Algorithms and Data Structures

The Sias leasing platform is a web-based leasing platform that shares many commonalities with popular e-commerce platforms, including recommendation algorithms, search algorithms and data structures. For example, when a user searches for a Canon digital camera product, its search algorithm can compare the relevance, recent hits, and seller's reputation to the order of the products; the recommendation algorithm can provide similar alternatives while you are browsing the product, such as Nikon camera; the data structure stores the attributes and relevance of the product.

In the product recommendation system of the Sias leasing platform, we mainly use the collaborative recom Collaborative filtering is mainly divided into two categories, one is based on users and the other is based on commodities:

User-based collaborative filtering recommendation: This method mainly focuses on users, and divides users into multiple categories according to the degree of similarity. The same user groups have higher similarities, so their needs are also very similar.

The other is product-based filtering recommendation: This is the recommendation system that we generally understand. When users search for something, we spread according to the products that users search for, and recommend more related products to users. At the same time, sort out these similar products or other related products. Then get the final recommendation. mendation system to recommend products.

In the collaborative filtering algorithm, the focus is on the basis of the calculation - the user or the product, will get different results; on the other hand, the similarity calculation method. The calculation methods of similarity mainly include the following: Consine similarity, Pearson correlation coefficient, and Jaccard similarity.

Cosine similarity: Let commodity A be represented as vector A, and commodity B be represented as vector B. Divide the inner product of vectors A and B by the modulus of A and the modulus of B, denoted as A\*B/|A|\*| B|, where vectors A and B are vectors in the space that can indicate the degree of correlation. The higher the degree of correlation, the closer the distance, the higher the Cosine similarity, and the highest value is 1.

Jaccard similarity: The modulus of the intersection of A and B divided by the union of A and B, similar to the confidence described above, except that the division of A and B is not a separate one. A or B, this operation expands the dividend and expands the representation range of the result, which in turn increases the robustness of the result.

Pearson correlation coefficient: The Pearson coefficient can be regarded as the first two sets of data after the Z-score processing, and then the product of the two sets of data divided by the number of samples Z-scores generally represent the normal distribution, the distance of the data from the center point. The variable is subtracted from the mean and divided by the standard deviation. The standard deviation is equal to the square of the variable minus the mean, divided by the number of samples, and finally squared.

In the Sias leasing platform, in order to accurately describe the differences between commodities, as well as to improve the association of tables in the database, and the query association between multiple tables, we have established the following data structure:

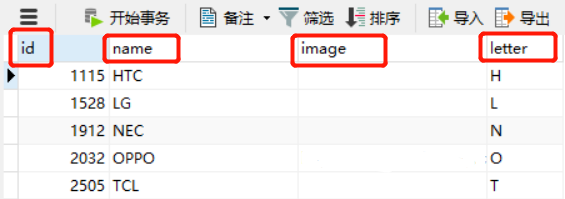
\_brand :brand table

Id : brand id (primary key)

Name : brand name

Image : Brand image address

Letter : the initials of the brand



\_categoory :product tax table

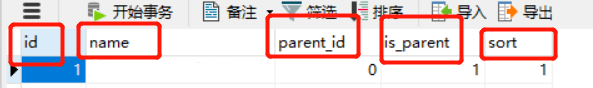
Id: product category id (primary key)

Name : the name of the product

Parent\_id : parent class id, top level field is filled in 0

Is\_parent : whether it is a parent node, 0 is no, 1 is yes

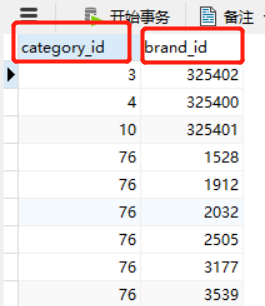
Sort : sort index, the smaller the lower the front



\_category\_brand : intermediate table of product categories and brands

Category\_id : product category id

Brand\_id : brand id



\_Sku (inventory unit): Each item of the SPU product set that is subdivided by specific characteristics

Id: sku table id (primary key)

Spu\_id : spu id

Title : the title of the item

Images: pictures of goods, multiple pictures separated by ','

Price : sales price

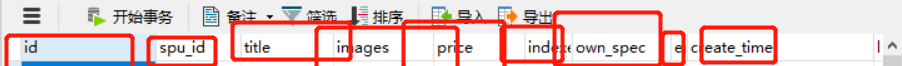
Index: the corresponding subscript combination of the unique specification attribute in the spu attribute template

Own\_spec: Sku's unique specification parameter key-value pair, json format, use linkedHashMap when deserializing, guarantee order

Enable : whether it is valid, 0 is invalid, 1 is valid

Create\_time : add time

Last\_update\_ time: last modified time



\_Spu (standard product unit): a set of products with common attributes

Id: spu id (primary key)

Title : title

Sub \_title : subtitle

Cid1: primary directory id

Cid2: secondary directory id

Cid3: three-level directory id

Brand\_id : the brand id of the product

Saleable : whether it is on the shelf, 0 off the shelf, 1 on the shelf',

Valid : whether it is valid, 0 has been deleted, 1 is valid

Create\_time : add time

Last\_update\_time : last modified time



The relationship between the table and the table:

There are multiple commodities (spu) under one brand, one-to-many relationship';

Categories and commodities (spu) are one-to-many relationships, and categories and brands are many-to-many relationships';