Machine Learned Ranking of Entity Facets

Roelof van Zwol
Yahoo! Research
roelof@yahoo-inc.com

Lluis Garcia Pueyo
Yahoo! Research
lluis@yahoo-inc.com

Mridul Muralidharan
Yahoo! Research
mridulm@yahoo-inc.com

Börkur Sigurbjörnsson
Yahoo! Research
borkur@yahoo-inc.com

ABSTRACT

The research described in this paper forms the backbone of a service that enables the faceted search experience of the Yahoo! search engine. We introduce an approach for a machine learned ranking of entity facets based on user click feedback and features extracted from three different ranking sources. The objective of the learned model is to predict the click-through rate on an entity facet. In an empirical evaluation we compare the performance of gradient boosted decision trees (GBDT) against a linear combination of features on two different click feedback models using the raw click-through rate (CTR), and click over expected clicks (COEC). The results show a significant improvement in ranking performance, in terms of discounted cumulated gain, when ranking entity facets with GBDT trained on the COEC model. Most notably this is true when evaluated against the CTR test set.

Categories and Subject Descriptors

H.3.3 [**Information Retrieval**]: Information Search and Retrieval; H.3.5 [**Information Retrieval**]: On-line Information Services

General Terms

Experimentation, Measurement, Performance

Keywords

ranking entity facets, click feedback, GBDT

1. ABOUT RANKING ENTITY FACETS

The major Web search engines are gradually changing the search experience. Most notably this is visible through the introduction of semantic search assistants, the enrichment of the search results shown to the user and other components that try to predict the user intent. Key to enriching the search experience is the wide-scale availability of user-generated content and other knowledge bases such as Wikipedia, the Internet Movie Database (IMDB), GeoPlanet^{TM}, or Freebase to name a few.

The research presented here is part of the faceted search experience of the Yahoo! Web and Image search engines [4].

SIGIR'10, July 19–23, 2010, Geneva, Switzerland.
ACM 978-1-60558-896-4/10/07.

Figure 1: Screen captions of entity facets show in the search engine interface.

Figure 1 shows a fragment of the search engine interface depicting the facets bar for the celebrity "Daniel Day-Lewis" and the location "Geneva, Switzerland".

To support the entity facet ranking application of Figure 1, we propose a machine learned ranking of entity facets based on user click feedback. Given an entity of interest, we have collected a large pool of related candidate facets, e.g. related entities. These entity facet pairs have been extracted from various knowledge bases such as Wikipedia, GeoPlanet^{TM}and other sources. Typically this provides us with a few hundred candidate facets for an entity. The task is then to rank the candidate facets related to each entity in our pool based on its relevance.

Facets are ranked using statistical features extracted from three different sources that contain entity information in the context of images: query terms entered by users in query logs, query session information from users in query logs, and tags provided by users annotating their photos in Flickr. Query term information captures entities that co-occur in a single user query, while query sessions provide information about entities that frequently co-occur in a user session. Flickr tags have a good coverage of travel and location related entities, as well as topics of a more general nature, but tend to be less focussed on, for instance, celebrity entities. For every source, we extract different unary, symmetric and asymmetric features such as query frequency, conditional probability, KL divergence, etc. [4]. For the initial launch of the faceted search experience, we constructed a ranking function that is a linear combination of the conditional probabilities extracted from the three ranking sources .

The main contribution of this paper is a machined learned approach for ranking entity facets based on user click feedback. We propose to learn a ranking using the full set of features extracted from the ranking sources that will predict the click-through rate (CTR) on an entity facet [1]. For that purpose we introduce two click models: raw click-through rate on the facets, and the click over expected click (COEC), which is claimed to be more robust towards the position-bias on a click as users tend to click more on those results shown high in the ranking [3].

The click-feedback is used as the ground truth for our training, development and test sets. We have experimented with various learners, but for the experiment reported here we limit ourselves to the discussion of the performance using stochastic gradient boosted decision trees (GBDT) [2]. We used least squares regression as our loss function.

Next we collected the user click feedback on the facets over a period of three months, based on which we compute the click-through rate and click over expected click for each entity facet pair that was shown at least 25 times to a user. The latter constraint is to ensure that the CTR and COEC values are stable enough to be used as the labels for our training, development, and test sets. We join the feature set with the CTR and COEC sets, using the entity facet pair as the key. Next we split the collection into training, development and test sets. When splitting, we ensure that an entity can only occur in one of the three collections.

