Learning To Rank在个性化电商搜索中的应用

吴晨 (搜索BG: Natural Artificial Intelligence)

2016.10.22

Outline

❶ Background

❷ Learning to Rank

❸ Personalized E-Commerce Search

❹ Summary

❺ Reference

Background

Predict relevance scores and re-rank products returned by an e-commerce search engine on the search engine result page (SERP)

- Data Using
  - Search, browsing, and transaction histories for all users and specifically the user interacting with the search engine in the current session
  - Product properties and meta-data
- Method Using
  - Machine Learning (e.g. RankSVM, LambdaMart)
  - Ranking Function (e.g. BM25, Cosine Similarity)

LEARNING TO RANK

Introduction

- Ranking Problem
  - Learning to Match?
- Methods
  - Pointwise
  - Pairwise
  - Listwise
- Theory (PAC)
  - Generalization
  - Stability
- Applications
  - Search
  - Recommender System
  - Question Answering
  - Sentiment Analysis

Formulation

- Machine Learning
  - Supervised learning with labeled data
- Ranking of objects by subject
  - Feature based ranking function
- Approach
  - Traditional
    - BM25 (Probabilistic Model)
  - New
    - Query and associated products form Group (Train Data)
    - Groups are i.i.d
    - Features (query and product) in Group are not i.i.d
    - Model is a function of features

Issues

- Data Labeling
  - Relevance metric (Point)
  - Ordered pairs
  - Ordered list
- Feature Extraction
  - Relevance (User/Query-Prod Feature)
  - Semantic (User/Query-Prod Feature)
  - Importance (Prod Feature)
- Learning Method
  - Model
  - Loss Function
  - Optimized Algorithms
- Evaluation Measure
  - NDCG@k

Methods

- Machine Learning
  - Classification
  - Regression
  - Ordinal Classification/Regression
- Ordinal Regression
  - Pointwise
    - Transfer ranking to regression
    - Ignore group info
- Learning to Rank
  - Pairwise
    - Transfer ranking to binary classification
  - Listwise
    - Straightforward represent learning

Pointwise Model

- McRank (2007)
- Ordinal Liner Regression
  - (Staged) Logistic Regression

Pairwise Model

- RankSVM (2000)
  - Pairwise classification
- IR SVM (2006)
  - Cost-sensitive Pairwise
  - Using modified hinge loss
- RankBoost (2003)
- RankNet (2005)
- LambdaMart (2008)

Listwise Model

- Plackett-Luce Model
- ListMLE (2008)
- ListNet (2007)
  - Parameterized Plackett-Luce Model
- AdaRank (2007)
- PermuRank (2008)
- SVM-Map (2007)

$$P(\pi) = \prod_{j=1}^{m} \frac{\phi(s_{\pi(j)})}{\sum_{l=j}^{m} \phi(s_{\pi(l)})}$$

Optimize

- Direct Optimization
  - AdaRank
  - SVM Map
- Approximation
  - Soft Rank
  - Lambda Rank
- Learning Framework
  - Data Representation
  - Expected Risk
  - Empirical Risk
  - Generalization Analysis
- Evaluation
  - Pairwise approach and Listwise approach perform better than Pointwise approach

Applications

- Search
  - Re-Ranking
- Recommender System
  - Collaborating Filter

PERSONALIZED E-COMMERCE SEARCH

Introduction

- Pertinence
  - Log Analysis
  - Conversion in E-commerce
    - give a greater score to clicks that eventually got converted into a sale
- Data
  - User info
  - List of the terms that forms the query
  - Displayed items and their domains
  - Items on which the user clicked
  - Timing of all of these actions
  - History Behaviors Day 28 to Day 30
- Ensemble Model
  - Boosting
  - Bagging
  - Stacking
- Trap
  - Position Bias

Related Work

- Clicks feedback
- When to do personalize ?
  - Long Term
  - Short Term
- Past interaction timescales
- Search behaviors beyond clicks
- Learning from all repeated results

Features

- Aggregate Features
  - User-specific / Anonymous
- Query features
- User click habits
  - Number of times the user clicked on the item in the past
- Session features
- Non-Personalized Rank
  - Read linearly
  - Computed with information
- Inhibiting/Promoting features
  - Query click entropy

