

**TravelSmart: Hybrid for Personalized Tourist Recommendations**

*Student*

**Khong Duy Tung**

**Nguyen Manh Hai**

**Dinh Duc Thanh**

*Instructor*

**Phan Duy Hung**

DAP391m: Artificial Intelligence - Data Science Project

FPT University

Hanoi, May 2024

***Abstract: In the age of digital tourism, providing tailored travel advice is essential to improving visitor experiences. In this work, we introduce TravelSmart, an intelligent travel recommendation system that uses collaborative filtering to provide personalized recommendations for tourist attractions.***

***Keywords:******travel recommendation system, collaborative filtering algorithms, Personalized tourist attraction recommendation.***

**1. Introduction**

***1.1. Problem and Motivation***

In recent years, the rapid advancement of artificial intelligence (AI) has dramatically transformed many different industries, including the travel sector. The large amount of travel information available online these days is both an advantage and a challenge: travelers are likely to become bored when looking through information on media platforms, which might prevent them from making the best decision. Recommendation systems have become an effective tool for improving user experience by offering customized recommendations based on unique tastes and behavior in order to address the problem. Based on the application of the big data recommendation system in the movie app [[1](https://ieeexplore.ieee.org/document/10102110)], the method of constructing a big data recommendation system suitable for the industry such as e-commerce, tourism, medical treatment, and education is summarized, and the significance of the research of the big data recommendation system is analyzed. The purpose of this paper is to demonstrate how recommender systems can automatically organize product arrangements based on users' browsing behaviors. These systems analyze patterns and preferences to tailor the display of products to individual users. Although recommender systems have achieved substantial progress and have become integral to many online platforms, there are still areas that need improvement. Enhancements are required to refine their accuracy, adaptability, and overall effectiveness to ensure they meet and exceed customer expectations, ultimately leading to higher levels of user satisfaction.

***1.2. Related Works***

The importance of this topic is highlighted by the significant growth in the tourism sector. In 2023, Vietnam's tourism industry aims to welcome 8 million international tourists. By August 2023, the number of international tourists had reached 7.8 million, nearly reaching the set target [[2](https://www.vietnam.vn/du-lich-viet-nam-nam-2023-vuot-xa-ky-vong/)]. Since July, the number of monthly international visitors has continuously exceeded 1 million.

The results of some research have been found such as: In Personalized Tourist Attraction Recommendation [[3](https://ieeexplore.ieee.org/document/8832961)], tourists face a large number of tourist attractions, and spend a considerable amount of time and energy to select satisfactory tourist attractions. By calculating the relationship between tourists and attractions tags, tourist attractions and attractions tags, a user interest model is constructed. Then, according to the user interest model, the interest degree of the new attraction to be recommended is predicted, and finally the tourist attraction recommendation set is generated. As for the data, in [[4](https://ieeexplore.ieee.org/document/8938546)] Collecting the Tourism, authors adopt the global Extract Transform Load (ETL) process. This research aims to show how the supporting data including the database structure, the data architecture, and the data representation in the tourism recommendation system are collected. The result of the research can be used for the need of further research - particularly on the tourism recommendation system considering the change of weather or traffic condition and its effect on making decisions when planning a travel or a tour.

The difference between the travel recommendation and traditional recommendation methods is that travel data has its own unique characteristics [[5](https://sci-hub.se/10.1109/icosec49089.2020.9215404)]: first, the cost of the travel varies greatly according to the range accepted by different people; second, the time limit of travel is also due to the nature of the work of different tourists and habits are also very different. Hence, based on the literature review, the current models can be summarized as follows

Summary, research on personalized attraction recommendations and data collection processes for tourism systems offers insights systems offers insights for optimizing recommendations.

***1.3. Contribution***

Future studies could explore adapting recommendations based on dynamic factors like weather and traffic conditions to enhance travel planning decisions in response to the industry’s rapid growth. This study aims to tackle the hurdles encountered in travel recommendation systems, aiming to deliver tailored and impactful suggestions. Reviewing existing literature underscores pervasive issues like information overload, standardized experiences, and shortcomings in current algorithms. Prominent research highlights the efficacy of hybrid filtering methods in achieving both precision and variety in recommendations. Collaborative filtering's potential [[6](https://www.researchgate.net/publication/373680697_Intelligent_Travel_Recommendation_Systems_for_Transforming_Nigeria%27s_Tourism)], content-based features, machine learning-based enhancements, and hybrid recommendation systems are investigated.

Despite the complexity of analyzing extensive travel data, this project will be instrumental in ushering in a new era of travel planning for those who want to travel independently. The main contribution of this paper is to create an app that focuses on collaborative filtering and tagging in a travel recommender system because [[7](https://ieeexplore.ieee.org/document/8832961)] collaborative filtering recommendation technology is the most popular and widely used recommendation technology in current recommendation algorithms. The main idea is to divide users into different user groups according to their similar interests and hobbies, and recommend a particular user according to the preference habits of other users in the user group to which he belongs. It provides personalized recommendations by analyzing user interactions and item attributes. This approach improves recommendation quality, addresses the cold start problem, and enhances serendipity. It also offers flexibility in recommendation strategies, handles data sparsity, and supports diverse user preferences. The process is provided into seven steps: Data collection, Preprocessing, Collaborative filtering, Tag-based filtering, Combine both approaches, Model training, Deployment and Evaluation. By following these steps, you can build an effective travel recommender system using both collaborative filtering and tags.

**2. Methodology**

***2.1. Method of data collection***

In this section, the recommender system with all three of these modules is also extensively used to promote tourism. In the tourism recommender system the main focus is towards the evaluation of user’s physical and psychological functionality levels. In tourist destination recommender, The user profile is used to create and recommend places based on previous visitor experiences using collaborative filtering. The personal digital service in the recommender system is used to explore the most current concept of user groups. People make decisions in social activities like traveling and shopping. The primary distinction between a single user-based suggestion and a group-based recommendation is the methods used to meet the targeted user's interests and preferences. When users travel in a group rather than individually, the trip recommender system learns the group preferences from the individual choices specified in their profiles.

***2.2 Personalized recommendation algorithm***

***2.2.1 Collaborative Filtering Recommendation Algorithm***

***2.2.1.1 Collaborative Filtering Theory***

Collaborative filtering recommendation algorithm discovers users' preferences by mining users historical behavior data, divides users into groups based on different recommended similar resources. Collaborative filtering recommendation algorithms are divided into two categories: user-based collaborative filtering and item-based collaborative filtering.

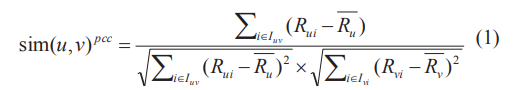
(1) User-based collaboratIve filtering: User-based collaborative filtering algorithm discovers user's preferences for goods or content (such as purchase, collection, content review or sharing) through user's historical behavior data, and measures and scores these preferences. The relationship between users is calculated according to their attitudes and preferences for the same goods or content, and Resource Recommendation is made among users with the same preferences. In short, if both user A and user B have purchased items x,y and z, then both user A and user B belong to the same category of users. Therefore, user B can also be recommended to buy items purchased by user A.

(2) Item-based collaborative filtering: Similar to the user-based collaborative filtering algorithm, the item-based collaborative filtering algorithm obtains the relationship between items by calculating the scores of different users for different items. Users are recommended similar items based on the relationship between items, and the score here represents the user's attitude and preference towards products. In short, if user A buys both item a and item b, then a and b are highly correlated. When user B also buys item A, it can be inferred that he also needs to buy item A.

***2.2.1.2 Collaborative Filtering: Traditional Similarity Calculation Method***

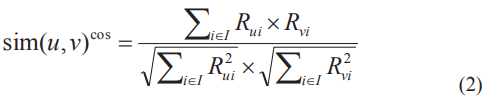
[[11](https://sci-hub.se/10.1109/iccsec.2017.8447000)]The similarity measures in the cooperative filtering algorithm are usually based on the Pearson correlation coefficient (PCC), cosine similarity (COS), Adjusted Cosine Correlation. Assume that the user-item scoring matrix is R(m,n),where m is the number of users and n is the number of items. R(ui) represents the user rating on item i.I(uv) is the user u and user v common evaluation of the project.

1. **Pearson Correlation Coefficient (PCC)**



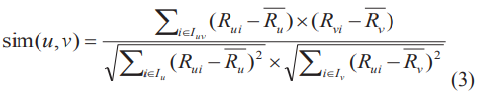
Where R(u) is the average score of user u in I(uv) and R(v) is the average score of user v in I(uv) .

1. **Cosine Similarity (COS)**



Where R(vi) represents the user’s rating on item i.

1. **Modified Cosine Correlation**

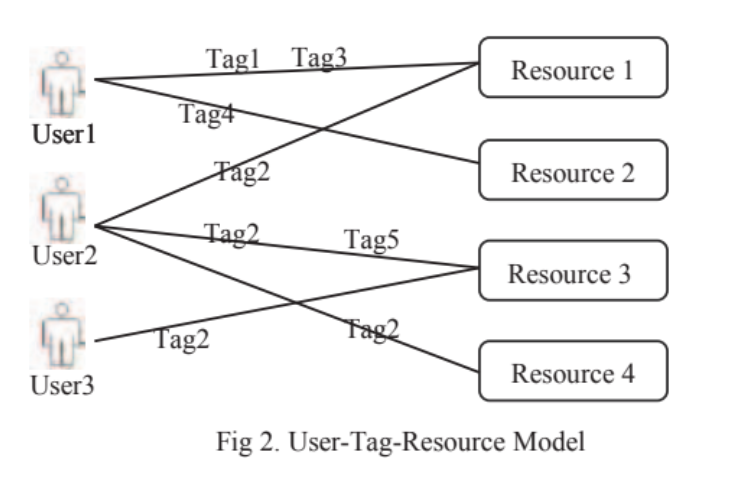


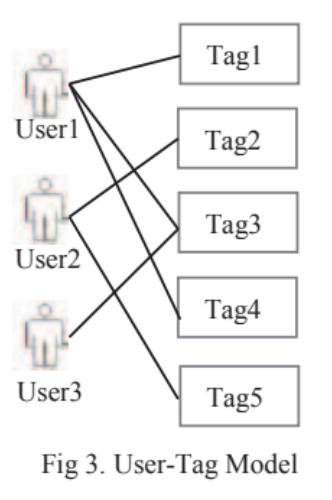
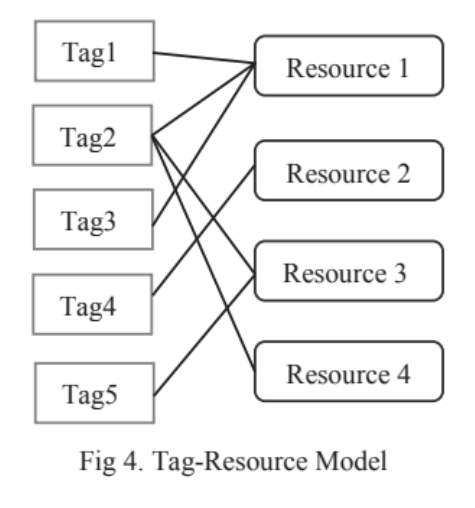
The relevant parameters are as described above.

