

# Recommender Systems Study Design (2)

Yaqi Zhou; Yanan Huo; Li Luo; Kejia Shi

March 17, 2016

# CHTC First Try

1. Learn to use R commands on CHTC platform
2. Try to perform the same cleaning method of the 100k file on CHTC

# From 100k to 20m

## 1. Only Movies/Ratings/Links/Tags datasets

no user information: no need to perform RForest

only focus on using recommenderlab (IBCF and UBCF)

## 2. Scales are different

1,682->27,278 movies and their links;

100,000->20,000,263 ratings;

extra: 465,564 tags;

## 3. Performed an initial data cleaning

ditched timestamps for simplicity;

rearrange genre into separate ones;