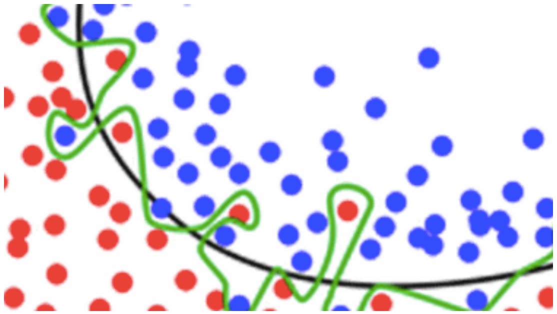
Methodology

- Classification will be used
  - Parameter of the classifier should be tuned to optimize the NDCG score on the cross validation set
- Query Full
  - SERPs returned in response to a query
- Query Less
  - SERPs returned in response to the user click on some product category

Ensemble Model

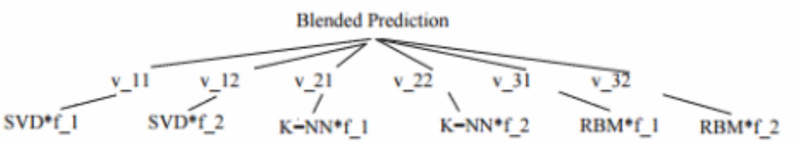
A very powerful technique to increase accuracy on a variety of ML tasks

- Boosting
- Bagging



Ensemble Correlation

- Voting
- Weighing
- Averaging
- Rank averaging

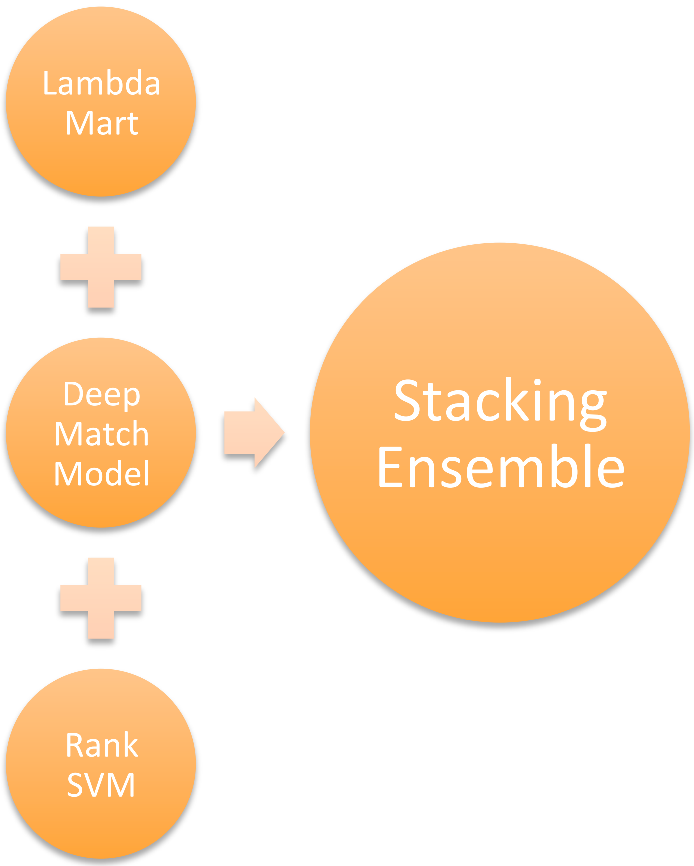


- Stacking / Blending
  - Split the train set in A and B
  - Fit a first-stage models on A and create predictions for B
  - Fit the same model on B and create predictions for A
  - Finally fit the model on the entire train set and create predictions for test set
  - Train a second-stage stacker model on the probabilities from the first-stage model(s).