***2.2.2 Content-Based Recommendation Theory***

A tag is a non-hierarchical structure used to describe information and can be used to describe the semantics of an item. Tourists can use the attraction tag to describe their views on tourist attractions, so the attraction tag is the link between tourists and tourist attractions, and is also an important data source for responding to tourists' interests. Each tourist can mark multiple tourist attractions at the same time, and the set of attractions tagged by each tourist can potentially show the interest and hobby of the tourist. An attractions tag can mark multiple tourist attractions at the same time, and the tag can reflect the common features of these attractions .An attractions tag can mark multiple tourist attractions, which can reflect the common characteristics of these tourist attractions. An attractions tag is used by multiple tourists, indicating that these tourists are interested in the same type of tourist attractions and have the same preferences.

A complete tag system usually consists of three elements: user, resource, and tag, as shown in Figure 2. In this paper, it is shown as the relationship among tourists, tourist attractions and attractions tag. Different from the collaborative filtering recommendation algorithm based on tourists' two-dimensional rating matrix for tourist attractions, the personalized recommendation algorithm based on tags relies on the three-dimensional relationship between tourists, tags and resources. It is necessary to decompose the three-dimensional matrix into two two-dimensional matrices, namely, the user-tag matrix and tag-resource matrix, as shown in Figure 3 and 4, and complete the recommendation process by replacing the tourist rating matrix with the importance data of tag.

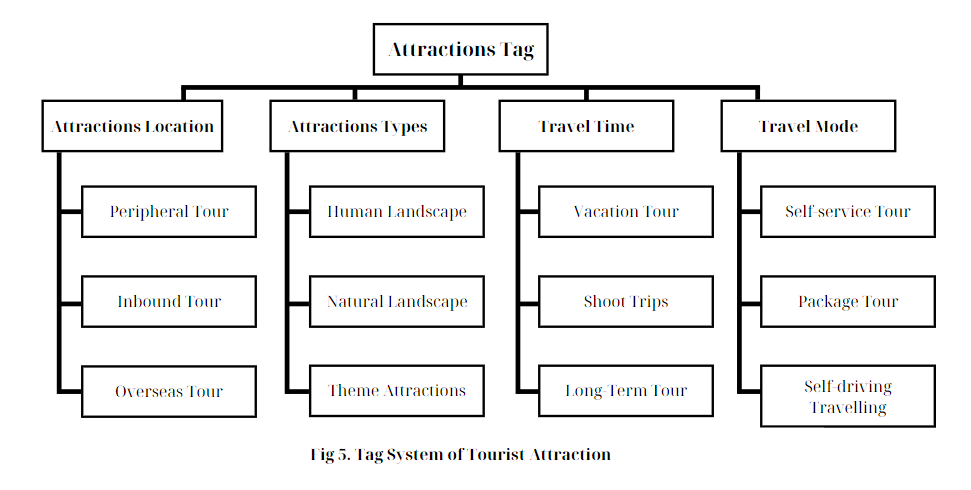


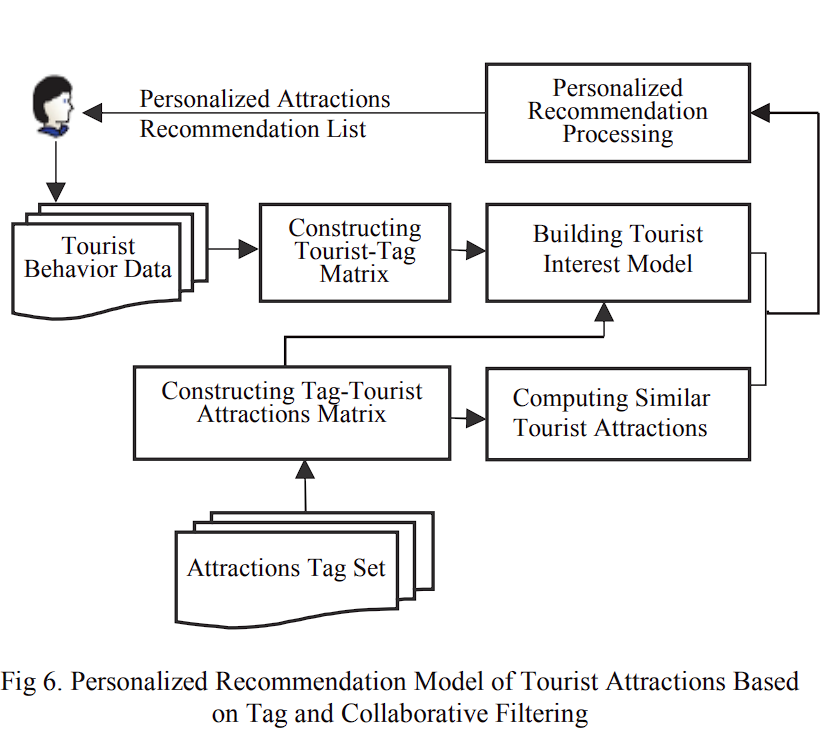
***2.3 Tag System of Tourist Attractions***

Another important part is the Tag System of Tourist Attractions. Tags typically consist of simple words or phrases, and the creation of a comprehensive tagging system generally falls into two categories. One approach involves utilizing social tags, which boast diversity and can capture users' personalized information. However, this method is susceptible to tagging errors and lacks semantic accuracy. Alternatively, experts can develop a tagging system for users to employ. In this paper, we employ the expert interview method to establish such a tagging system[[8](https://www.researchgate.net/publication/221140928_Personalized_recommendation_in_social_tagging_systems_using_hierarchical_clustering)].

[[9](https://sci-hub.se/10.1109/ccdc.2019.8832961)] This paper describes the development of an expert-designed tag system for categorizing tourist attractions. Through interviews with tourism enthusiasts, the researchers identified four main attributes related to evaluating attractions: location, type, time, and travel mode. Based on these factors, they constructed a 4-category, 12-tag system for classifying tourist destinations, as shown in Figure 5.

