

# Data

2013.4.25 백승환

- Data 파일의 용량이 매우 커서 기존의 viewer 로는 읽기도 시간이 많이 걸립니다. 저는 LTF Viewer 를 다운받아서 쓰고 있는데 data 를 읽기만 가능하고 그대신 빠른 것 같습니다. 단 user\_sns.txt 파일은 이 뷰어로 열리지 않는데 그 이유는 잘 모르겠습니다. 어쨌든 LTF Viewer 를 추천합니다.

이름	설명	Size	부가 설명
Item	person, organization, group 와 같이 user 에게 추천되는 항목들.	6k	<ul style="list-style-type: none"><li>• Category 로 분류됨. Category 는 hierarchical 한 관계를 지님. "."으로 구분됨. Ex) science-and-technology.internet.mobile</li><li>• Follow 한 사람이 좋아하는 category 를 좋아할 가능성이 높다.</li></ul>
Tweet	User 가 posting 한 것.		<ul style="list-style-type: none"><li>• tweet 을하면 자신을 follow 하고 있는 user 는 해당 tweet 을 보게 된다.</li></ul>
retweet	Tweet 을 다시 post 하는 것. Comment 를 붙일 수도 있다.		<ul style="list-style-type: none"><li>• 자신을 follow 하고 있는 user 에게도 보여줘서 더 많은 사람과 공유할 수 있게 해준다.</li></ul>
Comment	Tweet 에 댓글을 다는 것이다.		<ul style="list-style-type: none"><li>• follow 하고 있는 사람에게 알려지지 않는다.</li></ul>
Followee/follower	B ← A		

- **Rec\_log\_train.txt:** user 가 어떤 item 의 recommendation 을 accept 했는지 안했는지를 나타낸다. [(UserId)\t(ItemId)\t(Result)\t(Unix-timestamp)]

- **Rec\_log\_test.txt:** \_train.txt 와 format 은 같으나 accept 할지 안할지를 우리의 알고리즘으로 prediction 해야한다. [(UserId)\t(ItemId)\t(Result)\t(Unix-timestamp)]
- **User\_profile.txt:** user 에 대한 정보가 들어가있다. 이곳에 있는 정보를 모두 이용하는 것이 좋을 듯 하다. 이중 tag 는 user 의 관심사를 직접 자신이 적은 것으로 direct feedback 이라 볼 수 있다. Tag 는 0 개에서 많은 수까지 저장 가능하다. [(UserId)\t(Year-of-birth)\t(Gender)\t(Number-of-tweet)\t(tag-id1;tag-id2;...;tag-idN)]
- **Item.txt:** item 의 카테고리는 “.”으로 구분되며 해당 item 과 관련있는 keyword id 들의 list 가 나열된다. [(ItemId)\t(Item-Category)\t(id1;id2;...;idN)]
- **User\_action.txt:** @ (at) action 에 대한 정보가 담겨있다. Tweet, retweet, comment 같은 것을 알리고 싶을 때 @을 쓰면 해당 user 에게 알림이 간다. 그리고 retweet 에 대한 정보도 담겨있는데 해당 user 의 글을 retweet 한 횟수가 그것이고 comment 의 수는 그 유저의 글에 comment 를 단 횟수이다. 이들을 통해서 user 의 친밀도를 계산 할 수 있을 것 같다. Follow 정보 이외에도 user 의 친밀도에 대한 정보를 뽑을 수 있다. [(UserId)\t(Action-Destination-UserId)\t(Number-of-at-action)\t(Number-of-retweet )\t(Number-of-comment)]
- **User\_sns.txt:** user 의 follow history 를 제공한다. [(Follower-userid)\t(Followee-userid)]
- **User\_key\_word.txt:** user 가 가지고있는 weighted keyword list 를 나타낸다. [(UserId)\t(kw1:weight1;kw2:weight2;...kw3:weight3)]