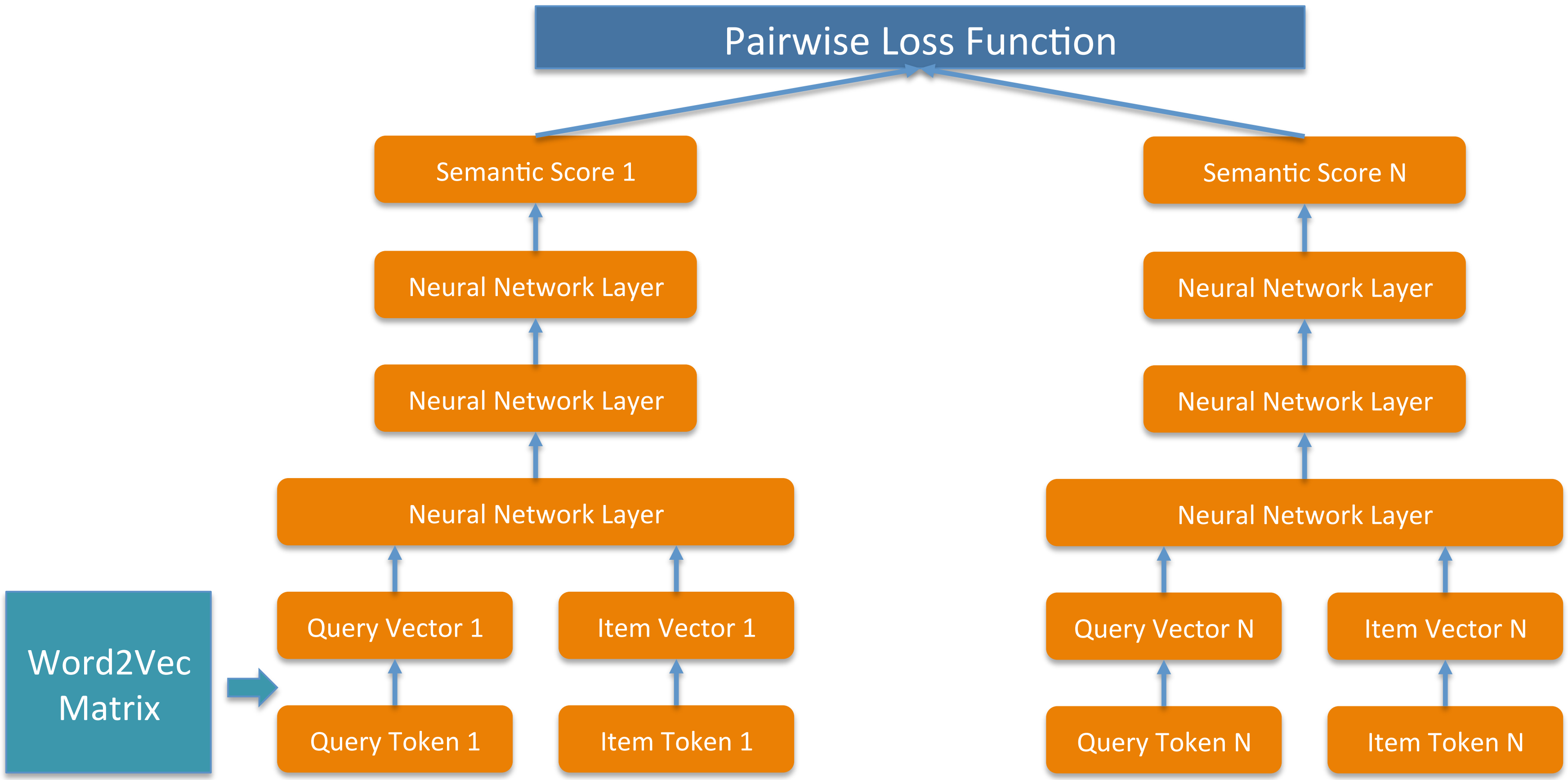
Stacking with logistic regression is one of the more basic and traditional ways of stacking