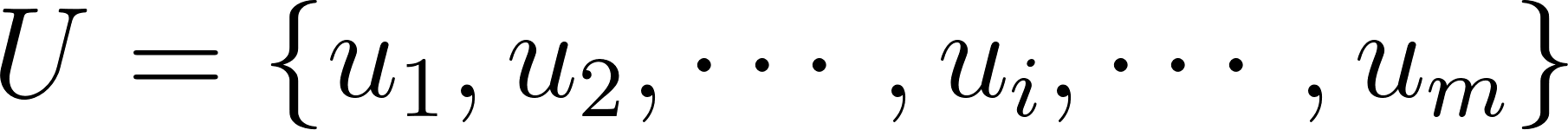
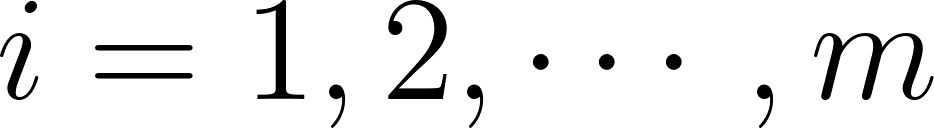
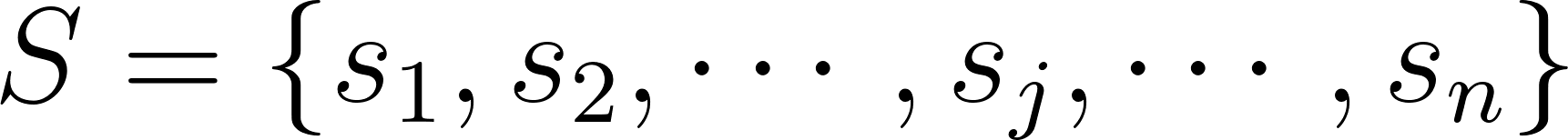
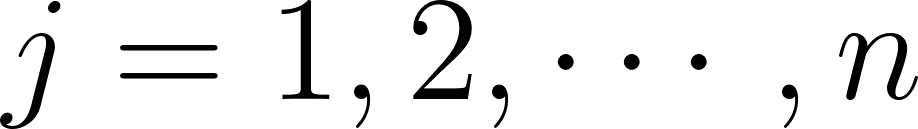
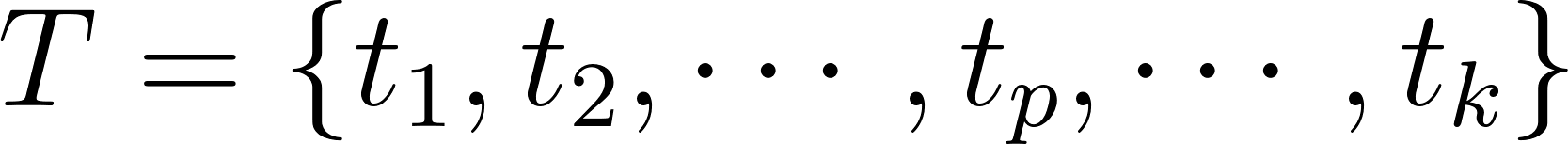
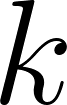
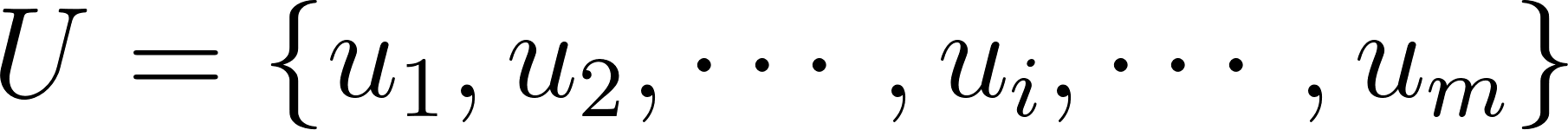
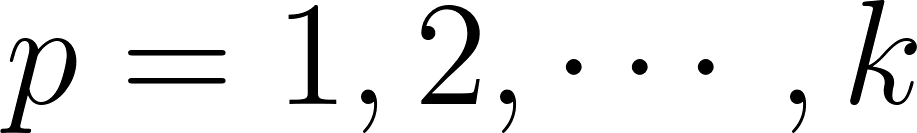
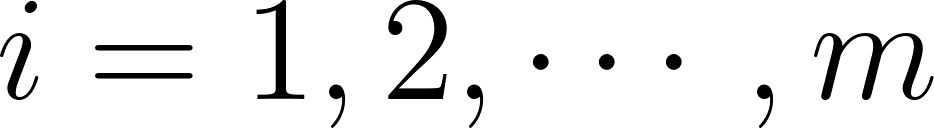
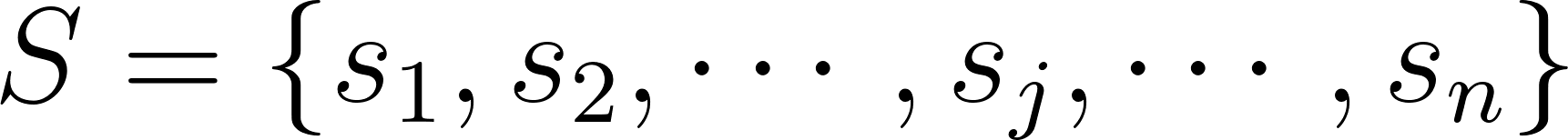
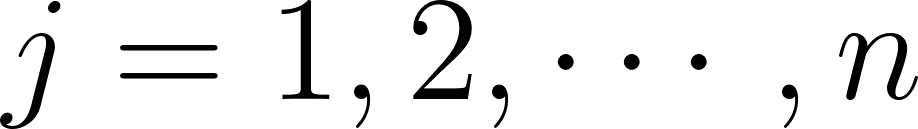
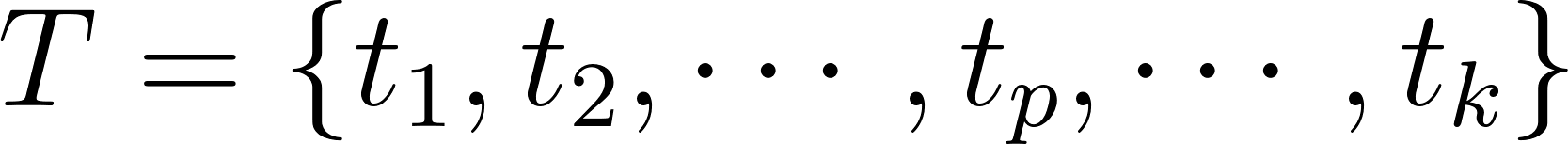
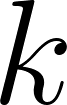
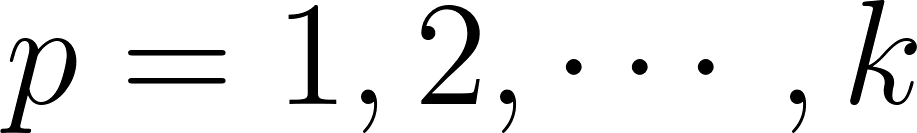
****

[[10](https://sci-hub.se/10.1109/ccdc.2019.8832961)] Different from the traditional collaborative filtering recommendation algorithm, the tag-based personalized recommendation of tourist attractions integrates the tag of tourist attractions into the personalized recommendation model, and uses the tag of tourist attractions as the intermediary to mine the interest of tourists from both aspects of tourists and tourist attractions. In order to get the TOP-N recommendation for tourists, it is necessary to calculate the preference degree and similarity degree of tourists for attractions tag. As shown in Figure 6.



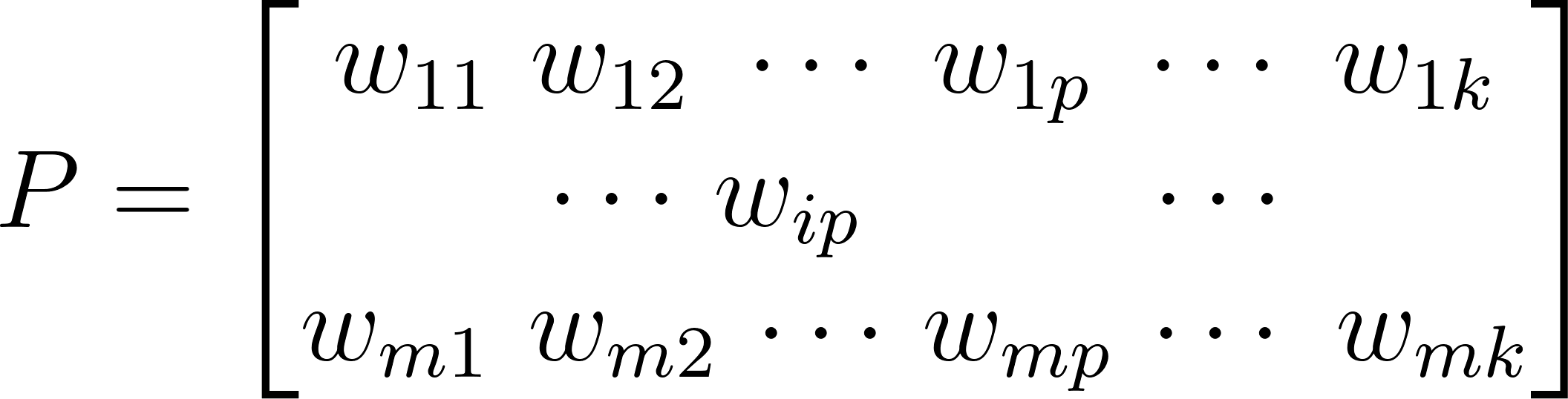
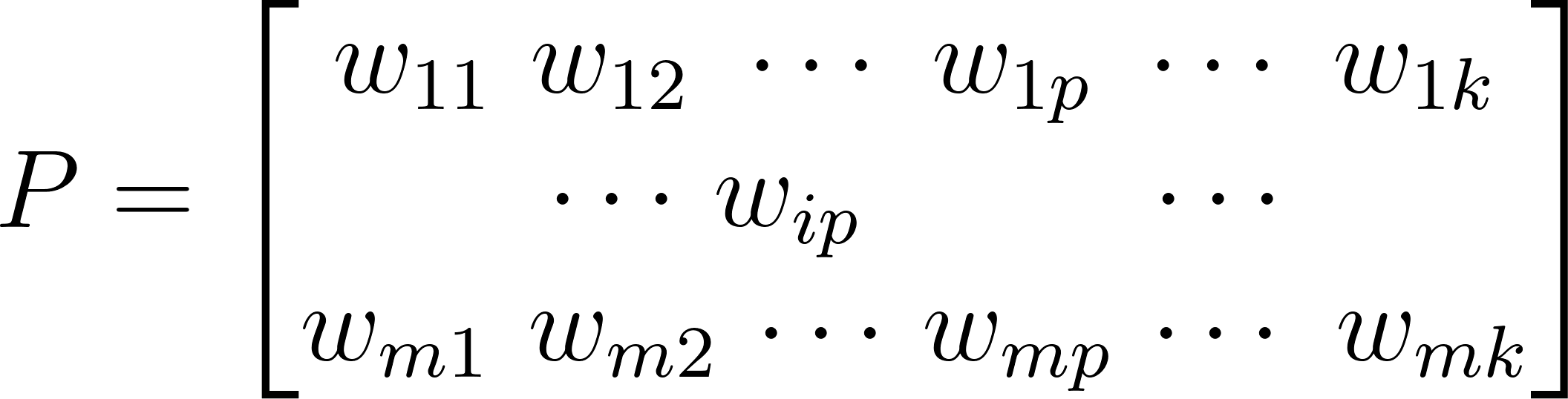
***2.4 Constructing the Model of Tourists' Interest in Tourist Attractions***

The tourist choice model for tourist attractions is derived from the multi-dimensional link among tourists, attraction tags, and tourist attractions. The tags linked to tourist attractions contribute to the features of the tourist preference model.