2. EVALUATION

The objective of the experiment is to measure the prediction accuracy based on the click-through rate of an entity facet pair. We first present the setup of the experiment, followed by a discussion of the results of the evaluation.

Ranking strategies. Central in the experiment are the three ranking strategies: (1) **Baseline**. A linear combination of the conditional probabilities. (2) **GBDT**$_{ctr}$. GBDT trained on the CTR click model and (3) **GBDT**$_{coec}$. GBDT trained on the COEC click model.

Test sets. For the experiment we use two test sets both containing the same 100 entities and their 10+ facets that have not been used for training or parameter tuning. For a fair comparison of the performance between queries we have normalized the CTR and COEC values for each of the facets of the 100 selected entities to be in the range of $[0, 1]$.

Evaluation metrics. To evaluate the performance on the test collection, we adopt *Discounted Cumulative Gain* (DCG) as our metric. DCG is an effectiveness measure that is used frequently for information retrieval tasks, and allows for the use of a graded relevance scale.

Results. The overall performance of the baseline and the two GBDT models is reported in Table 1. For each of the two test sets, CTR and COEC, the mDCG and mnDCG is included. The performance of all strategies, independent of the test set is good ($mnDCG > 0.67$). It can be clearly seen that on both the CTR and COEC test sets, the GBDT models outperform the baseline strategy. Using the normalized metric (mnDCG) we see that can better estimate the actual COEC using the GBDT$_{coec}$ model than predicting the raw CTR with our GBDT$_{ctr}$ model. This gives us a first indication that the COEC click model is more effective than the CTR click model when learning to rank entity facets.

Figure 2 plots the nDCG scores at different points in the ranking. This allows for a direct comparison of the different strategies across the two test sets. In addition to the strategies already discussed, we introduce a new variant where we evaluate the performance of the GBDT$_{coec}$ strategy, e.g. the GBDT model that was trained to optimized the click prediction on the COEC click model, against the CTR test set. As can be seen this gives a near optimal performance over the first ten positions in the ranking, and a perfect prediction of the most important facet for each of the 100 entities in our test set.

Table 1: Overall performance.

	CTR		COEC	
Run	**mDCG**	**mnDCG**	**mDCG**	**mnDCG**
Ideal	2.375	-	2.594	-
Baseline	1.728	0.709	1.812	0.677
GBDT$_{ctr}$	2.090	0.874	-	-
GBDT$_{coec}$	2.343	0.986	2.436	0.930

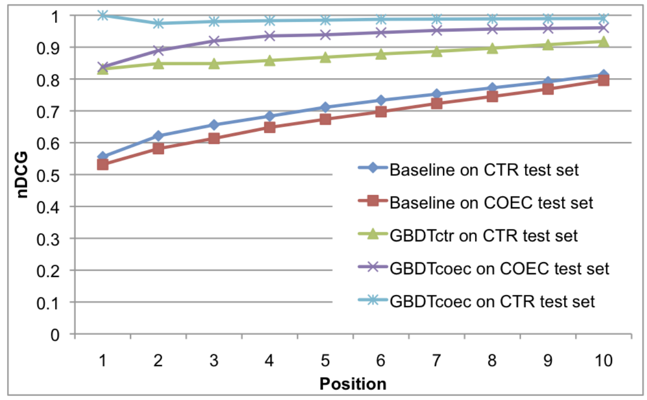

Figure 2: nDCG comparison of performance of baseline and GBDT on CTR vs COEC test sets.

3. REFERENCES

[1] N. Craswell, O. Zoeter, M. Taylor, and B. Ramsey. An experimental comparison of click position-bias models. In *WSDM '08: Proceedings of the international conference on Web search and web data mining*, pages 87–94, New York, NY, USA, 2008. ACM.

[2] J. H. Friedman. Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29:1189–1232, 2001.

[3] Y. Zhang and R. Jones. Comparing click logs and editorial labels for query rewriting. In *Query Log Analysis: Social And Technological Challenges*, 2007.

[4] R. van Zwol, B. Sigurbjörnsson,et al. Faceted Exploration of Image Search Results. In *WWW2010.* Raleigh, NC, USA. 2010.