We have just talked about Online Evaluation Techniques.

Let's now focus on Off-line Evaluation Techniques.

In order to analyze the off-line evaluation techniques, we have to mention some important aspects.

The task of the recommender system, the data set,

the partitioning of the data set, and the error metrics that will allow

us to estimate the goodness of our recommender system.

**Next Slide**

The first method we could use to accomplish the task of the off-line

evaluation is the rating prediction.

The goal is to go as near as possible to the true value of the rating.

As shown in the slide, the predicted value is 3.7 and

it could be considered a good recommendation since the true value is 4.

**Next Slide**

The second task is the top-N recommendation.

The goal of this type or

recommendation is to find an item which are relevant for the user.

The typical approaches to 1st rank items from the most relevant that is

positioned in the first position,

to the less relevant which will occupy the last position.

Later, we selected the first N items from the list that has been just created.

In this example, five movies have been ranked according to the taste of a user

from the one that is more likely to be liked by the user, to the less relevant.

**Next Slide:**

The evaluation data set represents all the information we have available to

make the proper recommendations.

When we speak about recommender systems, the data set is always represented by the URM.

A matrix that contains users on the rows, items on the columns, and the rating that a user has given to an item on the intersection.

Usually, we know a very little percentage of all the possible ratings.

The part that we know is called ground truth, and it is made up of all the nonzero ratings.

They can be divided into two main categories.

The relevant part contains all the positive opinions given by the users.

On the other hand, the non-relevant part contains all the negative opinions.

However, there are some other data that are not contained in the ground truth.

These data are the unknown ratings.

Typically, this set is made up of the interactions between users and

items in which the formers haven't rated the letters.