**Abstract**

One of the first things to do while planning a trip is to book a

good place to stay. Booking a hotel online can be an overwhelming task with

thousands of hotels to choose from, for every destination. Motivated by the

importance of these situations, we decided to work on the task of recommending

hotels to users. We used “515K Hotel Reviews Data In Europe”

dataset on Kaggle which has a variety of features that helped

us achieve a deep understanding of the process that makes a

user choose certain hotels over others.

One of the ﬁrst things to do while planning a trip is

to book a good place to stay. Booking a hotel online

can be an overwhelming task with thousands of ho-

tels to choose from, for every destination. Motivated

by the importance of these situations, we decided to

work on the task of recommending hotels to users

**Objectives**

A hotel recommendation aims to predict which hotel

a user is likely to choose from among all hotels. In

this project, the objective is to transform implicit

information provided by user into explicit features

for hotel recommendation system.

**Introduction**

Text is an effective and widely existing form of opinion

expression and evaluation by users, as shown by the large

number of online review comments over tourism sites, hotels

and services. As a direct expression of user’s needs and

emotions, text-based tourism data mining has the potential to

transform the tourism industry.

Content-based filtering is a common approach in

recommendation system. The features of the items previously

rated by users and the best-matching ones are recommended.

In our case, we will be transforming implicit information of

hotel attributes as features for this recommendation.

**Methodology**

In a general way

There are two main data selection methods:

Collaborative-filtering: In collaborative-filtering items are recommended, for example hotels, based on how similar your user profile is to other users’, finds the users that are most similar to you and then recommends items that they have shown a preference for. This method suffers from the so-called cold-start problem: If there is a new hotel, no-one else would’ve yet liked or watched it, so you’re not going to have this in your list of recommended hotels, even if you’d love it.

Content-based filtering: This method uses attributes of the content to recommend similar content. It doesn’t have a cold-start problem because it works through attributes or tags of the content, such as views, Wi-Fi or room types, so that new hotels can be recommended right away.

The point of content-based is that we have to know the content of both user and item. Usually you construct user-profile and item-profile using the content of shared attribute space. For example, for a movie, you represent it with the movie stars in it and the genres (using a binary coding for example).

There are a number of popular encoding schemes but the main ones are: One-hot encoding, Term frequency–inverse document frequency (TF-IDF) encoding, Word embeddings

In this project, we will be discussing content-based filtering of recommender engine, turning implicit attributes into explicit features for hotel recommender engine.

**Conclusion**

With the increasing of applications in the Internet, the source of data is getting more and more richer. Therefore, the various factors in the new data brings new challenges. It is also a chance to create novel methods to achieve better recommendation results. Social networks are still the focus of the recommendation research, integration methods and new algorithms will continue to appear in the future. The sound, location and other user preference information are received more and more attention. The future of the recommender system will be an area of innovation and research.

**Applications**

Jupiter Notebook

**References**

[1] Li, J.; Xu, L.; Tang, L.; Wang, S.; Li, L. Big data in tourism research: A literature review. Tour. Manag. 2018, 68, 301–323.

[2] Qin Li 1,2, Shaobo Li 3,4,\* , Sen Zhang 1,2, Jie Hu 5 and Jianjun Hu 3,A Review of Text Corpus-Based Tourism Big Data Mining