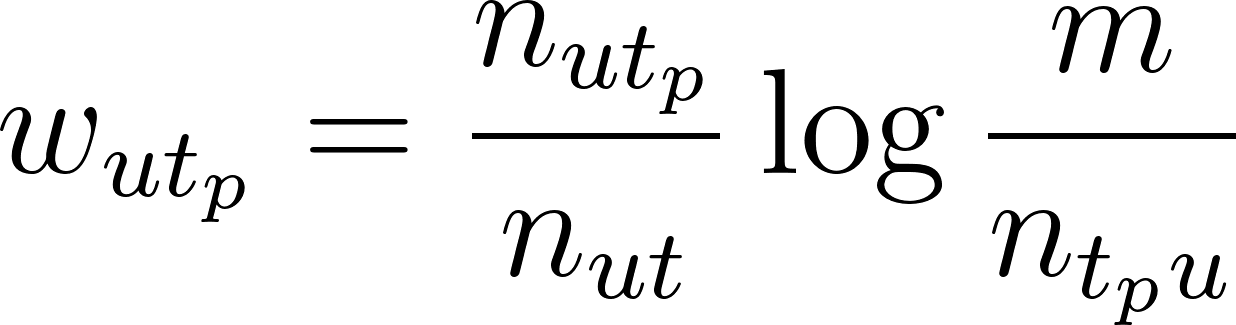
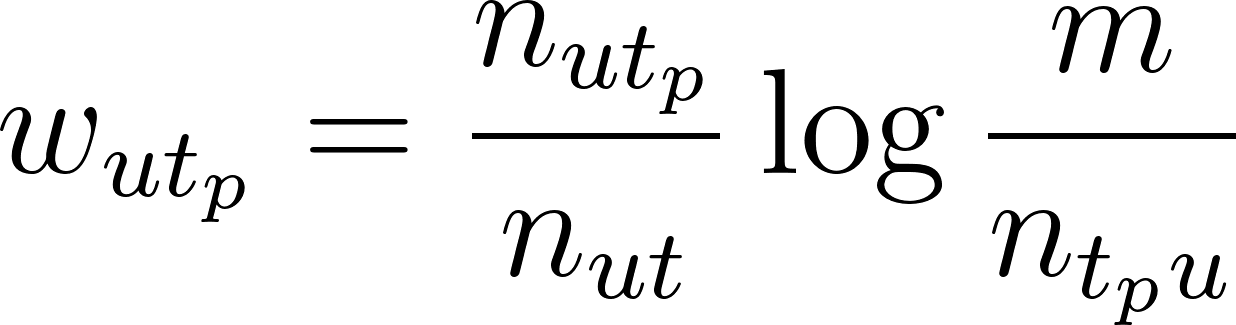
Establish tourist set [](https://www.codecogs.com/eqnedit.php?latex=U%3D%5C%7Bu_1%2C%20u_2%2C%20%5Ccdots%2C%20u_i%2C%20%5Ccdots%2C%20u_m%5C%7D#0), m is the total number of tourists, [](https://www.codecogs.com/eqnedit.php?latex=i%20%3D%201%2C%202%2C%20%5Ccdots%2C%20m#0). The set of all tourist attractions is [](https://www.codecogs.com/eqnedit.php?latex=S%20%3D%20%5C%7Bs_1%2C%20s_2%2C%20%5Ccdots%2C%20s_j%2C%20%5Ccdots%2C%20s_n%20%5C%7D#0), [](https://www.codecogs.com/eqnedit.php?latex=n#0) is total number of tourist attractions [](https://www.codecogs.com/eqnedit.php?latex=j%20%3D%201%2C2%2C%20%5Ccdots%2C%20n#0). And the set of all attractions tags is [](https://www.codecogs.com/eqnedit.php?latex=T%20%3D%20%5C%7Bt_1%2C%20t_2%2C%20%5Ccdots%2C%20t_p%2C%20%5Ccdots%2C%20t_k%20%5C%7D#0), where [](https://www.codecogs.com/eqnedit.php?latex=k#0) is the total number of tags, [, m is the total number of tourists, . The set of all tourist attractions is ,  is total number of tourist attractions . And the set of all attractions tags is , where  is the total number of tags, ](https://www.codecogs.com/eqnedit.php?latex=p%20%3D%201%2C2%2C%20%5Ccdots%2C%20k#0)

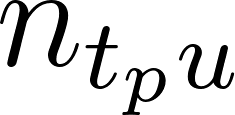
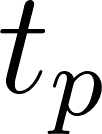
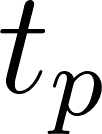
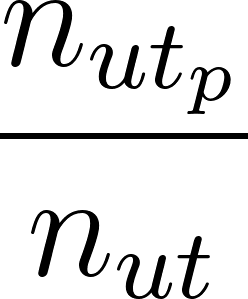
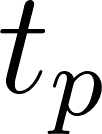
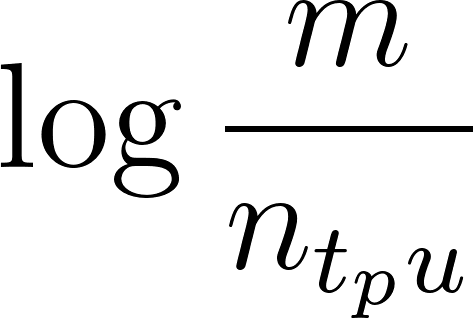
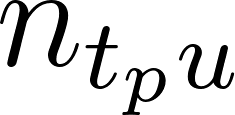
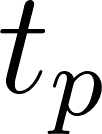
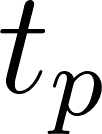
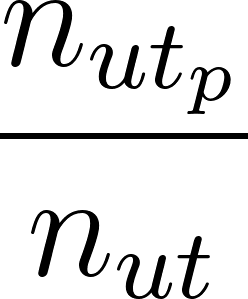
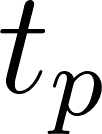
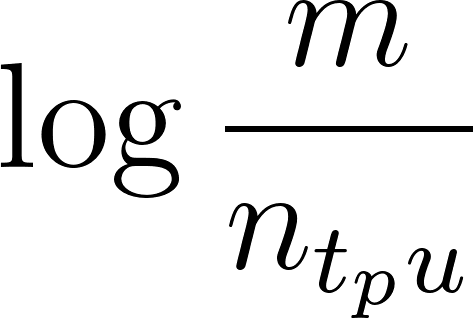
***2.4.1 Design the Tourist-Attractions Matrix***

Certain tourists have a preference for the tourist sites designated with the tag that they use more often the more times they use it. The tourist attraction tags' significance criteria were determined using the TF-IDF (Term Frequency-Inverse Document Frequency) approach. In addition to expressing the ability to categorize attractions, TF-IDF can partially represent the user's preference for tourism attractions. The tourist-attractions tag matrix is created using past visitor behavior data, which indicates the likelihood that visitors will select the tag or the significance of the tag to visitors. The matrix P is represented by formula (1), which indicates the significance of the attractions tag.

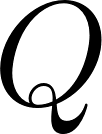
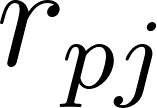
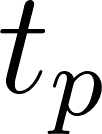
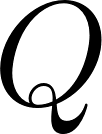
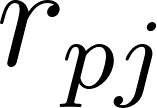
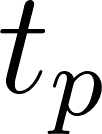
[](https://www.codecogs.com/eqnedit.php?latex=P%20%3D%20%5Cbegin%7Bbmatrix%7Dw_%7B11%7D%20%5C%2C%5C%2C%20w_%7B12%7D%20%5C%2C%5C%2C%20%5Ccdots%20%5C%2C%5C%2C%20w_%7B1p%7D%20%5C%2C%5C%2C%20%5Ccdots%20%5C%2C%5C%2C%20w_%7B1k%7D%20%5C%5C%5C%5C%20%7B%5Ccdots%7D%20%5C%2C%5C%2C%20w_%7Bip%7D%20%5Cquad%20%5Cquad%20%20%5C%2C%5C%2C%5C%2C%20%5Ccdots%20%5C%5C%5C%5C%20w_%7Bm1%7D%20%5C%2C%5C%2C%20w_%7Bm2%7D%20%5C%2C%20%5Ccdots%20%5C%2C%20w_%7Bmp%7D%20%5C%2C%20%5Ccdots%20%5C%2C%5C%2C%20w_%7Bmk%7D%20%5Cend%7Bbmatrix%7D#0)(1)(1)

