

1. Trans系列的实验部分：

所用的数据：FreeBase15K, WordNet18

数据格式：

entity2id, relation2id

train2id, valid2id: head tail relation

test2id: label head tail relation

评测方法：

先把待测试的head或者tail依次替换成所有entity，然后将他们按照打分函数排序

meanRank: 所有正确实体的平均排名

hit@10: 正确实体排在前10的数量占测试总数的百分比

TransD (unif)	242	229	79.2	92.5
results:				
left 637.296814 0.789600 0.534200 0.021800				
left(filter) 623.093628 0.919000 0.772000 0.046800				
right 669.684204 0.796400 0.566600 0.024200				
right(filter) 656.550171 0.917200 0.779000 0.049600				
results (type constraints):				
left 284.303406 0.801000 0.581800 0.236600				
left(filter) 270.100189 0.931000 0.833000 0.296800				
right 301.431396 0.811200 0.612200 0.264000				
right(filter) 288.297394 0.930200 0.837000 0.319000				
results (1-1):				
left 1039.738037 0.738095 0.404762 0.000000				
left(filter) 1039.476196 0.738095 0.404762 0.000000				
right 743.880981 0.714286 0.309524 0.000000				
right(filter) 743.595215 0.714286 0.333333 0.000000				
results (1-n):				
left 410.862488 0.939361 0.853276 0.054142				
left(filter) 410.754730 0.940444 0.860977 0.073091				
right 881.187317 0.602599 0.304819 0.005414				
right(filter) 846.423950 0.905252 0.753655 0.056849				
results (n-1):				
left 743.718323 0.608279 0.284200 0.004543				
left(filter) 700.675903 0.916709 0.764260 0.049975				
right 373.186768 0.944472 0.872287 0.056032				
right(filter) 373.120148 0.944977 0.881373 0.072186				
results (n-n):				
left 801.456665 0.864602 0.455752 0.000000				
left(filter) 800.229187 0.894690 0.640700 0.000000				
right 836.585815 0.856637 0.468142 0.000000				
right(filter) 835.419495 0.895575 0.657522 0.000000				

hit@10差不多，但是meanRank不稳定，好的情况和论文结果差不过。

2. 关于知识图谱用于做推荐的好处

1. 可以挖掘到更深的信息，因为链路可以很深，引入了更多的语义关系
2. 知识图谱提供了不同的关系链接种类，有利于推荐结果的发散
3. 可解释性

3. KKbox的数据集

members_v3/

msno

城市

年龄

性别

注册方式

注册时间

到期日期

transaction/

msno

购买方式

会员计划长度

计划定价

支付的实际金额

是否自动续订

交易日期

会员资格到期日期

是否取消

userlogs/

msno: 用户ID

日期: 格式%Y%m%d

num 25: 播放的歌曲数量少于歌曲长度的25%

num 50: 在歌曲长度的25%到50%之间播放的歌曲数量

num 75: 在歌曲长度的50%到75%之间播放的歌曲数量

num 985: 在歌曲长度的75%到98.5%之间播放的歌曲数量

num 100: 播放的歌曲数量超过歌曲长度的98.5%

num uniq: 播放的独特歌曲数量

总秒数: 总播放秒数

train/

msno: 用户ID

song_id: 歌曲id

source_system_tab: 触发事件的选项卡的名称。系统选项卡用于对KKBOX移动应用程序功能进行分类。例如, tab我的库包含操作本地存储的功能, 而tab搜索包含与搜索相关的功能。

source_screen_name: 用户看到的布局的名称。

source_type: 用户首先在移动应用上播放音乐的入口点。入口点可以是专辑, 在线播放列表, 歌曲.....等。

target: 这是目标变量。target = 1表示在用户的第一个可观察的侦听事件之后的一个
月内触发了重复的侦听事件, 否则为target = 0。

test/

id: 行id (将用于提交)

msno

song_id

source_system_tab

source_screen_name

source_type

songs/

song_id

song_length: in ms

genre_ids: 流派类别。有些歌曲有多种类型, 它们由|分隔

artist_name

composer
lyricist
anguage

song_extra_info
song_id
song name

isrc 国际标准录音代码，理论上可以用作歌曲的标识。但值得注意的是，提供商生成的ISRC尚未得到官方验证; 因此，ISRC中的信息，例如国家代码和参考年份，可能会产生误导/错误。多首歌曲可以共享一个ISRC，因为单个录制可以多次重新发布。

目的：预判用户之后还会听那些歌