pacifier is intended for children, so its characteristic of softness should be maintained. Some people also use it as a gift to give to the parents who need it. The accessories of the existing market for baby pacifiers are not as good as they should be. Both strap and bag have been criticized for being unusable and inconvenient. In addition, better materials should be used in making baby pacifiers.