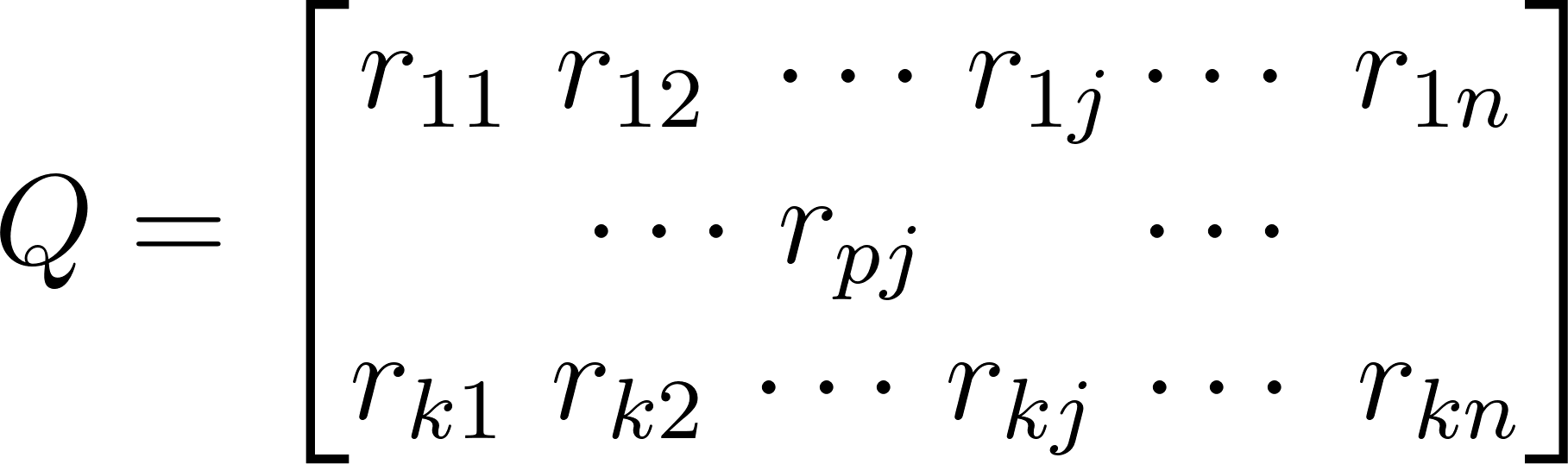
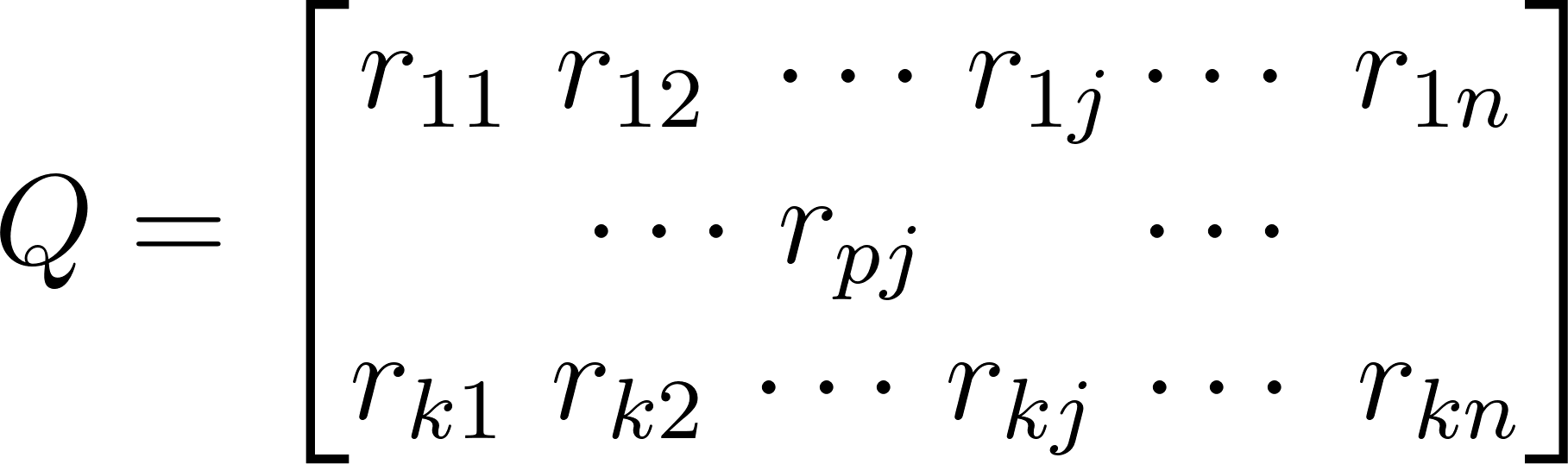
The TF-IDF method is used to determine the value of [](https://www.codecogs.com/eqnedit.php?latex=w#0). Reducing the weight of well-liked tags and resources and enhancing the originality and personality of recommendation results can be achieved by using TF-IDF to determine the significance of tags to users and resources. The formula displays the process of formula (2).. Reducing the weight of well-liked tags and resources and enhancing the originality and personality of recommendation results can be achieved by using TF-IDF to determine the significance of tags to users and resources. The formula displays the process of formula (2).

[](https://www.codecogs.com/eqnedit.php?latex=w_%7But_p%7D%20%3D%20%5Cfrac%7Bn_%7But_p%7D%7D%7Bn_%7But%7D%7D%5Clog%7B%5Cfrac%7Bm%7D%7Bn_%7Bt_p%20u%7D%7D%7D#0)

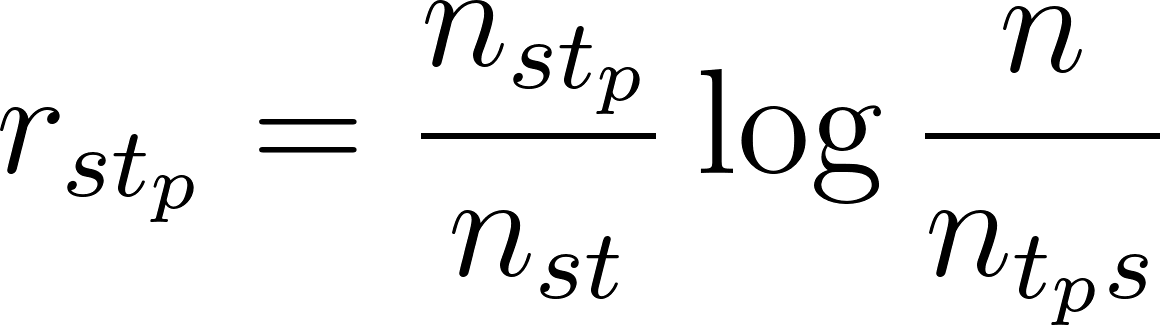
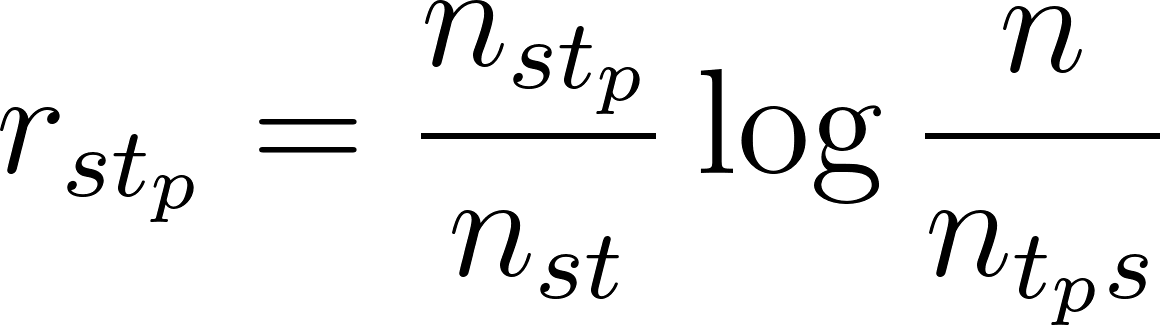
[In](https://www.codecogs.com/eqnedit.php?latex=w_%7But_j%7D%20%3D%20%5Cfrac%7Bn_%7But_j%7D%7D%7Bn_%7But%7D%7D%5Clog%7B%5Cfrac%7Bm%7D%7Bn_%7Bt_j%20u%7D%7D%7D#0) the formula, [](https://www.codecogs.com/eqnedit.php?latex=n_%7Bt_pu%7D#0) represents the number of users who use tag [](https://www.codecogs.com/eqnedit.php?latex=t_p#0), [](https://www.codecogs.com/eqnedit.php?latex=m#0) represents the total number of users, [](https://www.codecogs.com/eqnedit.php?latex=n_%7Bu%20t_p%7D#0) represents the number of tags [](https://www.codecogs.com/eqnedit.php?latex=t_p#0) that user [](https://www.codecogs.com/eqnedit.php?latex=u#0) uses, [](https://www.codecogs.com/eqnedit.php?latex=n_%7But%7D#0) represents the number of tags that user [](https://www.codecogs.com/eqnedit.php?latex=u#0) uses, [](https://www.codecogs.com/eqnedit.php?latex=%5Cfrac%7Bn_%7But_p%7D%7D%7Bn_%7But%7D%7D#0) represents the frequency that user [](https://www.codecogs.com/eqnedit.php?latex=u#0) uses tag [](https://www.codecogs.com/eqnedit.php?latex=t_p#0), and [](https://www.codecogs.com/eqnedit.php?latex=%5Clog%7B%5Cfrac%7Bm%7D%7Bn_%7Bt_p%20u%7D%7D%7D#0) represents the importance of tags among all tags of the user. represents the number of users who use tag ,  represents the total number of users,  represents the number of tags  that user  uses,  represents the number of tags that user  uses,  represents the frequency that user  uses tag , and  represents the importance of tags among all tags of the user.

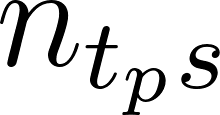
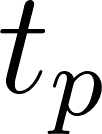
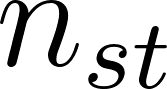
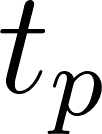
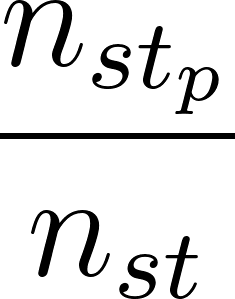
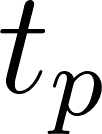
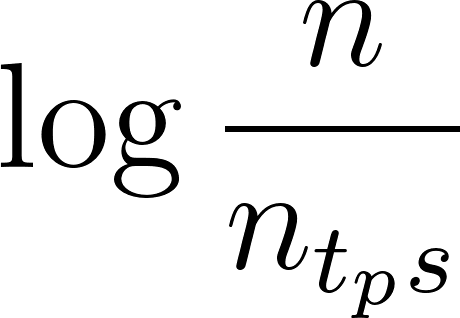
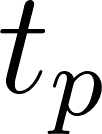
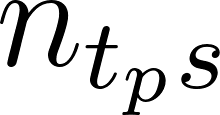
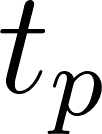
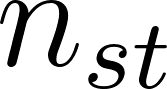
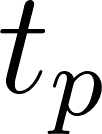
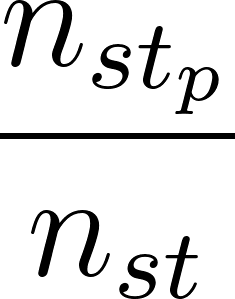
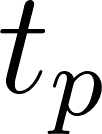
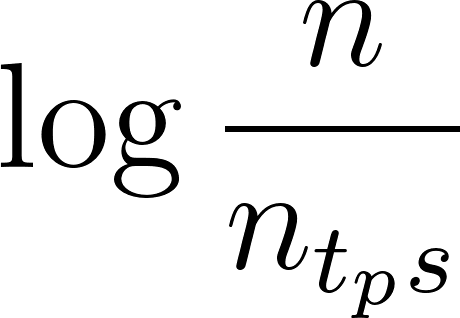
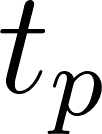
***2.4.2 Design Attractions Tag-Tourist Attractions Matrix***

A tourist attraction may have multiple users simultaneously tag it, and the tag may capture the essence of the attraction. A tourist attraction tag may represent the features of the attraction if it is more significant to the attraction. The matrix [](https://www.codecogs.com/eqnedit.php?latex=Q#0) (3) depicts the attractions tag-tourist attraction matrix, where [](https://www.codecogs.com/eqnedit.php?latex=r_%7Bpj%7D#0) denotes the importance of the attractions tag [](https://www.codecogs.com/eqnedit.php?latex=t_%7Bp%7D#0) to the tourist attraction [](https://www.codecogs.com/eqnedit.php?latex=j#0), or, put another way, the extent to which the tag characterizes the tourist attractions. (3) depicts the attractions tag-tourist attraction matrix, where  denotes the importance of the attractions tag  to the tourist attraction , or, put another way, the extent to which the tag characterizes the tourist attractions.