Scenario I: Query Full

Relevance Match + Semantic Match

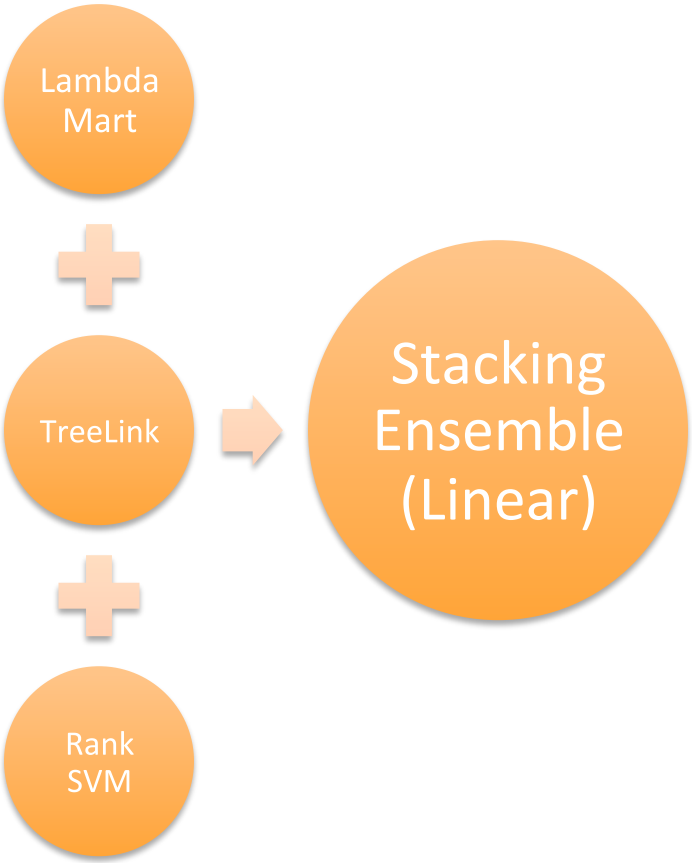


Scenario I: Query Full – DNN Model

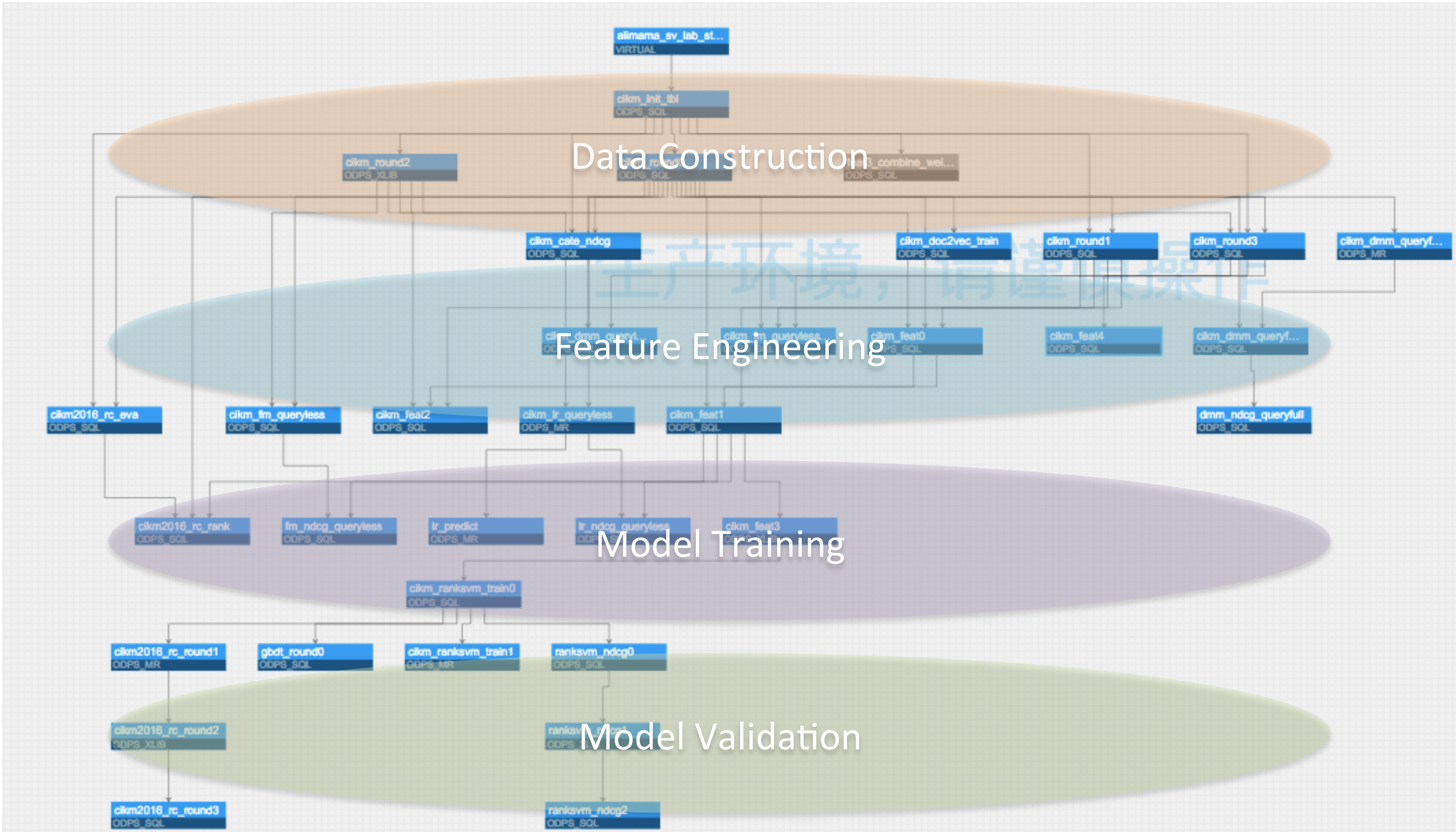


Scenario II: Query Less

Click Model + Recommender System



Work Flow



Data Construction
Feature Engineering
Model Training
Model Validation
alimama_sv_lab_st...
VIRTUAL
cikm_init_tbl
ODPS_SQL
cikm_round2
ODPS_XLIB
cikm_cate_ndcg
cikm_doc2vec_train
cikm_round1
cikm_round3
cikm_dmm_queryf...
ODPS_MR
cikm_feat0
cikm_feat4
cikm2016_rc_eva
cikm_fm_queryless
cikm_feat2
cikm_lr_queryless
cikm_feat1
dmm_ndcg_queryfull
cikm2016_rc_rank
fm_ndcg_queryless
lr_predict
lr_ndcg_queryless
cikm_feat3
cikm_ranksvm_train0
cikm2016_rc_round1
gbdt_round0
cikm_ranksvm_train1
ranksvm_ndcg0
cikm2016_rc_round2
cikm2016_rc_round3
ranksvm_ndcg2

ACM CIKM 2016 Competition

Improved the challenging non-personalized baseline by 21.28%

DIGINETICA
Retail Techology Company

CIKM Cup 2016 Track 2: Personalized E-Commerce Search Challenge

Organized by spirinus

In this challenge you are encouraged to build a personalized search ranking algorithm given large-scale data sets of search and ...

Aug 05, 2016-Oct 05, 2020

89 participants

	User	Team Name	FinalNDCG (weighted average)	SearchNDCG (query-full; textual queries)	CategoryNDCG (query-less; category facets)
1	minerva	Ali-Search	0.4262 (1)	0.5574 (1)	0.3935 (1)
2	Dmitrii_Nikitko		0.4149 (2)	0.5301 (2)	0.3861 (2)
