Assignment Weibo

Problem description

You want to investigate how a user's characteristics and posting behavior is related to his/her probability of being a spammer.

Three datasets of the Chinese microblogging website *Weibo* are available:

- weibo_user: contains information on the characteristics of 781 users.
- user_post: contains information on 40,443 posts of the users.
- labels: contains the response variable, which resolves to 1 if the user is detected as a spammer, 0 otherwise.

Variables related to message (or post) content are not required to be included into the analysis.

Data description

weibo_user has the following attributes:

- user_id: user ID in Weibo
- user_name: user nickname
- gender: user registration gender
- class: user level given by Weibo
- message: user registration location or other personal information
- post_num: number of posts of this user up to now
- follower_num: number of followers of this user
- followee_num: number of 'followees' of this user

user_post has the following attributes:

- post_id: post ID given by Weibo
- post_time: time post was shared
- content: content of the post (in Chinese)
- poster_id: user ID of user who shared this post
- poster_url: URL-link to the user
- repost_num: number of retweets by others
- comment_num: number of comments by others