[](https://www.codecogs.com/eqnedit.php?latex=Q%20%3D%20%5Cbegin%7Bbmatrix%7Dr_%7B11%7D%20%5C%2C%5C%2C%20r_%7B12%7D%20%5C%2C%5C%2C%20%5Ccdots%20%5C%2Cr_%7B1j%7D%20%5Ccdots%20%5C%2C%5C%2C%20r_%7B1n%7D%20%5C%5C%5C%5C%20%7B%5Ccdots%7D%20%5C%2C%5C%2C%20r_%7Bpj%7D%20%5Cquad%20%20%20%5C%2C%5C%2C%5C%2C%20%5Ccdots%20%5C%5C%5C%5C%20r_%7Bk1%7D%20%5C%2C%5C%2C%20r_%7Bk2%7D%20%5C%2C%20%5Ccdots%20%5C%2C%20r_%7Bkj%7D%20%5C%2C%20%5Ccdots%20%5C%2C%5C%2C%20r_%7Bkn%7D%20%5Cend%7Bbmatrix%7D#0)(3)(3)

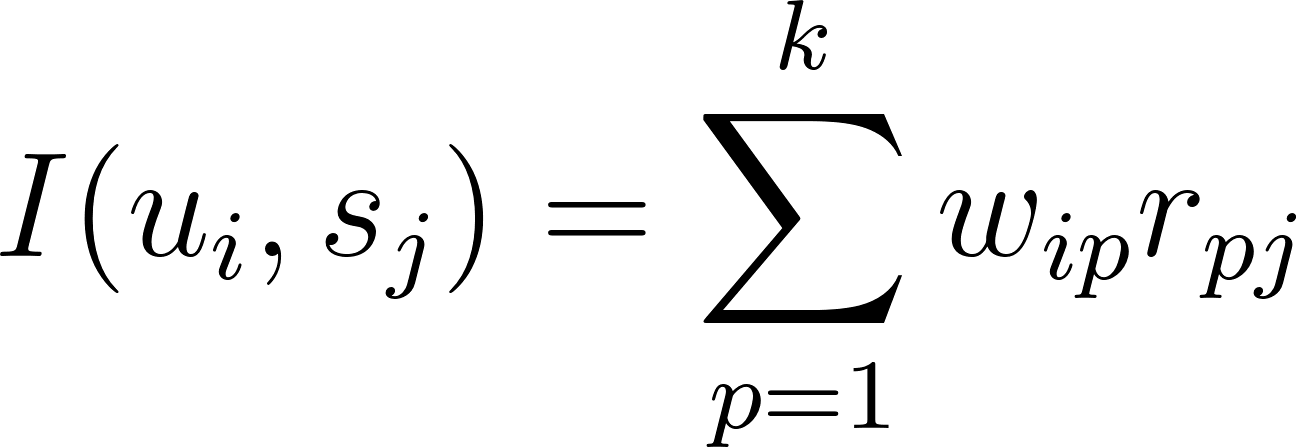
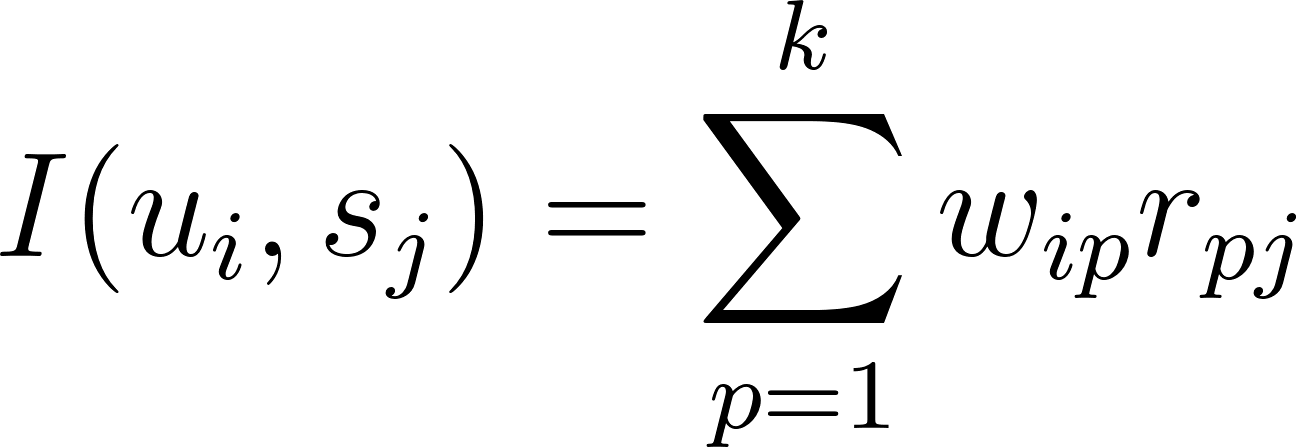
The TF-IDF method is used to calculate the importance of attractions tag to the attractions, as formula (4)

[](https://www.codecogs.com/eqnedit.php?latex=r_%7Bs%20t_p%7D%20%3D%20%5Cfrac%7Bn_%7Bs%20t_p%7D%7D%7Bn_%7Bs%20t%7D%7D%20%5Clog%7B%5Cfrac%7Bn%7D%7Bn_%7Bt_p%20s%7D%7D%7D#0)

[](https://www.codecogs.com/eqnedit.php?latex=n_%7Bt_p%20s%7D#0) represents the number of tourist attractions tagged by tag [](https://www.codecogs.com/eqnedit.php?latex=t_p#0), [](https://www.codecogs.com/eqnedit.php?latex=n#0) the total number of tourist attractions [](https://www.codecogs.com/eqnedit.php?latex=s#0). [](https://www.codecogs.com/eqnedit.php?latex=n_%7Bs%20t%7D#0) represents the number of tags marked on tourist attractions [](https://www.codecogs.com/eqnedit.php?latex=s#0), [](https://www.codecogs.com/eqnedit.php?latex=n_%7Bs%20t_p%7D#0) represents the number of tags [](https://www.codecogs.com/eqnedit.php?latex=t_p#0) for attraction [](https://www.codecogs.com/eqnedit.php?latex=s#0). [](https://www.codecogs.com/eqnedit.php?latex=%5Cfrac%7Bn_%7Bs%20t_p%7D%7D%7Bn_%7Bst%7D%7D#0) represents the frequency of tourist tag [](https://www.codecogs.com/eqnedit.php?latex=t_p#0) appears for attraction tourist [](https://www.codecogs.com/eqnedit.php?latex=s#0). And [](https://www.codecogs.com/eqnedit.php?latex=%5Clog%7B%5Cfrac%7Bn%7D%7Bn_%7Bt_p%20s%7D%7D#0) represents the importance of tag [](https://www.codecogs.com/eqnedit.php?latex=t_p#0) among all tags of all tourist attractions. represents the number of tourist attractions tagged by tag ,  the total number of tourist attractions .  represents the number of tags marked on tourist attractions ,  represents the number of tags  for attraction .  represents the frequency of tourist tag  appears for attraction tourist . And  represents the importance of tag  among all tags of all tourist attractions.

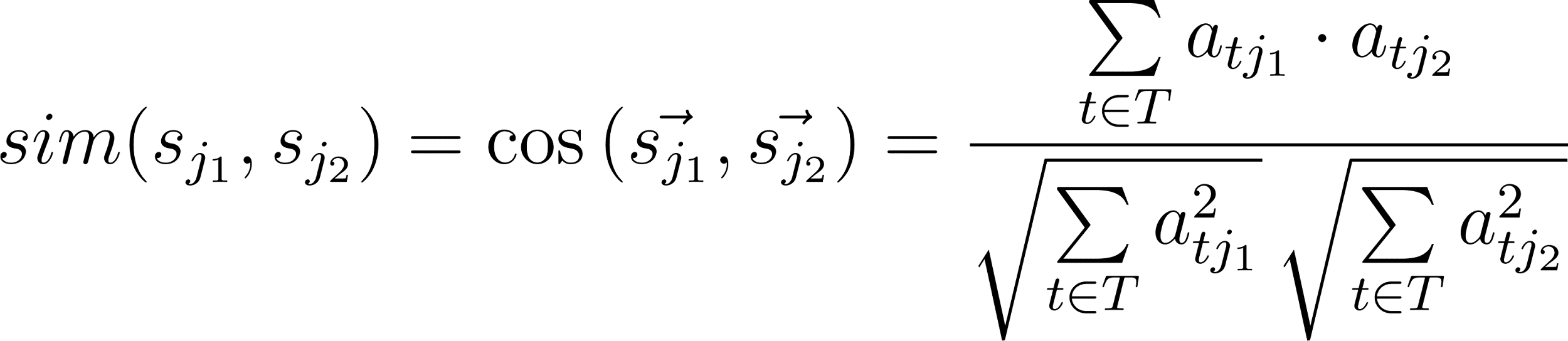
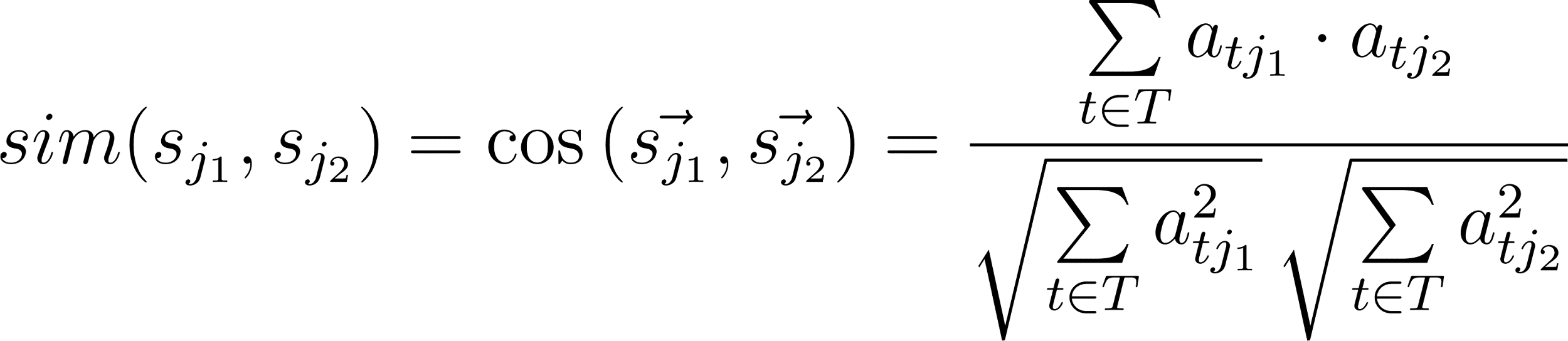
***2.4.3 Tourists’ Interest Model of Tourist Attractions***