3	tjy	red fruit yard	0.3916 (3)	0.4221 (5)	0.3840 (3)
4	wistuba		0.3769 (4)	0.4495 (4)	0.3588 (4)
5	joaopalotti		0.3712 (5)	0.4860 (3)	0.3425 (6)

Further work

- Learning from implicit data
  - Labeled Data Generate
- Model (Feature) learning
  - Model as Feature
- Scenario-dependent ranking

Summary

- Branch of Machine Learning
- Feature Extraction
- Ensemble Method
- Engineering
  - Dataflow
  - Workflow
- Production
  - New sort is greatly influenced by the initial sort.
  - Initial sort can probably be considered as not holding much pertinence information
  - Practical solution is zero all your rank feature before prediction

Reference

- C. J. C. Burges. From RankNet to LambdaRank to LambdaMART: An overview. Technical report, Microsoft Research, 2010.
- Hang Li. Learning to Rank. In ACML Tutorial, 2009
- Zhengdong Lu, Hang Li, A Deep Architecture for Matching Short Texts, In Proceedings of Neural Information Processing Systems 26 (NIPS), 1367-1375, 2013.
- Wei Wu, Hang Li, Jun Xu, Learning Query and Document Similarities from Click-through Bipartite Graph with Metadata, In Proceedings of the Sixth ACM International Conference on Web Search and Data Mining (WSDM), 687-696, 2013.
- 周志华. 机器学习, 171-190, 267-287, 2016
- Hang Li & Zhengdong Lu. Deep Learning for Information Retrieval. In SIGIR Tutorial, 2016
- http://mlwave.com/kaggle-ensembling-guide/
- http://machinelearningmastery.com/machine-learning-ensembles-with-r/

E-Mail: wuchen.wc@alibaba-inc.com