Opera Solutions, LLC
180 Maiden Lane
17th Floor
New York, NY 10038
+1 (646) 437 2100 telephone
+1 (646) 437 2101 facsimile
www.operasolutions.com

Personalize Expedia Hotel Searches – ICDM 2013

High-Level approach

Dec 6th,2013

Individual models

Performance check (model validation)

- 10% random subset as validation set fixed for ALL experiments

SGD model

- Own implementation
- Linear model on 19 numeric features (total model has only 19 parameters !)
- Trained with pairwise ranking updates for 50 epochs
- 0.50377 on leaderboard

Neural Network

- Own implementation
- Inputs: numeric + categoric features
- 1 hidden layer 30 neurons (tanh)
- Minibatch training with rank-updates
- Best individual: 0.5297 on leaderboard

GB - Gradient Boosted Decision Tree For Ranking

- Own implementation
- Base idea from C. Burges's LambdaMART
- Added: support for categoric features, random splits
- Best individual: 0.5256 on leaderboard

LambdaMART by RankLib

- → http://sourceforge.net/p/lemur/wiki/RankLib/
- Inputs: numeric features + average value of those per prop_id
- Best individual: 0.5338 on leaderboard

Model	NDCG valid	NDCG leaderboard
Random benchmark	0.348	0.3479
Position benchmark	0.4967	0.4999
valueBook sum	0.443	0.4418
SVD on prop_id / srch_dest_id	0.4618	0.46317

Good consistency of validation and leaderboard
Score by taking a random 10% data subset

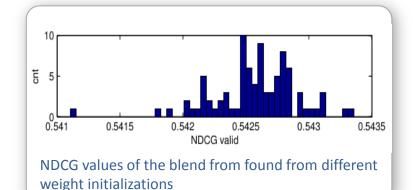## Blending

24 individual predictors (leaderboard scores)

- 6x Neural Net (0.511 … 0.529)
- 1x SGD model (0.52)
- 4x LambdaMART (0.510 … 0.533)
- 11x GB - Gradient Boosted Decision Tree For Ranking (0.49 … 0.526)

Linear Blender $s = \sum_{i=1}^{N} p_i w_i$

- Weights are search by parameter searcher APT1 (see Netflix Prize doc)
- Maximizing the NDCG on the validation set
- Take the best runs from 100 random initializations
  - Problems with many local minima
- Final solution: 0.5407 leaderboard score

Insights / Conclusion

- An ensemble of ranking models outperforms the best individual
- Dataset was large enough to use a single validation set
  (no cross-validation needed)
- Good data quality
- Time and price compare features does not help   Why ??
- Biggest improvements by adding features for each prop_id
  - Mean of numeric values
  - Stddev of numeric values
  - Median of numeric values

NDCG values of the blend from found from different weight initializations