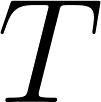
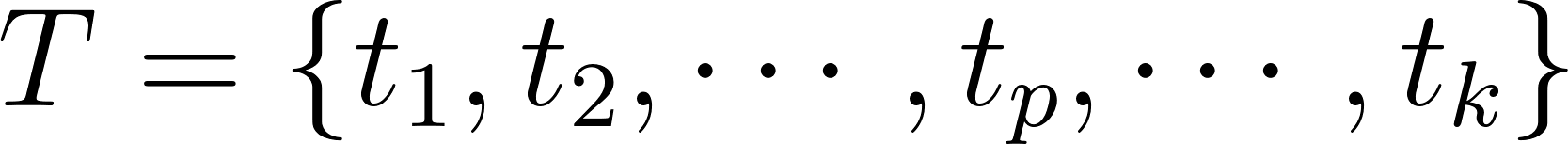
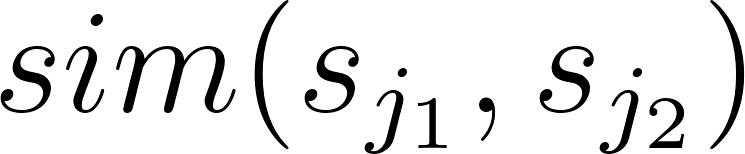
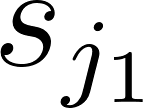
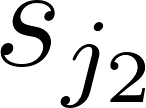
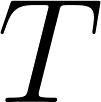
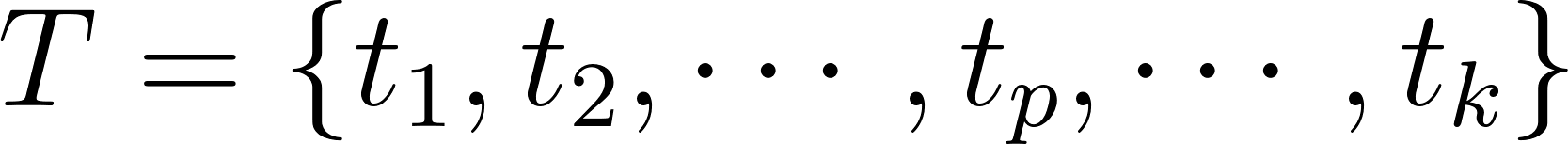
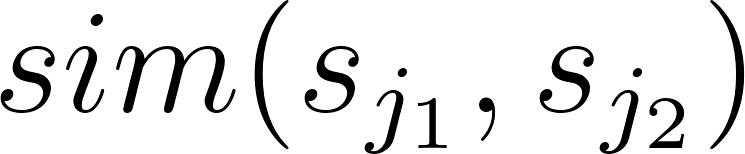
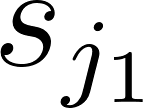
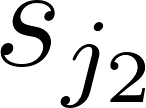
This formula expresses the interest model of tourists [](https://www.codecogs.com/eqnedit.php?latex=u_i#0) for tourist attractions [](https://www.codecogs.com/eqnedit.php?latex=s_j#0).  for tourist attractions .

[](https://www.codecogs.com/eqnedit.php?latex=%5Cqquad%20I(u_i%2C%20s_j)%20%3D%20%5Cdisplaystyle%20%5Csum_%7Bp%3D1%7D%5Ek%20w_%7Bip%7D%20%5C*%20r_%7Bpj%7D#0)

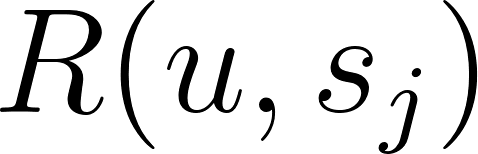
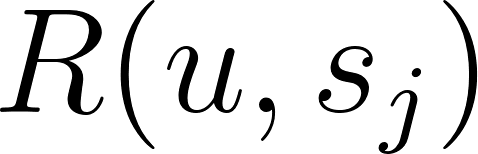
***2.4.4 Calculate the similarity of tourist attractions***

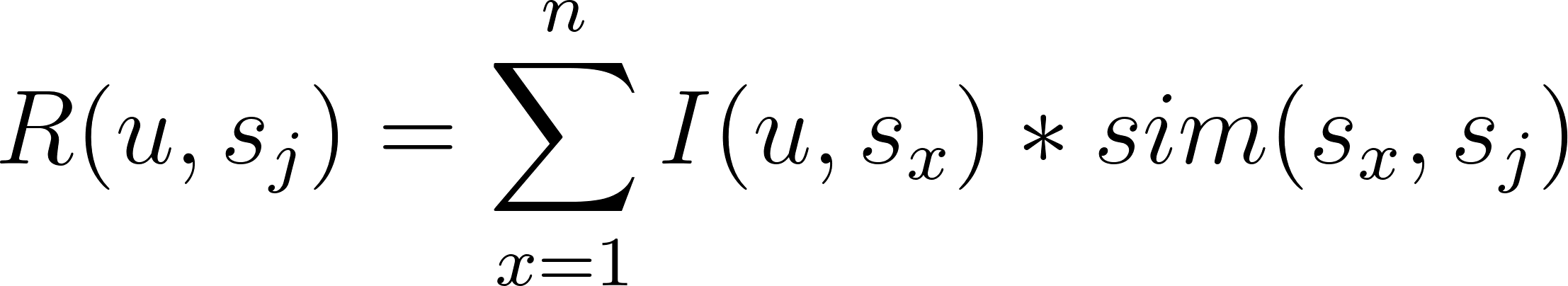
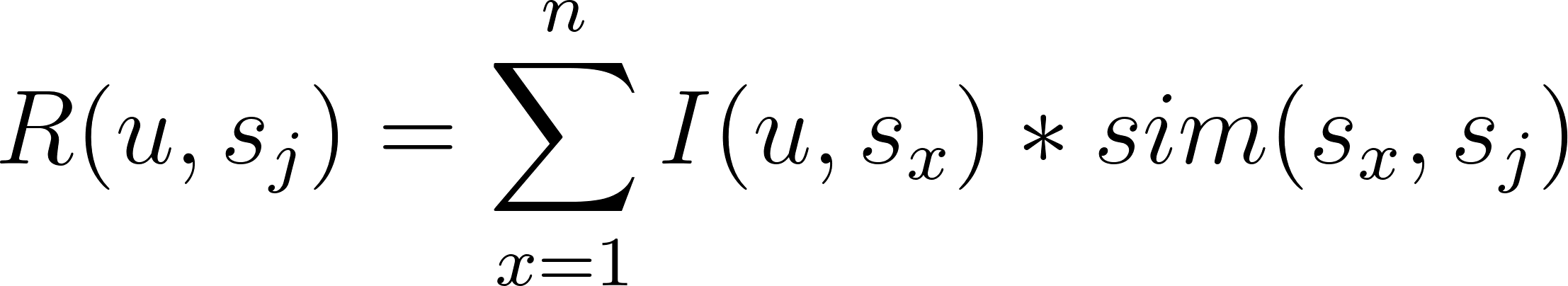
The degree of similarity between two distinct tourist attractions is indicated by their similarity. The conventional collaborative filtering algorithm overlooks the properties of resources themselves in favor of focusing on the quantity or rating of resources shared by various users. Because of this, it is also advantageous for users to find new resources by using the past user-marked resource behavior as characteristic information and applying similarity calculations to it. Cosine, modified cosine similarity, or Pearson similarity can all be used to determine how similar two resources are. It is discovered that the tourist rating data is sparser when compared to e-commerce user rating data, which lowers the similarity calculation's accuracy. E-commerce customers and lovers for toursis are not the same. They rarely repurchase items, and they frequently skip returning to the same tourist destinations after a brief absence. When it comes to recommending tourist destinations, recommendation diversity is quite crucial. Instead of using the conventional approach of determining the similarity by users' ratings of resources, the importance of attractions tag is employed in this research to compute the similarity between tourist attractions. Users can benefit from finding new resources as well. The key determinant is the importance of the tourist attractions tag; the more significant the tag, the more comparable the attractions [](https://www.codecogs.com/eqnedit.php?latex=i#0) and [](https://www.codecogs.com/eqnedit.php?latex=j#0). The formula (6) illustrates how the cosine similarity formula is used to determine how similar tourist destinations are. and . The formula (6) illustrates how the cosine similarity formula is used to determine how similar tourist destinations are.

