Summary

With the popularization of Internet technology, online shopping has gradually become an important way for people to consume. In Amazon's online marketplace, customers can rate and review after purchase, and these ratings and reviews help other customers make purchasing decisions.

First, we screened the indicators in the three data sets provided to obtain seven important indicators related to product review analysis. For the review indicators, we used a classifier based on Naive Bayesian theory to perform sentiment analysis on the review indicators. In order to achieve the purpose of quantifying indicators and considering the importance of review indicators, we further extracted the information in the reviews, and based on the reviews and its relationship with time, we obtained the review length, review density, and total review indicators.

Secondly, we use the analytic hierarchy process to establish an SRR evaluation model, and analyze the comprehensive score of a single product based on 9 related indicators such as star ratings and reviews. In the SRR evaluation model, these basic indicators are combined to obtain three effective indicators that reflect product quality: enthusiasm for comment, credibility of comment, and product popularity. Combined with time, a comprehensive evaluation level of the product is obtained. Using the SRR model, we can score a product.

Then, in order to reflect whether the rating of star ratings can lead to more reviews, we conducted an analysis of variance on the rating star index and the comment density index, and concluded that there is a significant relationship between the rating star rating and the comment density. Different star ratings have an impact on the number of reviews. Considering that triggering is a long process, we analyze the correlation between the amount of individual reviews and the total of reviews next month. It turns out that 1-star reviews can trigger more reviews, while 5-star reviews can't.

Then, we first extracted high-frequency words as features and calculated the TF-IDF of 12 words, of which 8 words were related to quality description and 4 were unrelated words. Then, a correlation analysis is performed between their TF-IDF and star ratings, and the results indicate that there is only a correlation between these specific descriptors and star indicators.

Finally, we have performed some analysis on the sensitivity and advantages and disadvantages of the model.

Keywords: sentiment analysis, analytic hierarchy process, analysis of variance, TF-IDF