推荐系统与机器学习 学习路径

一、工具使用

工欲善其事，必先利其器

- ✓ 熟练掌握 Python，上手使用 numpy, pandas 工具包
- ✓ 机器学习工具包 sklearn，LibFM，xlearn
- ✓ 深度学习工具包 Tensorflow, keras, pytorch （掌握其一即可）
- ✓ 自动机器学习工具：TPOT，auto-sklearn，autokeras
- ✓ 推荐系统工具：Surprise, lightfm
- ✓ 深度学习的 CTR 工具：DeepCTR
- ✓ 特征工程：FeatureTools
- ✓ Embedding：Gensim

二、算法模型：

器欲尽其能，必先得其法

Model Related：

- ✓ 分类算法：LR，Decision Tree，Naive Bayes，SVM，KNN
- ✓ 矩阵分解：ALS-WR，FunkSVD，BiasSVD，SVD++
- ✓ FM 模型：FM，FFM，DeepFM，NFM，AFM
- ✓ 树模型：GBDT，XGBoost，LightGBM，CatBoost，NGBoost
- ✓ Attention 模型：DIN，DIEN，DSIN，Transformer，BERT
- ✓ Embedding：Word2vec，Item2vec，Graph Embedding

Engineering Related：

- ✓ 特征工程：探索性数据分析 EDA，数据清洗，数据采样，数据规范化，特征选择，统计特征
- ✓ 推荐架构：大厂推荐架构，召回，排序，model serving
- ✓ 其他 Pipeline：数据抓取，可视化，可解释性

三、Project 实战

由浅入深，积累你的 Project 经验

1、MNIST 手写数字识别

2、Titanic 乘客生存预测

https://www.kaggle.com/c/titanic/data

3、如何对球队数据进行聚类分析

https://github.com/cystanford/Recommended_System/tree/master/L2/team_cluster

4、Declicious 基于标签的推荐系统

https://grouplens.org/datasets/hetrec-2011/

5、STEAM Video Games 游戏推荐系统

https://www.kaggle.com/tamber/steam-video-games

6、员工离职预测
https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset

7、santander 银行产品推荐
https://www.kaggle.com/c/santander-product-recommendation/overview

8、三国演义人物 embedding 及相似度计算
数据来自三国演义小说 txt

9、Netflix 提供推荐算法，1 亿评分数据（用户量 48 万，DVD 数 1.8 万）大小 680M
https://www.kaggle.com/netflix-inc/netflix-prize-data

10、CTR 广告点击率预测，Avazu CTR 数据集，数据集大小 1G，解压后 8G
https://www.kaggle.com/c/avazu-ctr-prediction

11、未来 6 个月房价走势，自己抓取数据，趋势预测分析

四、Reading List

熟读经典，如何提一个 Valuable 的问题，并给出 state of art 的解决方案

Deep Interest Network for Click-Through Rate Prediction,2018 （阿里提出的 DIN 模型）
https://arxiv.org/abs/1706.06978

Deep Interest Evolution Network for Click-Through Rate Prediction,2018（阿里提出的 DIEN 模型）
https://arxiv.org/abs/1809.03672

Deep Session Interest Network for Click-Through Rate Prediction, IJCAI 2019（阿里的 DSIN 模型）
https://arxiv.org/abs/1905.06482

Real-time Personalization using Embeddings for Search Ranking at Airbnb（2018kdd best paper）
https://www.kdd.org/kdd2018/accepted-papers/view/real-time-personalization-using-embeddings-for-search-ranking-at-airbnb

The Youtube video recommendation system, ACM Conference on Recommender Systems 2010
http://www.inf.unibz.it/~ricci/ISR/papers/p293-davidson.pdf

Latent dirichlet allocation, Journal of Machine Learning Research 2003
https://www.lri.fr/~sebag/COURS/BleiNgJordan2003.pdf

Matrix factorization techniques for recommender systems, [J]. Computer 2009

http://cseweb.ucsd.edu/classes/fa17/cse291-b/reading/Recommender-Systems.pdf

Community-based user recommendation in uni-directional social networks, CIKM 2013
http://dnslab.jnu.ac.kr/classes/old_courses/2015s_das/[CIKM_2013]%20Community-Based%20User%20Recommendation%20in%20Uni-Directional%20Social%20Networks.pdf

BPR: Bayesian personalized ranking from implicit feedback, UAI 2009
http://www.arxiv.org/ftp/arxiv/papers/1205/1205.2618.pdf

Sparse linear methods for top-n recommender systems, ICDM 2011
http://glaros.dtc.umn.edu/gkhome/fetch/papers/SLIM2011icdm.pdf

Collaborative Filtering Recommender Systems, AdaptiveWeb 2006
http://faculty.chas.uni.edu/~schafer/publications/CF_AdaptiveWeb_2006.pdf

Toward Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions, IEEE Transactions on Knowledge and Data Engineering 2005
http://www.ist.tugraz.at/felfernig/images/recommender_systems_future.pdf

Top-K Off-Policy Correction for a REINFORCE Recommender System, wsdm 2019
http://alexbeutel.com/papers/wsdm2019_reinforce_recs.pdf

Deep Learning Recommendation Model for Personalization and Recommendation Systems, 2019
https://arxiv.org/abs/1906.00091v1

AutoML A Survey of the State of the Art, KDD 2019
https://arxiv.org/abs/1908.00709

图书推荐：

《Recommender Systems Handbook》
《Recommender Systems: An Introduction》
《Programming Collective Intelligence》

技术专栏：

内容	Blog
协同过滤	https://www.ethanrosenthal.com/2015/11/02/intro-to-collaborative-filtering/
深度学习 in spotfiy	http://benanne.github.io/2014/08/05/spotify-cnns.html

五、相关学术会议

一篇顶级会议 paper，是进入大厂最直接的入场券

ACM RecSys : The ACM Conference Series on Recommender Systems

ACM SIGKDD : The ACM SIGKDD Conference on Knowledge Discovery and Data Mining

ACM SIGIR : The ACM International Conference on Research and Development in Information Retrieval

ICDM : The IEEE International Conference on Data Mining

ICML : The International Conference on Machine Learning