[](https://www.codecogs.com/eqnedit.php?latex=sim(s_%7Bj_1%7D%2C%20s_%7Bj_2%7D)%20%3D%20%5Ccos%7B(%5Cvec%7Bs_%7Bj_1%7D%7D%2C%20%5Cvec%7Bs_%7Bj_2%7D%7D)%7D%20%3D%20%5Cfrac%7B%5Cunderset%7Bt%20%5Cin%20T%7D%7B%5Csum%7D%20%5C%2C%20a_%7Bt%20j_1%7D%20%5Ccdot%20a_%7Bt%20j_2%7D%7D%7B%5Csqrt%7B%5Cunderset%7Bt%20%5Cin%20T%7D%7B%5Csum%7D%20%5C%2C%20a_%7Bt%20j_1%7D%5E2%7D%20%5C%2C%20%5Csqrt%7B%5Cunderset%7Bt%20%5Cin%20T%7D%7B%5Csum%7D%20%5C%2C%20a_%7Bt%20j_2%7D%5E2%7D%7D#0)(6)(6)

In the formula (6), [](https://www.codecogs.com/eqnedit.php?latex=T#0) is a set of tags for all tourist attractions. It is known that [](https://www.codecogs.com/eqnedit.php?latex=T%20%3D%20%5C%7Bt_1%2C%20t_2%2C%20%5Ccdots%2C%20t_p%2C%20%5Ccdots%2C%20t_k%5C%7D#0), [](https://www.codecogs.com/eqnedit.php?latex=a_%7Bt%20j%7D#0) are the importance of tag [](https://www.codecogs.com/eqnedit.php?latex=t#0) to tourist attractions [](https://www.codecogs.com/eqnedit.php?latex=s_j#0), and [](https://www.codecogs.com/eqnedit.php?latex=sim(s_%7Bj_1%7D%2C%20s_%7Bj_2%7D)#0) represents the similarity between the tourist attractions [](https://www.codecogs.com/eqnedit.php?latex=s_%7Bj_1%7D#0) and [](https://www.codecogs.com/eqnedit.php?latex=s_%7Bj_2%7D#0). is a set of tags for all tourist attractions. It is known that ,  are the importance of tag  to tourist attractions , and  represents the similarity between the tourist attractions  and .

***2.4.5 Personalized Tourist Attractions Recommendation Result Set***

According to tourist interest and similarity of tourist attractions, the preference degree [](https://www.codecogs.com/eqnedit.php?latex=R(u%2C%20s_j)#0) of tourist [](https://www.codecogs.com/eqnedit.php?latex=u#0) for tourist attractions [](https://www.codecogs.com/eqnedit.php?latex=s_j#0) is calculated, or as known as recommendation score, and the recommendation result set is formed, and Top\_N attractions are recommended to tourists. The calculation formula for recommendation score is as formula (7) of tourist  for tourist attractions  is calculated, or as known as recommendation score, and the recommendation result set is formed, and Top\_N attractions are recommended to tourists. The calculation formula for recommendation score is as formula (7)

[](https://www.codecogs.com/eqnedit.php?latex=R(u%2C%20s_j)%20%3D%20%5Csum_%7Bx%3D1%7D%5En%20I(u%2C%20s_x)%20*%20sim(s_x%2C%20s_j)#0) (7) (7)

***2.4.6 Application of personalized recommendation algorithm***

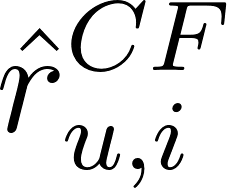
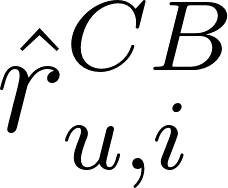
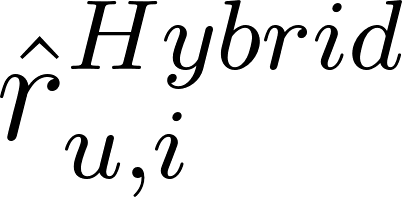
This paper selects 100 pieces of user data of a tourism website for experiment. This data includes tourist id, tourist attractions name, tourist attractions tag, tourist attractions rating and other information. According to the previous description of the algorithm, it is necessary to construct two matrices of tourists-tags and tag-attractions. The values of matrices express the preferences of tourists for a certain tag and the description degree of tags to attractions, respectively. The letter U is used to represent tourists and TA is used to represent tourist attractions. The data were divided into two groups, one group of training data and the other group of experimental data. According to the training data, the tourists’ preference matrix is calculated, and the similarity between the tourist attractions in the experimental data and the scored data is predicted, and the recommendation matrix is calculated.

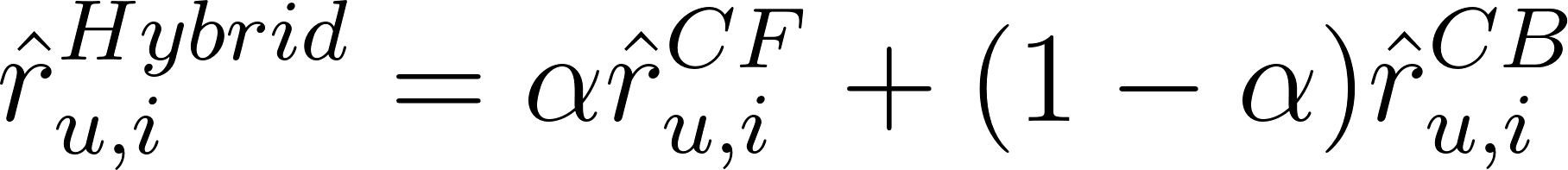
***2.5 Hybrid Recommendation Systems in Tourism***

To enhance the accuracy and relevance of recommendations in tourism, we employ a hybrid recommendation system that combines Collaborative Filtering (CF) and Content-Based (CB) filtering. This approach leverages the strengths of both methods while mitigating their individual weaknesses, leading to a more robust and comprehensive recommendation framework.

***2.5.1 Hybrid Recommendation Approach***

The hybrid recommendation system integrates CF and CB by computing separate recommendation scores from each method and then combining them. This combination is controlled by a parameter, typically denoted as α (alpha), which balances the influence of CF and CB in the final recommendation score.

Mathematically, let [](https://www.codecogs.com/eqnedit.php?latex=%5Chat%7Br%7D_%7Bu%2Ci%7D%5E%7BCF%7D#0)​ be the predicted rating of user [](https://www.codecogs.com/eqnedit.php?latex=u#0) for item [](https://www.codecogs.com/eqnedit.php?latex=i#0) obtained from the CF model, and [](https://www.codecogs.com/eqnedit.php?latex=%5Chat%7Br%7D_%7Bu%2Ci%7D%5E%7BCB%7D#0)​ be the predicted rating from the CB model. The hybrid recommendation score [](https://www.codecogs.com/eqnedit.php?latex=%5Chat%7Br%7D_%7Bu%2Ci%7D%5E%7BHybrid%7D#0) is calculated as:

[](https://www.codecogs.com/eqnedit.php?latex=%5Chat%7Br%7D_%7Bu%2Ci%7D%5E%7BHybrid%7D%20%3D%20%5Calpha%20%5Chat%7Br%7D_%7Bu%2Ci%7D%5E%7BCF%7D%20%2B%20(1-%5Calpha)%5Chat%7Br%7D_%7Bu%2Ci%7D%5E%7BCB%7D#0)

where [](https://www.codecogs.com/eqnedit.php?latex=%5Calpha#0) is a weighting parameter between [](https://www.codecogs.com/eqnedit.php?latex=0#0) and [](https://www.codecogs.com/eqnedit.php?latex=1#0). By adjusting [](https://www.codecogs.com/eqnedit.php?latex=%5Calpha#0), the system can be tuned to favor either CF or CB recommendations depending on the specific context or user preferences.

***2.5.2 Implementation Details***

The system begins by mapping place IDs to their respective names and calculating the mean ratings for each place. This helps in providing fallback recommendations for new or unknown users. For known users, the system retrieves predicted ratings generated by the CF model. These ratings are based on historical user behavior and the similarities between users. Simultaneously, the system calculates content-based scores by leveraging the similarity between places and the user’s past preferences. This involves performing a dot product between the place similarity matrix and the CF predicted ratings. The final recommendation scores are computed by linearly combining the CF and CB predictions using the [](https://www.codecogs.com/eqnedit.php?latex=%5Calpha#0) parameter. This ensures that the final recommendations consider both user behavior and item content. The system sorts the items based on the hybrid scores and selects the top k items to recommend to the user. This ranking mechanism ensures that only the most relevant items are presented.

In summary, the hybrid recommendation system offers a well-rounded and complete way to improve customer satisfaction and engagement. It is a sophisticated approach to tailored recommendations in the tourism industry.

***References:***  
*1. https://www.vietnam.vn/du-lich-viet-nam-nam-2023-vuot-xa-ky-vong/*

*2. Y. Cui, C. Huang and Y. Wang, "Research on Personalized Tourist Attraction Recommendation based on Tag and Collaborative Filtering," 2019 Chinese Control And Decision Conference (CCDC), Nanchang, China, 2019, pp. 4362-4366, doi: 10.1109/CCDC.2019.8832961.*

*3. R. Y. Saputra, L. E. Nugroho and S. S. Kusumawardani, "Collecting the Tourism Contextual Information data to support the tourism recommendation system," 2019 International Conference on Information and Communications Technology (ICOIACT), Yogyakarta, Indonesia, 2019, pp. 79-84, doi: 10.1109/ICOIACT46704.2019.8938546.*

*4.(... Failure to do so can result in user attrition, as users may abandon a platform in frustration if their desired items do not surface in initial search results (Nitin et al., 2021;Johnwendy, 2023). While contemporary recommender engines primarily rely on algorithms and data-driven methods to provide users with relevant recommendations, the seamless integration and synergy of search engines and recommender systems remain an unresolved challenge. …*

*5. L. Liu, "Intelligent Recommendation Framework for Tourist Attractions based on Data Mining Technology," 2020 International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2020, pp. 244-247, doi: 10.1109/ICOSEC49089.2020.9215404.*

*6. Y. Cui, C. Huang and Y. Wang, "Research on Personalized Tourist Attraction Recommendation based on Tag and Collaborative Filtering," 2019 Chinese Control And Decision Conference (CCDC), Nanchang, China, 2019, pp. 4362-4366, doi: 10.1109/CCDC.2019.8832961.*